Mathematical Physics 1: Linear algebra lecture notes

Govind S. Krishnaswami, Chennai Mathematical Institute, Oct 2009, updated 29 June 2016.

Contents

These notes are a short summary of the topics covered in this course for first year (first semester) B.Sc. students. They are not complete and are not a substitute for a text book (some suggestions may be found below). But they could be useful to you if you work out the examples, fill in the details and try to provide any missing proofs. Any corrections/comments may be sent to govind@cmi.ac.in

1 Some text books for linear algebra

- C. Lanczos, Applied analysis - chapter 2 on matrices and eigenvalue problems
- C. Lanczos, Linear differential operators, chapter 3 on matrix calculus
- T. M. Apostol, Calculus Vol 2, chapters 1-5
- Gilbert Strang, Introduction to linear algebra
- Gilbert Strang, Linear algebra and its applications
- Courant and Hilbert, Methods of mathematical physics, Vol 1
- Arfken and Weber, Mathematical methods for physicists
- Sheldon Axler, Linear algebra done right
- P.R. Halmos, Finite-dimensional vector spaces
- Serge Lang, Introduction to linear algebra
- Erwin Kreyszig, Advanced engineering mathematics

• K T Tang, Mathematical Methods for Engineers and Scientists 1: Complex Analysis, Determinants and Matrices

In addition, there are several books in the CMI library that cover linear algebra. Look under Dewey classification 512.5 for books from a mathematical viewpoint and under 530.15 for books on mathematical methods for physicists (eg. the book by Dettman or others such as 530.15 CAN, DEN, DET, COU, DAS, JEF)

2 Initial remarks

• Linear algebra is useful in many classical physics and engineering problems.

• Linear equations are a first approximation to more complicated and accurate non-linear equations. Near a point of equilibrium we can often linearize the equations of motion to study oscillations: vibrations of a solid or LC oscillations in an electrical circuit.

• Importance of linear algebra in physics is greatly amplified since quantum mechanics is a linear theory.

• Linear algebra is an example of a successful mathematical theory with many very satisfying theorems eg. Spectral decomposition.

• Linear algebra is also important in computer science e.g., web search, image (data) compression.

• Linear algebra is important in statistics: least squares fitting of data, regression.

• Linear algebra is important in electrical engineering, eg. the fast fourier transform.

• Linear algebra is fun and the basic concepts are not difficult. It has a nice interplay between algebra (calculation) and geometry (visualization). It may also be your first encounter with mathematical abstraction, eg. thinking of spaces of vectors rather than single vectors.

• The basic objects of linear algebra are (spaces of) vectors, linear transformations between them and their representation by matrices.

• Examples of vectors include position and momentum of a particle, electric and magnetic fields at a point etc.

• Examples of matrices include inertia $I_{ij} = \int \rho(\vec r) \left(r^2 \delta_{ij} - r_i r_j\right) d^3r$ tensor of a rigid body, stress tensor (momentum flux density) $S_{ij} = p\delta_{ij} + \rho v_i v_j$ of an ideal fluid, Minkowski metric tensor $\eta_{\mu\nu}$ of space-time in special relativity.

• Matrix multiplication violates the commutative law of multiplication of numbers $AB \neq BA$ in general. Also there can be non-trivial divisors of zero: matrices can satisfy $AB = 0$ with neither A nor B vanishing.

• Matrix departure from the classical axioms of numbers is as interesting as spherical geometry departure from the axioms of Euclidean geometry.

3 An initial view of linear equations

• Many equations of physics and engineering are differential equations. But a differential is a limit of a difference quotient if we discretize. This turns linear differential equations into linear algebraic equations. We will make this precise later on. Linear equations also arise for example in the relation of the angular momentum vector to the angular velocity vector $L = I\omega$, where I is the inertia matrix.

• A basic problem of linear algebra is to solve a system of linear algebraic equations.

$$\begin{aligned} x - 2y &= 1 \qquad \textit{or} \qquad & ax + by &= c \\ 3x + 2y &= 11 & dx + ey &= f. \end{aligned} \tag{1}$$

• Cayley recognized that these should be thought of as a single matrix equation $Ax = b$.

$$\begin{pmatrix} 1 & -2 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 11 \end{pmatrix} \tag{2}$$

• Matrix vector multiplication $A \begin{pmatrix} x \\ y \end{pmatrix}$ by rows: dot product of each row of A with the column vector $\begin{pmatrix} x \\ y \end{pmatrix}$.

• A column vector such as $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ can be viewed as an arrow from the origin to the point $(1, 2)$ of the plane.

• Index notation for components of a vector and a matrix: Rather than calling the coefficients $a, b, d, e \cdots$, we can label them $a_{11}, a_{12}, a_{21}, a_{22}$ and the RHS as b_1, b_2

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{3}$$

• a_{ij} are called the components or entries (or matrix elements) of the matrix A.

• Finding the right notation is part of the solution of a problem.

• The matrix A operates on a vector x to produce the output Ax. So matrices are sometimes called operators.

• In this section we will not be very precise with definitions, and just motivate them with examples.

• A way of looking at matrix vector multiplication: $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} a \\ c \end{pmatrix} + y \begin{pmatrix} b \\ d \end{pmatrix}$ as linear combinations of columns: $x\times$ first column plus $y\times$ second column.

• Input variables x, y or vector $\begin{pmatrix} x \\ y \end{pmatrix}$ live in the domain space, here it consists of two component column vectors.

• The RHS vector $b = \begin{pmatrix} 1 \\ 11 \end{pmatrix}$ lives in the codimain or target space. Here the target space consists of two component vectors.

• The possible outputs, vectors Ax form the range or image space. The image is necessarily a subset of the target.

• Row picture (domain picture): Each row $x - 2y = 1, 3x + 2y = 11$ defines a line (more generally a plane or hyperplane) and the solution is at the intersection of these lines. This is the domain space picture since it is drawn in the space of input variables x, y.

• Column picture (target space picture of $Ax = b$): Linear combinations of the column vectors of A to produce the column vector b. The columns are possible outputs (for the inputs $(x = 1, y = 0)$ and $(x = 0, y = 1)$). So the columns are in the range, which is in the target.

• The row and column pictures, though both drawn on a plane are not the same spaces, one is in the domain and other in the target.

• In this example, linear combinations of the columns span the whole plane, so we can get any output vector $\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$ we want. Such square matrices are invertible.

• Example: $\begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 4 & 8 & 16 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \\ 20 \end{pmatrix}$ or

$$\begin{aligned} x + 2y + 4z &= 5 \\ 2x + 4y + 8z &= 10 \\ 4x + 8y + 16z &= 20. \end{aligned} \tag{4}$$

• Notice that the RHS is the sum of the first and third columns, so one possible solution is $(x, y, z) = (0, 1, 1)$.

• The coefficient matrix in this case is symmetric $A^T = A$. Transpose of matrix exchanges rows with columns. What is the transpose of A from the first example?

• But this is not the only solution, since the latter two equations are just multiples of the first one. So we have a 2-parameter family of solutions $(5 - 2y - 4z, y, z)$ with y and z arbitrary.

• So though there were apparently 3 equations in 3 unknowns, in fact there is only one independent equation and the system is 'under-determined'.

• The same system of equations has no solution if the RHS is changed a little bit at random to say, $(5, 10, 16)$. This is because the last two columns are multiples of the first column so the only possible outputs are multiplies of the first column.

• So a system can go from infinitely many solutions to no solutions with a small change in the RHS!

• This matrix will turn out not to be invertible. Generally, invertible square matrices are those for which you can uniquely solve $Ax = b$ for *any* b.

• It seems natural to have as many equations as unknowns (even-determined systems). But it is necessary in many contexts to study a rectangular system.

• Newton's second order differential equation for the position of a particle, $m\frac{\partial^2 x}{\partial t^2} = f(t)$ needs two initial conditions $x(0), \dot{x}(0)$. But we may wish to study the solutions without specifying the initial conditions. Then we have two less equations than unknowns.

• $\nabla V = F$. Given the force F, these are 3 equations for one unknown potential V, so it is strongly over determined. Only conservative forces arise from a potential!

• $\nabla \cdot v = 0$ is one equation for three velocity components of an incompressible fluid flow, so strongly under determined. There are many such flows!

• Rectangular underdetermined example. Try to solve it, draw the row and column pictures.

$$\begin{aligned} x + 2y + 3z &= 6 \\ 2x + 5y + 2z &= 4 \end{aligned} \tag{5}$$

• Square incompatible system, this one has no solutions:

$$\begin{aligned} x + 3y + 5z &= 4 \\ x + 2y - 3z &= 5 \\ 2x + 5y + 2z &= 8 \end{aligned} \tag{6}$$

• Notice that the third row is the sum of the first two, but the same is not true of the RHS.

• Linear algebra develops a systematic theory to understand all these possible features of a system of m equations in n unknowns. It also provides algorithms for solving the equations!

4 Matrix Multiplication

• An $m \times n$ matrix is a rectangular array of numbers (real or complex) with m rows and n columns. If $m = n$ we have a square matrix. If $m = n = 1$ the matrix reduces to a number (scalar). A $1 \times n$ matrix is a row vector. An $m \times 1$ matrix is a column vector.

• Av multiplication of a column vector by a matrix from the left is a new column vector. It is a linear combination (specified by the components of v), of the columns of A

$$\begin{pmatrix} | & | & \cdot & \cdot & | \\ c_1 & c_2 & \cdot & \cdot & c_n \\ | & | & \cdot & \cdot & | \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{pmatrix} = v_1 \begin{pmatrix} | \\ c_1 \\ | \end{pmatrix} + v_2 \begin{pmatrix} | \\ c_2 \\ | \end{pmatrix} + \cdots v_n \begin{pmatrix} | \\ c_n \\ | \end{pmatrix} \tag{7}$$

• Row picture of multiplication of a row vector from right by a matrix xA. The result is a linear combination of the rows of A, i.e. a new row vector.

$$\begin{pmatrix} x_1 & x_2 & \cdot & \cdot & x_m \end{pmatrix} \begin{pmatrix} \text{row}_1 \\ \text{row}_2 \\ \cdot \\ \cdot \\ \text{row}_m \end{pmatrix} = x_1(\text{row}_1) + x_2(\text{row}_2) + \cdots + x_m(\text{row}_\text{m}) \tag{8}$$

• Suppose A is an $m \times n$ matrix and B is $n \times p$. Then C = AB is an $m \times p$ matrix.

• Matrix multiplication in components $\sum_{k=1}^n A_{ik} B_{kj} = C_{ij}$

• Summation convention: repeated indices are summed except when indicated

• Sometimes we write A^i_k for A_{ik}, with row superscript and column subscript. Then $A^i_k B^k_j = C^i_j$.

• The identity matrix I is the diagonal matrix with $1's$ along the diagonal. $Iv = v$ for every vector and $IA = AI = A$ for every matrix.

• Matrix multiplication is associative, can put the brackets anywhere $A(BC) = (AB)C \equiv ABC$

• To see this, work in components and remember that multiplication of real/complex numbers is associative

$$[A(BC)]_{il} = A_{ij}(BC)_{jl} = A_{ij}B_{jk}C_{kl} = [(AB)C]_{il} \tag{9}$$

• Matrix multiplication distributes over addition $A(B+C) = AB + AC$

• Addition of matrices is commutative $A + B = B + A$, we just add the corresponding entries.

• The zero matrix is the one whose entries are all $0's$. $A + 0 = A$ and $0A = 0$ for every matrix, and $0v = 0$ for every vector.

• Outer product of a column vector with a row vector gives a matrix: This is just a special case of matrix multiplication of $A_{m\times 1}$ with $B_{1\times n}$ to give a matrix $C_{m\times n}$. For example

$$\begin{pmatrix} x \\ y \end{pmatrix} \begin{pmatrix} z & w \end{pmatrix} = \begin{pmatrix} xz & xw \\ yz & yw \end{pmatrix} \tag{10}$$

• Four different ways to multiply matrices. All are a consequence of the basic formula $\sum_{k=1}^n A_{ik}B_{kj} = C_{ij}$. They are just different ways of interpreting this equation. It is always good to have such different algorithmic/physical/geometric interpretations of equations.

• Example to practice matrix multiplication for rectangular matrices. Do it each of the four ways described below.

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 & 2 \\ 1 & 0 & 1 & 2 \\ -1 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 4 & 4 & 6 \\ -2 & 1 & 0 & -2 \end{pmatrix} \tag{11}$$

• 1. Rows times columns: This is the traditional way. C_{ij} is the dot product of the $i^{\rm th}$ row of A with the $j^{\rm th}$ column of B. To see this from the formula, fix both i and j and consider $\sum_{k=1}^n A_{ik}B_{kj} = C_{ij}$. The sum over k signifies the dot product of the $i^{\rm th}$ row of A with the $j^{\rm th}$ column of B.

• 2. By columns: columns of C are linear combinations of columns of A. The $j^{\rm th}$ column of C is a linear combination of the columns of A, each column taken as much as specified by the $j^{\rm th}$ column of B. This again follows from $\sum_{k=1}^n A_{ik}B_{kj} = C_{ij}$, by fixing j, so that we are looking at the $j^{\rm th}$ column of C (i is arbitrary). Then think of the B_{kj} as coefficients in a linear combination of the columns of A.

• 3. By rows: rows of C are linear combinations of rows of B. The $i^{\rm th}$ row of C is a linear combination of the rows of B, each row taken as much as specified by the $i^{\rm th}$ row of A.

• 4. Columns times rows: As a sum of matrices which are *outer products* of the columns of A and rows of B. For example, consider $A_{m\times 2}$ and $B_{2\times p}$ matrices. The matrix elements of their product can be written as

$$C_{ij} = A_{ik}B_{kj} = A_{i1}B_{1j} + A_{i2}B_{2j} \tag{12}$$

The first term in the sum $A_{i1}B_{1j}$ is the outer product of the first column of A with the first row of B. The second term is the outer product of the second column of A with the second row of B. More generally for $A_{m\times n}B_{n\times p}$, we will have a sum of n matrices.

• Example to see elementary row operation. $C = AB$ means A multiplies B from the left. Now do the multiplication by rows, i.e. rows of C are linear combinations of rows of B, with weights specified by the corresponding row of A.

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ -1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ -1 & 2 & 1 \end{pmatrix} \tag{13}$$

• Observe that left multiplying by A does the following elementary row operation: 1^{st} and 3^{rd} rows of C are the same as those of B. But the second row of C is obtained by multiplying the first row of B by -2 and adding it to the second row of B.

• Example that shows a product of non-zero matrices can be zero

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{14}$$

• Multiplication of matrices is in general not commutative, i.e. AB need not equal BA. For example, check this for

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{15}$$

In this case you will find that $AB = -BA$. But this is not so in general, as the following example indicates

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 3 \\ -1 & -2 \end{pmatrix} \Rightarrow AB = \begin{pmatrix} 0 & -1 \\ -2 & 3 \end{pmatrix}, \quad BA = \begin{pmatrix} 11 & 16 \\ -7 & -8 \end{pmatrix}. \tag{16}$$

5 Linear combinations and Linear (in)dependence

• Let us formally define two concepts which we have already met a few times in the lectures.

• Given a collection of vectors $v_1, v_2, \cdots, v_n$, a linear combination is a weighted sum $a_1 v_1 + a_2 v_2 + \cdots a_n v_n$, where a_i are numbers (real or complex)

• For example, $3\hat{x} + 2\hat{y}$ is a linear combination of these two unit vectors.

• Vectors are linearly dependent if there is a non-trivial linear combination of them that vanishes. i.e. the vectors satisfy a linear relation.

• For example $\hat{x}$ and $3\hat{x}$ are linearly dependent since they satisfy the linear relation $3(\hat{x}) - 3\hat{x} = 0$

• More formally, $v_1, v_2, \cdots, v_n$ are linearly dependent if $\sum_{i=1}^n a_i v_i = 0$ for some real numbers a_i not all zero.

• On the other hand, there is no non-trivial linear combination of $u = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $v = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ that vanishes. We say that u, v are linearly independent.

• Definition: $v_1, v_2, \cdots, v_n$ are linearly independent if the following condition is satisfied: $a_1 v_1 + a_2 v_2 + \cdots a_n v_n = 0 \quad \Rightarrow \quad a_1 = a_2 = a_3 = \cdots = a_n = 0$.

6 Gaussian elimination

• Now let us get back to a systematic procedure to solve systems of linear equations.

• Diagonal example: notice how easy it is to solve a system associated to a diagonal matrix

$$\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -3 \end{array} \right) \left(\begin{array}{c} x \\ y \\ z \end{array} \right) = \left(\begin{array}{c} \pi \\ e \\ 21.3 \end{array} \right) \tag{17}$$

• It will be fruitful to try to diagonalize matrices. Notice that it was important that the diagonal entries were not zero in order to find a solution, i.e. for invertibility.

• Actually, it is almost as easy to solve an upper-triangular system

$$\begin{array}{rcl} x - y + 3z &=& 0 \\ y + 2z &=& 3 \\ z &=& 1 \end{array} \tag{18}$$

• Back-substitute starting with the last equation to get $\left(\begin{array}{c} x \\ y \\ z \end{array} \right) = \left(\begin{array}{c} -2 \\ 1 \\ 1 \end{array} \right)$

• So it is worth reducing a matrix to upper triangular form. This is what Gaussian elimination does.

• Gaussian elimination is a systematic way of solving linear systems. This is how you would do it in practice and also how computer programs work.

• Cramer's rule using determinants would take much longer, we mention it later.

• Forward elimination and back substitution

• Work out an example

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \\ 2 & 6 & 8 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}. \tag{19}$$

• Forward elimination by elementary row operations: We bring A to upper triangular form by subtracting a multiple of one equation from another.

• For this we must carry along the rhs, so first form the augmented coefficient matrix

$$(A|b) = \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 1 & 0 & 2 & | & 2 \\ 2 & 6 & 8 & | & 3 \end{pmatrix} \tag{20}$$

• $a_{11} = 1$ is called the first pivot, it is important that it is non-zero. If it were zero we would look for a row whose first entry is non-zero and swap it with the first row.

• Then we want all entries below the first pivot to be zero, so subtract row-1 from row-2. Then subtract twice row-1 from row-3

$$\begin{aligned}(A|b) \;&=\; \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 1 & 0 & 2 & | & 2 \\ 2 & 6 & 8 & | & 3 \end{pmatrix} \xrightarrow{E_{21}} \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 0 & -2 & -1 & | & 1 \\ 2 & 6 & 8 & | & 3 \end{pmatrix} \xrightarrow{E_{31}} \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 0 & -2 & -1 & | & 1 \\ 0 & 2 & 2 & | & 1 \end{pmatrix} \\ &\xrightarrow{E_{32}} \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 0 & -2 & -1 & | & 1 \\ 0 & 0 & 1 & | & 2 \end{pmatrix} = (U| \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}) \end{aligned} \tag{21}$$

• Operation E_{21} makes the $(2,1)$ entry zero, and similarly E_{31} makes the $(3,1)$ entry zero.

• Thus A has been transformed into an upper triangular matrix U, with zeros below the diagonal. U is called the row-echelon form of A.

• The word echelon refers to a staircase shape. Echelon should remind you of a hierarchy: 'higher echelons of government'. Echelon form is also a military term referring to an arrangement of soldiers in rows with each row jutting out a bit more than the previous one.

• The pivots are the first non-zero entries in each row, when a matrix has been brought to row echelon (upper triangular) form.

• If there is no pivot in a row, then that is a row of zeros. A row of zeros means that row of A was a linear combination of other rows.

• Here, the pivot positions are the diagonal entries from the upper left, the pivots are $1, -2, 1$.

• The rank r is the number of pivots. A has rank $r = 3$, which is the maximum possible for a 3×3 matrix.

• A square $n \times n$ matrix is invertible if and only if the rank $r = n$.

• The rank is the number of linearly independent rows. In echelon form, the linearly independent rows are precisely the ones with pivots.

• There is at most one pivot in each column. In echelon form, the linearly independent columns are precisely the columns with pivots. The rank is also the number of linearly independent columns.

• The method is simplest if we have a pivot in each row.

• If there is a zero in a pivot position, we exchange with a lower row if available.

• For a square matrix, if there is a zero in the pivot position which cannot be removed by row exchanges, then we get a row of zeros, and the matrix is called singular or not invertible. For example, if $A_{33} = 7$ in the above example, then the third row of U would be full of zeros. We will deal with such situations later on.

• More precisely, we can define an elementary row operation as that of subtracting a multiple l_{ji} of row i from row j. In other words, $r_j \mapsto r_j + l_{ji} r_i$ while the remaining rows are left unaltered.

• In row echelon form, the equations are $\begin{pmatrix} 1 & 2 & 3 \\ 0 & -2 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$. Now it is easier to solve the equations starting with $z = 2$ and back-substituting to get $y = -3/2, x = -2$.

• Better still, we could continue with elementary row operations upwards, to make all entries above the pivots zero as well, i.e. get a diagonal matrix

$$\begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 0 & -2 & -1 & | & 1 \\ 0 & 0 & 1 & | & 2 \end{pmatrix} \xrightarrow{r_2 \to r_2 + r_3} \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 0 & -2 & 0 & | & 3 \\ 0 & 0 & 1 & | & 2 \end{pmatrix} \xrightarrow{r_1 \to r_1 - 3r_3}$$
$$\begin{pmatrix} 1 & 2 & 0 & | & -5 \\ 0 & -2 & 0 & | & 3 \\ 0 & 0 & 1 & | & 2 \end{pmatrix} \xrightarrow{r_1 \to r_1 + r_2} \begin{pmatrix} 1 & 0 & 0 & | & -2 \\ 0 & -2 & 0 & | & 3 \\ 0 & 0 & 1 & | & 2 \end{pmatrix} = (D| \begin{pmatrix} -2 \\ 3 \\ 2 \end{pmatrix}) \tag{22}$$

and it is easy to read off the solution $x = -2, y = -3/2, z = 2$.

• The diagonal matrix D contains the pivots along the diagonal. However, in general these are *not* the eigenvalues of A! But they are related, we will see that the determinant of A, which is the product of its eigenvalues, is also (upto a possible sign) the product of the pivots! Finding the pivots is relatively easy for large matrices, but finding eigenvalues is much harder.

6.1 More examples of gaussian elimination

• Let us consider some examples where the rank is less than maximal.

• Example of singular matrix but compatible system with infinitely many solutions

$$2x - y = 1, \; 4x - 2y = 2; \quad \Rightarrow \quad \begin{pmatrix} 2 & -1 & | & 1 \\ 4 & -2 & | & 2 \end{pmatrix} \to \begin{pmatrix} 2 & -1 & | & 1 \\ 0 & 0 & | & 0 \end{pmatrix} \tag{23}$$

There is only one pivot, 2. Elimination leads to a row of zeros, so the matrix is not invertible. Its rank is 1 which is less than the dimension 2. A system is compatible if each row of zeros in echelon form is matched with a zero on the rhs. This is indeed the case here, indicating that the second equation was redundant. We have only one linearly independent equation and infinitely many solutions parameterized by one parameter, $(x = \frac{1}{2}(a+1), y = a)$.

• Example of singular matrix and incompatible system with no solutions

$$2x - y = 1; \quad 4x - 2y = 3 \quad \Rightarrow \quad \begin{pmatrix} 2 & -1 & | & 1 \\ 4 & -2 & | & 3 \end{pmatrix} \to \begin{pmatrix} 2 & -1 & | & 1 \\ 0 & 0 & | & 1 \end{pmatrix} \tag{24}$$

There is only one pivot, 2. Elimination leads to a row of zeros, so the matrix is not invertible. Its rank is 1 which is less than the dimension 2. A system is compatible if each row of zeros in echelon form is matched with a zero on the rhs. That is not the case here, so we have an incompatible system and no solutions.

6.2 Formulating elementary row operations using matrix multiplication

• Let us return to the above 3×3 example.

