Lineare Algebra

Contents

https://students.aiu.edu/submissions/profiles/resources/onlineBook/Y5B7M4_Introduction_to_Linear_Algebra-_Fourth_Edition.pdf

Übungsstunde Notizen:
- https://n.ethz.ch/~jamatter/
- https://www.felixgbreuer.com/linalg

Notation:
LGS - Lineare Gleichung System - linear system of equations
$\mathcal{A}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ - the corresponding Abbildung von the matrix A. Matrix multiplication is often referred to as a function

Vector Operations

Euclidean Norm

$$\boldsymbol{x} \in \mathbb{R}$$
$$|\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$$
$$\boldsymbol{x} \in \mathbb{C}$$
$$|\boldsymbol{x}| = \sqrt{|x_1|^2 + |x_2|^2 + ... + |x_n|^2}$$
$$\text{Where } |x_1| = \sqrt{\text{re}(x_1)^2 + \text{im}(x_1)^2}$$

Dot / skalar Product

$$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$
$$< \boldsymbol{x}, \boldsymbol{y} >= x_1 y_1 + x_2 y_2 + ... + x_n y_n = \boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{y}$$
$$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$$
$$< \boldsymbol{x}, \boldsymbol{y} >= \overline{x_1} y_1 + \overline{x_2} y_2 + ... + \overline{x_n} y_n = \boldsymbol{x}^{\boldsymbol{H}} \boldsymbol{y}$$

The dot product can be used to calculate the angle between two vectors:

$$< x, y >= |x||y| \cos\left(\widehat{(x, y)}\right)$$
$$\widehat{(x, y)} = \arccos \frac{< x, y >}{|x||y|}$$
$$x \perp y \text{ is Orthogonal} \Leftrightarrow < x, y >= 0$$

$\frac{<x,y>}{|x||y|} \in [-1, 1]$ is proven by the Cauchy–Schwarz inequality.

Thus it can also be used with a unit vector to calculate the projection of a vector in its direction:

$$|u| = 1$$
$$< x, u >= |x| 1 \frac{x_u}{|x|}$$

Cross / Vektor product

The cross product is only defined in $\mathbb{R}^3$ and results in a vector orthogonal to both inputs, whose direction can be determined with the right hand rule.

$$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$$
$$\boldsymbol{x} \times \boldsymbol{y} = \begin{vmatrix} \boldsymbol{e^x} & \boldsymbol{e^y} & \boldsymbol{e^z} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

It can also be used for determining the angle between two vectors in $\mathbb{R}^3$:

$$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$$
$$\boldsymbol{x} \times \boldsymbol{y} = |\boldsymbol{x}||\boldsymbol{y}| \boldsymbol{n} \sin\theta$$

($\boldsymbol{n}$ is the unit vector normal to the plane spanned by $\boldsymbol{x}, \boldsymbol{y}$)

Fundamentals

Vectors

Lineare kombination - Summe von skalierten Vektoren
Linearly dependent - When two vectors can be expressed as a linear combination of the other and thus doesn't add any information to a LGS. *Basis* - the set of linearly independent vectors $e_1 ... e_n$ that span all of space R^n

Vektoren werden immer als Spalten in diesem Kurs gezeichnet.

Matrix multiplication comes from the motivation for an efficient way of representing linear combinations / transformations of space.

Geometry of an LGS

An LGS can be viewed geometrically (2D/3D) in multiple different ways:

1. A linear combination of vectors (the columns of the matrix), where we are solving for the set of scalars where the superposition of the vectors is equal to the RHS. The columns of the matrix can be viewed as basis vectors of a custom coordinate system, in which we need to find the equivalent of the RHS vector.
2. Alternatively it can be viewed as a set of line / plane equations (where each row is the normal vector to the plane, unsure if the coefficients are meaningful in ax + by=c) and solutions are points / lines of intersection.
3. The LHS can also be viewed (usually in 3B1B videos) as a linear transformation of space, where the columns of the matrix are where the basis vector of each dimension lands after the transformation. The solution is therefore the vector which after being transformed results in the RHS vector.

NOTE: I will mostly think in terms of the *linear transformation of space* intuition, because the others are not very meaningful when considering inverse matrices.

Superposition

In this example, one of the LHS vectors is a linear combination of the other two. This results in the LGS only being able to express vectors in a single plane rather than the entire 3D space (it doesn't contain a 3rd base component).

Infinite solutions - if the RHS vector lays in the plane expressed by $\boldsymbol{a_{1-3}}$, any point in the positive / negative direction of the solution vector lays in the plane.
No solutions - the vector does not lay perfectly on the plane, the LHS vectors lack a component (not necessarily base unit vector) in its direction.

Line / Plane equations

The solution is the point at which the lines / planes represented by the horizontal equations intersect. There are many possible arrangements which we can visualise, especially in $3D$ space.

Unique solution - Common point of intersection of n non parallel lines / planes.
Infinite solutions - Sheaf of planes or if all lines are the same.
No solution - Not all lines / planes meet at a common point, which is more likely the more equations are introduced into the system. Examples: Parallel lines, trianglular prism from 3 planes.

Gaussian Elimination

Method for solving a $m \times n$ system of equations, easy to implement algorithmically and works for all dimensions.

Pivot - element on the diagonal of a matrix that has a non 0 coefficient

Rang / rank - number of non 0 pivots, ie (number of rows - number of Kompatibilitaetsbedingungen) - the number of linearly independent rows / columns - the number of dimensions of the output of a linear transformation.

Row Rank = Column Rank:

$$\text{Rank}(\boldsymbol{A}) = \text{Rank}(\boldsymbol{A}^{\boldsymbol{T}})$$

Kompatibilitaetsbedingungen - Empty rows at the bottom of the matrix (0 coefficients in one of the equations). If their result is not 0 then there are no solutions for the system. If their result is 0 and the number of equations $\leq$ the number of variables, there are infinite solutions.
Intuition: When thinking of the LGS as superposition, each LHS vector has a 0 component in this dimension, meaning that $\forall x \in \mathbb{R}$ scalar in the Lineare Kombination satisfies the system. Viewing the system with insufficient equations as a system of planes, two planes will intersect along an entire line. In 2D, there would just be a single line, which of course has solutions along its entirety.

Free Variables - Any variables not accounted for due to no pivot in their column are called *free variables.* These can be thought of as degrees of freedom, we are free to give them any arbitrary value and the other variables for that specific solution in the linear combination then depend on these.

"Order is half of the work in maths." - *Vasile Gradinaru*

Tips:
- Never divide / subtract in Gaussian elimination. Either multiply by $\frac{1}{x}$ or -1.
- Switch rows columns carefully **before** carrying out additions.
- **Only** add the row who's pivot is currently being considered! Otherwise it is difficult to capture the operation in the elimination matrix (more on this later).
- When switching rows to get pivots in the correct place, it is usually best to swap a line with zero pivot with the row that has the largest pivot in that place.

U - Upper (Deutsch: R - Rechts) Matrix - Matrix with 0s under the diagonal and any numbers above it
L - Lower Matrix - Matrix with 0s above the diagonal and any numbers below it
Identity Matrix - Matrix with 0s above and below the diagonal, which only contains 1s
Tridiagonal Matrix - Matrix with 3 diagonals, and otherwise 0s everywhere

Homogene LGS - $\boldsymbol{Ax} = 0$ hat eine triviale Loesung $\boldsymbol{x} = 0$, unless it has free variables.

Square Matrices ($m \times n$):
This only applies to square matrices

Regular Matrix, Rank = n, has exactly one solution for arbitrary RHS and only the trivial solution when homogenous
Singular Matrix (Single / peculiar), Rank < n, has infinite / no solutions and has infinite non trivial solutions when homogenous

$m > n$ - An overdetermined LGS only has solutions for specific RHS values (if the rows are not linearly dependent) and therefore has no inverse (singular).

Matrices

Transposed Matrix
For a matrix with notation:

$$i := 1, ..., m \text{ Zeilen}$$
$$j := 1, ..., n \text{ Spalten}$$
$$\boldsymbol{A} = \left[a_{ij}\right]$$
$$\text{Example: } \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

The transposed matrix A^T is:

$$\boldsymbol{A^T} = \left[a_{ji}\right]$$
$$\text{Example: } \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

This can be thought of as pinning the first elements of each row and letting the rest of the row swing down vertically.

Hermitian matrix - $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, $\boldsymbol{A^H}$ - The same procedure however each element becomes its complex conjugate $\overline{a}$, written as $\boldsymbol{A}^\dagger$ in physics.

LTD: Investigate thoroughly https://en.wikipedia.org/wiki/Hermitian_adjoint

Vectors may be treated like $\mathbb{R}^{n \times 1}, \mathbb{C}^{n \times 1}$ matrices and transposed in the same manner.