• Formulation of elementary row operations using matrix multiplication from left. The procedure of subtracting row-1 from row-2 to make the(21)-entry vanish, can be achieved by multiplying A by the elementary matrix E_{21}

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 2 \\ 2 & 6 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -2 & -1 \\ 2 & 6 & 8 \end{pmatrix} \tag{25}$$

• To see this recall how to multiply matrices row-by-row. That method pays off now!

• Similarly: $E_{31} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}$ subtracts row-1 from row-3 and makes the $(3,1)$ entry vanish.

• And $E_{32} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ adds row-2 to row-3 and makes the $(3,2)$ entry vanish.

• Notice that the elementary matrices are always lower triangular (have zeros above the diagonal) since we add a multiple of an earlier row to a later row. They are also called elimination matrices.

• More formally, an elementary row operation $r_i \mapsto r_i + m_{ij} r_j$, adds a multiple (m_{ji}) of row j to row i. It leaves all rows other than row j fixed. It is represented by left multiplication by the following elementary matrix E. E has 1's along the diagonal. All other entries vanish except for the (i,j)-entry, which is m_{ij}

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & m_{ij} & \cdots & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{26}$$

• In terms of elementary matrices we have $E_{32}(E_{31}(E_{21}A)) = U$. Using associativity to move the brackets around, we have $EA = U$ where $E = E_{32}(E_{31}E_{21})$ and U is upper triangular (echelon form).

• As a bonus we will get a decomposition of a matrix into a product of lower and upper triangular matrices $A = LU$. U is in row echelon form, with pivots as the first non-zero entry in each row, and L is a lower triangular matrix with $1's$ as the last non-vanishing entry in each row. This is a nice way of summarizing gaussian elimination. If A is square and invertible, then the pivots and 1's will be along the diagonal.

• From $E_{32}(E_{31}E_{21})A = U$, we move the elementary matrices E_{ij} to the RHS by inverting them: $A = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U = LU$.

• The inverse of an elementary matrix is easy: E_{21} subtracts the first row from the second. So $E_{21}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ adds the first row to the second.

• Similarly, $E_{31}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$ and $E_{32}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$.

• In particular, the inverse of a lower triangular elementary matrix is also lower triangular and elementary.

• Now $L = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix}$, notice that it is lower triangular.

• Check that $EL = I$ and that $A = LU = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -2 & -1 \\ 0 & 0 & 1 \end{pmatrix}$ as advertised.

• Show that the product of lower triangular matrices is again lower triangular.

6.3 Row exchanges and permutation matrices

• Consider the system

$$-y = 1, \quad 4x - 2y = 3, \Rightarrow \begin{pmatrix} 0 & -1 & | & 1 \\ 4 & -2 & | & 3 \end{pmatrix} \to \begin{pmatrix} 4 & -2 & | & 3 \\ 0 & -1 & | & 1 \end{pmatrix} \tag{27}$$

Since there was a zero in the $(1,1)$ position, we exchanged the first two rows. And now the matrix is already in row echelon form with two pivots $4, -1$

• More generally, if there was a zero in a pivot position and we exchanged rows to bring A to row echelon form U, then we have $PA = LU$, where P is the product of the row exchanges performed, P is a permutation matrix.

• For example, $P_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ permutes the first two rows of A when multiplied from the left PA. Check that it has the desired effect on the identity matrix. So the permutation matrix is obtained by applying the desired permutation to the identity.

• What is the permutation matrix that cyclically permutes the rows $1 \to 2 \to 3$?

• An invertible square matrix A has a unique decomposition $PA = LDU$ with L lower triangular with 1's along the diagonal, D diagonal with pivots along the diagonal and U upper triangular with 1's along the diagonal. P is a permutation matrix needed to bring A to echelon form.

6.4 Inverse of a square matrix

• A square matrix maps n-component column vectors in the domain to n-component column vectors in the target.

• The inverse of A (when it exists) must go in the opposite direction.

• Picture A and A^{-1} as maps between sets.

• The problem of inverting a matrix A is related to the problem of solving $Ax = b$ and expressing the answer as $x = Lb$. But for this to be the case, we need $LA = I$. This motivates the definitions that follow.

• If A has a left inverse $LA = I$ and a right inverse $AR = I$, then they must be the same by associativity (we can move brackets around)

$$(LA)R = L(AR) \quad \Rightarrow \quad IR = LI \quad \Rightarrow \quad R = L = A^{-1} \tag{28}$$

• An $n \times n$ square matrix is defined to be invertible if there is a matrix A^{-1} satisfying $A^{-1}A = AA^{-1} = I$. If not, A is called singular.

• In terms of maps, invertibility implies that A and A^{-1} must be 1-1. Moreover, the image of A must be the domain of A^{-1}, and the image of A^{-1} must equal the domain of A.

• When the inverse exists, it is unique by associativity. Suppose A has two inverses B and C, then by definition of inverse,

$$AB = BA = I, \qquad CA = AC = I. \tag{29}$$

Using associativity, $(CA)B = C(AB)$ but this simplifies to $B = C$.

• A real number is a 1×1 matrix. It is invertible as long as it is not zero. Its inverse is the reciprocal.

• A 2×2 matrix $A = (ab|cd)$ is invertible iff the *determinant* $ad - bc \neq 0$. Its inverse is

$$A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{30}$$

To get this result, use Gaussian elimination for the system

$$ax + by = f, \qquad cx + dy = g, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{31}$$

The augmented matrix in row echelon form and finally diagonal form are

$$\begin{pmatrix} a & b & | & f \\ 0 & d - \frac{bc}{a} & | & g - \frac{cf}{a} \end{pmatrix} \to \begin{pmatrix} a & 0 & | & f - \frac{ba}{ad-bc}(g - \frac{cf}{a}) \\ 0 & \frac{ad-bc}{a} & | & g - \frac{c}{a}f \end{pmatrix} \tag{32}$$

• The solution may be read off as $x = \frac{df-bg}{ad-bc}$, $y = \frac{-cf+ag}{ad-bc}$. For the solution to exist for arbitrary data f, g, we need $ad - bc \neq 0$. We can write the solution in matrix form and then read off the inverse

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} f \\ g \end{pmatrix} = A^{-1}\begin{pmatrix} f \\ g \end{pmatrix} \tag{33}$$

• We come to a very useful criterion for invertibility.

• A matrix is invertible iff $A_{n\times n}$ does not annihilate any non-zero vector. Vectors annihilated by A are called its zero-modes and they form its kernel $ker(A)$ or null space $N(A)$.

• Indeed, if A is invertible, then $Ax = 0$ implies $x = A^{-1}0 = 0$, so A has a trivial kernel.

• The converse is harder to prove and can be skipped. The idea is that if the kernel of A is trivial, then the columns of A are linearly independent. So the image of $x \mapsto Ax$ is the whole of the target space of n-component vectors (i.e. A maps onto the target). Then we can use the method given below to find the left inverse $LA = I$ of A. Now if the columns of a square matrix A are linearly independent, it also follows (say using gaussian elimination) that the rows are linearly independent. So acting on row vectors yA, A maps onto the space of n-component row vectors. Then we can use the method below to find the right inverse $AR = I$ of A. Finally by an earlier lemma, we know that $L = R = A^{-1}$, so we have shown that A is invertible.

• Remark: The kernel being trivial is equivalent to saying that the columns are linearly independent. Indeed $Av = v_1c_1 + v_2c_2 + \cdots v_nc_n$ is a linear combination of the columns of A. So if the columns are linearly independent, this vanishes only for the zero coefficients $v_i \equiv 0$ and so there is no vector annihilated by A.

• The point about invertibility of A is that it *guarantees unique* solutions to the $n \times n$ systems $Ax = b$ and $yA = c$ for *any* column n-vector b and any row n-vector c

$$x = A^{-1}b, \quad \text{and} \quad y = cA^{-1} \tag{34}$$

• Conversely, suppose we can solve $Ax = b, yA = c$ for any b, c. Then the left inverse $LA = I$ is the matrix L whose columns are the solutions $\mathbf{x}$ of $Ax = b$ for the the n cartesian basis vectors $(b_i)_j = \delta_{ij}$. $L = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n,)$. Similarly, the right inverse R has rows given by the solutions $\mathbf{y}$ of $yA = c$ for each of the cartesian basis vectors $(c_i)_j = \delta_{ij}$. We have already seen that the left and right inverses are the same, so $A^{-1} = L = R$ and A is invertible.

• But in practice inverting a matrix is not an efficient way of solving a particular system of equations (i.e. for a specific b or c). Elimination is the way to go.

• An $n \times n$ matrix A is invertible iff row elimination produces n (non-zero) pivots.

• A is not invertible iff elimination produces a row of zeros.

• Square A is invertible iff the columns of A are linearly independent. (could also use rows).

• Example: Inverse of a diagonal matrix $A = \text{diag}(\lambda_1, \cdots, \lambda_n)$ is again diagonal with entries given by the reciprocals, $A^{-1} = \text{diag}(\lambda_1^{-1}, \cdots, \lambda_n^{-1})$.

• Example of a singular matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$

• This matrix annihilates the vector $x = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. It has a row of zeros. It has only one pivot. Its determinant vanishes. And finally, we can't solve $Ax = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ for instance. What are the only $b's$ for which we can solve $Ax = b$?

• $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ is also singular. The second row is twice the first. Check the other equivalent properties.

• Elementary matrices are invertible because their columns are linearly independent.

• The inverse of an elimination matrix is easy to find. Suppose A is the matrix that subtracts twice the first row from the second row of a 2×2 matrix. Then its inverse must add twice the first row to the second.

$$A = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \quad \Rightarrow \quad A^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \tag{35}$$

• Example: $A = \begin{pmatrix} -1 & 2 & 0 \\ 3 & -4 & 2 \\ 6 & -3 & 9 \end{pmatrix}$. This matrix has a non-trivial kernel. Notice that the third column is twice the first added to the second. So any vector of the form $c \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$ is annihilated by A. So it is not invertible.

• The inverse of a product is the product of inverses in the reversed order, when they exist. To see why, draw a picture of the maps.

$$(AB)^{-1} = B^{-1}A^{-1} \quad \text{since} \quad B^{-1}A^{-1}AB = I \tag{36}$$

• The sum of invertible matrices may not be invertible, e.g. $I - I = 0$ is not invertible.

• Gauss-Jordan elimination is a systematic procedure to find the inverse of an $n \times n$ matrix. It is described in a later section.

• We will get a formula for the inverse after studying determinants.

• Remark: If A is an $n \times n$ matrix, we can express its inverse (when it exists) using its minimal polynomial, which is a polynomial of minimal degree $p(x) = p_0 + p_1 x + \cdots p_k x^k$ such that $p(A) = 0$. A matrix is invertible iff $p_0 \neq 0$. In that case, $A^{-1} = -p_0^{-1}(p_1 + p_2 A + \cdots p_n A^{n-1})$. $p(x)$ may have degree less than n and need not be the same as the characteristic polynomial, though it is always a factor of the characteristic polynomial $\det(A - xI) = 0$.

6.5 Transpose of a Matrix

• Transpose of an $m \times n$ matrix is the $n \times m$ matrix whose rows are the columns of A (in the same order).

• In components, $(A^T)_{ij} = A_{ji}$

• Transpose of a column vector is a row vector.

• $(A^T)^T = A$

• $(AB)^T = B^T A^T$

• Fundamental transposition formula $(x^T A y)^T = y^T A^T x$

• A square matrix which is its own transpose $A^T = A$ is called symmetric.

• Real symmetric matrices are a particularly nice class of matrices and appear in many physics and geometric problems. They appear in quadratic forms defining the kinetic energy of a free particle or a system of free particles.

• Real symmetric matrices behave a lot like real numbers.

• The operations of transposition and inversion commute $(A^{-1})^T = (A^T)^{-1}$.

• Suppose A is an invertible square matrix (i.e., has two-sided inverse $AA^{-1} = A^{-1}A = I$). Then A^T is also invertible and $(A^T)^{-1} = (A^{-1})^T$. To see this just take the transpose of $A^{-1}A = AA^{-1} = I$ to get $A^T(A^{-1})^T = (A^{-1})^T A^T = I$. But this is saying that $(A^{-1})^T$ is the inverse of A^T. In other words $(A^{-1})^T = (A^T)^{-1}$.

• The inverse of a symmetric invertible matrix $A^T = A$ is again symmetric. Suppose $B = A^{-1}$

$$AB = BA = I \Rightarrow B^T A^T = A^T B^T = I \Rightarrow B^T A = AB^T = I \tag{37}$$

So B^T is also the inverse of A and by uniqueness of the inverse, $B^T = B$.

• The inverse of a permutation matrix is its transpose. Show this first for row exchanges and then for products of row exchanges. Work with the 3×3 case first.

7 Vector space, span, subspace, basis, dimension

1. The basic operation defining a vector space is that of taking linear combinations of vectors $av + bw$. a, b are called scalars and v, w vectors.

2. A vector space is a space of vectors that is closed under linear combinations with scalar coefficients.

3. The multiplication by scalars distributes over addition of vectors $a(v + w) = av + aw$.

4. The scalars a, b that we can multiply a vector by are either real or complex numbers and give rise to a real or complex vector space. More generally, they can come from a field.

5. Examples of vector spaces: $R^2, R^3, R, C^2, R^n, C^n$

6. Non-examples: the following are not closed under linear combinations

   - A line not passing through the origin.
   - A half plane or quadrant or the punctured plane.
   - Unit vectors in R^2

7. So a vector space is also called a linear space, it is in a sense flat rather than curved.

Span

• Given vectors v, w, say in R^3, we can form all possible linear combinations with real or complex coefficients, $\{av + bw | a, b \in \mathbf{R} \text{ or } \mathbf{C}\}$. This is their (real or complex linear) span. For example, $3v - w$ is a linear combination. Unless otherwise specified, we will use real coefficients.

• *span*(v, w) is a two dimensional plane provided v and w were linearly independent. It is a vector space by itself.

• Eg $a\hat{x} + b\hat{y}$ is the span of the unit vector in the x and y directions. Geometrically, we go a units in the horizontal direction and b units in the vertical direction.

• For example, the span of the unit vector $\hat{x}$ is the x-axis while the span of $(1, 0, 0)$ and $(0, 1, 0)$ is the whole $x - y$ plane R^2 contained inside R^3

Subspace

• A subspace W of a vector space V is a subset $W \subseteq V$ that forms a vector space by itself under the same operations that make V a vector space.

• The span of any set of vectors from a vector space forms a vector space. It is called the subspace spanned by them.

• e.g., Any line or plane through the origin is a subspace of R^3. So is the point $(0, 0, 0)$.

• On the other hand, notice that $u = (1, 0, 0), v = (0, 1, 0), w = (1, 2, 0)$ span the same $x - y$ plane. There is a redundancy here, we don't need three vectors to span the plane, two will do.

• In other words, $w = (1, 2, 0)$ already lies in the span of $u = (1, 0, 0)$ and $v = (0, 1, 0)$, since $w - u - 2v = 0$.

• We say that u, v, w are linearly dependent if there is a non-trivial linear combination that vanishes.

• On the other hand, u and v are linearly independent and they span the plane.

• We say u, v are a basis for the plane.

Basis

• A basis for a vector space is a linearly independent collection of vectors $\{v_1, v_2, \cdots, v_n\}$ which span the space.

• $\hat{x}, \hat{y}$ is the standard basis for R^2, but $3\hat{x} + 2\hat{y}, \hat{y}$ is also a basis. Notice that bases have the same number of vectors (cardinality).

• The standard basis for R^n is the Cartesian one $(e_i)_j = \delta_{ij}$

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \; e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \; \cdots ; e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{38}$$

• Every vector can be *uniquely* written as a linear combination of basis vectors $x = x_i v_i$. We say that we have decomposed x into its components x_i in the basis. Proof: Suppose x has two *different* decompositions $x = x_i v_i$ and $x = x_i' v_i$, then $0 = x - x = (x_i - x_i') v_i$. But then we have a linear combination of basis vectors that vanish, which is not possible since v_i were linearly independent. So $x_i = x_i'$.

Dimension

• The dimension of a vector space is the cardinality of any basis. Equivalently, it is the maximal number of linearly independent vectors in the space.

• The dimension d of a subspace of an n-dimensional space must satisfy $0 \le d \le n$. The difference $n - d$ is called the co-dimension of the subspace.

• The dimension of C^n as a complex vector space is n. But it is also a real vector space of dimension $2n$

• Note that the dimension of a vector space should not be confused with the number of vectors in the space. The number of vectors is 1 for the trivial vector space and infinite otherwise.

• $\{(0)\}$ is not a basis for the 'trivial' vector space consisting of the zero vector alone. This is because the zero vector does not form a linearly independent set, it satisfies the equation $5(0) = 0$ for instance. The dimension of the trivial vector space is zero.

7.1 Intersection, Sum, Direct sum and Quotient

• The *intersection* of a pair of sub-spaces is again a sub-space $U \cap V$. It is the largest subspace contained in both.

• The set theoretic *union* $U \cup V$ of two sub spaces is not necessarily a vector space. When is it?

• On the other hand we can ask for the smallest sub-space that contains both U and V. This is the sum $U+V = \{u+v : u \in U, v \in V\}$. It is the smallest subspace containing both.

• *Dimension formula* for sum and intersection:

$$\dim(U+V) = \dim(U) + \dim(V) - \dim(U \cap V) \tag{39}$$

• This suggests a particularly interesting construction based on the case where $U \cap V$ is trivial.

• The *direct sum* of vector spaces $U \oplus V$ is the vector space of pairs $(u,v), u \in U, v \in V$ with sum defined componentwise $(u,v)+(u',v') = (u+u', v+v')$ and $c(u,v) = (cu, cv)$. Every vector in the direct sum is uniquely expressible as such a pair. Often we write $u+v$ for (u,v).

• R^3 is the direct sum of the $x-y$ plane and the z axis.

• The *direct sum* of the x-y plane and the x axis is a 3-dimensional vector space.

• The *sum* of the x-y plane and the x-axis is again the x-y plane.

• If U and V are sub-spaces of the same space, then the direct sum is isomorphic to the sum provided the intersection is trivial.

• We cannot in general subtract vector spaces, but the quotient and orthogonal complement come close to this notion.

• *Quotient* of a vector space V by a subspace W is a new vector space V/W. Its vectors are equivalence classes of vectors in V under the relation $v \sim v'$ if $v - v' \in W$. Notice that vectors in W are all in the equivalence class of the zero vector. V/W is neither in V nor W.

• The dimension of the quotient is the difference of dimensions.

• We will see later that with a dot or inner-product, we can take the *orthogonal complement* of W in V, i.e. $W^\perp \subseteq V$. The orthogonal complement is isomorphic to the quotient V/W.

• Eg. $R^2/span\{(0,1)\}$ is isomorphic to R^1 under the equivalence relation $(x,y) \sim (0,y)$ for all $x \in R$.

7.2 More examples of vector spaces

• The space consisting of just the zero vector is a zero dimensional space. It is called the trivial vector space.

• Consider the set of 2×2 *real matrices.* We can add matrices and multiply them by real numbers and the results are again 2×2 real matrices. So this is a real vector space $M_2(R)$. More generally we have the real vector space $M_n(R)$

• Note that if we consider the same set of 2×2 real matrices, it fails to be a complex vector space. Multiplication by an imaginary number takes us out of the set.

• The dimension of $M_2(R)$ is 4. What is a basis?

• The dimension of $M_n(R)$ is n^2

• On the other hand, the set of 2×2 complex matrices forms a complex vector space of complex dimension 4

• Interestingly $M_2(C)$ can also be thought of as a real vector space of twice the dimension, i.e., 8. Give a basis.

• What is the dimension of the space of real symmetric $n \times n$ matrices, and what is its codimension as a subspace of $M_n(R)$?

• The vector space of solutions of a *homogeneous linear differential equation*: For example consider the differential equation for the motion of a free particle on a line $x(t) \in R$

$$m\frac{\partial^2 x}{\partial t^2} = 0 \tag{40}$$

If $x(t)$ and $y(t)$ are solutions, then so is any real linear combination of them. This is a two dimensional real vector space, spanned by 1 and t.

• We should think of the differential operator $m\frac{\partial^2}{\partial t^2}$ as a 2×2 matrix acting on this space. It is just the zero matrix.

• Vector space spanned by the words in an alphabet: Given the English alphabet of 26 letters, we can form all words (with or without meaning) by stringing letters together. Now consider all real linear combinations of these words, such as the vectors

$$\begin{aligned} v &= 10 \text{ a} + 23 \text{ cat} - \pi \text{ xyz} + \text{dog} \\ w &= \text{pig} - 7 \text{ xyz} + 4\text{dog} \end{aligned} \tag{41}$$

Then $v + w = 10 \text{ a} + 23 \text{ cat} - (7+\pi) \text{ xyz} + 5 \text{ dog} + \text{pig} - 7 \text{ xyz}$ This is a real vector space. But it is infinite dimensional since there are an infinite number of (largely meaningless!) words. A basis consists of all possible words.

• We see that vector spaces are often specified either by giving a basis or as the solution space to a system of linear equations.

• A geometric example of a vector space is the space vectors tangent to a curve or surface at a point. For example, the tangent space to the sphere at the north pole is a two dimensional real vector space.

8 Linear transformations between Vector spaces, Isomorphism

• Given vector spaces D and T, a linear transformation from domain D to codomain or target T is a linear map

$$L : D \to T, \qquad L(au + bv) = aL(u) + bL(v) \tag{42}$$

• You can either form linear combinations before applying L or afterwards, the result is the same. So L morphs D into T. Sometimes we say that L is a linear morphism.

• Importantly, $L(0) = 0$.

• Both D and T must be real vector spaces or both must be complex vector spaces, no mixing up.

• A pair of vector spaces D and T are said to be isomorphic if there is a linear map L between them that is *invertible* (1-to-1 and onto). Then we say that $D \cong T$ are the same abstract vector space and that L is an isomorphism.

• A basic result is that any real vector space V of (finite) dimension n is isomorphic to R^n. The isomorphism maps a basis of V to the standard basis of R^n and extends to all of V by linearity. Write this out in symbols.

• All complex vector spaces of dimension n are isomorphic to C^n.

• An example of a linear transformation from R^2 to R^2 is $R_{\pi/2}$ a clockwise rotation by a right angle. Check this.

• $R_{\pi/2}$ is an isomorphism, its inverse is a counter-clockwise rotation by a right angle.

• But a translation of every vector to the right by two units is not a linear transformation, since $L(0) \neq 0$

• A linear map from $V \to V$ is called an endomorphism. If it is invertible, then it is called an automorphism.

• A reflection about any line through the origin is an automorphism of R^2

• A projection that projects every vector in the plane to its horizontal component is an endomorphism of R^2. It is not 1-1 or invertible, since all vertical vectors are annihilated.

• Composition: Given $L_1 : U \to V$ and $L_2 : V \to W$ we can compose these to get a linear transformation $(L_2 \circ L_1) : U \to W$. It is important that the target of L_1 is the same as the domain of L_2. L_2 follows (acts after) L_1.

• Note that this composition is not in general a commutative operation, and indeed we cannot even define $L_1 \circ L_2$ if the target of L_2 is not the same as the domain of L_1.

8.1 Matrix of a Linear map

• An $m \times n$ matrix A defines a linear transformation from $R^n \to R^m$, since we know it acts linearly on n-vectors to produce m-vectors $A(bv + cw) = bAv + cAw$

• The converse is also true.

• Every linear transformation $L : D \to T$ between finite dimensional (real or complex) vector spaces can be represented by a (real or complex) matrix. But there may be more than one such matrix representation.

• For this we need to pick a basis for the domain and target spaces. Different bases may give different matrix representations.

• But notice that we specified all the above linear transformations without any matrix or basis. So a linear transformation exists as a geometric entity, and a matrix is just a very useful algebraic representation of it.

• So let $e_1 \cdots e_n$ be a basis for D and $f_1 \cdots f_m$ be a basis for T.

• To specify a linear map, it suffices to say how it acts on each of the basis vectors for the domain space, $L(e_1), L(e_2) \cdots, L(e_n)$. We can extend it to the rest of the vectors in D by taking linear combinations of the e_i

• Now $L(e_1)$ is a vector in the target T, so it must be a linear combination of the $f_1, \cdots f_m$.

• So suppose

$$\begin{aligned} L(e_1) &= a_{11} f_1 + a_{21} f_2 + \cdots a_{m1} f_m \\ L(e_2) &= a_{12} f_1 + a_{22} f_2 + \cdots a_{m2} f_m \\ &\vdots \end{aligned}$$

$$L(e_n) \quad = \quad a_{1n}f_1 + a_{2n}f_2 + \cdots a_{mn}f_m. \tag{43}$$

• We can already see a matrix emerging.

• The matrix of L in this pair of bases is then an $m \times n$ matrix

$$L = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{44}$$

• The a_{ij} are called the matrix elements of L in this basis.

• $\vec{e}_i$ are vectors. They are geometric entities in their own right. Now e_1 can be written in any convenient basis. But since e_i themselves form a basis, it is most convenient to write e_1 in the e_i basis, where $e_1 = 1e_1 + 0e_2 + \cdots 0e_n$. Therefore, in the e_i basis, e_1 can be regarded as a vector with components $(1, 0, \cdots 0)$. Doing the same for the others, we may write the basis vectors e_i in *their own* basis as the following column vectors of length n,

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^e_{n\times 1}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^e_{n\times 1}, \cdots e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}^e_{n\times 1}. \tag{45}$$

Similarly the $\vec{f}_j$ in the the f-basis are columns of length m

$$f_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^f_{m\times 1}, f_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}^f_{m\times 1}, \cdots f_m = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}^f_{m\times 1}. \tag{46}$$

Note that there is no special choice being made here, these are forced upon us if we want to express a basis vector in the self-same basis.

• We see that Le_1 picks out the first column of the matrix as desired

$$Le_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{pmatrix} = a_{11}f_1 + a_{21}f_2 + \cdots a_{m1}f_m \tag{47}$$

• In short, the i^{th} column of L is the image of the i^{th} basis vector e_i.

• The matrix of a linear transformation will generally be different in different bases. Only very special linear maps have the same matrix in all bases, these are the identity map $L(e_i) = e_i$ and the zero map $L(e_i) = 0$.

• The transformation of matrix elements under changes of bases is treated later on.

• The composition of linear maps corresponds to multiplication of their matrix representatives.

• In the special case where the domain and target are the same vector space, then L is a square $n \times n$ matrix, and it is convenient to use the same basis for both domain and target $e_i = f_i$.

$$
\begin{aligned}
L(e_1) &= a_{11}e_1 + a_{21}e_2 + \cdots a_{n1}e_n \\
L(e_2) &= a_{12}e_1 + a_{22}e_2 + \cdots a_{n2}e_n \\
&\vdots \\
L(e_n) &= a_{1n}e_1 + a_{2n}e_2 + \cdots a_{nn}e_n.
\end{aligned}
\tag{48}
$$

• In this case, if it is invertible, L can itself be regarded as a change of basis: from the e_i basis to the $L(e_i)$ basis whose basis vectors are the columns of L.

• The inverse of a linear map is represented by the inverse matrix.

• We see that L basically scrambles up the basis vectors into linear combinations. A major goal of linear algebra is to find bases adapted to the linear map that make the rhs look simpler, especially those in which as many as possible matrix elements vanish.

• A particularly convenient basis (when it exists) is the eigenvector basis. This is a basis that belongs to an endomorphism, and is usually kept well-hidden. A major goal of linear algebra is to determine it when it exists. In the eigenvector basis,

$$
L(e_1) = \lambda_1 e_1, \quad L(e_2) = \lambda_2 e_2, \quad \cdots, \quad L(e_n) = \lambda_n e_n \tag{49}
$$

so that L is the diagonal matrix

$$
L = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix} \tag{50}
$$

• The diagonal elements are called the *eigenvalues* or characteristic values.

• Example: Find the matrix elements of the counter clockwise rotation by a right angle in the standard basis for R^2

8.2 Matrix of a linear transformation in different bases

• Example: The projection $P : R^2 \to R^2$ that projects every geometric vector to its horizontal component. Check that this is a linear transformation. Here the domain and target are the same vector space, so we can use a single basis. If f_1 and f_2 are the standard cartesian basis vectors in the horizontal and vertical directions, then $Pf_1 = f_1$ and $Pf_2 = 0$. In the f-basis, the columns of the matrix representation of P are the images of f_1 and f_2, so

$$
f_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}_f, \quad f_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}_f, \quad P_f = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{51}
$$

Since P is diagonal in the f-basis, we say that the f-basis is an eigenbasis for P. f_1, f_2 are eigenvectors of P with eigenvalues 1 and 0.

• Notice that $P_f^2 = P_f$, this is common to all projection matrices: projecting a vector for a second time does not produce anything new.

• But we are not obliged to work in the standard cartesian basis. So let us pick another basis consisting of $e_1 = f_1$ and $e_2 = f_1 + f_2$. So geometrically, e_1 is the standard cartesian horizontal basis vector, but e_2 is a vector that points north-east. In the f-basis we have

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}_f, \quad e_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}_f \tag{52}$$

But $\{e_1, e_2\}$ are also a basis in their own right. So we can also write e_1, e_2 in the e-basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}_e, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}_e \tag{53}$$

So we see that the same geometric vector may have different representations in different bases! Now the matrix of the projection P in the e-basis is the matrix whose columns are the images of e_1 and e_2 in the e-basis. Since $Pe_1 = e_1$ and $Pe_2 = e_1$, we have

$$P_e = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \tag{54}$$

P is not diagonal in the e-basis, so the e_i are not an eigenbasis for P. Nevertheless, the e-basis is a legitimate basis to use.

• Moreover, even in the e-basis, we see that $P_e^2 = P_e$

• We see that the same linear transformation P can have different matrix representations in different bases. However, P_e and P_f are related by a *change of basis*. First observe that the two bases are related by $e_1 = f_1, \quad e_2 = f_1 + f_2$ which may be written in matrix form as

$$e \equiv \begin{pmatrix} \vec{e}_1 \\ \vec{e}_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \vec{f}_1 \\ \vec{f}_2 \end{pmatrix} \equiv S^T \begin{pmatrix} \vec{f}_1 \\ \vec{f}_2 \end{pmatrix} \quad \text{where } S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{55}$$

In short $e = S^T f$. Calling it S^T is a matter of convenience so that the columns (rather than rows) of S are the components of e_i in the f-basis. S is called a change of basis. Notice that S is invertible, which is guaranteed since its columns form a basis and so are linearly independent.

• Now we can state the change of basis formula for a matrix: $P_e = S^{-1}\, P_f\, S$, which can be checked in our case

$$S^{-1}\, P_f\, S = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = P_e \tag{56}$$

• But you may ask, *Why does a matrix transform in this way?* To understand this, begin with a geometric vector x. It may be represented in the e-basis as well as in the f-basis, and in general has the components $x(e)_i$ and $x(f)_j$ (the notation is not optimal, but we are continuing with the notation used in the example above, for clarity)

$$x = x(e)_i\, \vec{e}_i \;=\; x(e)^T e, \qquad x \;=\; x(f)_j\, \vec{f}_j \;=\; x(f)^T f \tag{57}$$

But since $e = S^T f$, we have

$$x = x(f)^T f = x(e)^T e = x(e)^T S^T f = (Sx(e))^T f \Rightarrow x(f) = Sx(e) \tag{58}$$

So we have derived the change of basis formula for the components of a vector. If basis vectors transform according to $e = S^T f$ then the components of any vector x transform according to $x(e) = S^{-1}\, x(f)$.

• Now let us see how a matrix transforms. Suppose in the f-basis a linear map A acts on vectors according to $A_f x(f) = b(f)$, where $x(f)$ and $b(f)$ are components of $\vec{x}$ and $\vec{b}$ in the f-basis. Then,

$$A_f \; S \; x(e) = S \; b(e) \quad \Rightarrow \quad (S^{-1} A_f S) x(e) = b(e) \tag{59}$$

We define A_e by $A_e \; x(e) = b(e)$ so we conclude that

$$A_e = S^{-1} A_f S, \quad \text{as advertised!} \tag{60}$$

We say that the transformation from A_f to $A_e = S^{-1} A_f S$ is a 'similarity' or general linear transformation. We say that we conjugate A_f by S to get A_e

• In particular, $P_e = S^{-1} P_f S$ for the projection of the previous example.

9 Gauss-Jordan elimination to find A^{-1}

• Gauss-Jordan elimination is a systematic procedure to find the inverse of an $n \times n$ matrix.

• Recall that A^{-1} allows us to solve $Ax = b$ for any rhs. So to find A^{-1} we will basically solve $Ax = b$ for any target b. But in fact, it suffices to do it for a target space basis, such as

$$\{e_i\} = \{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \} \tag{61}$$

For, suppose we had solved $A\vec{x}_i = \vec{e}_i$. Then we can decompose/analyse the desired output b in the e_i-basis $\vec{b} = b_i \vec{e}_i$. Finally, the solution to $Ax = b$ is given by $\vec{x} = b_i \vec{x}_i$. Here, we sum over the repeated index i.

• These $e_i's$ are the columns of I. We will augment A by these columns and form the augmented matrix $(A|I)$. The word augment means to add on.

• Then we perform elementary row operations and row exchanges to the augmented matrix till A is reduced to echelon form and then eliminate upwards to bring it to reduced row echelon form (diagonal form) and finally divide by the pivots to reach the identity I. Automatically, I will be turned into A^{-1}.

• In effect we are multiplying $(A \mid I)$ by A^{-1} from the left to get $(A^{-1}A \mid A^{-1}I) = (I \mid A^{-1})$.

• Note that we must use elimination to clear the entries both above and below the diagonal to reach the reduced row echelon form and finally divide by the pivots to reduce A to I

• If there aren't n pivots, then elimination will produce a row of zeros and A is not invertible.

• Example, perform Gauss-Jordan elimination to find the inverse of the tridiagonal (discretized second derivative) matrix (Notice that the inverse is also symmetric.)

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix} \tag{62}$$

• Example find the inverse of the matrix Q. How is Q^{-1} related to Q^T?

$$Q = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{63}$$

• Find the inverse of the cyclic permutation matrix and comment on its relation to the transpose

$$\sigma = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{64}$$

• Find the inverse of the elementary matrix

$$A = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \quad \Rightarrow \quad A^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \tag{65}$$

10 Vector spaces associated to a matrix $A_{m\times n}$

10.1 Column space

• An $m \times n$ real matrix is a linear transformation from $R^n \to R^m$, from the domain space to the target space.

• Consider the 4×3 matrix A as a linear transformation from $R^3 \to R^4$.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 5 & 1 & 6 \\ 3 & 2 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{66}$$

• The column space $C(A)$ is the space of linear combinations of the columns of A.

• Recall that Ax is a linear combination of the columns of A. So $C(A)$ is the space of all possible outputs Ax.

• $C(A)$ is a subspace of the target space. It is also called the range or image of the linear map.

• The significance of the column space: If $b \in C(A)$, then we can solve $Ax = b$.

• So if $b \in R^m$ does not lie in the Column space, we cannot solve the equation $Ax = b$

• The dimension of $C(A)$ is the number of linearly independent columns, which is also the number of pivots or the rank of A, $\dim C(A) = \text{ rank } = r$. In this example, it is 2. Since $C(A)$ is a subspace of the target, we must have $\dim C(A) \leq m$. And since there are only n columns, we also have $\dim C(A) \leq n$.

• If the rank is equal to the dimension of the target, $r = m$ then we can solve $Ax = b$ for any $b \in R^m$.

• Elementary row operations change the column space in general, but not its dimension, since elementary row transformations are invertible. For example, A and A' below have different 1-dimensional column spaces.

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \xrightarrow{r_2 \mapsto r_2 - r_1} A' = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{67}$$

10.2 Row space

• The row space of a matrix $A_{m\times n}$ is the space spanned by the rows.

• $R(A)$ is a subspace of the domain $R(A) \subseteq R^n$

• The dimension of $R(A)$ is the number of independent rows. It is the same as the number of pivots in the row echelon form of A

• Thus $\dim R(A) = r =$ rank of the matrix $= \dim C(A)$ and $0 \leq r \leq n$ and $0 \leq r \leq m$.

• Elimination (elementary row operations) does not change the row space of a matrix.

• In echelon form, the row space is spanned by the pivot rows. Eg, the first two rows of U in this example:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -6 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{68}$$

• Here the row space is a 2 dimensional subspace of R^4

• Observe that $R(A)$ is the same as the column space of A^T. So $R(A)$ is the space of possible outputs of the transposed matrix, $A^T y$ as y ranges over R^m.

• The dimension of the row space is the number of truly independent equations in $Ax = 0$.

• Later, we will see that the row space and the kernel have only the zero vector in common. Moreover, every vector in the domain is either in the row space or in the kernel.

10.3 Null space or kernel or space of zero modes of $A_{m\times n}$

• An important vector space is the space of solutions of the system of homogeneous linear equations $Ax = 0$. If x, y are solutions, then so is any linear combination.

• This space is called the null space $N(A)$ or the kernel $ker(A)$. It is the space of vectors that are annihilated by A.

• $N(A) \subseteq R^n$ is a subspace of the domain, i.e., the inputs.

• The kernel contains vectors that are annihilated by A.

• The dimension of the kernel is called the number of zero modes in physics.

• If the kernel is zero dimensional we say it is trivial.

• For an $n \times n$ matrix, recall that A is invertible iff the kernel is trivial. This is one reason it is very important.

• Though the kernel was defined using the homogeneous equation $Ax = 0$, it is very useful to solve inhomogeneous equations $Ax = b$. More precisely, any two solutions of $Ax = b$ differ by a vector in the null space.

$$Ax = b, \; Ay = b \quad \Rightarrow \quad A(x-y) = 0 \quad \Rightarrow \; x - y \in ker(A) \tag{69}$$

• If we had one solution to $Ax = b$ we can produce a new one by adding any vector in the kernel. So suppose x_{par} is any particular solution of $Ax = b$. Then the 'general' solutions are given by $x = x_{par} + x_{null}$ where x_{null} is any vector in the null space.

• Unlike the solutions to a homogeneous equation, the solutions to the inhomogeneous equation $Ax = b \neq 0$ do not form a vector space. To start with, the zero vector is not a solution. And we can't add solutions x, y to get a new solution $A(x+y) = 2b \neq b$.

• Nevertheless, the solutions to $Ax = b$ form what is called an affine space, eg. a line or a plane that does not pass through the origin.

• We will show later that $\dim N(A) = n - r$ and that $N(A) \cap R(A)$ is the trivial vector space.

10.4 Left null space or $N(A^T)$ and cokernel

• The left null space is the space of solutions to $yA = 0$.

• Transposing we see that it is essentially the same as the kernel of the transpose $A^T x = 0$ where we denote $y^T = x$. Here we are identifying the row vectors in the left null space with their transposes which are the columns vectors annihilated by A^T

• $N(A^T)$ is a subspace of the Target space, R^m.

• We will show later that $\dim N(A^T) = m - r$. We will also see that the column space and the left null space intersect at the zero vector and moreover, every vector in the target is either in the column space or in $N(A^T)$.

• A space closely related (isomorphic) to $N(A^T)$ is the co-kernel of A. It is the quotient of codomain by image. i.e. Target$/C(A)$ i.e. $R^m/C(A)$.

10.5 Dimension of the kernel and rank-nullity theorem

• To actually find $N(A)$ we must find all solutions to $Ax = 0$. We do this by elimination.

• Let us illustrate this by an example

$$A_{3\times 4} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 4 & 6 & 8 & 10 \end{pmatrix} \tag{70}$$

• Here $m = 3, n = 4$

• Elimination does not change the null space or the solutions of a linear equation.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 4 & 6 & 8 & 10 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -6 \\ 0 & -2 & -4 & -6 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -6 \\ 0 & 0 & 0 & 0 \end{pmatrix} = U \tag{71}$$

• The equation in echelon form is $Ux = 0$

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -6 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{72}$$

• We have only two pivots, 1 and -2, so the rank $r = 2$. We have only two rows with pivots, so we have only two independent equations.

• The row of zeros means that the third equation was a linear combination of the earlier ones.

• The first two ($= r$) columns are pivot columns and the last two ($=4 - r$) are called free columns (they are free of pivots). The free columns are linear combinations of earlier columns. The equations in echelon form are

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 &= 0 \\ -2x_2 - 4x_3 - 6x_4 &= 0 \end{aligned} \tag{73}$$

• The $r = 2$ variables in the pivot columns are called the pivot variables (x_1 and x_2).

• The $n - r = 4 - r = 2$ variables in the pivot-free columns are the free variables x_3 and x_4.

• Notice that no matter what value we assign to the free variables, we can always solve for the pivot variables uniquely.

• Pivot variables are constrained variables and we can solve for them uniquely.

• Free variables are unconstrained. So we can assign any value to them. Here we have two free variables which can be assigned any real values. So let us pick the standard basis for them $\begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Any other values for the free variables are linear combinations of these.

• Solving for the pivot variables for each of these choices of the free variables gives the two linearly independent 'special' solutions, which lie in the kernel

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix} \tag{74}$$

• The null space is the space spanned by these special solutions.

$$N(A) = \ker(A) = \{a \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 2 \\ -3 \\ 0 \\ 1 \end{pmatrix} | a, b \in \mathbf{R}\} \tag{75}$$

Check that A annihilates the vectors in $N(A)$.

• The key point is that every variable is either a free variable or a pivot variable. The number of pivot variables is the rank. The number of free variables is the dimension of the null space ($n - r$): it is the difference between the dimension of the domain and the rank. This is the rank-nullity theorem $rank(A) + \dim N(A) = n$

• Replacing $A_{m\times n}$ with $(A^T)_{n\times m}$ we find that $\dim N(A^T) = m - r$.

• Let us summarize the list of dimensions of the 4 subspaces associated to a rank r matrix $A_{m\times n}$

$$\begin{aligned} \dim C(A) = r, &\qquad \dim N(A^T) = \dim \operatorname{coker}(A) = m - r \\ \dim R(A) = \dim C(A^T) = r, &\qquad \dim N(A) = \dim \ker(A) = n - r \end{aligned} \tag{76}$$

• The rank-nullity theorem and these results are sometimes included in the fundamental theorem of linear algebra.

11 Inner product, norm and orthogonality

• The standard inner or dot product on R^n is $x \cdot y = (x,y) = x^T y = \sum_i x_i y_i$. Here we think of x, y as a column vectors.

• The inner product is symmetric $(x,y) = (y,x)$.

• The norm or length of a vector $||x||$ is the square-root of its inner product with itself.

• $||x||^2 = x^T x = x_1^2 + x_2^2 + \cdots x_n^2$. Notice that this is automatically positive being the sum of squares.

• The **norm** is the positive square-root, $||x|| = (x^T x)^{1/2}$, which is seen to be the usual Euclidean length of the vector.

• Suppose x and y are a pair of vectors at right angles. The hypotenuse of the right triangle formed by them has length $||x+y||$, so $||x+y||^2 = ||x||^2 + ||y||^2$. The LHS-RHS must vanish,

$$||x+y||^2 - ||x||^2 - ||y||^2 = (x+y)^T(x+y) - x^T x - y^T y = x^T y + y^T x = 2(x,y) = 0 \qquad (77)$$

• So if a pair of vectors are orthogonal, their inner product vanishes $x \cdot y = (x,y) = x^T y = 0$. The converse is also true, $a^2 + b^2 = c^2$ implies that a, b, c are the lengths of the sides of a right triangle. This follows from the cosine formula in trigonometry: $a^2 + b^2 - 2\ a\ b\ cos\theta = c^2$, where a, b, c are the lengths of the sides of a triangle.

• So a pair of vectors are orthogonal iff their inner product vanishes.

• Cauchy-Schwarz Inequality: For a pair of n-dimensional vectors x, y, the Cauchy-Schwarz inequality is

$$|(x,y)|^2 \le (x,x)(y,y) \quad \text{or} \quad |(x,y)| \le ||x||\ ||y|| \qquad (78)$$

It merely says that the cosine of the angle between a pair of vectors is of magnitude ≤ 1. The angle between a pair of vectors is

$$\cos\theta = \frac{(x,y)}{||x||\ ||y||} \qquad (79)$$

• The triangle inequality states that $||x+y|| \le ||x|| + ||y||$. It says that the length of a side of a triangle is always $\le$ the sum of the lengths of the other two sides. Draw a picture of this. We have equality precisely if $x = \lambda y$ (i.e. they are collinear).

• The proof of the triangle inequality uses the Cauchy-Schwarz inequality. Begin by considering

$$\begin{aligned} ||x+y||^2 &= (x+y)^T(x+y) = x^T x + y^T y + x^T y + y^T x = ||x||^2 + ||y||^2 + 2x \cdot y \\ &\le ||x||^2 + ||y||^2 + 2||x||\ ||y|| = (||x|| + ||y||)^2 \end{aligned} \qquad (80)$$

Taking the square root, $||x+y|| \le ||x|| + ||y||$.

• For complex vectors in C^n, the standard (hermitian) inner product is $(z,w) = \bar{z}^T w = z^\dagger w$, where $\bar{z}$ denotes the complex conjugate vector.

• For a complex number $z = x + iy$ with real x, y, the complex conjugate $\bar{z} = z^* = x - iy$. The absolute value of a complex number is its length in the complex plane $|z| = \sqrt{|\bar{z}z|} = \sqrt{x^2 + y^2}$

• The notation $\bar{z}$ is more common in the mathematics literature while z^* is more common in physics to denote the complex conjugate.

• The complex conjugate transpose, $z^\dagger$ is called the (Hermitian) adjoint of the vector z. For complex vectors, the hermitian adjoint plays the same role as the transpose does for real vectors.

• This is the appropriate inner product since it ensures that $(z,z) = ||z||^2 = z^\dagger z = |z_1|^2 + \cdots |z_n|^2$ is real and non-negative and so its positive square-root $(z^\dagger z)^{1/2}$ is the length of the vector z.

• The only vector with zero norm is the zero vector.

• The hermitian inner product is not symmetric but satisfies $(z,w)^* = (w,z)$.

• $(z,w) = z^\dagger w$ is linear in the second entry and anti-linear in the first: $(\lambda z, \mu w) = \lambda^* \mu (z,w)$ and $(z+z', w) = (z,w) + (z',w)$, $(z, w+w') = (z,w) + (z,w')$.

• A pair of vectors are orthogonal if their inner product vanishes $z^\dagger w = 0$

• In the language of quantum mechanics, a vector is a possible state of a system and a (hermitian) matrix is an observable.

• Expectation value of a matrix observable A in the state x is defined as the complex number $x^\dagger A x$

• A vector space with an inner product is also called a Hilbert space.

11.1 Orthonormal bases

• A basis $\{q_i\}_{i=1}^n$ for a vector (sub)space is orthogonal if the basis vectors are mutually orthogonal, $q_i \perp q_j$ or $q_i^T q_j = 0$ for $i \neq j$.

• In addition it is convenient to normalize the basis vectors to have unit length, $||q_i|| = 1$. Then we say the basis q_i is orthonormal or o.n.

• A convenient way of packaging an orthonormal basis is to collect the basis vectors as the columns of a matrix Q. Then the columns of Q are orthonormal

$$q_i^T q_j = \delta_{ij} \quad \text{or} \quad Q^T Q = I \tag{81}$$

• Example, the standard cartesian $x-y$ basis is o.n. But so is any rotated version of it. The columns of Q and Q' below are both o.n. bases for R^2

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad Q' = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{82}$$

• Check the o.n. conditions.

• The basis $(1,0,0)$ and $(0,1,0)$ is an orthonormal basis for the $x-y$ plane contained in R^3. In this case Q is a rectangular 3×2 matrix,

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{83}$$

yet it satisfies $Q^T Q = I_{2\times 2}$. Note that $QQ^T \neq I$, in fact it is a projection matrix!