Matrix addition / scalar multiplication is carried out in the same way as vectors.

$$(\boldsymbol{AB})^{\boldsymbol{T}} = \boldsymbol{B^T}\boldsymbol{A^T}$$
$$(\boldsymbol{A} + \boldsymbol{B})^{\boldsymbol{T}} = \boldsymbol{A^T} + \boldsymbol{B^T}$$
$$(\boldsymbol{A} + \boldsymbol{B})^{\boldsymbol{H}} = \boldsymbol{A^H} + \boldsymbol{B^H}$$

The rank of a matrix is the same as its tranpose:

$$\text{Rank}(\boldsymbol{A}) = \text{Rank}(\boldsymbol{A^T})$$

Matrix Symmetry
Symmetrical - $\boldsymbol{A^T} = \boldsymbol{A}$
Antisymmetrical - $\boldsymbol{A^T} = -\boldsymbol{A}$
Hermetian Symmetry - $\boldsymbol{A^H} = \boldsymbol{A}$

Matrix Multiplication
Can be thought as the combination of transformations of space.

Two matrices may only be multiplied if they have the same inner dimensions:

$$\boldsymbol{A}_{\boldsymbol{X} \times \boldsymbol{Y}} \times \boldsymbol{B}_{\boldsymbol{Y} \times \boldsymbol{Z}} = \boldsymbol{C}_{\boldsymbol{X} \times \boldsymbol{Z}}$$

Several LGS with the same LHS can be solved simultaneously with matrix multiplication:

$$\boldsymbol{X} = \left[\overrightarrow{X_1}, ..., \overrightarrow{X_n}\right], \boldsymbol{B} = \left[\overrightarrow{B_1}, ..., \overrightarrow{B_n}\right]$$
$$\boldsymbol{A^{-1}X} = \boldsymbol{B}$$

$\text{Rank}(\boldsymbol{AX}) = \min(\text{Rank}(\boldsymbol{A}), \text{Rank}(\boldsymbol{X}))$

Matrix multiplication is usually not commutative, however always associative and distributive.

Inverse

The inverse of a matrix $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^{-1}$, which reverses the transformation of space represented by matrix $\boldsymbol{A}$. Therefore $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$.

The inverse can be used to solve a LGS for arbitrary RHS vectors.

Regulaer, invertierbar und voller Rang sind synonyme dafuer, dass eine Matrix einen Inverse hat. Therefore here are some equivalent conditions which show that a matrix $\boldsymbol{A}$ is regular:
- $\boldsymbol{A}$ is invertierbar
- $\text{Rang}(\boldsymbol{A}) = n$
- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is solvable for any $\boldsymbol{b}$
- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ only has the trivial solution $x = 0$

Identities:

$$(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$
$$(\boldsymbol{A}^{\boldsymbol{T}})^{-1} = (\boldsymbol{A}^{-1})^{\boldsymbol{T}}$$

Eliminationsmatrix (aka Protokolmatrix)

Matrix used for tracking the process of Gaussian elimination. The LHS / RHS multiplied by the elimination matrix results in the current state of the elimination!

It starts as the identity matrix, then the scalar by which another row was multiplied by before adding is written in the position of the currently eliminated variable of the row it was added to.

Important: Keep the elimination matrix lower! This means that for the current column, only the current row with 1 in the diagonal may be added to other rows. If this doesn't work, use a permutation.

Caution: when swapping rows, do NOT forget adjusting the Elimination Matrix accordingly, by simply swapping all non diagonal values in the rows (this is done in a mathematical manner with Permutation matrices later).

Properties of elimination matrices:
- The inverse of the elimination matrix is itself, but non diagonal values become negative. This makes sense intuitively, as $\boldsymbol{E}\boldsymbol{E}^{-1} = \boldsymbol{I}$ so for $\boldsymbol{E}_{\boldsymbol{ij}} + \boldsymbol{E}_{\boldsymbol{ij}}^{-1} = \boldsymbol{0}$ they must have opposite polarities.
- Two lower elimination matrices (with no overlapping elements!) multiplied together is the identity matrix with the combination of both lower elements. This means we can chain steps of Gaussian elimination together nicely.

Permutation Matrix

Orthogonal Matrix used to track the permutation of rows in LU-Zerlegung. This is simply the identity matrix with the corresponding rows swapped.

Properties of permutation matrices:

$$P_{13} := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$
$$P^{-1} = P^T$$
$$\text{Row permutation: } P_{13} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 7 & 8 & 9 \\ 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix}$$
$$\text{Column permutation: } \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} P_{13} = \begin{pmatrix} 3 & 2 & 1 \\ 6 & 5 & 4 \\ 9 & 8 & 7 \end{pmatrix}$$

Calculating the Inverse

The inverse can be calculated through Gaussian elimination (full Gaussian elimination, ie. with back substitution already carried out so the LHS matrix is the identity matrix) with a RHS of $b = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}$ and then finding which X results in $X^{-1}b =$ our eliminated original matrix (by simply reading the coefficients of each component of b).

This can be simplified as the so-called **Gauss-Jordan Elimination**. This can be described as the following transformation through regular Gaussian elimination. All operations happen on both sides in both matrices, unlike LU decomposition.

$$[A \mid I] \rightsquigarrow [I \mid A^{-1}]$$
$$\left(\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{4} \end{array}\right)$$

The elimination matrix can be used with any b to apply the steps of elimination, for example when the first row was multiplied by 2 and added to the 2nd row:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ 2b_1 + b_2 \\ b_3 \end{pmatrix}$$

Cool Fact: Just before each row was multiplied to make the pivots 1, the pivots of the LHS multiplied together is equal to the determinant, which is how computers calculate it for very large matrices. This can also be related to the fact, that a matrix is only invertable if its determinant is non 0 (therefore there are no empty rows / pivots).

LU Lower Upper (LR Left Right) Zerlegung

A matrix can be decomposed into an upper and lower matrix, such that:

$$A = LU$$
$$PA = LU$$

This can be used to decouple the factorization phase from the actual solving phase in Gaussian elimination. When the number of RHS we need to solve for is relatively small and the A is extremely large, it is more efficient to carry out LU Zerlegung and the additional steps to solve each system separately rather than to calculate the inverse through Gauss-Jordan elimination.

Proof of $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$:
For steps $1, 2, 3, ..., n$ of Gaussian elimination, the end result of the LHS is an upper matrix.

$$\boldsymbol{E}_{\boldsymbol{n}}\boldsymbol{E}_{\boldsymbol{n}-\mathbf{1}}\boldsymbol{E}_{\boldsymbol{n}-\mathbf{2}}...\boldsymbol{E}_{\mathbf{1}}\boldsymbol{A} = \boldsymbol{U}$$

The chain of steps (elimination matrices) can be moved to the RHS by multiplying both sides by their inverses due to $\boldsymbol{E}\boldsymbol{E}^{-\mathbf{1}} = \boldsymbol{I}$:

$$\boldsymbol{A} = \boldsymbol{E}_{\boldsymbol{n}}^{-\mathbf{1}}\boldsymbol{E}_{\boldsymbol{n}-\mathbf{1}}^{-\mathbf{1}}\boldsymbol{E}_{\boldsymbol{n}-\mathbf{2}}^{-\mathbf{1}}...\boldsymbol{E}_{\mathbf{1}}^{-\mathbf{1}}\boldsymbol{U}$$

The chain of elimination steps can of course be represented as a single lower matrix $\boldsymbol{E}$, therefore:

$$\boldsymbol{A} = \boldsymbol{E}^{-\mathbf{1}}\boldsymbol{U} = \boldsymbol{L}\boldsymbol{U}$$

Where E^{-1} is very easy to find (non diagonal elements simply $\times -1$ as mentioned earlier).

Using a combination of Elimination matrices and row + column permutations (these are needed to preserve the diagonal 1s of the resulting elimination matrix), the entire Gaussian elimination process can be encoded as one L matrix.

This is very powerful as the inverses of E and P matrices are easy to find and apply in reverse to the RHS in order to solve the LGS.

Using the LU Decomposition
The decomposed system can then be used to solve for $\boldsymbol{x}$ in the following way:
1. $Ax = b$
2. Decompose into the form $PA = LU$
3. Replacing A as PLU, $Lc = Pb$. c can be solved easily thanks to the form of L
4. Based on the above rearrangement, $c = Ux$, in which x can also be solved easily with backsubstitution

LTD: Other decomposition methods (Cholesky etc)

Orthogonale Matrizen

Orthogonal matrix - A square matrix whose columns are perpendicular to each other (dot product 0) and their Euclidean Norms are 1. They do not change lengths or angles - ie they only rotate / reflect space.