• But if $Q_{n\times n}$ is a square matrix, then $Q^TQ = I$ implies that Q has a left inverse. Does it have a right inverse? Being a basis, we know that the columns of Q are linearly independent. Being square, the rows must also be linearly independent as the rank is n. But if the rows are linearly independent, it means the rows span the domain or equivalently, $c = yQ$ has a unique solution for any c. This means Q has a right inverse. By the equality of left and right inverses, we conclude that $Q^{-1} = Q^T$ and that $QQ^T = Q^TQ = I$. Such a matrix is called an orthogonal matrix.

11.2 Orthogonality of subspaces:

• A pair of subspaces $V, W \subseteq U$ are orthogonal if every vector in V is perpendicular to every vector in W

• Orthogonal complement of $span(a)$ is the space of all vectors b that are orthogonal to it: $b^Ta = 0$. Check that it is a subspace.

• More generally the orthogonal complement $V^\perp$ of a subspace V is the space of all vectors that are orthogonal to every vector in V

• For example, the orthogonal complement of the x-axis in R^2 is the y axis. The orthogonal complement of the $x-y$ plane in R^3 is the z-axis.

• Note that the orthogonal complement of a subspace is quite different from its complement as a subset.

Orthogonal direct sum

• If U, V are orthogonal complements of each other as subspaces of W ($U^\perp = V, V^\perp = U$), then every vector in W can be written uniquely as a sum $w = u + v$ with $u \in U$ and $v \in V$. We say that W is the orthogonal direct sum of U and V. $W = U + V$ and $W \cong U \oplus V$.

11.3 Components of a vector in an orthonormal basis

• A vector $x \in R^n$ can be decomposed or analyzed or expanded in any basis

$$x = x_1q_1 + \cdots x_nq_n = \sum_j x_jq_j \tag{84}$$

• The components x_i are uniquely determined. Indeed suppose $x = x_iq_i = x_i'q_i$, then $\sum_i(x_i - x_i')q_i = 0$ which is possible iff $x_i = x_i'$ since q_i (being a basis) are linearly independent.

• But what are the components x_i? If q_i form an orthonormal basis ($(q_i, q_j) = \delta_{ij}$), we can find the x_i easily by taking inner products with the basis vectors. Indeed, $(x, q_i) = x_i$ so

$$x = \sum_j (x, q_j)q_j \tag{85}$$

11.4 General inner products on vector spaces

• More generally an inner product on a vector space is a way of speaking of lengths and angles between vectors, but having the same basic properties as the standard inner product.

• These basic properties are the axioms of an inner product

• Inner product on a real vector space (x,y) is a symmetric, non-degenerate bilinear form. Symmetry means $(x,y)=(y,x)$. Non-degenerate means $(x,y)=0$ for all y implies $x=0$. This is saying that the only vector that is perpendicular to all vectors is the zero vector. Bilinear means $(ax+by,u)=a(x,u)+b(y,u)$ and similarly for linear combinations in the second entry.

• In the complex case (z,w), bilinearity is replaced with linearity in the second entry w and anti-linearity in the first entry due to the complex conjugation $(\lambda z+\zeta, \mu w+\omega)=\bar{\lambda}\mu(z,w)+(\zeta,\omega)$.

• Also for complex vector spaces, we speak of a hermitian inner product in the sense that symmetry is replaced by $(z,w)^* = (w,z)$

• Example, given any symmetric strictly positive matrix A, we can define the inner product $(x,y)=x^T A y$. The standard case arises from the choice $A=I$.

• Given a hermitian strictly positive matrix H, we get a hermitian inner product $(z,w)=z^\dagger H w$. The standard choice is $H=I$

• A vector space with an inner product is called a Hilbert space. Hilbert spaces are the basic playground of quantum mechaincs.

11.5 Norm of a matrix

• Just as you can assign a length to a vector, one can also assign a length to a matrix.

• For $A: V\to W$, the operator norm measures how much A magnifies a unit vector: $||A|| = \sup_{||x||=1} ||Ax||$

• The Hilbert-Schmidt norm of a matrix is obtained from the sum of the squares of all its entries

$$||A||_{H-S} = (\sum_{i,j} |a_{ij}|^2)^{1/2} \tag{86}$$

11.6 Orthogonality of Row space and Null space and of $Col(A)$ and $N(A^T)$

• The row space $R(A)=C(A^T)$ is orthogonal to the null space $N(A)$

• Suppose $x\in N(A)$, then $Ax=0$, but this equation just says that every row of A has zero dot product with x

$$\begin{pmatrix} --r_1-- \\ --r_2-- \\ \vdots \\ --r_m-- \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} r_1\cdot x \\ r_2\cdot x \\ \vdots \\ r_m\cdot x \end{pmatrix} = 0 \tag{87}$$

• Thus the null space is orthogonal to the row space.

• The column space $C(A)$ is orthogonal to $N(A^T)$

• This follows by applying the previous argument to A^T, since the columns of A^T are the rows of A. $A^T y=0$ implies y is orthogonal to the rows of A^T and hence to the columns of A.

• Thus, the row space and the null space are orthogonal subspaces of the domain space R^n. But we have already seen that their dimensions $\dim R(A)=r$, $\dim N(A)=n-r$ add up to that of the the domain. So in fact $N(A)$ and $R(A)$ are orthogonal complements of each other in the the domain R^n

• $R(A) \cap N(A) = \{0\}, \qquad R(A) \perp N(A)$

• Similarly, $C(A)$ and $N(A^T)$ are orthogonal subspaces of the target space R^m. But we have already seen that their dimensions $\dim C(A) = r$, $\dim N(A^T) = m - r$ add up to that of the the domain. So in fact $N(A^T)$ and $C(A)$ are orthogonal complements of each other in the the target R^m

• $C(A) \cap N(A^T) = \{0\}, \qquad C(A) \perp N(A^T)$

• Draw the picture of the orthogonal decomposition of the domain R^n and range R^m.

• These facts are sometimes included in the fundamental theorem of linear algebra.

12 Compatibility and general solution of $Ax = b$

12.1 Compatibility of $Ax = b$ and the adjoint equations

• Consider a system of m equations in n unknowns $Ax = b$. A is an $m \times n$ matrix and x is an n component column vector and b is an m component column vector.

• The vector on the rhs b is often called the data or the inhomogeneity.

• A commonly used terminology is

1. Under-determined: less equations than unknowns $m < n$
2. Over-determined: more equations than unknowns $m > n$
3. Even determined or balanced: same number of equations and unknowns $m = n$.

• One should bear in mind that this terminology may bear little relation to the actual number of solutions. But it is still a reasonable terminology. Generically, an under-determined system has infinitely many solutions, an over-determined system no solutions and an even-determined system a unique solution. What does 'generically' mean?

• An under-determined system may have no solutions as below:

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ 2x_1 + 2x_2 + 2x_3 &= 0 \end{aligned} \tag{88}$$

But by making an arbitrarily small change to the matrix we get a new system

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ (2.0001)x_1 + 2x_2 + 2x_3 &= 0 \end{aligned} \tag{89}$$

whose matrix A now has rank 2 rather than 1 since the rows are independent. The new system now has a 1-parameter family of solutions, i.e. infinitely many solutions as is typical of under-determined systems.

• An over-determined system may have infinitely many solutions as these three equations in two unknowns shows. Find the solutions

$$x_1 + x_2 = 3$$

$$
\begin{aligned}
2x_1 + 2x_2 &= 6 \\
3x_1 + 3x_2 &= 9
\end{aligned} \tag{90}
$$

But by making a small change, say in the rhs $9 \to 9.01$, the system ceases to have any solution, which is typical of over determined systems.

• An even-determined system may have no solutions:

$$x_1 - x_2 = 1 \quad \text{and} \quad 2x_1 - 2x_2 = 3 \quad \text{has no solution} \tag{91}$$

But by making a small change $-1 \to -1.001$ the system has a unique solution, which is typical. An even-determined system may have infinitely many solutions

$$6x + 8y = 0 \quad \text{and} \quad 3x + 4y = 0 \;\; \text{has many solutions} \tag{92}$$

But by a small change $3 \to 3.1$ it becomes a system with just one solution, which is typical.

• The actual number of solutions is determined by the rank r of A in relation to m and n as well as the data on the rhs b

• $Ax = b$ is called incompatible or inconsistent if the system admits no solution for the data b.

• $Ax = b$ is called compatible if it admits 1 or more solutions. A necessary and sufficient condition for compatibility is for b to lie in the column space. But this is not a useful condition in practice since it is just a restatement of the definition.

• A more useful compatibility condition is formulated in terms of $N(A^T)$, in fact we already have it. We found that $C(A)$ and $N(A^T)$ are orthogonal and that their sum is the target space R^m. So a vector b in the target is in $C(A)$ iff it is orthogonal to every vector in $N(A^T)$. So $Ax = b$ is compatible iff

$$y^T b = 0 \;\; \text{for every solution to } A^T y = 0 \tag{93}$$

• For every linearly independent solution of $A^T y = 0$, i.e. for every independent vector in $N(A^T)$ we have one compatibility condition $y^T b = 0$. Since $\dim N(A^T) = m - r$, we have $m - r$ compatibility conditions. So there are $m - r$ compatibility conditions to be checked. If they are all satisfied, then and only then, $Ax = b$ is a consistent system and will have at least one solution.

• Remark: A homogeneous system $Ax = 0$ is always compatible since $b = 0$ has zero dot product with any y. Indeed, $x = 0$ is always a solution.

12.2 General solution to inhomogeneous system $A_{m\times n} x_{n\times 1} = b_{m\times 1}$

• The most general solution to $Ax = b$ is of the form $x = x_{particular} + x_{null}$, where x_{null} is an arbitrary vector in the null space $N(A)$ and $x_{particular}$ is any specific solution of $Ax = b$. This is because the difference between any two solutions is a solution of the homogeneous system.

• $N(A^T)$ provides the compatibility conditions $b \cdot N(A^T) = 0$. Once the compatibility conditions are satisfied, there will be at least one (particular) solution.

• A convenient choice for the particular solution is the one obtained by setting all the free variables to zero and solving for the pivot variables. To this we must add an arbitrary vector in the null space to get the most general solution.

• The kernel parameterizes the space of solutions. If the kernel $N(A)$ is trivial, there is a unique solution. In general there is an $n - r(= \dim N(A))$ parameter family of solutions.

• Some special cases:

• For an $n \times n$ system $Ax = b$ with $rank(A) = n$ maximal, there is a unique solution $x = A^{-1}b$.

• For an $n \times n$ system $Ax = b$ with $rank(A) = r < n$, there are $n - r$ compatibility conditions to be satisfied. If they are satisfied, there is then an $n-r$ parameter family of solutions otherwise there are no solutions.

• If $A_{m\times n}$ has full column rank, $r = n < m$, there are no free variables, $N(A)$ is trivial, so there is at most one solution. The solution exists provided the $m - r(= \dim N(A^T))$ compatibility conditions $b \cdot N(A^T) = 0$ are satisfied.

• If $A_{m\times n}$ has full row rank, $r = m < n$, $N(A^T)$ is trivial, so there are no compatibility conditions. There is an $n - r = \dim N(A)$ parameter of solutions (i.e. infinitely many).

• If $A_{m\times n}$ has less than maximal rank, $r < n, r < m$, then there are $m - r$ compatibility conditions. If they are satisfied, then there is an $n - r$ parameter family of solutions.

• In practice, to solve $Ax = B$ we use elimination to reduce the augmented matrix $(A|x)$ to reduced row echelon form $(R|d)$. In this form we have used elimination to eliminate entries both above and below the pivots and finally divided by the pivots. Let us suppose the first r columns are the pivot columns and the remaining $n - r$ columns are the free columns for simplicity (this is not always the case, though with the help of row exchanges and columns exchanges in conjunction with reordering of variables we can reach this form)

$$Rx = \begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{pivot} \\ x_{free} \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix} \tag{94}$$

We identify the null space matrix, whose columns are a basis for $N(A)$.

$$N = \begin{pmatrix} -F_{r\times n-r} \\ I_{n-r\times n-r} \end{pmatrix} \tag{95}$$

Check that $RN = -IF + FI = 0$. This means R annihilates each column of N. The columns of N are linearly independent since the lower block contains the identity matrix. So the columns of N form a basis for $N(A)$.

• Then we find the particular solution obtained by setting all the free variables to zero, this is very easy in reduced row echelon form, the pivot variables are just the components of d.

$$x_{particular} = \begin{pmatrix} d \\ 0 \end{pmatrix} \tag{96}$$

• Then the general solution is the particular solution plus any linear combination of the columns of N

$$x_{n\times 1} = \begin{pmatrix} d_{r\times 1} \\ 0_{n-r\times 1} \end{pmatrix} + N_{n\times n-r} \quad x_{free_{n-r\times 1}} \tag{97}$$

• The $n - r$ free variables are not determined and parameterize the space of solutions.

• Work out the example

$$\begin{pmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \\ 7 \end{pmatrix}. \tag{98}$$

13 Projection matrices

• Projections are an important class of matrices, not least because the density matrix of a pure state of a quantum system is a projection matrix.

Orthogonal projection onto a line through the origin

• Example of projection of a vector onto a rectangular coordinate axis

• A line through the origin is just a 1-d vector space spanned by a vector a.

• We seek to project a vector v onto the span of a. Let us call the projection $Pv = a\xi$, since it must be a multiple of a.

• Then the orthogonality of the projection means that the difference between v and its projection Pv, i.e. the error vector $e = v - Pv$ must be perpendicular to a

$$e \perp a \Rightarrow a^T e = 0 \Rightarrow a^T(v - Pv) = 0 \Rightarrow a^T v = \xi a^T a \Rightarrow \xi = \frac{a^T v}{a^T a} \tag{99}$$

• So $Pv = a\xi = \frac{aa^T}{a^T a} v$

• Another way to find the projection $P_a v$ is to observe that $Pv = \xi a$ is the vector along a that is closest to v. So ξ must be chosen so that the error vector $e = v - Pv$ has minimal length.

$$||e||^2 = (v - \xi a)^T (v - \xi a) = v^T v - 2\xi a^T v + \xi^2 a^T a \Rightarrow \frac{\partial ||e||^2}{\partial \xi} = -2a^T v + 2\xi a^T a = 0 \Rightarrow \xi = \frac{a^T v}{a^T a} \tag{100}$$

• Projection map $v \mapsto Pv$ is a linear transformation, since it is linear in v.

• The matrix of the projection onto the subspace spanned by a is

$$P_a = \frac{aa^T}{a^T a} \quad \text{or} \quad P_{ij} = \frac{a_i a_j}{\sum_k a_k a_k} \tag{101}$$

• The product of a column vector by a row vector with the same number n of components is called the outer product, it is an $n \times n$ matrix. So P_a is the outer product of a with itself divided by the inner product of a with itself.

• Notice that $P_a a = 1$, i.e., the projection of $\vec{a}$ onto itself is the identity matrix.

• Notice that if $v \perp a$, then $P_a v = 0$.

• It is easy to check that P_a satisfies the following two properties: it is symmetric and squares to itself $P^2 = P, \quad P^T = P$. We will see that more general projections also satisfy these properties and they can be taken as the defining properties of projections. (caution: $P^T = P$ is true only in orthonormal bases)

• Notice that $I - P_a$ also satisfies these conditions. It is the projection onto the orthogonal complement of a. Indeed, it is just the error vector $(I - P_a)v = v - P_a v$, which we know to be orthogonal to $\vec{a}$.

• For example, the projection matrix onto the line spanned by the unit column vector $a = (1, 0, 0)$ is

$$P_a = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{102}$$

• From this example, we see that P_a is a rank-1 matrix, since it has a single pivot, or equivalently, a single independent column or row.

• However, not all rank-1 matrices are projections.

• A rank-1 matrix can always be be written as an outer product $A = uv^T$. Multiplying by columns we see that uv^T is the matrix whose columns are $(v_1 u, v_2 u, \cdots, v_n u)$, so it has only one linearly independent column. Conversely, any matrix with only one linearly independent column is of this form.

• Among rank-1 matrices, uv^T, only if u, v point in the same direction and have reciprocal lengths is the result a rank-1 projection.

• Notice also that the trace $\text{tr } P_a = 1$ is its rank. This is generally true of projections onto 1-dimensional subspaces.

• Consider another example, projection onto $a = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

$$P_a = \frac{1}{5}\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \qquad \text{tr } P_a = 1 \tag{103}$$

• Since $P_a = P_{\lambda a}$, P_a is independent of the particular vector a. P_a only depends on the subspace spanned by a.

• If $a \perp b$, i.e. $b^T a = 0$, then $P_a P_b = P_b P_a = 0$ as can be seen from the formula. Projections to orthogonal directions commute.

• Projection to orthonormal basis vectors: A virtue of orthonormal bases is that it is very easy to find the projection onto a basis vector in an orthonormal basis,

• If $\vec{x} = \sum_i x_i \vec{b}_i$ where $\vec{b}_i$ are an o.n. basis, then $P_{b_i} x = x_i \vec{b}_i$ (no sum on i) where $x_i = (x, b_i)$ are the components.

• To see this use the above formula and orthonormality $\vec{b}_i^T \vec{b}_j = \delta_{ij}$

$$P_{b_i} x = \frac{b_i b_i^T}{b_i^T b_i} x = b_i b_i^T x = b_i x_i \quad \text{(no sum on } i) \tag{104}$$

• In particular, any vector can be expanded in an orthonormal basis $\vec{b}_i$ as $a = \sum_i P_{b_i} a$

13.1 Orthogonal projection to a subspace

• More generally we can consider orthogonal projection of v onto a subspace of a vector space.

• A subspace is often specified by a collection of basis vectors $a_1 \cdots a_d$. So it is convenient to think of these as the columns of a matrix $A_{n\times d}$. Then our problem is to find the orthogonal projection onto the column space of A.

• Pv is a linear combination of columns of A, $Pv = a_1\xi_1 + a_2\xi_2 + \cdots + a_d\xi_d$, so let $Pv = A\xi$ for some column vector ξ.

• The error vector $e = Pv - v = A\xi - v$ must be perpendicular to every vector in the subspace and so $A^T e = 0$, i.e., $e \in N(A^T)$ or equivalently, $e \perp C(A)$. Thus

$$A^T(A\xi - v) = 0 \;\Rightarrow\; A^T A\xi = A^T v \;\Rightarrow\; \xi = (A^T A)^{-1} A^T v \tag{105}$$

• So $Pv = A\xi = A(A^TA)^{-1}A^Tv$. Thus the projection matrix is $P = A(A^TA)^{-1}A^T$.

• We have used the fact that A^TA is invertible if the columns of A are linearly independent (columns of A are linearly independent as they are a basis for the subspace onto which we wish to project).

• Proof: Let us show that A^TA has trivial kernel. Suppose there was a vector $x \neq 0$ such that $A^TAx = 0$. Columns of A are linearly independent, so $Ax = 0$ iff $x = 0$. So let $Ax = y \neq 0$. We see that $y \in C(A)$. Now we know that $N(A^T)$ is orthogonal to $C(A)$. So y could not be in the kernel of A^T, so $A^Ty \neq 0$. In other words, $A^TAx \neq 0$, and this contradicts the assumption. So A^TA must have trivial kernel and therefore be invertible.

• On the other hand, if A was square, and invertible, then $P = I$. In this case, the columns of A span the whole space and $Pv = v$ for every vector.

• If v lies in $C(A)$, we expect $Pv = v$, Indeed, $v = A\eta$ for some column vector η. So $Pv = A(A^TA)^{-1}A^TA\eta = A\eta = v$.

• We check that $P^2 = P$ and $P^T = P$

• Invariance of P_A under change of basis for $C(A)$. The basis given by the columns of A was merely a convenient way to specify the subspace. The projection onto a subspace should depend only on the subspace and not the particular basis we choose for it.

• Suppose we choose a different basis whose basis vectors $b_1, \cdots, b_n$ are some (invertible) linear combinations of the columns of A. The new basis vectors can be assembled in the columns of a new matrix

$$\begin{aligned} b_1 &= c_{11}a_1 + c_{21}a_2 + \cdots + c_{d1}a_d \\ b_2 &= c_{12}a_1 + c_{22}a_2 + \cdots + c_{d2}a_d \\ &\vdots \end{aligned} \tag{106}$$

$$\begin{pmatrix} b_1 & b_2 & . & . & b_d \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & . & . & a_d \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} & . & . & c_{1d} \\ c_{21} & c_{22} & . & . & c_{2d} \\ & & \vdots & & \\ c_{d1} & c_{d2} & . & . & c_{dd} \end{pmatrix} \tag{107}$$

• So the transformation of bases is $B = AC$. Note that A and B are both $n \times d$ matrices and C is a $d \times d$ matrix.

• C is invertible. Why? Observe that the i-th column of C is just the vector b_i expressed in the a-basis. Since the b_i are a basis, they are linearly independent and so the columns of C are linearly independent. So C has trivial kernel and is invertible.

• Now we show that $P_B = P_A$ provided C is invertible. We use the fact that A^TA is invertible:

$$\begin{aligned} P_B &= B(B^TB)^{-1}B^T = AC((AC)^TAC)^{-1}(AC)^T = AC(C^TA^TAC)^{-1}C^TA^T \\ &= ACC^{-1}(A^TA)^{-1}(C^T)^{-1}C^TA^T = A(A^TA)^{-1}A^T = P_A \end{aligned} \tag{108}$$

• Example, projection onto the $x-y$ plane in R^3. In this case it is convenient to take the usual cartesian o.n. basis for the $x-y$ plane so that

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad A^TA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad P_A = AA^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{109}$$

• Notice that P_A projects to a 2-dimensional subspace and its rank and trace are also 2.

• In general $I - P_A$ is the projection to the orthogonal complement of the subspace spanned by the columns of A.

• The rank of a projection matrix d is the dimension of the subspace to which it projects and this equals its trace. (But the rank is not related to the trace in this way for an arbitrary matrix.)

• To see this, just pick an o.n. basis of d vectors for the subspace as in the above example and in this basis, P_A is a block matrix with d pivots all equal to unity. Its trace is also d

$$P_A = \begin{pmatrix} I_{d\times d} & 0 \\ 0 & 0 \end{pmatrix} \tag{110}$$

• Density matrix in quantum mechanics of fermions (e.g. electrons) is a projection matrix.

• The number of fermions is the rank of the projection.

13.2 Best possible solution of overdetermined systems

• When an $m \times n$ system $Ax = b$ has no solution (incompatible equations) we can still look for the vector $\hat{x}$ that comes closest to being a solution.

• The point about an incompatible system is that b does not lie in the column space of A.

• To make it compatible, we must replace b by a vector in $C(A)$, not *any old* vector but the one closest to b. But this is precisely the projection of b on to $C(A)$.

• So we replace our system with a new system $A\hat{x} = P_A b$, where $P_A = A(A^TA)^{-1}A^T$.

• Thus we must solve $A\hat{x} = A(A^TA)^{-1}A^Tb$ which is the same as

$$A^TA\hat{x} = A^Tb \tag{111}$$

• This is the equation obtained by multiplying the incompatible one by A^T from the left. We can forget how we derived it and consider the equation $A^TA\hat{x} = A^Tb$ in its own right. This equation is called the normal equation.

• The $n \times n$ matrix A^TA is obviously symmetric.

• The remarkable thing about the normal equation is that it always has at least one solution, no matter how over determined the original system $Ax = b$ was. Indeed, A^TA may not even be invertible, it may have a non-trivial kernel and the formula for P_A may not make sense, but the normal equations always have a solution!

• Proof: We want to show that $A^TA\hat{x} = A^Tb$ always has a solution, which is the same as showing that $A^Tb \in C(A^TA)$. But this is the same as showing that $A^Tb \perp N((A^TA)^T)$, which is the same as $A^Tb \perp N(A^TA)$. But $N(A^T)$ is orthogonal to $C(A)$. $SoN(A^TA) = N(A)$. So we need to show that $A^Tb \perp N(A)$ which is the same as $C(A^T) \perp N(A)$, i.e. $R(A) \perp N(A)$, which is true!

• Moreover, A^TA is a positive (semi-definite) matrix in the sense that all expectation values are ≥ 0

$$x^TA^TAx = (xA)^TAx = ||Ax||^2 \geq 0 \tag{112}$$

• $A^T A$ is invertible iff A has linearly independent columns.

• Proof: $A^T A$ invertible $\Rightarrow$ it has trivial kernel. But its kernel is the same as that of A since $N(A^T)$ is orthogonal to $C(A)$. In more detail, the only ways $x \neq 0$ can lie in $N(A^T A)$ is for x to be annihilated by A or for A^T to annihilate Ax. The second possibility cannot happen since $N(A^T) \perp C(A)$. So $N(A^T A) = N(A)$. So the rectangular matrix A must have trivial kernel, which means its columns are linearly independent.

• Conversely, if A has independent columns, then its kernel is trivial and since $N(A^T) \perp C(A)$, $A^T A$ also has trivial kernel and being square, is invertible.

• Alternate proof of converse: If A has linearly independent columns, A has trivial kernel. Now suppose $A^T A x = 0$ for some $x \neq 0$. For this x, $Ax \neq 0$ since $N(A) = \{0\}$ and so $||Ax|| > 0$, i.e., $x^T (A^T A x) > 0$. But this contradicts the assumption that $A^T A x = 0$. So $A^T A$ has trivial null space and is therefore invertible and also strictly positive.

• Also draw the picture with $\dim(C(A)) = r = n$ and $\dim(R(A)) = r = n$, $\dim(N(A)) = 0$, $\dim(N(A^T)) = m - r$.

13.3 Example of least-squares fitting

• Suppose we apply a potential difference V across a wire or other circuit element and measure the current I that flows through it to get the table

$$\begin{pmatrix} Voltage(V) & V_1 & V_2 & V_3 & \cdots & V_m \\ Current(Amps) & I_1 & I_2 & I_3 & \cdots & I_m \end{pmatrix} \tag{113}$$

Ohm's law says that for some wires, the current generated is proportional to the applied voltage $I = \frac{1}{R} V$, where the proportionality constant is called the conductance $G = \frac{1}{R}$ and R is called the resistance. Of course, the wire may not obey Ohm's law exactly and there could be deviations. We want to fit a curve to the data and to allow for some simple deviations from Ohm's law let us try to fit a straight line $I(V) = GV + C$ or a parabola $I = FV^2 + GV + C$ passing through the data points

$$\begin{aligned} I_1 &= FV_1^2 + GV_1 + C \\ I_2 &= FV_2^2 + GV_2 + C \\ &\vdots \\ I_m &= FV_m^2 + GV_m + C \end{aligned} \tag{114}$$

We want to find the values of F, G, C. Of course, if we find that F and C are very small, we would say that the wire obeys Ohm's law closely. The above can be written as a matrix equation $Ax = b$ for $x = (F, G, C)^T$

$$\begin{pmatrix} V_1^2 & V_1 & 1 \\ V_2^2 & V_2 & 1 \\ \vdots & & \\ V_m^2 & V_m & 1 \end{pmatrix} \begin{pmatrix} F \\ G \\ C \end{pmatrix} = \begin{pmatrix} I_1 \\ I_2 \\ \vdots \\ I_m \end{pmatrix} \tag{115}$$

For $m > 3$ this is an over-determined system and generically will not have any solution F, G, C. In other words, the column of currents b will not be in the column space of A.

The problem then is to find the values of F, G, C that best describe the data. By best fit we mean that we want to minimize the error, $E = \sum_{j=1}^{m} (I_j - I(V_j))^2$, which is the square of the difference between the measured current and that given by the curve. This is called least-squares fitting. But this error is just the norm of the error vector $e = b - Ax$ i.e., $E(x) = e^T e$.

So we will replace the system $Ax = b$ which has no solution with a new system where b is replaced by a vector in $C(A)$. Minimizing the norm of the error vector means we must replace b with its orthogonal projection to $C(A)$. This leads to the new system of 'normal equations'

$$A\hat{x} = P_A b, \quad \text{or} \quad A\hat{x} = A(A^T A)^{-1} A^T b \quad \text{or} \quad A^T A\hat{x} = A^T b \tag{116}$$

The solution of the normal equations $\hat{x} = (\hat{F}, \hat{G}, \hat{C})^T$ always exists and minimizes the error $E_{min} = E(\hat{x}) = ||b - A\hat{x}||^2$.

• In the special case where we wish to fit a straight line ($F = 0$),

$$A^T A = \begin{pmatrix} \sum_i V_i^2 & \sum_i V_i \\ \sum_i V_i & m \end{pmatrix}, \qquad A^T b = \begin{pmatrix} \sum_j V_j I_j \\ \sum_j I_j \end{pmatrix} \tag{117}$$

where m is the number of data points. Notice that $A^T A$ is symmetric. The normal equations are below, solve them in a specific example

$$\begin{pmatrix} \sum_i V_i^2 & \sum_i V_i \\ \sum_i V_i & m \end{pmatrix} \begin{pmatrix} \hat{G} \\ \hat{C} \end{pmatrix} = \begin{pmatrix} \sum_j V_j I_j \\ \sum_j I_j \end{pmatrix} \tag{118}$$

14 Operators on inner-product spaces

• An *inner product space* is a vector space V with an inner product (x, y) for $x, y \in V$. For example, R^n with the standard dot product $(x, y) = x^T y$ is an inner product space. Inner product spaces are also called *Hilbert spaces* and are the arena for geometric discussions concerning lengths and angles.

• Suppose $A : U \to U$ is a linear transformation from the inner product space U to itself, then we call A an operator on the inner product space U. This concept also applies to $A : U \to V$.

• Dirac Bra-Ket notation: Suppose e_i are a basis for a vector space, say R^n. Think of these as column vectors. Dirac's notation for them is $|e_i\rangle$. Indeed any column vector x is called a ket-vector, and may be written as a linear combination $|x\rangle = \sum_{i=1}^{n} x_i |e_i\rangle$. On the other hand, the basis of row vectors e_i^T are denoted $\langle e_i|$. Any row vector y is a linear combination $\langle y| = \sum_i y_i \langle e_i|$.

• Moreover, the inner product is written as $(x, y) = \langle x|y\rangle = \sum_{i,j} x_i y_j \langle e_i|e_j\rangle$. If e_i are an orthonormal basis, then $\langle e_i|e_j\rangle = \delta_{ij}$, and $\langle x|y\rangle = \sum_i x_i y_i$.

• The matrix elements A_{ij} of a linear transformation $A : V \to V$ in the basis e_i is given by

$$A_{ij} = e_i^T A e_j = (e_i, A e_j) = \langle e_i|A|e_j\rangle \tag{119}$$

To see this note that Ae_j is the j^{th} column of A and $e_i^T A$ is the i^{th} row of A or equivalently, the i^{th} column of A^T. Combining these, $e_i^T A e_j$ is the entry in the i^{th} row and j^{th} column. Alternatively, write e_j in the e-basis as the column vector with zeros everywhere except for a

1 in the j^{th} slot and similarly e_i^T as the row vector with a 1 in the i^{th} slot and zeros elsewhere and perform the matrix multiplication.

• More generally, A could be rectangular. Suppose $A : U \to V$, then the matrix element A_{ij} in the e_j basis for U and f_i basis for V is given by $A_{ij} = f_i^T A e_j = (f_i, Ae_j) = \langle f_i A e_j \rangle$.

14.1 Orthogonal Transformations

• A rotation of the plane about the origin is a linear transformation that preserves distances and angles. A reflection about a line through the origin also preserves lengths and angles. Orthogonal transformations generalize this concept to other dimensions. Transformations that preserve inner products are also called isometries.

• An orthogonal transformation on, say a real vector space R^n with inner product is one which preserves the inner product, i.e. $(u, v) = (Qu, Qv)$ for all u, v. The reason it is called orthogonal is because it is represented by an orthogonal matrix.

• Bear in mind that to define an orthogonal transformation our vector space must have an inner product.

• In particular, an orthogonal transformation $u \to Qu$ preserves the length of u: $(u, u) = ||u||^2 = (Qu, Qu) = ||Qu||^2$ and the angle between u and v: $\frac{(u,v)}{||u||||v||} = \frac{(Qu,Qv)}{||Qu||||Qv||}$

• For the standard inner product $(u, v) = u^T v$, we have $u^T v = u^T Q^T Q v$. Since this is true for all u and v, it follows that $Q^T Q = I$.

• In more detail, take u and v to be any orthonormal basis $e_i^T e_j = \delta_{ij}$, then $u^T v = u^T Q^T Q v$ becomes $e_i^T Q^T Q e_j = e_i^T e_j = \delta_{ij}$. This merely says that the matrix elements of $Q^T Q$, $(Q^T Q)_{ij} = e_i^T Q^T Q e_j$ are the same as the matrix elements of the unit matrix:

• So an orthogonal matrix is an $n \times n$ matrix that satisfies $Q^T Q = I$. In other words, the columns of Q are orthonormal.

• So the left inverse of Q is Q^T. But we showed in an earlier section that if the columns of Q are orthonormal, then the right inverse is also Q^T. In other words $QQ^T = Q^T Q = I$. This means the rows of Q are also orthonormal.

• The inverse and transpose of an orthogonal matrix are also orthogonal.

• Check that the product of two orthogonal matrices is also orthogonal.

• The identity matrix and $-I$ are obviously orthogonal.

• The reflection in the x axis in R^2 is orthogonal

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{120}$$

• A 2×2 real matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is orthogonal provided the rows are orthonormal: $a^2 + b^2 = c^2 + d^2 = 1$ and $ac + bd = 0$. These conditions can be 'solved' in terms of trigonometric functions.

• 2×2 orthogonal matrices are either rotations by θ

$$Q = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{121}$$

or rotations by θ composed with a reflection $(x,y) \to (x,-y)$

$$Q = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \tag{122}$$

• Permutation matrices are matrices obtained from permutations of the columns (rows) of the identity matrix. But permuting the columns (rows) does not change the fact that the columns (rows) of I are orthonormal. So permutation matrices are also orthogonal.

$$Q_{(132)} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{123}$$

• So the inverse of a permutation matrix is just its transpose $Q_{132}^T Q_{132} = I$

14.2 Unitary Transformations

• A unitary transformation preserves the inner product on a complex vector space $(z,w) = (Uz, Uw)$ for all z,w. For the standard hermitian inner product on C^n, $(z,w) = z^\dagger w$ this becomes $z^\dagger w = (Uz, Uw) = z^\dagger U^\dagger U w$. Repeating the steps used for orthogonal matrices, unitary matrices are those square matrices that satisfy

$$U^\dagger U = U U^\dagger = I \tag{124}$$

Here the hermitian adjoint of any matrix or vector is the complex conjugate transposed: $A^\dagger = (A^T)^*$. Notice that $(z, Aw) = z^\dagger A w = (A^\dagger z)^\dagger w = (A^\dagger z, w)$ where we used $(A^\dagger)^\dagger = A$.

• For a general inner product space the adjoint $A^\dagger$ of a matrix A is defined through its matrix elements using the above relation $(A^\dagger z, w) \equiv (z, Aw)$.

• We notice that the inverse of a unitary matrix U is its adjoint $U^\dagger$.

• All real orthogonal matrices are automatically unitary, since complex conjugation has no effect.

• A 2×2 complex matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is unitary provided the rows are orthonormal $|a|^2 + |b|^2 = |c|^2 + |d|^2 = 1$, $a\bar{c} + b\bar{d} = 0$.

• Define the matrix exponential as the matrix $e^{Ax} = \sum_{n=0}^\infty \frac{A^n x^n}{n!}$. The sum is absolutely convergent for any square matrix and defines e^{Ax}. We can use it to find more unitary matrices.

• Example: $\sigma_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the first Pauli matrix, it is hermitian. It turns out that $U = e^{i\sigma_1 x}$ is a unitary matrix for any real x. To see this, use the formula for the matrix exponential to show that $U = e^{i\sigma_1 x} = I\cos x + i\sigma_1 \sin x$. It follows that $U^\dagger = I\cos x - i\sigma_1 \sin x$ and that $U^\dagger U = U U^\dagger = I$.

15 Gram-Schmidt orthogonalization and QR decomposition

• We have seen that orthonormal bases $q_i^T q_j = \delta_{ij}$ are very convenient, The components of any vector in an orthonormal basis are just its inner products with the basis vectors

$$x = x_i q_i \Rightarrow x_i = (q_i, x) \tag{125}$$

• So given any basis, it is useful to convert it into an orthonormal basis.

• This is what the Gram-Schmidt procedure of successive orthogonalization does.

• Another reason to be interested in it is the following.

• It is generally hard to invert a matrix A. But there are two classes of matrices that are fairly easy to invert.

• Orthogonal matrices are trivial to invert. $Q^T Q = QQ^T = I \Rightarrow Q^{-1} = Q^T$.

• And inverting a triangular matrix with non-zero diagonal elements is also quite simple (use successive eliminations (since it is already in echelon form). The inverse of a triangular matrix is again triangular.

• For example

$$\begin{pmatrix} a & 0 & \mid & 1 & 0 \\ c & d & \mid & 0 & 1 \end{pmatrix} \to \begin{pmatrix} a & 0 & \mid & 1 & 0 \\ 0 & d & \mid & -c/a & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & \mid & 1/a & 0 \\ 0 & 1 & \mid & -c/ad & 1/d \end{pmatrix} \tag{126}$$

• So it is very interesting that any matrix can be factorized as a product $A = QR$ of an orhogonal matrix and an upper triangular matrix. Upper triangular is also called right triangular, so the letter R.

• The successive orthogonalization procedure actually produces the QR decomposition of a matrix.

• It begins with independent vectors $a_1, a_2 \cdots a_n$ which are the columns of A. From them, it produces an orthonormal basis for $C(A)$, $q_1, q_2, \cdots q_n$.

• Suppose first that the a_i are orthogonal but not necessarily of length 1. Then we can get an orthonormal basis by defining $q_i = \frac{a_i}{||a_i||}$. So the key step is to get an orthogonal basis of vectors.

• To start with, let $q_1 = a_1/||a_1||$. The next vector is a_2, but it may not be orthogonal to a_1, so we subtract out its projection on a_1, and then normalize the result. We continue this way:

$$\begin{array}{rclr} \tilde{q}_1 & = & a_1, & q_1 = \tilde{q}_1/||q_1|| \\ \tilde{q}_2 & = & a_2 - P_{q_1} a_2, & q_2 = \tilde{q}_2/||q_2|| \\ \tilde{q}_3 & = & a_3 - P_{q_1} a_3 - P_{q_2} a_3, & q_3 = \tilde{q}_3/||q_3|| \\ & \vdots & & \\ \tilde{q}_n & = & (1 - P_{q_1} - P_{q_2} - \cdots - P_{q_{n-1}}) a_{n-1}, & q_n = \tilde{q}_n/||q_n|| \end{array} \tag{127}$$

• By construction, for each r, q_r is orthogonal to all the q's before it, and it is normalized. So we have an orthonormal system of vectors which may be assembled as the columns of an orthogonal matrix $Q = (q_1 q_2 \cdots q_n)$, $Q^T Q = I$

• But we also see the triangular character of the construction. a_1 is along q_1, a_2 is a combination of q_1 and q_2, a_r is a combination of $q_1 \cdots q_r$ etc. But precisely which combinations?

• To find out, we just reap the benefit of our construction. Since q_i are an orthonormal basis, the components of any vector in this basis are just the inner products:

$$\begin{array}{rcl} a_1 & = & (q_1, a_1) q_1 \\ a_2 & = & (q_1, a_2) q_1 + (q_2, a_2) q_2 \\ a_2 & = & (q_1, a_3) q_1 + (q_2, a_3) q_2 + (q_3, a_3) q_3 \end{array}$$

$$\vdots$$
$$a_n = (q_1, a_n)q_1 + (q_2, a_n)q_2 + \cdots + (q_n, a_n)q_n \tag{128}$$

• In matrix form this is $A = QR$

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} = \begin{pmatrix} q_1 & q_2 & \cdots & q_n \end{pmatrix} \begin{pmatrix} q_1^T a_1 & q_1^T a_2 & \cdots & q_1^T a_n \\ 0 & q_2^T a_2 & \cdots & q_2^T a_n \\ 0 & 0 & \cdots & \cdots \\ 0 & 0 & \cdots & q_n^2 a_n \end{pmatrix} \tag{129}$$

• As an example, let us find the orthonormal basis arising from and the corresponding QR decomposition

$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, a_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, a_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \tag{130}$$

In this case you can guess the answer easily.

• Apply the Gram-Schmidt procedure to the following basis for R^3

$$a_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, a_2 = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}, a_3 = \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix} \tag{131}$$

• Use the QR decomposition to invert A.

• 2 dimensional example

$$a_1 = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix}, a_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{132}$$

• Consider the vector space of real polynomials in one variable $-1 \le x \le 1$ with the inner product $(f, g) \int_{-1}^{1} f(x)g(x)dx$. A basis is given by the monomials $1, x, x^2, x^3, \ldots$. However the basis is not orthogonal or even normalized, for example $(1, 1) = 2$. Use the Gram-Schmidt procedure to convert it to an orthonormal basis. The corresponding polynomials are the Legendre polynomials.

16 Invariance of matrix equations under orthogonal/unitary and general linear changes of basis

• Suppose we have an o.n. basis e_i, $e_i^T e_j = \delta_{ij}$ and a vector $\vec{x}$ with components x_i in this basis $\vec{x} = x_i e_i = x^T e$

• Then we make a change of basis to a new orthonormal system $\bar{e}_i$:

$$\begin{aligned} \bar{e}_1 &= q_{11}e_1 + q_{21}e_2 + \cdots + q_{n1}e_n \\ \bar{e}_2 &= q_{12}e_1 + q_{22}e_2 + \cdots + q_{n2}e_n \\ &\cdots \\ \bar{e}_n &= q_{1n}e_1 + q_{2n}e_2 + \cdots + q_{nn}e_n \end{aligned} \tag{133}$$

• Let $e = \begin{pmatrix} e_1 \\ e_2 \\ \cdots \\ e_n \end{pmatrix}$ denote the column vector whose rows are the e_i and similarly $\bar{e} = \begin{pmatrix} \bar{e}_1 \\ \bar{e}_2 \\ \cdots \\ \bar{e}_n \end{pmatrix}$ denote column vector whose rows are $\bar{e}_i$. In matrix form $\bar{e} = Q^T e$ or $\bar{e}^T = e^T Q$ where $Q = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{pmatrix}$. The columns of Q are the components of the new basis in the old basis. Since the new basis is orthonormal, these columns are orthonormal and so Q is an orthogonal matrix $Q^T Q = QQ^T = I$.

• So we can write $\bar{e} = Q^T e$ or $e = Q\bar{e}$.

• The physical vector $\vec{x}$ of course does not change, but it has new components in the new basis $\vec{x} = x^T e = \bar{x}^T \bar{e}$. Thus

$$\vec{x} = x^T Q\bar{e} = (Q^T x)^T \bar{e} \Rightarrow \bar{x} = Q^T x. \tag{134}$$

• So the new and old components are related by $x = Q\bar{x}$ or $\bar{x} = Q^T x$.

• Now suppose we had a matrix equation $Ax = b$. Since both x and b are vectors, they transform in the same way

$$AQ\bar{x} = Q\bar{b} \quad \Rightarrow \quad Q^T AQ\bar{x} = \bar{b} \tag{135}$$

• Thus the equation takes the same form in the new reference frame if we let $\bar{A} = Q^T AQ$. This is the transformation rule for a matrix under an orthonormal change of basis.

• It follows that $\bar{A} + \bar{B} = Q^T(A + B)Q$ and $\bar{A}\bar{B} = Q^T ABQ$. So any polynomial (algebraic function) in matrices transforms in the same way as a single matrix.

$$F(\bar{A}, \bar{B}, \cdots, \bar{P}) = Q^T F(A, B, \cdots, P)Q \tag{136}$$

• So if we have an algebraic relation among matrices $F(A, B, \cdots P) = 0$ then we have the same algebraic relation among the orthogonally transformed matrices

$$F(\bar{A}, \bar{B}, \cdots, \bar{P}) = 0 \tag{137}$$

• Thus we have the invariance of matrix equations under orthogonal transformations.

• Moreover, the inverse of an (invertible) matrix transforms in the same way $\bar{A}^{-1} = Q^T A^{-1} Q$.

• Furthermore, the transpose of a matrix transforms in the same way: $\bar{A}^T = Q^T A^T Q$. So any algebraic matrix equation involving matrices, their inverses and their transposes is invariant under orthogonal transformations.

• For example $A - A^T = 0$ becomes $\bar{A} - \bar{A}^T = 0$. So a matrix that is symmetric in one o.n. frame is symmetric in any other o.n. frame.

• For example, if R is orthogonal in one orthonormal frame, $R^T R = I$, then the transformed matrix $\bar{R} = Q^T RQ$ is also orthogonal $\bar{R}^T \bar{R} = I$.

• If we replace orthogonal by unitary and transpose by conjugate-transpose $\dagger$, then all of the above continues to hold. So a matrix that is hermitian in one o.n. frame is hermitian in every other o.n. basis for C^n.

• While components of vectors and matrices generally transform as above, some special vectors and matrices, have the same components in every o.n. frame. These are the zero vector, zero matrix and identity matrix.

• The angle between two vectors, length of a vector and inner product of a pair of vectors are also invariant under orthogonal and unitary transformations as discussed earlier.

• The trace and determinant of a matrix are also orthogonally and unitarily invariant as discussed in the next section.

• Under a general linear transformation S (invertible but not necessarily orthogonal or unitary), under which the basis vectors transform as $\bar{e} = S^T e$, the components of a vector transform as $x = S\bar{x}$ or $\bar{x} = S^{-1}x$ and those of a matrix transform as $\bar{A} = S^{-1}AS$ or $A = S\bar{A}S^{-1}$. A and $\bar{A}$ are called similar matrices.

• Algebraic equations in matrices (not involving the transpose) are again invariant under general linear transformations. General linear transformations are also called similarity transformations.

17 Determinant and Trace

17.1 Invertibility and Volume

• Consider the 2×2 system

$$ax + by = f, \quad cx + dy = g, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} \tag{138}$$

• To be solvable for arbitrary data f, g, we need the solution $x = \frac{fd-bg}{ad-bc}$, $y = \frac{ag-cf}{ad-bc}$ to exist, i.e., $ad - bc \neq 0$. The augmented matrix in row echelon form is

$$\begin{pmatrix} a & b & | & f \\ 0 & d - \frac{bc}{a} & | & g - \frac{cf}{a} \end{pmatrix} \tag{139}$$

• So for a 2×2 matrix $(ab|cd)$ to be invertible we need its determinant $ad - bc \neq 0$.

• Notice that the determinant is the product of pivots a, $d - \frac{bc}{a}$

• We seek the analogue of this number that determines when an $n \times n$ matrix is invertible.

• Recall that an $A_{n \times n}$ is invertible iff it has independent rows (columns). This suggests a geometric interpretation of the rows (columns) of A, think of them as edges of a parallelepiped that emanate from a vertex that is located at the origin.

• Then the rows (columns) are independent iff the volume of the parallelepiped is non-zero. If they are dependent, the parallelepiped becomes degenerate and lies within a hyperplane and does not occupy any n-dimensional volume.

• So the volume of the parallelepiped is a natural candidate for the determinant of a matrix.