In other words, the columns are rows of an orthogonal matrix are orthonormal to each other (see Gram-Schmidt). TODO: Learn how to link to reference typst

The inverse of a rotation / reflection of space is logically its transposition (consider the rotation of base vectors to different axis):

$$\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{Q} = \boldsymbol{I} : \text{Q is Orthogonal}$$
$$\boldsymbol{Q}^{\boldsymbol{H}}\boldsymbol{Q} = \boldsymbol{I} : \text{Q is Unitary}$$
$$\text{Sei P Orthogonal}$$
$$\boldsymbol{Q}\boldsymbol{P} \text{ und } \boldsymbol{P}\boldsymbol{Q} \text{ sind auch Orthogonal}$$

An alternative intuition for this is considering matrix multiplication with the Hermetian as dot products between columns / rows, showing orthonormality (1 dot product pair with length 1, the others $0 \Rightarrow$ orthogonal).

Orthogonal matrix multiplication is still generally not commutative.

LTD: Investigate unitary vs orthogonal applications

Properties of orthogonal matrices

Let $\boldsymbol{Q}$ be an orthogonal matrix:

- Preserves lengths:

$$< \boldsymbol{Qx}, \boldsymbol{Qy} >=< \boldsymbol{x}, \boldsymbol{y} >$$
$$|\boldsymbol{Qx}| = |\boldsymbol{x}|$$

- Preserves angles $\widehat{\boldsymbol{Qx}, \boldsymbol{Qy}} = \widehat{\boldsymbol{x}, \boldsymbol{y}}$
- Are always regular ⇔ columns are linearly independent

Rotation Matrix

Matrix that rotates space by α degrees anticlockwise:[1]

$$R(\alpha) = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

Due to the identity $\boldsymbol{Q^{-1}} = \boldsymbol{Q^T}$ rotation anti-clockwise is:

$$R(\alpha)^T = R(-\alpha) = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

For a rotation in a certain plane of $\mathbb{R}^n$, simply keep all columns and rows unaffected by the rotation as the identity matrix. For example a anti-clockwise rotation in the $x \times y$ plane of $\mathbb{R}^3$:

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

$x \times z$ plane:

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

and so on in higher dimensions...

Givens Rotation $G(\varphi) = R(-\varphi)$ - Simply a rotation by the angle φ in the anti-clockwise direction

Rotations which are not confined to a plane can be achieved through a series of plane rotations multiplied together, still resulting in an orthogonal matrix.

Reflection (Householder) Matrix

The orthogonal Householder matrix Q represents the reflection of space (a vector $\boldsymbol{x}$ which does not lie on the plane) over an arbitrary plane with normal **unit** vector $\boldsymbol{u}$:

$$|\boldsymbol{u}| = 1$$
$$\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{2uu^T}$$

"Fuer die Computer sind alle Zahlen schön" - *Vasile Gradinaru*

QR Zerlegung (QU Zerlegung)

QR Decomposition is a different approach to solving LGSs, where the matrix Q is orthogonal.

[1] Derivation through trigonometry in script by Dr. V Gradinaru

Likewise to LU decomposition, it can be used to solve LGSs:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \Leftrightarrow \boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{b} \Leftrightarrow \boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{b}$$

Advantage:

- Reduced rounding errors due to the way computers represent floating point numbers (fractions)

Disadvantage:

- 3 times as inefficient as LU-Zerlegung

The goal is to find a series of orthogonal transformations, which transform $\boldsymbol{A}$ into an upper matrix when multiplying it. Since orthogonal matrices multiplied together result in an orthogonal matrix, they can easily be combined into one and used to solve the LGS as described above.

Givens rotations

Considering a vector, for example $\binom{x_1}{x_2}$. We can transform one of its components into 0 by rotating it onto an axis, resulting in $\binom{r}{0}$ the following applies:

$$\begin{aligned} r &= \sqrt{x_1^2 + x_2^2} \\ \cos(\theta) &= \frac{x_1}{r} \\ \sin(\theta) &= \frac{x_2}{r} \end{aligned}$$

This works in $\mathbb{R}^n$ dimensions. TODO: Complex as well?

Using these formulas with a Givens (clockwise) rotation, we can simply replace cos and sin in the matrix with the fractions, targeting the row we don't mind changing and the element we currently want to transform into 0.

By applying a series of Givens rotations, we can carry out the full Gaussian elimination, ending up with an Upper / Right matrix.

Householder reflection

The goal is to find a Householder matrix that reflects a column of A onto the axis of the row we are **not** trying to eliminate. As we know the resulting vector, for example $|\boldsymbol{x}| \cdot e_x$, the unit vector u is simply the normed vector between the original and resulting column of A:

$$\begin{aligned} \boldsymbol{x} &\in \mathbb{R}^n \\ \boldsymbol{v} &= \boldsymbol{x} - |\boldsymbol{x}| \cdot e_n \\ \boldsymbol{u} &= \frac{\boldsymbol{v}}{|\boldsymbol{v}|} \end{aligned}$$

The Householder matrix can then be found as before:

$$\boldsymbol{Q} = \boldsymbol{I} - \boldsymbol{2}\boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}}$$

When targeting inner columns, the reflection matrix should only be in the bottom right corner of the current Q, to avoid ruining our previous progress (the rest of it stays as I. This means that we ignore the upper row(s).

Which method to choose?

- Givens rotation is great for targeting specfic elements to turn into 0, and it needs a series of rotations to reduce several dimensions at once. It is ideal if there are already several 0s in the column.

- Householder reflections have the power to turn all except one element of a column into 0s at once (reflect a vector in $\mathbb{R}^3$ directly onto the x-axis for example)

Linear Vector Spaces

"Es macht Spaß" - *Vasile Gradinaru*

LTD: Hilbert Space

Linear - Lines are mapped to lines after the transformation

$\mathbb{R}^n$ and $\mathbb{C}^n$ are only two examples of many possible vector spaces. Considering the vector space V, the following operations / axioms are defined:

Definition 2.1.0.1. Vektorraum

Ein reeller *Vektorraum / linearer Raum* V ist eine Menge mit zwei Operationen:

Addition von Elementen aus V $(+)$:

$$a, b \in V : a + b \in V,$$

Multiplikation mit Skalaren $(\cdot)$:

$$\alpha \in \mathbb{R}, a \in V : \alpha \cdot a \in V.$$

Zusätzlich müssen folgende Eigenschaften gelten:

1. Kommutativitätsgesetz: $a + b = b + a$ für alle $a, b \in V$.
2. Assoziativitätsgesetz: $(a + b) + c = a + (b + c)$ für alle $a, b, c \in V$.
3. Es gibt $0 \in V$, sodass $a + 0 = a$ für alle $a \in V$ (dieses $0 \in V$ heisst Nullvektor).
4. Für jedes $a \in V$ gibt es ein $-a \in V$, so dass $a + (-a) = 0$.
5. Kompatibilität mit der Multiplikation mit Skalaren:
   $\alpha(\beta a) = (\alpha\beta)a$ für alle $\alpha, \beta \in \mathbb{R}, a \in V$.
6. Die Addition der Skalare ist distributiv gegen die Multiplikation mit Elementen aus V:
   $(\alpha + \beta)a = \alpha a + \beta b$ für alle $\alpha, \beta \in \mathbb{R}, a \in V$;
   Die Adition in V ist distributiv gegen die Multiplikation mit Skalare:
   $\alpha(a + b) = \alpha a + \alpha b$ für alle $\alpha \in \mathbb{R}, a, b \in V$.
7. Neutralelement für die Multiplikation mit Skalaren: $1 \cdot a = a$ für alle $a \in V$.

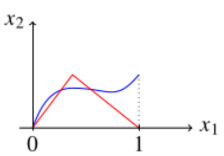
Two linear spaces can proved to be equal to one another by proving $A \subseteq B \wedge B \subseteq A \therefore A = B$

Continuous Differentiable Functions

$$C^s[a, b] := \{f : [a, b] \to \mathbb{R}, f \text{ is continuous and has } s \text{ derivatives}\}$$

Continuous - The function is continuous between [a, b], meaning there are no gaps or jumps in this interval.
Has s derivatives - The derivative exists at all points in the interval $[a, b]$ (ie. it is never a vertical line with an infinite gradient) and so on s times.

The following red function is a member of $C^1[0, 1]$ but not $C^2[0, 1]$, as its first derivative jumps from a positive to a negative value and its derivative therefore doesn't exist at the point of the jump (non-continuous).