• Check that area of a parallelogram with vertices at (a, b) and (c, d) has area $ad - bc$. First consider a rectangle, then a parallelogram with base along the x-axis. In general,the area of a parallelogram is the base times the height. For simplicity both points are in the first quadrant

$$Area = \sqrt{a^2 + b^2} \times h, \quad \sin\theta = \frac{h}{\sqrt{c^2 + d^2}},$$

$$\cos\theta = \frac{ac+bd}{\sqrt{a^2+b^2}\sqrt{c^2+d^2}}, \quad \sin\theta = \sqrt{1-\cos^2\theta} = \frac{ad-bc}{\sqrt{a^2+b^2}\sqrt{c^2+d^2}} \tag{140}$$

So $Area = ad-bc$. We could also use the formula that the magnitude of the area is the norm of the cross-product of the two vectors, which would lead to $|Area| = \sqrt{a^2+b^2}\sqrt{c^2+d^2}|\sin\theta| = |ad-bc|$

17.2 Postulates or axioms of determinant

• Axioms of $\det A$ motivated via (signed) volume of parallelepiped

1. $\det I = 1$, since the volume of a cube is 1.

2. $\det A \to -\det A$ if a pair of different rows of A are exchanged. For example, the signed-area of a parallelogram flips sign if we change the orientation. The rows form either a right handed or left handed coordinate system in three dimensions.

• So we now know the determinant of permutation matrices, for example $\det\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -1$.

• Recall that a permutation matrix is obtained by permuting the rows of the identity matrix. Every permutation is a product of row exchanges (exchanges are sometimes called transpositions, nothing to do with transpose!). The determinant of a permutation P is 1 if an even number of row exchanges are made and -1 if an odd number of exchanges are made. $\det P$ is also sometimes called the sign of a permutation $\operatorname{sgn} P$ or the parity of the number of transpositions that P is a product of.

3. $\det A$ is linear in each row separately (holding other rows fixed). eg. the volume doubles if we double the length of one edge.

• This means for instance that for an $n\times n$ matrix, $\det rA = r^n \det A$ and for 2×2 matrices,

$$\det\begin{pmatrix} \lambda a & \lambda b \\ c & d \end{pmatrix} = \lambda \det\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det\begin{pmatrix} a+e & b+f \\ c & d \end{pmatrix} = \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \det\begin{pmatrix} e & f \\ c & d \end{pmatrix} \tag{141}$$

• Note that $\det(A+B) \neq \det A + \det B$ in general.

• For example, in the 2×2 case, we can get the formula $\det A = ad - bc$ using these three axioms.

• These three axioms define the determinant uniquely, i.e., for any matrix, there is only one number $\det A$ satisfying these properties. This can be shown using elimination and the pivot formula (see next section).

• One should really think of the determinant as a function of the rows $\det A = \det(r_1, r_2, \cdots, r_n)$.

• It is the unique complex-valued anti-symmetric multilinear function of the rows that is normalized to $\det I = 1$. Later we will see that 'rows' can be replaced with 'columns'.

17.3 Properties of determinants

• To get a formula it helps to derive some properties of the determinant from these axioms

• If a row of A vanishes, $\det A = 0$

• If two different rows are the same, then $\det A = 0$

• $\det A$ is unchanged if we subtract a multiple of a row from another row. This follows from linearity.

$$\det \begin{pmatrix} a & b \\ c - \lambda a & d - \lambda b \end{pmatrix} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{142}$$

• So $\det A$ does not change under elementary row operations (without row exchanges).

• Recall that $PA = LU$ summarizes gaussian elimination, where P is a permutation of the rows, L is lower triangular with 1's on the diagonal and U is upper triangular (row echelon form) with the pivots in the diagonal. All matrices here are square.

• If U is an upper (or lower) triangular matrix, $\det U$ is the product of its diagonal entries.

• Proof: If the pivots are non-zero, we can use elimination to eliminate all non-diagonal entries of U and then reduce to the identity by dividing by the pivots. In case a diagonal entry of U is zero, then U is singular and we can use elimination to get a row of zeros.

• So $\det E = 1$ for any elementary (elimination) matrix E that implements row elimination.

• So $\det A$ is the product of its pivots times $(-1)^{\text{number of row exchanges}}$ which is the sign of the permutation (row exchanges) that is needed to bring A to echelon form.

• In particular for a diagonal matrix, $\det \mathrm{diag}(\lambda_1, \cdots, \lambda_n) = \lambda_1 \lambda_2 \cdots \lambda_n$.

• If A is singular, then $\det A = 0$, since elimination will produce a row of zeros. So if A has non-trivial kernel, then its determinant vanishes.

• If A is invertible, then $\det A \neq 0$, since elimination will produce n non-vanishing pivots whose product is non-zero.

• Uniqueness: There is a unique function $D(A)$ that satisfies the axioms of a determinant. Sketch of Proof: Suppose there were two functions D_1 and D_2 both satisfying the axioms of a determinant. Suppose further that there is a square matrix A with $D_1(A) \neq D_2(A)$, in other words, the determinant is not unique. We will end up in a contradiction. We showed above that row operations do not change the determinant as defined by the three axioms (except that row exchanges change the sign). So suppose R is the row echelon form of A. Then $D_1(R) = D_2(R)$. Suppose A is singular. Then R has a row of zeros and the axioms imply that $|D_1(A)| = |D_1(R)| = 0$ and $|D_2(A)| = |D_2(R)| = 0$. This is a contradiction. So either the assumptions are wrong or A must not be singular. If A is not singular, R has n non-zero pivots along the diagonal and zeros elsewhere. The axioms imply that $D_1(A) = (-1)^e D_1(R)$ and $(-1)^e D_2(R) = D_2(A)$, both being the product of pivots (multiplied by ± 1 where used e row exchanges). Again we have a contradiction. So our assumption that the determinant is not unique is false and $\det A$ is unique.

• In many applications, the most important property of the determinant is the product rule $\det AB = \det A \det B$.

• Check the product rule for 2×2 and 3×3 matrices by explicit calculation.

• It follows that $\det A^p = (\det A)^p$

• $\det AB = \det A \det B$. If $\det B = 0$, then AB and B are singular and this is trivially true.

If $\det B \neq 0$, then we check that $D(A) = \frac{\det AB}{\det B}$ satisfies the same 3 axioms as $\det A$ and by uniqueness, $D(A) = \det A$

• Caution: The pivots of AB need not equal the products of the pivots of A and those of B. For example, $A = B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ each have a single pivot equal to 1. But $AB = 0$ has no pivots. So one has to be careful when multiplying pivots of matrices. Also, if R_A is the row echelon form of A and R_B is the echelon form of B, $AB \neq R_A R_B$ in general! For example, let $A = B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then $AB = \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix}$. But $R_A = R_B = \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix}$ and $R_A R_B = \begin{pmatrix} 1 & -2 \\ 0 & 4 \end{pmatrix} \neq AB$. Nevertheless, $\det AB = \det R_A R_B$.

• $\det A^{-1} = 1/\det A$ since $\det(AA^{-1}) = \det I = 1$

• The determinant of a permutation P is the same as that of its transpose P^T. This is because permutations are orthogonal $PP^T = I$. Applying the product rule, and using $\det P = \pm 1$ implies $\det P^T = \det P$.

• $\det A^T = \det A$. If A is singular, so is A^T and then both sides vanish. Otherwise use elimination to write $PA = LU$ and transpose to $A^T P^T = U^T L^T$ and use the product formula. U is upper triangular with non-zero pivots along the diagonal. L is lower triangular with all pivots equal to 1. Note that the determinant of such a triangular matrix only involves the diagonal elements (the other entries can be killed off by elimination) which are unchanged by transposition. So $\det L = \det L^T$, $\det U = \det U^T$. Moreover, $\det P = \det P^T$. So we conclude that $\det A^T = \det A$.

• $\det A^* = (\det A)^*$

• $\det A^\dagger = (\det A)^*$

• The determinant of an orthogonal matrix $Q^T Q = I$ is ± 1

$$1 = \det Q^T Q = (\det Q)^2 \tag{143}$$

This is a generalization of what we found for permutation matrices $\det P = \pm 1$.

• The determinant of a unitary matrix $U^\dagger U = I$ is a complex number of unit magnitude

$$1 = \det U \det U^\dagger = \det U (\det U)^* = |\det U|^2 \tag{144}$$

• Basis independence of determinant. The determinant is unchanged by a similarity transformation $A' = SAS^{-1} \Rightarrow \det A' = \det SAS^{-1} = \det S \det A \det S^{-1} = \det A$

• Orthogonal and unitary transformations are special cases of similarity transformatons. So the determinant is invariant under unitary and orthogonal transformations

$$\det(U^\dagger H U) = \det U \det H \det U^\dagger = \det H |\det U|^2 = \det H, \quad \text{and} \quad \det A = \det Q^T A Q \tag{145}$$

• The determinant of a projection matrix (other than the identity) is 0. This is because a projection onto a proper subspace always has less than maximal rank and therefore is not invertible.

• Consider the homogeneous $n \times n$ system $Ax = 0$. We have seen that the following are equivalent

1. There is a non-trivial solution x
2. A has non-trivial kernel
3. A is not invertible
4. $\det A = 0$
5. Rows of A are dependent
6. Columns of A are dependent
7. A has a zero eigenvalue

• The formula for the cross product of vectors in R^3 can be formally expressed in terms of a determinant using the usual cartesian components

$$\vec{b} \times \vec{c} = \det \begin{pmatrix} \hat{i} & \hat{j} & \hat{k} \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \tag{146}$$

• Triple product of vectors $a \cdot (b \times c)$ is the volume of a parallelepiped determined by the vectors a, b, c, it may be expressed as a determinant, which makes clear that $a \cdot (b \times c) = -b \cdot (a \times c)$

$$a \cdot (b \times c) = \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \tag{147}$$

17.4 Formulas for determinants of $n \times n$ matrices

• **Pivot formula:** $\det A$ is the product of its pivots times the sign of the permutation (row exchange) that is needed to bring A to echelon form. If A does not have n pivots, $\det A = 0$. This is how determinants are calculated numerically.

• But there is no explicit formula for the pivots in terms of the original matrix elements a_{ij}. If we want an explicit formula for $\det A$ we proceed as below.

• **Sum over permutations:** This basically reduces the problem to determinants of permutation matrices but involves $n!$ terms.

• Repeatedly use linearity to write $\det A$ as the sum of $n!$ determinants of matrices which have at most one non-zero element in each row and each column. But the latter are just multiples of permutation matrices whose determinant are just the signs of the permutations.

• For example in the 2×2 case (also work out the 3×3 case.)

$$\begin{aligned} \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \det \begin{pmatrix} a & 0 \\ c & d \end{pmatrix} + \det \begin{pmatrix} 0 & b \\ c & d \end{pmatrix} \\ &= \det \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} + \det \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} + \det \begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix} + \det \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} \\ &= \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} ad + \det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} bc \end{aligned}$$

$$= \quad A_{11}A_{22} \text{ sgn } (1)(2) + A_{12}A_{21} \text{ sgn } (12) \tag{148}$$

• Here the permutation matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ exchanges the first two rows, so we call it the permutation (12) and its determinant is denoted $\text{sgn } (12) = -1$. The identity matrix is the permutation $(1)(2)$ sending the first row to itself and the second to itself, $\text{sgn } (1)(2) = +1$. (1342) is the permutation $P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$.

• To begin with there are $n^n = 2^2$ terms in the sum (linearity in each row produces n terms). But those terms that have a column of zeros do not contribute. We are left with terms that have precisely one non-zero entry in each row and precisely one non-zero entry in each column. Their determinants are multiplies of the determinant of permutation matrices. There are $n!$ permutation matrices and so we have a sum of $n! = 2$ terms. Each of these surviving terms is a multiple of the determinant of a permutation matrix. Here the identity is also a permutation, which does not permute anything.

• Thus we have a sum over all permutations of the rows (or columns).

$$\det A = \sum_{\text{permutations} \quad P} \det(P) A_{1P(1)} A_{2P(2)} \cdots A_{nP(n)} \tag{149}$$

Here we define $P(i) = j$ if the i^{th} row is sent to the j^{th} row by the permutation matrix P.

• So any determinant is a linear combination of determinants of permutation matrices. Half $(= n!/2)$ the $\det P$'s are $+1$ and half -1.

• **Laplace's cofactor expansion:** This expresses $\det A$ as an n-term linear combination of determinants of $n-1 \times n-1$ matrices, which are the minors of A.

• In the 2×2 case we have seen that

$$\left|\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right| = \left|\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}\right| + \left|\begin{pmatrix} 0 & b \\ c & 0 \end{pmatrix}\right| = ad\left|\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right| + bc\left|\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right| = a\det(d) - b\det(c) \tag{150}$$

• It is instructive to consider the 3×3 case. By linearity and avoiding columns of zeros we get

$$\begin{aligned}
\left|\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\right| &= \left|\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix}\right| + \det\begin{pmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{pmatrix} + \det\begin{pmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{pmatrix} \\
&= a_{11}\left[a_{22}a_{33}\det\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + a_{23}a_{32}\det\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\right] \\
&\quad + a_{12}\left[a_{21}a_{33}\det\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + a_{23}a_{31}\det\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}\right] \\
&\quad + a_{13}\left[a_{21}a_{32}\det\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} + a_{22}a_{31}\det\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}\right] \\
= a_{11}\det M_{11} \quad &- \quad a_{12}\det M_{12} + a_{13}\det M_{13} \;=\; a_{11}C_{11} - a_{12}C_{12} + a_{13}C_{13}
\end{aligned} \tag{151}$$

• More generally, expanding along the i^{th} row of an $n \times n$ matrix,

$$\det A = \sum_j A_{ij} C_{ij}, \quad \text{no sum on } i. \tag{152}$$

where the cofactor $C_{ij} = (-1)^{i+j} \det M_{ij}$ is the signed determinant of the minor M_{ij}. The minor M_{ij} is the sub-matrix obtained by omitting the row and column containing A_{ij}.

• It can be obtained by judiciously collecting terms in the sum over permutations formula involving each entry in any given row (or column). For example, expanding along the 1^{st} row,

$$\det A = \sum_j A_{1j} C_{1j} \tag{153}$$

This comes from writing

$$\begin{aligned}\det A &= \sum_P (\det P)\; A_{1P(1)} A_{2P(2)} \cdots A_{nP(n)} = A_{11} \sum_{P:P(1)=1} (\det P)\; A_{2P(2)} \cdots A_{nP(n)} \\ &\quad + A_{12} \sum_{P:P(1)=2} (\det P)\; A_{2P(2)} \cdots A_{nP(n)} + \cdots \\ &\quad + A_{1n} \sum_{P:P(1)=n} (\det P)\; A_{2P(2)} \cdots A_{nP(n)} \end{aligned} \tag{154}$$

We observe that the coefficient of A_{11} is itself a determinant, indeed it is the determinant of the minor M_{11}. Similarly the coefficient of A_{12} is a sum of products of the entries of the minor M_{12} and some work shows that it is in fact equal to $-\det M_{12}$. Proceeding this way we get the cofactor expansion formula.

• Applying the cofactor formula to A^T and using $\det A = \det A^T$ we see that we could expand along columns rather than rows.

• Example: Determinant of a tri-diagonal matrix: useful in calculating determinants of differential operators.

• **Product of eigenvalues:** If A has eigenvalues $\lambda_1 \cdots \lambda_n$ listed according to algebraic multiplicity then $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$ is the product of eigenvalues.

• This comes from the constant term in the characteristic polynomial

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) = 0 \tag{155}$$

17.5 Cramer's rule for solving $n \times n$ linear systems

• When A is invertible, $Ax = b$ has solution $x = A^{-1} b$. This can be expressed as a ratio of determinants. Below, B_i is the matrix A with the i^{th} column replaced by b, $(B_i)_{kl} = A_{kl}$ if $l \neq i$ and $(B_i)_{ki} = b_k$.

$$x_i = \frac{\det B_i}{\det A} \tag{156}$$

• There is a trick to get this formula. Let us illustrate it in the 3×3 case. We being with the identity obtained by multiplying by columns

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix} \tag{157}$$

Taking determinants, $\det A \cdot x_1 = \det B_1$, so that $x_1 = \frac{\det B_1}{\det A}$. Similarly,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 1 & x_1 & 0 \\ 0 & x_2 & 0 \\ 0 & x_3 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{pmatrix} \tag{158}$$

taking determinants, $(\det A)x_2 = \det B_2$. In this manner we find the solution of $Ax = b$

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^T = \frac{1}{|A|} \begin{pmatrix} \det B_1 & \det B_2 & \cdots & \det B_n \end{pmatrix}^T \tag{159}$$

• Cramers rule is not a computationally efficient way of solving a system, elimination is much faster.

17.6 Formula for the inverse

We can write the above formula for the solution of $Ax = b$ in terms of cofactors. We expand $\det B_i$ along column i using the cofactor formula $\det B_i = (B_i)_{ji} C_{ji} = b_j C_{ji}$. Thus

$$x_i = \frac{\det B_i}{\det A} = \frac{1}{\det A} b_j C_{ji} \quad \Rightarrow \quad x = \frac{C^T b}{\det A} \tag{160}$$

But we know that $x = A^{-1} b$. So we conclude that if A is invertible, then a formula for the inverse in terms of the transposed cofactor matrix C is

$$A^{-1} = \frac{1}{\det A} C^T \tag{161}$$

• This formula is not a computationally efficient way of inverting a matrix, Gauss-Jordan elimination is quicker.

17.7 Volume element: Change of integration variable and Jacobian determinant

• An important application of determinants is in the change of volume element when (non-linearly) changing integration variables in multi-dimensional integrals.

• An invertible square matrix A can be regarded as a linear change of variable from the standard o.n. basis $(x_i)_j = \delta_{ij}$ to a new basis $\mathbf{y}_i$ given by the columns of A:

$$I = \begin{pmatrix} | & \cdots & | \\ x_1 & \cdots & x_n \\ | & \cdots & | \end{pmatrix}; A = \begin{pmatrix} | & \cdots & | \\ y_1 & \cdots & y_n \\ | & \cdots & | \end{pmatrix}; \tag{162}$$

$\mathbf{y}_i = A\mathbf{x}_i$ or $(y_i)_j = A_{jk}(x_i)_k = A_{ji}$. (Thus A is the derivative of y with respect to x evaluated at $(x_i)_j = \delta_{ij}$: $A_{jk} = \frac{\partial (y_i)_j}{\partial (x_i)_k}$). Under this change of variable, the unit hypercube (whose edges are $\mathbf{x}_i$) is transformed into a parallelepiped whose edges are the columns $\mathbf{y}_i$ of A. So the volume of the parallelepiped formed by the basis vectors is multiplied by $\det A$.

• Now we would like to apply this idea to differentiable non-linear changes of variable. This is given by a function from $R^n \to R^n$: $(x_1 \cdots x_n) \mapsto (y_1(\mathbf{x}), \cdots y_n(\mathbf{x}))$

• A non-linear change of variable can be approximated by an affine (linear + shift) one in a small neighbourhood of any point x', $y_i(x) = y_i(x') + J_{ij}(x-x')_j + \cdots$. Up to an additive constant shift, this linear transformation is the *linearization* of y, or the Jacobian matrix $J_{ij} = \frac{\partial y_i}{\partial x_j}$ where the derivatives are evaluated at $x = x'$. So near each point, the unit hyper cube is transformed to a parallelepiped whose volume is $\det J$

• The Jacobian matrix is $J_{ij}(\mathbf{x}) = \frac{\partial y_i}{\partial x_j}$ and the Jacobian determinant is $\det J_{ij}$

• The change of variable formula for volume elements is

$$|\det J| \; dx_1 \cdots dx_n = dy_1 \cdots dy_n \tag{163}$$

So that

$$\int dy_1 \cdots dy_n \; f(\mathbf{y}) \;\; = \;\; \int dx_1 \cdots dx_n \;\; |\det J(\mathbf{x})| \;\; f(\mathbf{y}(\mathbf{x})) \tag{164}$$

• Jacobian determinant for transformation from cartesian to polar coordinates on the plane $x = r\cos\theta, y = r\sin\theta$

$$dx\; dy \;\; = \;\; dr\; d\theta\; \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = dr\; d\theta\; \det \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} = r\; dr\; d\theta. \tag{165}$$

• Ex. Work out the Jacobian determinant for transformation from cartesian to spherical polar coordinates. $z = r\cos\theta, \quad x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi$.

• Note: The Jacobian matrix of the gradient of a function is the Hessian matrix which we will encounter when studying positive definite matrices.

17.8 Trace

• The trace of a matrix is the sum of its diagonal entries, $\text{tr } A = A_{ii}$

• The trace is cyclic: $\text{tr } AB = \text{tr } BA$, since $\text{tr } AB = A_{ij}B_{ji} = B_{ji}A_{ij} = \text{tr } AB$, $\text{tr } ABC = \text{tr } CAB = \text{tr } BCA$.

• Basis independence of trace under similarity transformation: $\text{tr } S^{-1}AS = \text{tr } SS^{-1}A = \text{tr } A$. In particular the trace is invariant under orthogonal and unitary transformations $\text{tr } Q^TAQ = \text{tr } A$, $\text{tr } A = \text{tr } U^\dagger AU$.

18 Diagonalization: Eigenvalues and Eigenvectors

• For an $n \times n$ matrix, the domain and target space are both R^n or both C^n and may be identified. So $x \mapsto Ax$ transforms $x \in C^n$ to another vector in C^n. The vectors that behave in the simplest manner are those x sent to a multiple of themselves. i.e., $Ax = \lambda x$ does not change the direction of x.

• The subspace spanned by x is called an invariant subspace under A. This is a particularly useful feature if we want to apply A again, for then $A^2x = \lambda^2 x, A^3 x = \lambda^3 x, \cdots$. In a sense x does not mix with other vectors under application of A.

• This is very useful in solving time-evolution problems. Eg systems of differential equations $\frac{\partial u}{\partial t} = Au$, where we need to apply A repeatedly to evolve $u(t)$ forward in time.

• For an $n \times n$ matrix, the equation $Ax = \lambda x$ is called the eigenvalue problem.

• The scalars $\lambda_1, \lambda_2, \cdots, \lambda_n$ for which the eigenvalue problem can be solved non-trivially are called the eigenvalues and the corresponding non-zero vectors $x_1, x_2, \cdots x_n$ are the eigenvectors or principal axes. The zero vector $x = 0$ is not considered an eigenvector of any matrix, since it trivially solves $Ax = \lambda x$ for any λ.

• Eigen-vector is a German word meaning own-vector, the eigenvectors of a matrix are like its private property.

• Note that if x is an eigenvector of A with eigenvalue λ, $Ax = \lambda x$, then so is any non-zero multiple, $A(cx) = \lambda(cx)$. So eigenvectors are defined up to an arbitrary normalization (scale) factor. Often, it is convenient to normalize eigenvectors to have length one, $||x|| = 1$.

• Consider $Ax = \lambda x$ which is the homogeneous system $(A - \lambda I)x = 0$. We know that a non-trivial solution (eigenvector) exists iff $\det(A - \lambda I) = 0$.

• So the eigenvalues λ_i are precisely the solutions of $\det(A - \lambda I) = 0$.

$$\det(A - \lambda I) = \det \begin{pmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \alpha_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn-\lambda} \end{pmatrix} = 0 \tag{166}$$

• This is an n^{th} order polynomial equation in λ. It is called the characteristic equation.

• For example, the characteristic equation of the real symmetric matrix $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ is

$$\det \begin{pmatrix} 1 - \lambda & 2 \\ 2 & 4 - \lambda \end{pmatrix} = (1 - \lambda)(4 - \lambda) - 4 = \lambda^2 - 5\lambda = \lambda(\lambda - 5) = 0 \tag{167}$$

The eigenvalues are $\lambda = 0, 5$ and the corresponding eigenvectors are $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Notice that the eigenvalues are real, we will see that had to be the case because A is symmetric. The determinant is $1 \times 4 - 2 \times 2 = 0$ which is the same as the product of eigenvalues, 5×0. Notice that the trace is $1 + 4 = 5$ which is the same as the sum of eigenvalues.