The trigonometric functions have infinite continuous derivatives:

$$\{\sin, \cos\} \in C^\infty[a, b]$$

The larger the value of s, the smaller the set. All functions in C^2 also belong in C^1 and so on...

$$C^\infty \subset \ldots \subset C^3 \subset C^2 \subset C^1 \subset C^0$$

This set is a linear vector space, as addition and / or scalar multiplication result in a member of the same set, for example two continuous functions in an interval $[a, b]$ added together still result in a continuous function throughout the same interval.

Polynomials

$$\mathcal{P}_n := \{\text{polynomials of degree} \leq n - 1\}$$
$$\mathcal{P}_1 \subset \mathcal{P}_2 \subset \mathcal{P}_3 \subset \ldots \subset \mathcal{P}_\infty$$

For example the polynomial $x^2 + 1$ with degree 2 is a member of $\mathcal{P}_3$ but also all higher sets such as $\mathcal{P}_4$ as the coefficient of x^3 is simply 0.

Addition and **scalar** multiplication are indeed valid operations that result in a member of the same set, therefore it is a linear vector space.

The Taylor Series in Analysis demonstrates how any continuous function in C^0 can be approximated using $\mathcal{P}_n$.

Other Linear Spaces

- $L^2[a, b] := \left\{f : [a, b] \to \mathbb{R}, \int_a^b |f|^2 dt\right\}$- Space of quadratically integrable functions
- $\ell := \{a_n \text{ converges}\}$- Space of convergent sequences
- $\ell := \{a_n \text{ converges}\}$- Space of convergent sequences

Linear Subspace

Sei V ein linearer Raum mit $U \subseteq V$ und $U \neq \emptyset$. Dann heisst U *Unterraum* von V, falls gilt:

1. Wenn $x, y \in U$ dann auch $x + y \in U$.
2. Wenn $\alpha \in \mathbb{R}, x \in U$ dann auch $\alpha x \in U$.

A linear subspace must also include a zero vector.

$$U := \left\{\boldsymbol{x} \in \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \in \mathbb{R}^2\right\}$$
$$U' := \left\{\boldsymbol{x} \in \begin{pmatrix} x_1 \\ 5 \end{pmatrix} \in \mathbb{R}^2\right\}$$

U' is not a subspace, as vector addition leads to $\begin{pmatrix} x_1{}' \\ 10 \end{pmatrix}$ which has left the space.

A single contradictory example must be found to prove that something is not a linear subspace.

Span

The set of all possible linear combinations of elements of a linear space. For example the span of basis vectors of $\mathbb{R}^3$ is the entire $\mathbb{R}^3$, as any point in 3D space can be expressed as linear combination of the basis.

The range of the transformation of space $\mathcal{A}(x) = \boldsymbol{A}\boldsymbol{x}$ can be expressed as the span of its columns:

$$\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$$
$$\text{Range } \mathcal{A} = \text{Span}\{A_1, A_2, \ldots, A_n\}$$

Erzeugendensystem - Set of vectors which span a vector space V.

Image / Column Space / Range / Codomain - of a transformation can be found by capturing any compatibility conditions in a vector (if there are any), for example of $b_3 - b_1 - b_2 = 0$, the range can be expressed as:

$$\left\{ b_1, b_2 \in \mathbb{R}, \begin{pmatrix} b_1 \\ b_2 \\ b_1 + b_2 \end{pmatrix} \right\}$$

This can also be expressed as a span by breaking it into a linear combination:

$$b_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + b_2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \text{Span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$$

Monome - $p_{n(t)} = t^n$, for example $p_0(t) = 1, p_1(t) = t, p_2(t) = t^2, ...$
The set of monomes is a linear subspace of the polynomials $\mathcal{P}_n$, which can be spanned by $\text{Span}\{p_k \mid k < n\}$

Kernel (Null-Space / Kern)
Let $\boldsymbol{A} := m \times n$. The kernel is the set of vectors that is transformed to 0 by $\boldsymbol{A}$.

The kernel can be found by solving the LGS $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ with Gaussian Elimination. This provides infinite solutions for a singular matrix or the null vector for a regular system.

Each linear space already has two simple subspaces:
- The kernel of its matrix
- The range of it as a transformation

Basis
Linearly Dependant - A vector in the set is linearly dependent if other vectors in the set can express it as a linear combination. In the case of two elements:

$$\alpha \in \mathbb{R}$$
$$\boldsymbol{v_1} = \alpha \boldsymbol{v_2}$$

This check can be restructured as a linear combination of several vectors $\boldsymbol{v_n}$, where we find the kernel of the matrix $\boldsymbol{A}$, meaning the set of scalars which would express each column of the matrix as a linear combination of the other columns:

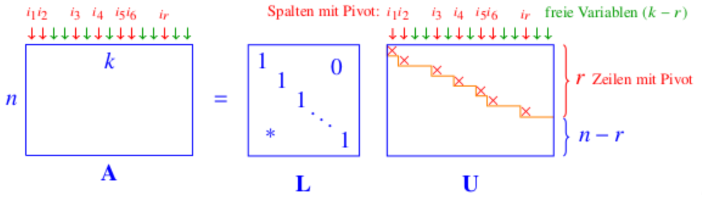
$$\begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$
$$x_1\boldsymbol{v_1} + x_2\boldsymbol{v_2} + x_3\boldsymbol{v_3} + x_4\boldsymbol{v_4} + ... = 0$$
$$-\boldsymbol{v_1} = \frac{x_2}{x_1}\boldsymbol{v_2} + \frac{x_3}{x_1}\boldsymbol{v_3} + \frac{x_4}{x_1}\boldsymbol{v_4} + ...$$
$$\boldsymbol{A} = (\boldsymbol{v_1} \ \ \boldsymbol{v_2} \ \ \cdots \ \ \boldsymbol{v_n})$$
$$\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$$

If only the trivial solution exists for $\boldsymbol{x}$ (A is regular), all vectors $\boldsymbol{v_n}$ are linearly independent.

Basis - A set of linearly independent vectors that spans the entire linear space (minimal Erzeugenden system) and stay completely within the linear space (not allowed to span a parent space as well). There can be several independent bases in a space, but all bases have the same number of elements.

Any element in the linear space can be determined by a **unique** linear combination of the basis. Proof in script.

Canonical basis - Bases taken "as canon" (accepted standard) for each space, for example $e_x, e_y, e_z \in \mathbb{R}^3$



The basis of an erzeugenden system can be found through LU Zerlegung, where the columns with a pivot are linearly independent of one another and therefore the corresponding columns in A are a basis of the range / image (span of the erzeugenden system). The other columns are so called free variables because they can take

Basis of $\mathcal{P}_n$ are the monomes
Basis of $n \in \mathbb{N}, \mathcal{P}_n = \{p_i = t^i |\ t \in \mathbb{R}, i \in \mathbb{N}_0 < n\}$ - The monomes are linearly independent (proof in script) and span the entire polynomial space.

A trick for proving linear independence of certain items, for example polynomials $q_1, q_2, \ldots$ is expressing them in terms of another linear space that is known to be linearly independent, for example monomes and rearranging the basis check equation as follows:

$$\begin{aligned} q_1 &= p_0 + p_2 \\ q_2 &= p_0 - p_2 \\ q_n &= \ldots \end{aligned}$$

$$\begin{aligned} x_1 q_1 + x_2 q_2 + \ldots &= 0 \\ x_1(p_0 + p_2) + x_2(p_0 - p_2) + \ldots &= 0 \\ (x_1 + x_2)p_0 + (x_1 - x_2)p_2 + \ldots &= 0 \\ \begin{pmatrix} 1 & 1 & \ldots \\ 1 & -1 & \ldots \\ \ldots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ldots \end{pmatrix} &= 0 \end{aligned}$$

Since the monomes have already been proved to be linearly independent, we can construct the coefficients containing x as an LGS with right hand side 0 and show that this matrix indeed only has the trivial solution (it is regular).

Dimensions
The number of elements in a basis of a space.

Finite Dimensional Vector Space - A finite set of basis vectors spans the entire space.

Examples:
- Finite dimensional: $\mathbb{R}^n, \mathcal{P}_n$ - Dimension = n. (Hence the awkward degree $\leq n - 1$ notation for $\mathcal{P}_n$)
- Non-finite dimensional: $\mathcal{P}, C^k, L^2$

Sometimes non-finite dimensional spaces can be approximated using finite dimensional spaces (Taylor series).

Satz 2.3.0.10. Grösse von erzeugenden Systemen in einem endlichdimensionalen Raum

Sei V ein linearer Raum mit Dimension n, dann:

1) *sind mehr als n Elemente von V linear abhängig.*
2) *sind weniger als n Elemente von V nicht erzeugend.*
3) *sind n Elemente von V linear unabhängig genau dann, wenn sie auch erzeugend sind.*

Fundamental Theorem of Linear Algebra (Gilbert Strang)

For a $n \times k$ matrix $\boldsymbol{A}$ with rank r:

$$\dim(\mathrm{Im}(\boldsymbol{A})) = r$$
$$\dim(\mathrm{Kernel}(\boldsymbol{A})) = k - r = \text{Number of free variables}$$

Transposing the matrix leads to ($\mathrm{Rank}(\boldsymbol{A}) = \mathrm{Rank}(\boldsymbol{A^T})$):

$$\dim(\mathrm{Im}(\boldsymbol{A^T})) = r$$
$$\dim(\mathrm{Kernel}(\boldsymbol{A^T})) = n - r$$

For example, considering a 3x3 matrix P with rank 2:
- The transformation maps any x onto a 2D plane due to only 2 independent vectors in the basis, the image has dimension 2. A solution only exists if the RHS vector is in this plane.
- The kernel of the matrix $Px = 0$ includes an entire line of vectors, meaning that there is 1 free variable (the non-pivot column) that scales along this line.

LTD: Review these proofs

Orthogonal Spaces - Two linear spaces are orthogonal $U \perp V$ to one another when any two vectors in the spaces are always orthogonal to one another $:< \boldsymbol{u}, \boldsymbol{v} >= 0$

$$\mathrm{Kernel}(A) \perp \mathrm{Im}(A^T)$$
$$\mathrm{Im}(A) \perp \mathrm{Kernel}(A^T)$$

This law can be used to easily decompose any linear space into two orthogonal elements.

Coordinates

The coordinates of a specific element $x \in \mathbb{R}^n$ formed using the basis $\mathcal{B} = \{v_1, v_2, v_3, ...\}$ are written as:

$$x = x_1\boldsymbol{v_1} + x_2\boldsymbol{v_2} + x_3\boldsymbol{v_3} + ...$$
$$\not{k}_{\mathcal{B}} : \boldsymbol{V} \to \mathbb{R}^3 := \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \end{pmatrix}$$

This mapping transforms a tuple of coordinates into the resulting point in $\mathbb{R}^n$ using an abstract basis $\mathcal{B}$.

Coordinates can of course be transformed to coordinates of another basis in the same linear space as follows:
- Both coordinates are unique for their underlying basis and both result in an element in the same linear space.

- Therefore each basis vector in the target basis can be represented as a linear combination of the original basis.
- We can represent this mapping as a matrix $\boldsymbol{C}$ where $\boldsymbol{C}\tilde{\boldsymbol{x}} = \boldsymbol{x}$

$\boldsymbol{C}$ is found by:

1. Represent each vector in the new basis as a coordinate in the old basis.
2. These are now the columns of $\boldsymbol{C}$.
3. Solve $\boldsymbol{C}\tilde{\boldsymbol{x}} = \boldsymbol{x}$
4. The matrix can then be used to change the basis of any vector, as well as calculating the inverse (in case it is easier than finding a new suitable $\boldsymbol{C}$)

1) Wir schreiben jedes Element aus der neuen Basis $\tilde{\mathcal{B}}$ als eine lineare Kombination von Elementen aus der alten Basis $\mathcal{B}$:

$$q_1 = 1p_0 + 1p_2 + 0p_4 \implies \text{Spalte} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

$$q_2 = 1p_0 - 1p_2 + 0p_4 \implies \text{Spalte} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

$$q_3 = 1p_0 + 1p_2 + 1p_4 \implies \text{Spalte} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

2) Wir nehmen die Vektoren, die wir im Schritt 1) erhalten haben und schreiben diese **spaltenweise** in einer Matrix:

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

3) Wir benutzen diese Matrix um das lineare Gleichungssystem $\mathbf{C}\tilde{\mathbf{x}} = \mathbf{x}$ nach $\tilde{\mathbf{x}}$ zu lösen:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \implies \qquad \begin{cases} \tilde{x}_1 &= -\frac{1}{2}x_1 + \frac{1}{2}x_2 - x_3 \\ \tilde{x}_2 &= \frac{1}{2}x_1 - \frac{1}{2}x_2 \\ \tilde{x}_3 &= x_3 \end{cases}.$$

LTD: is $\tilde{x}$ the point we want to transform in terms of the new basis?

Linear Transformations

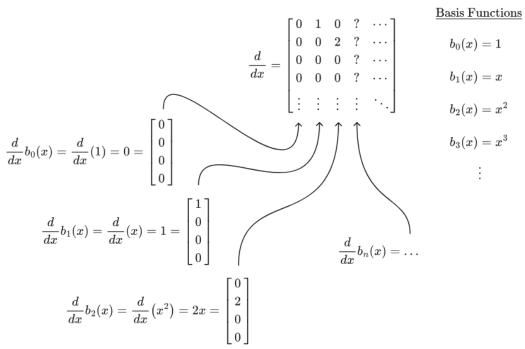
Useful resource: https://www.3blue1brown.com/lessons/abstract-vector-spaces

Here is the formal definition of a linear transformation:

$$\mathcal{F} : X \to Y$$
$$\mathcal{F}(x_1 + x_2) = \mathcal{F}(x_1) + \mathcal{F}(x_2) \forall x_1, x_2 \in X$$
$$\alpha\mathcal{F}(x) = \mathcal{F}(\alpha x) \forall x \in X, \alpha \in \mathbb{R}$$

This therefore means that the origin stays fixed.

For example, derivation is a linear transformation:



- Every linear transformation in a finite dimensional space can be represented as matrix multiplication.
- *Isomorphism* - A structure-preserving mapping, ie a bijective linear transformation is a called an isomorphism. The inverse is clearly also an isomorphism. A non-square matrix changes dimensions and thus is not bijective and isomorphic - no inverse.
- *Automorphism* - If the two sets the transformation maps between are the same.
- A linear transformation $\mathcal{F}$ is injective when $\text{Kernel}(\mathcal{F}) = \{0\}$.
- The same properties regarding dimensions sizes regarding basis apply to linear transformations:

Definition 3.2.0.7. Kernel und Bild

Sei $\mathcal{F}$ eine Abbildung, dann können folgende zwei Mengen definiert werden:

Kernel von $\mathcal{F}$:
$\text{Kern}(\mathcal{F}) = \{x \in X, \mathcal{F}x = 0\}$.
Wenn $\mathcal{F}$ eine lineare Abbildung ist, dann ist $\text{Kern}(\mathcal{F})$ ein linearer Unterraum von X.

Bild von $\mathcal{F}$:
$\text{Bild}(\mathcal{F}) = \{y \in Y$, so dass $x \in X$ mit $\mathcal{F}x = y\}$.
Wenn $\mathcal{F}$ eine lineare Abbildung ist, dann ist $\text{Bild}(\mathcal{F})$ ein linearer Unterraum von Y.

Satz 3.2.0.8. Kern einer linearen Injektion

Wenn $\mathcal{F} : X \longrightarrow Y$ eine lineare Abbildung ist, dann gilt:
$\mathcal{F}$ ist injektiv genau dann, wenn $\text{Kern}(\mathcal{F}) = \{0\}$.

Satz 3.2.0.9. Dimensionssatz

Wenn $\mathcal{F} : X \longrightarrow Y$ eine lineare Abbildung zwischen den beiden endlichdimensionalen Räumen X und Y ist, dann gilt:

$$\dim(\text{Kern}(\mathcal{F})) + \dim(\text{Bild}(\mathcal{F})) = \dim(X) .$$

Definition 3.2.0.10. Rang einer linearen Abbildung

Der Rang einer linearen Abbildung $\mathcal{F}$ zwischen zwei endlichdimensionalen Räumen ist

$$\text{Rang}(\mathcal{F}) = \text{Rang}(\mathbf{F}) = \dim(\text{Bild}(\mathbf{F})) .$$

A linear transformation is independent of the basis / coordinates used and can be represented with respect to many different bases in the same space. The following diagram highlights these relationships well:

Sei X ein linearer Raum mit den beiden Basen $\mathscr{B}$ und $\widetilde{\mathscr{B}}$, sei Y ein anderer linearer Raum auch mit zwei Basen $\mathscr{C}$ und $\widetilde{\mathscr{C}}$ und sei $\mathscr{F} : X \to Y$ eine lineare Abbildung. Das Diagramm dazu ist:

$$
\begin{tikzcd}
\widetilde{\mathscr{B}} : & \mathbb{R}^n \arrow[r, "\mathbf{B}"] \arrow[d, shift left, "k^{-1}_{\widetilde{\mathscr{B}}}"] & \mathbb{R}^m \arrow[d, shift left, "k^{-1}_{\widetilde{\mathscr{C}}}"] & : \widetilde{\mathscr{C}} \\
& X \arrow[r, "\mathscr{F}"] \arrow[u, shift left, "k_{\widetilde{\mathscr{B}}}"] \arrow[d, shift left, "k_{\mathscr{B}}"] & Y \arrow[u, shift left, "k_{\widetilde{\mathscr{C}}}"] \arrow[d, shift left, "k_{\mathscr{C}}"] & \\
\mathscr{B} : & \mathbb{R}^n \arrow[r, "\mathbf{A}"] \arrow[u, shift left, "k^{-1}_{\mathscr{B}}"] & \mathbb{R}^m \arrow[u, shift left, "k^{-1}_{\mathscr{C}}"] & : \mathscr{C}
\end{tikzcd}
$$