• The characteristic polynomial $\det(A - \lambda I)$ has n complex roots. These are the n eigenvalues of any $n \times n$ matrix. Some of them may be repeated roots, which should be counted with multiplicity.

• So the characteristic polynomial may be written as

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) \tag{168}$$

• Actually, it is convenient to multiply by $(-1)^n$ so that the polynomial is monic, i.e. coefficient of λ^n is 1. Expanding out the product, the characteristic equation may be written as

$$(-1)^n \det(A - \lambda I) = \lambda^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \cdots c_1\lambda + c_0 = 0 \tag{169}$$

• Setting $\lambda = 0$ we see that the constant term is the determinant upto a possible sign and this may also be identified with the product of eigenvalues

$$(-1)^n \det A = c_0; \qquad \det A = \lambda_1 \lambda_2 \cdots \lambda_n \tag{170}$$

• Moreover $-c_{n-1}$, the coefficient of $-\lambda^{n-1}$ is the sum of the eigenvalues $\lambda_1 + \lambda_2 + \cdots + \lambda_n$. It turns out that this is the trace of A.

• A polynomial equation of $n^{\rm th}$ order has n (in general complex) roots. The zeros of the characteristic polynomial are the n eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Generically, they are distinct. But it may happen that some of the eigenvalues coincide.

• The eigenvalues are also called characteristic values.

• The eigenvalues of A^T are the same as the eigenvalues of A. This is because $\det(A^T - \lambda I) = \det(A - \lambda I)$. So A and A^T have the same characteristic polynomial.

• To any given eigenvalue λ_1, there is a solution to the eigenvalue problem $A\vec{u}_1 = \lambda_1 \vec{u}_1$, giving the eigenvector $\vec{u}_1 = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$. Thus the spectrum consists of

$$\begin{array}{rl} \text{eigenvalues :} & \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n \\ \text{eigenvectors :} & \vec{u}_1,\ \vec{u}_2,\ \cdots,\ \vec{u}_n \end{array} \tag{171}$$

• For example, for the 3×3 identity matrix, the roots of the characteristic equation $(\lambda - 1)^3 = 0$ are $\lambda = 1, 1, 1$, and we would say that 1 is an eigenvalue with (algebraic) multiplicity three. We also say that 1 is an eigenvalue with degeneracy 3

• If eigenvalue λ has multiplicity 1 we say it is a non-degenerate eigenvalue.

• The identity matrix $I_{n\times n}$, satisfies $Ix = x$ for every vector. So every non-zero vector is an eigenvector. The characteristic equation is $(\lambda - 1)^n = 0$, so the only eigenvalue is 1, with an algebraic multiplicity n. Moreover, since every non-zero vector is an eigenvector, there are n-linearly independent eigenvectors corresponding the the eigenvalue 1.

• The space spanned by the eigenvectors corresponding to a given eigenvalue is called the λ-eigenspace of A. This is because it is closed under linear combinations and forms a vector space $Ax = \lambda x, Ay = \lambda y \Rightarrow A(cx + dy) = \lambda(cx + dy)$.

• For the identity matrix $I_{n\times n}$, the 1-eigenspace is the whole of R^n.

• The dimension of the λ-eigenspace is called the geometric multiplicity of eigenvalue λ. It is always $\leq$ algebraic multiplicity. For the identity matrix, the algebraic and geometric multiplicities of eigenvalue 1 are both equal to n.

• A matrix is deficient if the geometric multiplicity of some eigenvalue is strictly less than its algebraic multiplicity. This means it is lacking in eigenvectors. Analysis of such matrices is more involved. They will be dealt with later. Fortunately, the matrices whose eigen-systems we encounter most often in physics ((anti)symmetric, orthogonal, (anti)hermitian and unitary) are not deficient.

• The eigenvectors of non-deficient $n \times n$ matrices span the whole n-dimensional vector space.

• An example of a deficient matrix is

$$N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \Rightarrow \quad \det(N - \lambda I) = \lambda^2 = 0 \quad \Rightarrow \quad \lambda_1 = 0, \quad \lambda_2 = 0 \tag{172}$$

The eigenvectors are then the non-trivial solutions of $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$. So there is only one independent eigenvector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. So the 0-eigenspace is 1-dimensional, though the eigenvalue 0 has algebraic multiplicity two. In this case, the eigenvectors do not span the whole of R^2.

• On the other hand, the eigenvectors corresponding to a pair of distinct eigenvalues are always linearly independent.

• Proof: So we are given $Ax = \lambda x$ and $Ay = \mu y$, with $\lambda \neq \mu$ and eigenvectors $x, y \neq 0$. Now suppose x, y were linearly dependent, i.e. $cx + dy = 0$ with $c, d \neq 0$. We will arrive at a contradiction. Applying A,

$$cAx + dAy = 0 \quad \Rightarrow \quad c\lambda x + d\mu y = 0 \;\Rightarrow\; \lambda(cx + dy) + (\mu - \lambda)dy = 0 \quad \Rightarrow \quad (\mu - \lambda)dy = 0 \tag{173}$$

But $\mu \neq \lambda$ and $d \neq 0$, so $y = 0$, which contradicts the fact that y is a non-zero vector. So we conclude that eigenvectors corresponding to a pair of distinct eigenvalues are always linearly independent.

• This can be extended to any number of distinct eigenvalues: Eigenvectors corresponding to a set of distinct eigenvalues are linearly independent. One can prove this inductively.

• It follows that if an $n \times n$ matrix has n distinct eigenvalues, then the corresponding n eigenvectors are linearly independent and span the whole vector space.

• So matrices with n distinct eigenvalues are not deficient.

• When eigenvalues coincide, their corresponding eigenvectors may remain independent or become collinear. Deficiencies arise in the latter case.

18.1 More examples of eigenvalues and eigenvectors

• The zero matrix $0_{n\times n}$ annihilates all vectors $0x = x$, so every non-zero vector is an eigenvector with eigenvalue 0. The characteristic equation is $\lambda^n = 0$, so 0 is an eigenvalue with multiplicity n.

• Consider the diagonal matrix $D = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. Let us take $n = 3$ for definiteness. The eigenvalue equation becomes

$$\begin{pmatrix} \lambda_1 x_1 \\ \lambda_2 x_2 \\ \lambda_3 x_3 \end{pmatrix} = \lambda \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{174}$$

The solutions are $\lambda = \lambda_1$ with $x_2 = x_3 = 0$ and x_1 arbitrary (in particular we could take $x_1 = 1$ to get an eigenvector of length 1) and similarly two more. So the eigenvectors can be taken as $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ with eigenvalue λ_1, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ with eigenvalue λ_2 and finally $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ with eigenvalue λ_3. Notice that the normalized eigenvectors are just the columns of the identity matrix. The characteristic equation is $(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) = 0$. So the eigenvalues of a diagonal matrix are just its diagonal entries, and the eigenvectors are the corresponding columns of the identity matrix. The determinant is just the product of the diagonal elements.

• The eigenvalues are not always real, consider the rotation matrix

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \det(A - \lambda I) = \det \begin{pmatrix} \cos\theta - \lambda & \sin\theta \\ -\sin\theta & \cos\theta - \lambda \end{pmatrix} = \lambda^2 - 2\lambda\cos\theta + 1 = 0 \tag{175}$$

The roots of the characteristic polynomial are $\lambda = \cos\theta \pm i\sin\theta = e^{\pm i\theta}$, which are generally complex, but lie on the unit circle.

• The set of eigenvalues is called the spectrum of the matrix. It is a subset of the complex plane.

• Consider the projection from R^3 to the sub-space spanned by the vector $a = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, i.e. to the x-axis. $P_a = aa^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Geometrically, $Px = x$ for precisely those vectors along the x-axis. So a is itself a normalized eigenvector with eigenvalue 1. The 1-eigenspace of P is one-dimensional. Only vectors v orthogonal to the x-axis are annihilated $Pv = 0$. So non-zero vectors in the y-z plane are the eigenvectors with eigenvalue 0. So the 0-eigenspace of A consists of all vectors orthogonal to a. Of course, P_a is a diagonal matrix, so we could have read off its eigenvalues: $\{1,0,0\}$.

• The characteristic equation for P_A is $\det(P - \lambda I) = 0$, or $\lambda^2(\lambda - 1) = \lambda(\lambda^2 - \lambda) = 0$. Recall that for a projection matrix, $P^2 = P$. So we make the curious observation that P satisfies its own characteristic equation $P(P^2 - P) = 0$.

18.2 Cayley Hamilton Theorem

• One of the most remarkable facts about matrices, is that every matrix satisfies its own characteristic equation. This is the Cayley-Hamilton theorem.

• Let us first check this in the above example $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$. The characteristic equation is $\lambda^2 - 5\lambda = 0$. The Cayley-Hamilton theorem says that $A^2 - 5A = 0$. It is easy to check that $A^2 = \begin{pmatrix} 5 & 10 \\ 10 & 20 \end{pmatrix} = 5A$.

• Any matrix $A_{n\times n}$ satisfies its own characteristic equation

$$(A - \lambda_1)(A - \lambda_2)\cdots(A - \lambda_n) \equiv 0 \tag{176}$$

• Proof of the Cayley-Hamilton theorem. We will indicate the proof only for non-deficient matrices, i.e., those whose eigenvectors span the whole n-dimensional space. This is the case for matrices with n distinct eigenvalues.

• Essentially, we will show that every vector is annihilated by the matrix given by the characteristic polynomial $P(A) = (A - \lambda_1)(A - \lambda_2)\cdots(A - \lambda_n)$. It follows that $P(A)$ is the zero matrix.

Now $(A - \lambda_1)$, annihilates the first eigenvector x_1, $(A - \lambda_1)x_1 = 0$. Now consider $(A - \lambda_2)(A - \lambda_1)$, this matrix annihilates any linear combination of the eigenvectors x_1 and x_2 since the first factor annihilates x_2 and the second annihilates x_1 (the order of the factors does not matter, they commute). Continuing this way

$$P(A)(a_1x_1 + a_2x_2 + \cdots + a_nn_n) = 0. \tag{177}$$

But for a non-deficient matrix, the eigenvectors span the whole space, so $P(A)$ annihilates every vector and must be the zero matrix.

• The Cayley-Hamilton theorem states that a matrix satisfies an n^{th} order polynomial equation

$$A^n + c_{n-1}A^{n-1} + c_{n-2}A^{n-2} + \cdots + c_1 A + c_0 = 0. \tag{178}$$

In other words, we can express A^n in terms of lower powers of A. Similarly any power A^k with $k \geq n$, can be reduced to a linear combination of $I, A, A^2, \cdots, A^{n-1}$.

• Returning to the **example** $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, let us use the Cayley-Hamilton theorem to calculate A^{20}. Here the characteristic equation satisfied by A reads $A^2 = 5A$. This implies $A^3 = 5A^2 = 5^2 A$, $A^4 = 5^2 A^2 = 5^3 A$, $A^n = 5^{n-1} A$ for $n \geq 2$. Thus we have without having multiplied 20 matrices,

$$A^{20} = 5^{19} A = 5^{19} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \tag{179}$$

18.3 Diagonalization of matrices with n distinct eigenvalues

• If $A_{n\times n}$ is not deficient (as when it has n distinct eigenvalues), by a suitable invertible change of basis, we can bring it to diagonal form Λ with the diagonal entries of Λ given by the eigenvalues λ_i.

$$A = S\Lambda S^{-1} \quad \textit{or} \quad S^{-1}AS = \Lambda. \tag{180}$$

This process is called the diagonalization of the matrix. The invertible change of basis is called a general linear transformation S. If A is symmetric or hermitian, it turns out that the change of basis can be chosen to be an orthogonal or unitary transformation, which are special cases of general linear transformations.

• It is important to emphasize that the resulting diagonal matrix of eigenvalues Λ is in general different from the diagonal matrix D that might be obtainable through row elimination in the case when A has n (non-zero) pivots. The pivots are in general different from the eigenvalues. Row elimination involves left multiplication of A by elementary matrices while diagonalization involves left and right multiplication of A by S^{-1} and S.

• We can collect the n eigenvalues of A in the diagonal matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \tag{181}$$

• And collect the corresponding n eigenvectors x_i satisfying $Ax_i = \lambda_i x_i$ as the columns of a matrix S

$$S = \begin{pmatrix} | & | & \cdots & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & \cdots & | \end{pmatrix}. \tag{182}$$

Then notice that multiplying by columns

$$AS = \begin{pmatrix} Ax_1 & Ax_2 & \cdots & Ax_n \end{pmatrix}, \quad \text{and} \quad S\Lambda = \begin{pmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \end{pmatrix}. \tag{183}$$

Then the n solutions of the eigenvalue problem, may be summarized as

$$AS = S\Lambda. \tag{184}$$

Similarly we can consider the left eigenvalue problem for A, $y^t A = \mu y^t$ with row eigenvectors y^t. But taking the transpose, this is just the eigenvalue problem for the transpose $A^t y = \mu y$.

• But we know that the eigenvalues of A^t are the same as those of A, so we can write $A^t y_i = \lambda_i y_i$ for the n eigenvectors of A^T. The eigenvectors of A and A^t are in general different, but we will see that they are related. Let us collect the eigenvectors of A^t as the columns of a matrix $T = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$. Then

$$A^t\, T = T\Lambda \qquad \text{and} \qquad AS = S\Lambda. \tag{185}$$

Taking the transpose of these, we can calculate $T^t AS$ in two different ways to get

$$T^t AS = \Lambda T^t S \qquad \text{and} \qquad T^t AS = T^t S\Lambda. \tag{186}$$

Now let $W = T^t S$, then combining, we conclude that W commutes with Λ

$$\Lambda W = W\Lambda. \tag{187}$$

In other words,

$$\begin{pmatrix} 0 & (\lambda_1-\lambda_2)w_{12} & \cdots & (\lambda_1-\lambda_n)w_{1n} \\ (\lambda_2-\lambda_1)w_{21} & 0 & \cdots & (\lambda_2-\lambda_n)w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ (\lambda_n-\lambda_1)w_{n1} & (\lambda_n-\lambda_2)w_{n2} & \cdots & 0 \end{pmatrix} = 0. \tag{188}$$

Now since the λ's are distinct, we must have $w_{ij} = 0$ for $i \neq j$. Thus $W = T^t S$ is the diagonal matrix

$$W = \begin{pmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & w_{nn} \end{pmatrix}. \tag{189}$$

But $W = T^t S$ is merely the matrix of dot products of the eigenvectors of A^t and A, $w_{ij} = y_i^t x_j$. So we have shown that the left and right eigenvectors of A corresponding to distinct eigenvalues are orthogonal! We say that the x_i and y_j are in a bi-orthogonal relation to each other.

• But the normalization of the eigenvectors was arbitrary. By rescaling the $x_i \mapsto \frac{x_i}{w_{ii}}$ we can make W the identity matrix.

$$W = T^t S = I, \quad y_i^t x_j = \delta_{ij}. \tag{190}$$

Now we showed earlier that if A has distinct eigenvalues, its eigenvectors form a linearly independent set. So the columns of S are linearly independent and it is invertible. The same holds for T. So with this normalization, we find that $T^t = S^{-1}$. Putting this in the formula for $T^t AS =$ we get

$$S^{-1}AS = \Lambda \quad \text{or} \quad A = S\Lambda S^{-1}. \tag{191}$$

In other words, A may be diagonalized by the general linear transformation (similarity transformation) given by the invertible matrix S whose columns are the (appropriately normalized) eigenvectors of A!

• Now suppose $A^t = A$ is a symmetric matrix. Then there is no difference between left and right eigenvectors and $S = T$. But since $T^t S = I$, we must have $S^t S = I$ i.e., S is an orthogonal matrix. In other words, a symmetric matrix may be diagonalized by an orthogonal transformation. But the columns of an orthogonal matrix are orthonormal, so we conclude that the eigenvectors of a symmetric matrix may be chosen orthonormal. (Actually we have only proved this if the eigenvalues are distinct, though the result is true even if the symmetric matrix has repeated eigenvalues)

• Similarly, a hermitian matrix H may be diagonalized by a unitary transformation U whose columns are the eigenvectors of H. Moreover the eigenvectors are orthogonal and may be taken orthonormal by rescaling them

$$H = U\Lambda U^\dagger, \quad \text{with} \qquad U^\dagger U = I. \tag{192}$$

• More generally, a normal matrix is one that commutes with its adjoint, $A^\dagger A = AA^\dagger$ or $[A^\dagger, A] = 0$. Essentially the same proof as above can be used to show these two statements: If the eigenvectors of a matrix A with distinct eigenvalues are orthogonal, then A is a normal matrix. Conversely, the eigenvectors of a normal matrix with distinct eigenvalues may be taken orthonormal. In fact, more is true A may be diagonalized by a unitary transformation iff A is normal. Examples of normal matrices include but are not restricted to (anti)-symmetric, orthogonal, (anti)-hermitian and unitary matrices.

• A matrix A is diagonalizable if there is a basis where it is diagonal. In other words, it may be diagonalized by some similarity transformation S, i.e. $S^{-1}AS = \Lambda$, where Λ is the diagonal matrix with eigenvalues for the diagonal entries. The columns of S are then n linearly independent eigenvectors.

• If a matrix is diagonalizable, the basis in which it is diagonal is called the eigen-basis. The eigen-basis consists of n linearly independent eigenvectors. We have shown above that every matrix with n distinct eigenvalues is diagonalizable. Every hermitian or symmetric matrix is diagonalizable. For example $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ is diagonalizable. Find its eigenvalues and eigenvectors and the unitary transformation that diagonalizes it.

• Deficient matrices are not diagonalizable. Proof: Suppose a deficient matrix N were diagonalizable, $S^{-1}NS = \Lambda$. Then the columns of S would be n linearly independent eigenvectors of N. But a deficient matrix does not possess n linearly independent eigenvectors! Contradiction. Eg: The matrix $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is not diagonalizable. 0 is an eigenvalue with algebraic multiplicity 2 but geometric multiplicity one. N has only one eigenvector.

• **Simultaneous diagonalizability:** A pair of matrices $A, B : V \to V$ are said to be simultaneously diagonalizable if the same similarity transformation S diagonalizes them both i.e. $S^{-1}AS = \Lambda_A$ and $S^{-1}BS = \Lambda_B$. Here Λ_A and Λ_B are the diagonal matrices with eigenvalues of A and B along the diagonal respectively. Now the invertible matrix S contains the eigenvectors of A and B, so A and B share the same eigenvectors (though they may have different eigenvalues). Since S is invertible, the eigenvectors span the whole vector space V.

• If A and B are simultaneously diagonalizable, then they commute. $S^{-1}AS = \Lambda_A$ and $S^{-1}BS = \Lambda_B$. Now $[\Lambda_A, \Lambda_B] = 0$ as can be checked using the fact that they are diagonal. By the invariance of matrix equations under similarity transformations we conclude that $[A, B] = 0$. *If they commute in one basis, they commute in any other basis.*

• Sufficient **criterion for simultaneous diagonalizability**. Suppose A has n distinct eigenvalues and that a matrix B commutes with A, $[A, B] = 0$. Then B and A are simultaneously diagonalizable.

• Proof: Suppose x is an eigenvector of A with eigenvalue λ, $Ax = \lambda x$. Then we will show that x is also an eigenvector of B. Consider λBx, which can be written as $\lambda Bx = BAx = ABx$. So $A(Bx) = \lambda(Bx)$. x was already an eigenvector of A with eigenvalue λ. Now we found that Bx is also an eigenvector of A with eigenvalue λ. Since A has distinct eigenvalues, its eigenspaces are one dimensional and therefore Bx must be a multiple of x, i.e., $Bx = \mu x$. So we have shown that any eigenvector of A is also an eigenvector of B. Since the eigenvectors of A span the whole vector space we conclude that A and B have common eigenvectors and are simultaneously diagonalizable.

• Remark: We can replace the assumption that A have n distinct eigenvalues with some other hypotheses. For example we could assume that A and B both be hermitian and commuting. Then it is still true that they are simultaneously diagonalizable.

• Eg: Pauli matrices do not commute and they are not simultaneously diagonalizable. For example $[\sigma_2, \sigma_3] = i\sigma_1$ with $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Check that the unitary transformation that makes σ_2 diagonal forces σ_3 to become non-diagonal.

• Suppose A is invertible (in particular 0 is not an eigenvalue of A). Then eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A. This is why:

$$Ax = \lambda x \quad \Rightarrow \quad A^{-1}Ax = \lambda A^{-1}x \quad \Rightarrow \quad A^{-1}x = \frac{1}{\lambda}x \tag{193}$$

In fact, this shows that the eigenvector corresponding to the eigenvalue $\frac{1}{\lambda}$ of A^{-1} is the same as the eigenvector x of A corresponding to the eigenvalue λ. They have the same corresponding eigenvectors. In particular, if A was diagonalizable, then A^{-1} is diagonalizable simultaneously.

• Caution: An invertible matrix may not be diagonalizable. For example $N = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is invertible but not diagonalizable. It has only one linearly independent eigenvector, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ corresponding to the twice repeated eigenvalue $\lambda = 1$. $\lambda = 1$ has algebraic multiplicity two but geometric multiplicity only one. N is deficient. There is no basis in which N is diagonal.

18.4 Quadratic surfaces and principle axis transformation

• There is a geometric interpretation of the diagonalization of a symmetric matrix. It is called the principal axis transformation.

• In analytic geometry, the equation for an ellipse on the plane is usually given as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{194}$$

In this form, the major and minor axes are along the cartesian coordinate axes. Similarly, the equation of an ellipsoid embedded in 3d Euclidean space is often given as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{195}$$

Since it is defined by a quadratic equation, the ellipsoid is called a quadratic surface. The lhs involves terms that are purely quadratic in the variables. Such an expression (lhs) is called a quadratic form.

• More generally, an ellipsoid in n-dimensional space with axes along the cartesian coordinate axes is given by

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 = 1 \tag{196}$$

This can be regarded as a matrix equation $x^T \Lambda x = I$ for the column vector $x = (x_1, x_2, \cdots, x_n)^t$ and diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. $x^t \Lambda x$ is called the quadratic form associated to the matrix Λ.

However, often we are confronted with quadratic surfaces that are not aligned with the coordinate axes, but are in an arbitrarily rotated position. The equation for such a surface is again quadratic but with cross-terms of the form $x_i x_j$. For example

$$ax^2 + by^2 + cxy + dyx = 1 \tag{197}$$

But since $xy = yx$, only $c + d$ contributes, so we could have taken the coefficients of xy and yx to both equal $\frac{c+d}{2}$. More generally we have a quadratic equation

$$x_i A_{ij} x_j = 1 \quad \text{or} \quad x^t A x = 1 \tag{198}$$

where we may assume that $A_{ij} = A_{ji}$ is a real symmetric matrix.

• At each point P on the surface we have a normal direction to the surface, one that is normal (perpendicular) to the tangential plane to the surface through P.

• There is also the radius vector x of the point P.

• In general, the position vector and normal do not point along the same direction.

• The principal axes are defined as those radius vectors which point along the normal.

• In general, the normal to the surface at x points along Ax.

• To see this we first observe that if x lies on the surface, then a neighboring vector $x + \delta x$ also lies approximately on the surface if $(x + \delta x)^t A (x + \delta x) = 1$ up to terms quadratic in δx. In other words, $x^t A \delta x + \delta x^T A x = 0$, or $\delta x^t A x = 0$. Such δx are the tangent vectors to the surface at x. But this is just the statement that δx must be normal to Ax. So the normal vector must be along Ax.

• So the condition for x to be a principal axis is that it must be proportional to the normal Ax, or $Ax = \lambda x$, which is just the eigenvalue equation.

• Moreover, the eigenvalue has a geometric interpretation. Suppose x is a principal axis of A, then $x^t A x = \lambda x^t x = 1$ So $x^t x = \frac{1}{\lambda}$. But $x^t x$ is the square of the length of the position vector. So $\frac{1}{\lambda}$ is the square of the length of the semi-axis through P.

• Since A is symmetric, from the last section, we know that its eigenvectors are orthogonal. In other words, the principal axes are orthogonal. However, the principal axes may not point along the original cartesian coordinate axes. But if we take our new coordinate axes to point along the principal axes, then A is diagonal in this new basis. More precisely, A is diagonalized by an orthogonal transformation

$$Q^t A Q = \Lambda \tag{199}$$

where the columns of Q are the eigenvectors, $Q^t Q = I$ and Λ is the diagonal matrix of eigenvalues. So if we let $y = Q^t x$ then the equation of the surface $x^T A x = 1$ becomes $x^T Q \Lambda Q^T x = 1$ or simply $y^T \Lambda y$.

• In this geometric interpretation, we have implicitly assumed that the eigenvalues are real and that the eigenvectors are real vectors (for a real symmetric matrix). This is indeed true, as we will show in the next section.

• Finally, we point out the geometric meaning of coincidence of eigenvalues. Suppose $n = 2$, and suppose we have transformed to the principal axes. Then we have an ellipse $\lambda_1 x^2 + \lambda_2 y^2 = 1$ whose principal axes are along the x and y axes. Now if the eigenvalues gradually approach each other, $\lambda_1, \lambda_2 \to \lambda$ the ellipse turns into a circle. At the same time the diagonal matrix $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ tends to the multiple of the identity $\Lambda \to \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$. But every vector is an eigenvector of λI. In particular, we are free to pick any pair of orthogonal vectors and call them the principal axes of the circle.

• So when eigenvalues of a symmetric matrix coincide, the matrix does not become deficient in eigenvectors. It still possesses a system of n orthogonal eigenvectors, but some of them are no longer uniquely determined.

18.5 Spectrum of symmetric or hermitian matrices

• A real symmetric matrix is a real matrix $A : R^n \to R^n$ which equals its transpose $A = A^T$

• A hermitian matrix is a complex matrix $H : C^n \to C^n$ whose transpose is its complex conjugate: $(H^T)^* = H$, also written as $H^\dagger = H$.

• A special case is a real symmetric matrix. So every real symmetric matrix is also hermitian.

• Example: The Pauli matrix $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ is hermitian but not symmetric.

• The Pauli matrix $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is hermitian and symmetric.

• The diagonal matrix elements of H in any basis are real. In other words, let $z \in C^n$ be any vector, then $(z, Hz) = z^\dagger H z \in R$.

• To see this, take the complex conjugate of $z^T H z$, which is the same as the hermitian adjoint of the 1×1 matrix $z^\dagger H z$,

$$(z^\dagger H z)^* = (z^\dagger H z)^\dagger = z^\dagger H^\dagger z = z^\dagger H z \tag{200}$$

So $z^\dagger H z$ is a number that equals its own complex conjugate. So it must be real!

• In quantum mechanics, $\frac{(z,Hz)}{(z,z)}$ is called the normalized expectation value of H in the state z.

• Example: The three dimensional representation of angular momentum matrices in quantum mechanics are these hermitian matrices

$$L_x = \frac{1}{2}\begin{pmatrix} 0 & \sqrt{2} & 0 \\ \sqrt{2} & 0 & \sqrt{2} \\ 0 & \sqrt{2} & 0 \end{pmatrix}; \quad L_y = \frac{1}{2i}\begin{pmatrix} 0 & \sqrt{2} & 0 \\ -\sqrt{2} & 0 & \sqrt{2} \\ 0 & -\sqrt{2} & 0 \end{pmatrix}; \quad L_z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{201}$$

• The eigenvalues of a hermitian matrix are real. Suppose z is an eigenvector with eigenvalue λ, i.e., $Hz = \lambda z$. Taking the inner product with z,

$$(z, Hz) = z^\dagger H z = z^\dagger \lambda z = \lambda ||z||^2 \Rightarrow \lambda = \frac{(z, Hz)}{||z||^2} \tag{202}$$

$z^\dagger z = |z_1|^2 + \cdots |z_n|^2$ is real. Being the ratio of two real quantities, the eigenvalue λ is also real.

• Example check that the eigenvalues of σ_2 are real.

• Eigenvectors of a hermitian matrix corresponding to distinct eigenvalues are orthogonal. We have shown this previously but there is no harm in proving it again more directly.

• Proof: Suppose z, w are two eigenvectors, $Hz = \lambda z$ and $Hw = \mu w$, with eigenvalues $\lambda \neq \mu$, which are necessarily real. Then $w^\dagger H z = \lambda w^\dagger z$ and $z^\dagger H w = \mu z^\dagger w$. But the lhs are complex conjugates of each other, $(w^\dagger H z)^* = (w^\dagger H z)^\dagger = z^\dagger H w$. So $\lambda w^\dagger z = (\mu z^\dagger w)^*$. Or we have $w^\dagger z(\lambda - \mu) = 0$. By distinctness, $\lambda \neq \mu$, so $w^\dagger z = 0$ and w, z are orthogonal.

• Find the eigenvectors of σ_2 and show they are orthogonal.

• More generally, even if H has a repeated eigenvalue, we can still choose an orthogonal basis for the degenerate eigenspace so that eigenvectors of a hermitian matrix can be chosen orthogonal.

• Eigenvectors of a real symmetric matrix may be chosen real. This is important for the geometric interpretation of the eigenvectors as the principal axes of an ellipsoid. We will assume that the eigenvalues are distinct.

• Proof: We are given a real ($A^* = A$) symmetric matrix, so its eigenvalues are real. Suppose z is a possibly complex eigenvector corresponding to the eigenvalue $\lambda = \lambda^*$, i.e., $Az = \lambda z$. Taking the complex conjugate, $A^* z^* = \lambda^* z^*$ or $Az^* = \lambda z^*$, so z^* is also an eigenvector with the same eigenvalue. So $x = z + z^*$ is a real eigenvector with eigenvalue λ. So for every eigenvalue we have a real eigenvector. (Note: Moreover, the eigenspaces of A are one dimensional since we have n distinct eigenvalues and the corresponding eigenvectors must be orthogonal. So z and z^* are (possibly complex) scalar multiples of x.)

• Remark: Check that if H is hermitian, iH is anti-hermitian.

18.6 Spectrum of orthogonal and unitary matrices

• Orthogonal matrices are those real matrices that satisfy $Q^T Q = Q Q^T = I$.

• The columns of an orthogonal matrix are orthonormal.

• Unitary matrices are complex matrices satisfying $U^\dagger U = 1$. If a unitary matrix happens to be real, then it is necessarily orthogonal.

• The columns of a unitary matrix are orthonormal.

• A rather simple example of an orthogonal matrix is a *reflection* in the x axis, $Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

This happens to be diagonal. so the eigenvalues are $+1$ and -1, and the corresponding eigenvectors are the columns of the 2×2 identity matrix.

• Another example of an orthogonal matrix is the *rotation* matrix

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \det(A-\lambda I) = \det\begin{pmatrix} \cos\theta-\lambda & \sin\theta \\ -\sin\theta & \cos\theta-\lambda \end{pmatrix} = \lambda^2 - 2\lambda\cos\theta + 1 = 0 \quad (203)$$

The roots of the characteristic polynomial are $\lambda = \cos\theta \pm i\sin\theta = e^{\pm i\theta}$, which are generally complex, but lie on the unit circle.

• Eigenvalues of orthogonal and unitary matrices lie on the unit circle in the complex plane.

• This follows from the fact that orthogonal $Q^TQ = I$ and unitary $U^\dagger U = I$ matrices are isometries. They preserve the lengths of vectors: $||Qx|| = ||x||$ and $||Ux|| = ||x||$. So if we consider an eigenvector $Qv = \lambda v$, we have $||\lambda v|| = ||v||$ or $|\lambda|\ ||v|| = ||v||$, which implies $|\lambda| = 1$. The same works for unitary matrices as well.

• To see that orthogonal transformations are isometries starting from $Q^TQ = I$, consider $||Qx||^2 = (Qx)^TQx = x^TQ^TQx = x^Tx = ||x||^2$ since $Q^TQ = I$. Taking the positive square root, $||Qx|| = ||x||$ for all vectors x.

• Eigenvectors of unitary matrices corresponding to distinct eigenvalues are orthogonal.

• Proof: Suppose z, w are eigenvectors corresponding to distinct eigenvalues $\lambda \neq \mu$, $Uz = \lambda z$ and $Uw = \mu w$. Then we want to show that $z^\dagger w = 0$. So take the adjoint of the first equation $z^\dagger U^\dagger = \lambda^* z^\dagger$ and multiply it with the second and use $U^\dagger U = I$

$$z^\dagger U^\dagger U w = \lambda^*\mu z^\dagger w \quad \text{or} \quad (1-\lambda^*\mu) z^\dagger w = 0 \tag{204}$$

But since $\lambda^*\lambda = 1$ and $\lambda \neq \mu$ we have that $\lambda^*\mu \neq 1$. So the second factor must vanish, $z^\dagger w = 0$ and z and w are orthogonal.

• Remark: If H is hermitian, $U = e^{iH}$ is unitary.

18.7 Exponential and powers of a matrix through diagonalization

• Powers of a matrix are easily calculated once it is diagonalized. If $A = S\Lambda S^{-1}$, and $n = 0, 1, 2, \ldots$ is a positive integer

$$A^n = (S\Lambda S^{-1})^n = S\Lambda^n S^{-1} \tag{205}$$

Moreover, Λ^n is just the diagonal matrix with the n^{th} powers of the eigenvalues along its diagonal entries.

• Exponential of a matrix through diagonalization. If a matrix can be diagonalized by a similarity transformation $A = S\Lambda S^{-1}$, then calculating its exponential e^A is much simplified

$$e^A = e^{S\Lambda S^{-1}} = \sum_{n=0}^{\infty} \frac{(S\Lambda S^{-1})^n}{n!} = \sum_n \frac{S\Lambda^n S^{-1}}{n!} = Se^\Lambda S^{-1} \tag{206}$$

So we just apply the similarity transformation to e^Λ to get e^A. Moreover, since Λ is a diagonal matrix, its exponential is easy to calculate. If $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$, then

$$e^\Lambda = diag(e^{\lambda_1}, e^{\lambda_2}, e^{\lambda_3}, \cdots, e^{\lambda_n}) \tag{207}$$

18.8 Coupled oscillations via diagonalization

• Small displacements of a system about a point of stable equilibrium typically lead to small oscillations due to restoring forces. They are described by linearizing the equations of motion, assuming the departure from equilibrium is small. Hookes law for a slightly elongated spring is an example. If δx is the small displacement, Newton's law in Hooke's approximation says $m\ddot{\delta x} = -k\ \delta x$. This is a linear equation for one unknown function $\delta x(t)$.

• Similarly, suppose we have a pair of equally massive objects in one dimension connected by a spring to each other and also by springs to walls on either side in this order: wall spring mass spring mass spring wall. Let $\delta x_1, \delta x_2$ be small displacements of the masses to the right. Draw a diagram of this configuration. Newton's equations in Hooke's approximation (when the springs have the same spring constant k) are

$$\begin{aligned} m\ddot{\delta x}_1 &= -k\delta x_1 + k(\delta x_2 - \delta x_1) \\ m\ddot{\delta x}_2 &= -k\delta x_2 - k(\delta x_2 - \delta x_1) \end{aligned} \tag{208}$$

This is a pair of coupled differential equations; it is not easy to solve them as presented. But we can write them as a single matrix differential equation $\ddot{x} = Ax$ were $x = \begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix}$

$$\frac{d^2}{dt^2}\begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix} = \frac{k}{m}\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} \delta x_1 \\ \delta x_2 \end{pmatrix}. \tag{209}$$

Let $A = \frac{k}{m}\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$. The off-diagonal terms in A are responsible for the coupled nature of the equations. But A is real symmetric, so it can be diagonalized, which will make the equations uncoupled. Upon performing the principal axis transformation, $A = Q\Lambda Q^T$ where $\Lambda = \frac{k}{m}\begin{pmatrix} -1 & 0 \\ 0 & -3 \end{pmatrix}$ is the diagonal matrix of eigenvalues and Q is the orthogonal eigenvector matrix (which is independent of time, since A is). The equations become

$$\ddot{x} = Q\Lambda Q^T x \ \Rightarrow\ Q^T\ddot{x} = \Lambda Q^T x \tag{210}$$

So let $y = Q^T x$, then we get $\ddot{y} = \Lambda y$ which are the pair of uncoupled equations

$$\ddot{y}_1 = -(k/m)y_1, \qquad \ddot{y}_2 = -3(k/m)y_2 \tag{211}$$

If the initial condition was that the masses started from rest, then $\dot{y}(0) = 0$ and the solutions are

$$y_1(t) = y_1(0)\cos(\sqrt{\frac{k}{m}}t)y_2(t) = y_2(0)\cos(\sqrt{\frac{3k}{m}}t) \tag{212}$$

The method of solving these differential equations will be treated in the second part of this course. To get back $x(t)$ we just use $x(t) = Qy(t)$. So it only remains to find the eigenvector matrix Q, of A, which is left as an exercise.

19 Hilbert spaces and Dirac bra-ket notation

• A finite dimensional Hilbert space $\mathcal{H}$ is a finite dimensional vector space with an inner product (u,v) that is linear in v and anti-linear in u satisfying

$$(u,v) = (v,u)^* \quad \text{and} \quad (u,u) > 0, \quad \text{for} \quad u \neq 0 \tag{213}$$

• We will work with the example C^n with the standard inner product $(z,w) = z^\dagger w$. This is a Hilbert space. Notice that $(z,w) = (w,z)^*$. Moreover, for scalars a,b, $(az,w) = \bar{a}(z,w)$ while $(z,bw) = b(z,w)$. Finally, $(z,w+u) = (z,w) + (z,u)$. These properties ensure linearity in the second entry and anti-linearity in the first.

• Dirac notation: If we think of $V = C^n$ as made of column vectors, we denote the column vector v as the ket-vector $|v\rangle$. The space of ket-vectors form the vector space V. Similarly the n-component row vectors with complex entries are called the bra-vectors $\langle v|$. Moreover, $\langle v| = |v\rangle^\dagger$ and $\langle v|^\dagger = |v\rangle$ are adjoints of each other. For example

$$|v\rangle = \left(\begin{array}{c} 1 \\ i \\ -2i+3 \end{array} \right), \quad \langle v| = |v\rangle^\dagger = \left(1 \quad -i \quad 2i+3 \right) \tag{214}$$

The space of bra-vectors form a so-called dual space V^* to V. V and V^* are isomorphic vector spaces. Indeed any row vector $\langle w|$ defines a linear function on V, given by

$$f_{\langle w|}(|v\rangle) = \langle w|v\rangle \tag{215}$$

The dual space V^* is defined as the space of linear functions on V. $\langle w|v\rangle$ is called the pairing between the dual spaces.

• If $|v\rangle = \sum_i v_i |\phi_i\rangle$ is expressed as a linear combination of $|\phi_i\rangle$, then $\langle v| = |v\rangle^\dagger = \sum_i \langle \phi_i| v^*$.

• If e_i are a basis, $v \in \mathcal{H}$ a vector and $A : \mathcal{H} \to \mathcal{H}$ a linear transformation, then we can write $Ae_j = \sum_i A_{ij} e_i$ and $v = \sum_j v_j e_j$ and $Av = \sum_j v_j A e_j = \sum_{ij} v_j A_{ij} e_i = \sum_{ij} (A_{ij} v_j) e_i$. In other words $(Av)_i = A_{ij} v_j$. Now let us assume that e_i are an orthonormal basis, so $\langle e_i | e_j \rangle = \delta_{ij}$. Then we have

$$A|e_j\rangle = A_{ij}|e_i\rangle \quad \Rightarrow \quad \langle e_k|A|e_j\rangle = \sum_i A_{ij} \langle e_k|e_i\rangle = \sum_i A_{ij}\delta_{ki} = A_{kj} \tag{216}$$

We conclude that $A_{ij} = \langle e_i|e_j\rangle$ in any orthonormal basis $\{e_i\}$.

• Similarly, in an orthonormal basis, $|v\rangle = \sum_i v_i |e_i\rangle$ implies that $v_j = \langle e_j|v\rangle$.

• In a finite dimensional Hilbert space, we have seen that any vector can be decomposed in an o.n. basis as $|v\rangle = \sum_i \langle e_i|v\rangle |e_i\rangle$ or rearranging, $|v\rangle = \sum_i |e_i\rangle\langle e_i|v\rangle$. So we see that the linear transformation $\sum_i |e_i\rangle\langle e_i|$ takes every vector to itself, in other words, it must be the identity transformation, which is represented by the identity matrix in any basis. So

$$\sum_i |e_i\rangle\langle e_i| = I \tag{217}$$

This is called the completeness relation or property. We see that it is the sum of outer products of the orthonormal basis vectors $e_i e_i^\dagger = I$. It says that the sum of the projections to the one-dimensional subspaces spanned by the o.n. basis vectors e_i is the identity. We say that e_i are a complete o.n. basis.

• For example, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ form a complete o.n. basis for R^2. One checks that the completeness relation is satisfied

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{218}$$

• For a finite dimensional Hilbert space, every o.n. basis is complete.

• More precisely, a sequence of vectors $u_i \in \mathcal{H}$ is complete if there is no non-zero vector in $\mathcal{H}$ that is orthogonal to all of them.

• Similarly, for the bra-vectors, completeness of the o.n. basis e_i allows us to write

$$\langle v| = \sum_i \langle v|e_i\rangle\langle e_i| = \sum_i v_i^* \langle e_i| \tag{219}$$

• Let is see some more uses of the completeness relation of an orthonormal basis

$$\langle v|w\rangle = \sum_i \langle v|e_i\rangle\langle e_i|w\rangle = \sum_i v_i^* w_i \tag{220}$$

We say that we have inserted the identity between $\langle v|$ and $|w\rangle$.

• $\langle v|v\rangle = ||v||^2 = \sum_i \langle v|e_i\rangle\langle e_i|v\rangle = \sum_i \langle v|e_i\rangle\langle v|e_i\rangle^* = \sum_i |\langle v|e_i\rangle|^2$. This expresses the norm^2 of v as the sum of the absolute squares of its components in a complete o.n. basis.

• Note that for brevity, sometimes the basis-kets are denoted $|i\rangle$ instead of $|e_i\rangle$.

• We can recover the formula for matrix multiplication: $(AB)_{ij} = \langle i|AB|j\rangle = \sum_k \langle i|A|k\rangle\langle k|B|j\rangle = \sum_k A_{ik}B_{kj}$.

• The completeness relation says $I = \sum_i |i\rangle\langle i| = \sum_i P_i$ where $P_i = |i\rangle\langle i|$ (no sum on i) is the projection to the subspace spanned by $|i\rangle$.

• $P_i P_j = |i\rangle\langle i|j\rangle\langle j| = |i\rangle\delta_{ij}\langle j| = \delta_{ij} P_j$ (no sum on i or j). This says for instance that projections to orthogonal subspaces is zero $P_1 P_2 = 0$ while $P_1 P_1 = P_1$.

• A hermitian matrix $H^\dagger = H$ is also called self-adjoint. $(H^\dagger)_{ij} = H_{ij}$ can be written as $\langle i|H^\dagger|j\rangle = \langle i|H|j\rangle$. Now notice that $\langle j|H|i\rangle^* = \langle j|H|i\rangle^\dagger = \langle i|H^\dagger|j\rangle$. So the condition of hermiticity can be expressed

$$\langle i|H|j\rangle = \langle j|H|i\rangle^* \tag{221}$$

19.1 Function spaces and Hilbert spaces

• An important example of a vector space is the set of functions, say real-valued, on an interval $f : [0,1] \to R$. We can add such functions and multiply them by real constants to get back such functions $(f+g)(x) = f(x)+g(x)$ and $(cf)(x) = cf(x)$. Being closed under linear combinations, they form a vector space called a function space $\mathcal{F}$. These formulae are to be compared with $(v+w)_i = v_i + w_i$ and $(cv)_i = cv_i$ for vectors in R^n. So the value of a function f at the point x is analogous to the i^{th} component of the vector v.

• For example the functions $f(x) = 23$ and $g(x) = x$ are linearly independent elements of $\mathcal{F}$. This is because the only way in which $af + bg = 0$, i.e. $a\,23 + b\,x = 0$ for all $x \in [0,1]$ is for $a = b = 0$

• On the other hand, $f = x$ and $g = 2x$ are linearly dependent since $2f + g = 0$ for every $x \in [0,1]$.

• In this manner we see that the functions $1, x, x^2, x^3, \cdots$ are linearly independent elements of $\mathcal{F}$. So this function space is infinite dimensional!

• One can regard this infinite dimensional vector space as obtained by a limiting procedure applied to n-dimensional vector spaces of increasing dimension. These finite dimensional vector spaces could consist of vectors whose components f_i are the values $f(x_i)$ on an increasingly finer mesh of points lying in the interval $[0,1]$, such as $x_i = i/(n-1)$ for $i = 0, \cdots, n-1$.

• Another example of an infinite dimensional vector space is the space of complex-valued functions $f(x) : [0,\pi] \to \mathbf{C}$ satisfying $f(0) = f(\pi) = 0$. This is a vector space, since it is closed under linear combinations. We can turn it into a Hilbert space by specifying an inner product such as

$$(f, g) = \int dx \; \overline{f(x)} \; g(x) \tag{222}$$

This formula is to be compared with $(v, w) = \sum_i \overline{v_i} w_i$. Sum over i is replaced with integration. However, we have to be careful to admit only those functions for which these integrals are finite. Functions for which $\int |f(x)|^2 \, dx < \infty$ are called square integrable. Also, we need to admit all such functions while also preserving closure under linear combinations. This is a subtle issue in an infinite dimensional Hilbert space and beyond the scope of this course. We merely mention the definition of a (possibly infinite dimensional) Hilbert space. A vector space with an inner product is called an inner product space.

• Definition: A Hilbert space $\mathcal{H}$ (possibly infinite dimensional) is defined as a complete inner product space. Completeness means that every Cauchy sequence of vectors in $\mathcal{H}$ converges to a vector in $\mathcal{H}$. A sequence of vectors $f_1, f_2, f_3, \cdots$ is called a Cauchy sequence if $||f_m - f_n|| \to 0$ if $m, n \to \infty$. This condition ensures that $\mathcal{H}$ is not missing any vectors that the process of taking linear combinations may generate.

• A sequence of vectors $\{f_i\}_{i=1}^{\infty}$ in $\mathcal{H}$ is orthonormal if $(f_i, f_j) = \delta_{ij}$. An o.n. sequence is called complete if there is no non-zero vector in $\mathcal{H}$ that is orthogonal to all the f_i. A complete orthonormal sequence plays the same role as an o.n. basis does in a finite dimensional Hilbert space. This is because of the following theorem

• $\{f_i\}$ is a complete o.n. sequence in $\mathcal{H}$ iff any one of the following are satisfied

1. $f = \sum_{i=1}^{\infty} (\phi_i, f) f_i$ for every $f \in \mathcal{H}$
2. $(f, g) = \sum_{i=1}^{\infty} (f, f_i)(f_i, g)$ for all $f, g \in \mathcal{H}$
3. $||f||^2 = \sum_{i=1}^{\infty} |(f, \phi_i)|^2$ for all $f \in \mathcal{H}$

• Example: Let us return to the Hilbert space of square integrable functions on $[0,\pi]$ vanishing at $0, \pi$. Consider the sequence of vectors in $\mathcal{H}$ given by $f_n(x) = \sqrt{\frac{2}{\pi}} \sin(mx)$, for $m = 1, 2, 3, \ldots$. We see that the f_n satisfy the condition of vanishing at $0, \pi$ and are also square integrable as they are bounded functions. Moreover, a trigonometric calculation shows that $\int_0^{\pi} \; dx \; f_n^*(x) f_m(x) = \delta_{n,m}$. Thus f_n is an orthonormal sequence. The factor of $\sqrt{\frac{2}{\pi}}$ is needed to ensure that $||f_n|| = 1$, since the average value of the square of the sine function over a period is one half.