Es gibt also eine neue Matrix $\mathbf{B}$, welche auch die Abbildung $\mathscr{F}$ darstellt, und zwar mit den neuen Basen $\widetilde{\mathscr{B}}$, $\widetilde{\mathscr{C}}$. Da dies ein kommutatives Diagramm ist, können wir die Matrix $\mathbf{B}$ mit den internen Basistransformationen und $\mathbf{A}$ ausrechnen:

$$
\mathbf{B} = \underbrace{k_{\widetilde{\mathscr{C}}} k^{-1}_{\mathscr{C}}}_{\mathbf{S}^{-1}} \, \mathbf{A} \, \underbrace{k_{\mathscr{B}} k^{-1}_{\widetilde{\mathscr{B}}}}_{\mathbf{T}} = \mathbf{S}^{-1}\mathbf{A}\,\mathbf{T}
$$
$$
\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\,\mathbf{T} \iff \mathbf{A} = \mathbf{S}\,\mathbf{B}\,\mathbf{T}^{-1}.
$$

$\mathbf{T}$ realisiert den Wechsel von der Basis $\widetilde{\mathscr{B}}$ in die Basis $\mathscr{B}$ im linearen Raum X. D.h. $\mathbf{T}\widetilde{\mathbf{x}} = \mathbf{x}$ für den Koordinatenvektor $\mathbf{x}$ von $x \in X$ in der alten Basis $\mathscr{B}$ und den Koordinatenvektor $\widetilde{\mathbf{x}}$ von x in der neuen Basis $\widetilde{\mathscr{B}}$.
$\mathbf{S}$ realisiert den Wechsel von der alten Basis $\mathscr{C}$ in die neue Basis $\widetilde{\mathscr{C}}$ im linearen Raum Y.

Anti-linear (Conjugate-Linear) Transformation

$$
\mathcal{A} : X \to Y
$$
$$
\mathcal{A}(x_1 + x_2) = \mathcal{A}(x_1) + \mathcal{A}(x_2) \forall x_1, x_2 \in X
$$
$$
\alpha \mathcal{A}(x) = \mathcal{A}(\overline{\alpha} x) \forall x \in X, \alpha \in \mathbb{C}
$$

Linear Form (Linear Functional)

Simply any **linear** transformation that maps a linear space $V \to \mathbb{R}^1 \vee \mathbb{C}^1$.

LTD: Quaternions

This can be something as simple as $\varphi(x) \to 0$ or the dot product.

Riesz Representation Theorem

Let $\mathbb{V}$ be a linear space with a dot product and linear form φ. There exists a unique vector $w \in \mathbb{V}$ (called the Riesz representation) such that:

$$
\varphi(v) = \ < w, v > \forall v \in \mathbb{V}
$$

which can be represented in terms of any orthonormal basis $\{b_1, b_2, ..., b_n\} \in \mathbb{V}$:

$$
w = \overline{\varphi(b_1)} b_1 + \overline{\varphi(b_2)} b_2 + ... + \overline{\varphi(b_n)} b_n
$$

whereby the conjugation can be ignored if dealing with real linear forms.

This is a cool way of representing any linear form, essentially a regular linear transformation mapping to 1 dimension, as a single vector; because any linear transformation can be represented as a matrix. LTD: Chicken vs the egg?

Dual Space

The linear vector space V' of all linear forms of a linear vector space V, such that:

$$\varphi, \psi \in V'$$
$$x \in V, \alpha \in \mathbb{R}/\mathbb{C}$$
$$(\varphi + \psi)(x) = \varphi(x) + \psi(x)$$
$$\alpha\varphi(x) = \varphi(\alpha x)$$
$$\alpha\psi(x) = \psi(\alpha x)$$

Each element of V' can be represented as a vector thanks to the Riesz representation theorem.

TODO: Adjungierte Abbildung

Norms

A norm is a function that transforms any element in a linear space to a positive real number. It must respect the following properties:

Definition 4.1.0.1. Norm

Sei V ein linearer Raum.
Die Funktion $\|\cdot\| : V \longrightarrow [0, \infty[$ heisst Norm in V, falls sie folgende Eigenschaften erfüllt:

(N1) Aus $\|v\| = 0$ folgt $v = 0$.

(N2) Sei α ein Skalar und $v \in V$ beliebig. Dann gilt: $\|\alpha v\| = |\alpha|\|v\|$.

(N3) Für beliebige $v, w \in V$ gilt: $\|v + w\| \leq \|v\| + \|w\|$ (Dreiecksungleichung).

Definition 4.1.0.2. Normierter linearer Raum

Ein linearer Raum V, welcher eine Norm $\|\cdot\|$ besitzt, heisst normierter linearer Raum.

Satz 4.1.0.6. Äquivalenz der Normen

Alle Normen in $\mathbb{R}^d$ sind äquivalent. In anderen Worten:
Seien $\|\cdot\|$ und $|||\cdot|||$ Normen in $\mathbb{R}^d$, dann gibt es eine Konstante $C \geq 1$, abhängig von der Dimension d, so dass

$$\frac{1}{C}\|\mathbf{v}\| \leq |||\mathbf{v}||| \leq C\|\mathbf{v}\| \text{ für alle } \mathbf{v} \in \mathbb{R}^d .$$

A basic example of a norm in $\mathbb{R}$ is the absolute function:

$$x \in \mathbb{R}$$
$$|x|$$

This is the only universal norm in one dimensional spaces. TODO: How is this an isomorphism? https://en.wikipedia.org/wiki/Norm_(mathematics)

Taxicab Norm (L^1)

The name comes from the fastest zig zag path a taxi has to drive through in Manhattan's grid based street layout:

$$\|x\|_1 := \sum_{i=1} |x_i|$$

Euclidean Norm (L^2)

This ubiquitous norm represents the length of a vector and can alternatively be written in vector notation:

$$x \in \mathbb{R}^d$$
$$\|x\|_2 := \sqrt{x^T x}$$
$$x \in \mathbb{C}^d$$
$$\|x\|_2 := \sqrt{x^H x}$$

p-Norm

$$p \in \mathbb{R} \mid p \geq 1$$
$$\|x\|_p := \left(\sum_{k=1} |x_k|^p\right)^{\frac{1}{p}}$$

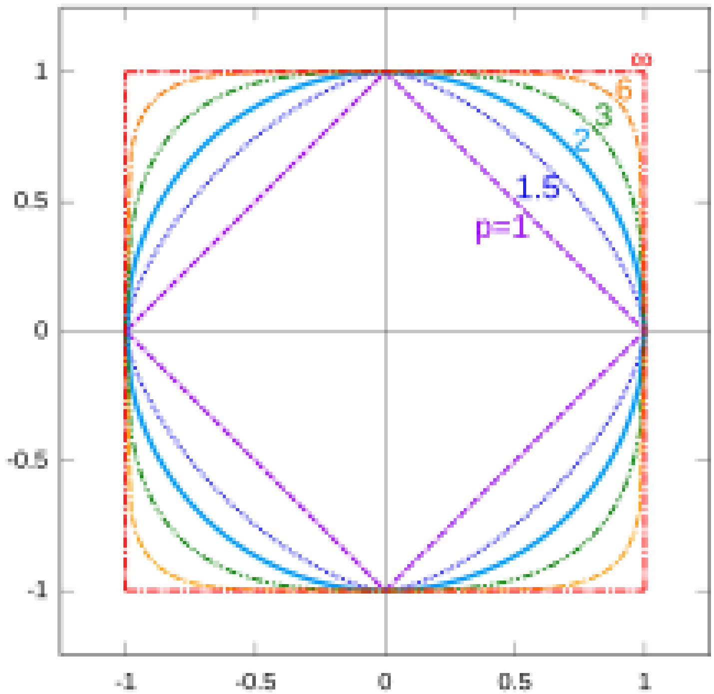
All L^n norms can be generalised as the p-norm.

Maximum Norm (L^∞)
A special case of the p-Norm as p approaches ∞:

$$\|\boldsymbol{x}\|_\infty := \max(|\boldsymbol{x_1}|, |\boldsymbol{x_2}|, \ldots)$$

The intuition behind this is that the largest element in the vector dominates the sum and therefore this norm effectively zooms in on the largest element.

Here is the unit circle as defined by various p-norms:



Cauchy-Bunjakovski-Schwarz Inequality
This is often regarded as the most important inequality in mathematics, it can be written in many different forms, the 3rd one is very intuitive:

$$\boldsymbol{v}^T\boldsymbol{w} \leq \|\boldsymbol{v}\|\|\boldsymbol{w}\|$$
$$< \boldsymbol{v}, \boldsymbol{w} > \quad \leq \quad < \boldsymbol{v}, \boldsymbol{v} > \cdot < \boldsymbol{w}, \boldsymbol{w} >$$
$$\|\boldsymbol{v}\|\|\boldsymbol{w}\| \cos(\theta) \leq \|\boldsymbol{v}\|\|\boldsymbol{w}\|$$

Where $\|\cdot\|$ is the Euclidean norm. This can be proven by induction.

Both sides are equal $\Leftrightarrow v, w$ are linearly dependent. Due to this, the angle definition of the dot product is actually an application of the inequality used throughout mathematics:

$$\cos(\theta) = \frac{< v, w >}{\|v\|\|w\|}$$

Matrix Norms

Every norm in $\mathbb{R}^n$ defines a corresponding matrix norm, which is a measure of how much that norm is affected after the matrix is applied as a linear transformation:

$$\boldsymbol{A} : m \times n$$
$$\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_p = 1$$
$$\|\boldsymbol{A}\|_p := \max\left(\|\boldsymbol{A}\boldsymbol{x}\|_p\right)$$

Here are some examples:

Maximum Column Sum ($\|\cdot\|_1$)

In this case $x = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \\ \dots \end{pmatrix}$ so that $\|x\|_1 = 1$.

Ax results in each column at a time. When the 1 norm is applied to each column, the sum of the column's elements is returned.

Therefore this simply returns the maximum column sum:

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & -3 \\ 2 & 3 & -1 \end{bmatrix} \implies \|\mathbf{A}\|_1 = \max\{|1| + |2|,\ |-2| + |3|,\ |-3| + |1|\} = 5\,.$$

Maximum Row Sum ($\|\cdot\|_\infty$)

The maximum transformation allowed is when $x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \dots \end{pmatrix}$.

Ax results in the sum of each row:

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & -3 \\ 2 & 3 & -1 \end{bmatrix} \implies \|\mathbf{A}\|_\infty = \max\{|1| + |-2| + |-3|,\ |2| + |3| + |-1|\} = 6\,.$$

LTD: Spectral norm after learning eigenvalues

Inner Products

In Euclidean space, the dot product is an inner product.

Definition 4.2.0.1. Skalarprodukt

Sei V ein linearer Raum.
Ein *Skalarprodukt* $\langle\cdot,\cdot\rangle$ ist eine Funktion von zwei Variablen in diesem Raum V

$$\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$$

welche die folgenden drei Eigenschaften erfüllt:

(S1) Die Funktion ist *linear im zweiten Argument*:

$$\langle x, ay + bz\rangle = a\langle x, y\rangle + b\langle x, z\rangle$$

für alle $x,\ y,\ z \in V$ und $a,\ b$ Skalare;

(S2) Die Funktion ist *symmetrisch*:

$$\langle x, y\rangle = \langle y, x\rangle \text{ über } \mathbb{R}$$
$$\langle x, y\rangle = \overline{\langle y, x\rangle} \text{ über } \mathbb{C}\,;$$

(S3) Die Funktion ist *positiv definit*:

$$\langle x, x\rangle \geq 0 \text{ für alle } x \in V$$
$$\langle x, x\rangle = 0 \implies x = 0\,;$$

falls nur der erste dieser beiden letzten Punkte erfüllt ist,
so sprechen wir von *positiv semi-definit*.

A *positiv semi-definit* operation is not an inner product.

Any operation that satisfies these properties leads to its corresponding norm being defined as:

Definition 4.2.0.4. Norm aus einem Skalarprodukt

Man sagt, dass *eine Norm* $\|\cdot\|$ *aus einem Skalarprodukt* $\langle\cdot,\cdot\rangle$ *kommt*, falls

$$\|x\| = \sqrt{\langle x, x\rangle}$$

für alle Elemente x eines linearen Raumes X.

Definition 4.2.0.5. Orthogonalität von Elementen eines linearen Raumes mit Skalarprodukt

Sei $x,\ y \in V$, wobei V ein linearer Raum mit einem Skalarprodukt $\langle\cdot,\cdot\rangle$ ist.
Wir sagen, dass x und y orthogonal ($x \perp y$) sind, falls $\langle x, y\rangle = 0$.

$$u \perp v \Leftrightarrow < u, v >= 0$$

Projectors

Projector - A linear transformation such that repeated application has no effect:

$$\boldsymbol{P_y P_y \ldots = P_y}$$

Thus, projection is idempotent (can be applied arbitrarily many times without changing after the first application).

Orthogonal Projection

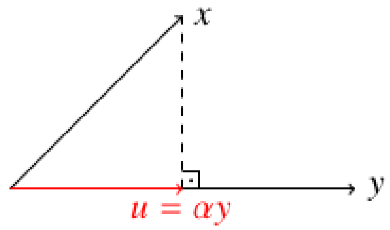
Definition 4.4.0.2. Orthogonale und schiefe Projektoren

Der Projektor $\mathscr{P} : V \to V$ im linearen Raum V mit dem Skalarprodukt $\langle\cdot,\cdot\rangle$ heisst *orthogonaler Projektor*, falls

$$x - \mathscr{P}x \perp \text{Bild}\,\mathscr{P} \text{ für alle } x \in V\,,$$

sonst heisst er *schiefer Projektor*.

Grahically, the projection of $\boldsymbol{x}$ onto $\boldsymbol{y}$ looks like this:



Since $(\boldsymbol{x} - \boldsymbol{u}) \perp \boldsymbol{y}$, we can derive the scalar α as:

$$\alpha = \frac{< \boldsymbol{y}, \boldsymbol{x} >}{< \boldsymbol{y}, \boldsymbol{y} >}$$

Thus, we can express the orthogonal projection of a vector onto $\boldsymbol{y}$ as the following function:

$$P_y : \mathbb{V} \to \mathbb{V}$$
$$P_y(x) := \frac{< \boldsymbol{y}, \boldsymbol{x} >}{< \boldsymbol{y}, \boldsymbol{y} >} \boldsymbol{y}$$

Projectors

Projector - Projection represented in matrix form, for example the orthogonal projection in the previous example can also be calculated using a matrix:

$$\begin{aligned} P_y(x) &= \frac{\boldsymbol{y}\boldsymbol{y}^{\boldsymbol{H}}}{\|\boldsymbol{y}\|^2} \boldsymbol{x} \\ &= \boldsymbol{P}_{\boldsymbol{y}} \boldsymbol{x} \end{aligned}$$

Projectors have the following properties:
- Every "successful" projection lies along the target $\mathrm{Im}(\boldsymbol{P}_{\boldsymbol{y}}) = \mathrm{Span}\{\boldsymbol{y}\}$.
- The kernel of a projector is the span of orthogonal vectors $\mathrm{Kern}(\boldsymbol{P}_{\boldsymbol{y}}) = \mathrm{Span}\{\boldsymbol{x} \mid \boldsymbol{x} \perp \boldsymbol{y}\}$

Satz 4.4.0.4. Orthogonale und schiefe Projektoren

Die Matrix $\mathbf{P}$ *ist ein orthogonaler Projektor genau dann, wenn*

$$\mathbf{P}^2 = \mathbf{P} \quad \textit{und} \quad \mathbf{P}^H = \mathbf{P}\,.$$

Falls nur die Bedingung $\mathbf{P}^2 = \mathbf{P}$ *erfüllt ist, so ist* $\mathbf{P}$ *ein schiefer Projektor.*

Unit Vectors

Definition 4.2.0.14. Einheitsvektor

Sei $v \in V$, mit V ein linearer Raum mit der Norm $\|\cdot\|$. Falls $\|v\| = 1$, dann heisst v *Einheitsvektor*.

Satz 4.2.0.15. Orthogonale Vektoren sind linear unabhängig

Seien $e_1, e_2, \ldots, e_n$ *Einheitsvektoren in einem linearen Raum* V *mit einer Norm* $\|\cdot\|$, *die aus dem Skalarprodukt* $\langle \cdot, \cdot \rangle$ *kommt.*
Falls die Einheitsvektoren $e_1, e_2, \ldots, e_n$ *paarweise orthogonal sind, so sind sie auch linear unabhängig.*

In a linear space with dimensions n, n orthogonal unit vectors therefore form a basis in this space, because they are all linearly independent.

The dot product of any vector $\boldsymbol{x}$ with a unit vector is simply the component of $\boldsymbol{x}$ in the direction of the unit vector, which is very useful throughout physics when dealing with components that are not along a predefined axis:

$$\boldsymbol{x} \in \mathbb{R}^n$$
$$< \boldsymbol{e}, \boldsymbol{x} >= \|\boldsymbol{x}\| \|\boldsymbol{e}\| \cos(\theta) = \|\boldsymbol{x}\| \cos(\theta)$$

Furthermore, this implies that any point can be represented as the sum of its projections onto a basis $\mathbb{B}$. This is essentially what the coordinate function defined earlier does:

$$\begin{aligned} \boldsymbol{x} &= \sum_{e \in \mathbb{B}} P_e(x) \\ &= \sum_{e \in \mathbb{B}} \frac{< \boldsymbol{e}, \boldsymbol{x} >}{< \boldsymbol{e}, \boldsymbol{e} >} \boldsymbol{e} \end{aligned}$$

If the basis B uses unit vectors (normal), this simplifies to:

$$= \sum_{e \in \mathbb{B}} \frac{< \boldsymbol{e}, \boldsymbol{x} >}{1} \boldsymbol{e} = \|\boldsymbol{x}\| \cos(\theta) \vec{\boldsymbol{e}}$$

This can also be used to project a vector into a subspace through a smaller dimensional identity matrix looking ahh:

TODO: Lecture 14 & 15 Gilbert Strang, unsure what script means here

Gram-Schmidt

Orthonormal - Normalised orthogonal vectors; ie. a set of vectors that are orthogonal between one another and have length (norm) 1.

Orthonormality can be tested with dot products:
- $< u, v >= 0$ - Orthogonal
- $< u, u >= 1$ - Normal

LTD: Does it matter which dot product you use, do they all work in the same way?

Theorem

The span of any set of linearly independent vectors can be spanned by a set of orthonormal vectors.

In other words, any basis can be converted to an orthonormal basis.

It is very computationally advantageous to use an orthonormal basis whenever possible. Many commonly used operations are simplified, for example when projecting an element onto an orthonormal basis as mentioned in the previous chapter.

Method

We would like to transform the basis $\mathbb{V} = \{v_1, v_2, ..., v_n\} \to \mathbb{U} = \{e_1, e_2, ..., e_n\}$ where $\mathbb{U}$ is orthonormal.

This is carried out in 2 steps:
1. Orthogonalise the vectors by recursively subtracting the previous n elements of the resulting basis, the first vector can remain unchanged:

$$
\begin{aligned}
u_1 &= v_1 \\
u_2 &= v_2 - P_{u_1}(v_2) \\
u_3 &= v_3 - P_{u_1}(v_3) - P_{u_2}(v_3) \\
&\dots \\
u_k &= v_k - \sum_{i=1}^{k-1} P_{u_i}(v_k)
\end{aligned}
$$

2. The resulting vectors are then normalised (this does not affect their orthogonality):

$$
e_k = \frac{u_k}{\|u_k\|}
$$

The process is visualised in 3D here, showing how subtracting the projected vector results in the orthogonal "missing piece of the triangle" for each projection: https://upload.wikimedia.org/wikipedia/commons/e/ee/Gram-Schmidt_orthonormalization_process.gif

If the **0** vector is output at any stage, the input vectors are linearly dependent.

Stability

When calculating the recursive method, the resulting vectors are usually not quite orthogonal due to rounding errors, which becomes worse after each iterative subtraction of projections.

This can be alleviated by instead projecting the "orthogonaler" u_k at each stage of the iterative subtraction:

$$
\begin{aligned}
\mathbf{u}_k^{(1)} &= \mathbf{v}_k - \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}_k), \\
\mathbf{u}_k^{(2)} &= \mathbf{u}_k^{(1)} - \mathrm{proj}_{\mathbf{u}_2}\left(\mathbf{u}_k^{(1)}\right), \\
&\vdots \\
\mathbf{u}_k^{(k-2)} &= \mathbf{u}_k^{(k-3)} - \mathrm{proj}_{\mathbf{u}_{k-2}}\left(\mathbf{u}_k^{(k-3)}\right), \\
\mathbf{u}_k^{(k-1)} &= \mathbf{u}_k^{(k-2)} - \mathrm{proj}_{\mathbf{u}_{k-1}}\left(\mathbf{u}_k^{(k-2)}\right), \\
\mathbf{e}_k &= \frac{\mathbf{u}_k^{(k-1)}}{\left\|\mathbf{u}_k^{(k-1)}\right\|}
\end{aligned}
$$

This of course results in more computation but significantly reduces rounding errors. Other orthogonalisation algorithms seen in chapter **QR decomposition** using Householder matrices / rotations can have reduced error rates.

Legendre Polynomials

The most straight-forward bases for $\mathcal{P}_n$ are the monomes, but they are not orthonormal. Using the Gram-Schmidt method, the **Legendre Polynomials** can be derived from them, allowing us to represent any polynomial through an orthonormal basis (see `./exercises/legendre-polynomials.pdf`)

There are many other such sets of polynomials with special names, for example Hermite and Laguerre polynomials.

QR Decomposition

TODO: Consider moving to QR section, linear vector spaces, norms, dot products, orthogonal, gram-schmidt, qr, The matrix $\boldsymbol{A}$ with columns $\{v_1, v_2, ..., v_n\}$ can be decomposed into matrices $\boldsymbol{QR}$ because the Gram-Schmidt is an orthogonalization process just like rotation / reflection:

1. Calculate the corresponding orthonormal vectors $\{e_1, e_2, ..., e_n\}$ using the Gram-Schmidt process
2. Each column in $\boldsymbol{A}$ can now be expressed as a linear combination of the new orthonormal basis as projections onto the unit vectors:

$$v_k = \sum_{i=1}^{k} P_{e_i}(v_k) = \sum_{i=1}^{k} < e_i, v_k > e_i$$

3. Therefore this set of linear combinations can be represented as matrix multiplication $\boldsymbol{QR}$ where:

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{QR} \\ &= (e_1 \;\; e_2 \;\; ... \;\; e_n) \begin{pmatrix} < e_1, v_1 > & < e_1, v_2 > & ... & < e_1, v_n > \\ 0 & < e_2, v_2 > & ... & < e_2, v_n > \\ 0 & 0 & ... & < e_3, v_n > \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & < e_n, v_n > \end{pmatrix} \end{aligned}$$

Advantages:

- Straightforward implementation and slightly more computationally efficient

Disadvantages:

- Very numerically unstable compared to other orthogonalization methods

Eigenwerte und Eigenvektoren

Eigenvectors - Vectors which a linear transformation scales onto the original line it spans.
Eigenvalues - The corresponding scalar by which a set of eigenvectors is multiplied by.

These are very important when analyzing the effects of linear transformations without a standard basis. For example one can find the axis of a rotation of an arbitrary 3D rotation (vectors with eigenvalue 1). LTD: Possible applications

Some transformations, for example rotations in 2D, have no non-zero eigenvalues.

Finding Eigenvalues / Vectors

The eigenvalues λ and eigenvectors $\boldsymbol{v}$ of a matrix $\boldsymbol{A}$ relate as follows:

$$\begin{gathered} \boldsymbol{Av} = \lambda \boldsymbol{v} \\ \boldsymbol{Av} = \lambda \boldsymbol{Iv} \\ (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = 0 \end{gathered}$$

The infinite non-zero solutions (eigenvectors) for $\boldsymbol{v}$ are only possible when $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$ (most of space is projected onto a single line / plane; down a dimension). This results in a polynomial of degree n TODO: Order/ trace of the matrix? which can be solved by factorising.

Once we find the eigenvalues of the matrix, the matrix $\boldsymbol{A} - \lambda \boldsymbol{I}$ can be computed for each one and finding each set of eigenvectors simply becomes the task of finding the different nullspaces:

$$\begin{gathered} \boldsymbol{A}_{\boldsymbol{\lambda_1}} \boldsymbol{v} = 0 \\ \boldsymbol{A}_{\boldsymbol{\lambda_2}} \boldsymbol{v} = 0 \\ ... \end{gathered}$$

LTD: Implications of complex eigenvalues?

TODO: Diagonalising matrices, diagonal entries are the eigenvalues, eigenbasis

Upcoming

Determinant - The factor by which a linear transformation (usually represented as a matrix) changes any area / volume in space. Can only be computed for square matrices.

Non-Zero determinant - No information is lost, there is precisely one transformation which reverses the effects on space (inverse matrix)

Next 3B1B Video - Dot products and duality

LTD:
- 4D, Flatland trick, hyperplane, cube etc