Linear algebra

This book was originally written for my teaching at Charles University in Prague. It covers the basic linear algebra topics including linear systems, vectors, matrices, linear mappings and dot products. Contains many relations between them and geometric interpretations.

Contents

Linear Algebra OrgPage **1**

Filtering . . . 1

Systems of linear equations **2**

Systems of linear equations: Ax=b . . . 2

Matrix notation . . . 2

History . . . 3

Analysis of dynamical systems . . . 3

Operations . . . 4

Gaussian elimination . . . 5

Examples of elimination . . . 6

Geometry of elementary operations . . . 6

Geometry of Gaussian elimination . . . 7

Iterative methods . . . 8

Geometric structure of the solution set . . . 9

Row interpretation . . . 10

Column interpretation . . . 11

Vector spaces **12**

Geometry of vectors . . . 12

Properties of vectors in $\mathbb{R}^n$ . . . 12

Groups . . . 13

Fields . . . 14

Finite fields . . . 15

Vector spaces . . . 16

Vector subspaces . . . 16

Affine subspaces . . . 17

Linear combinations . . . 17

Linear span and spanning sets . . . 18

Linear dependence and independence . . . 19

Bases . . . 19

Constructing a basis . . . 20

Exchange lemma . . . 21

Steinitz exchange theorem . . . 21
Dimension . . . 22
Why use different bases? . . . 22
Quotient spaces . . . 23
Polynomial interpolation . . . 24
Lagrange interpolation . . . 24

Matrices **26**

Matrices . . . 26
Matrix multiplication . . . 26
Matrices of row operations . . . 27
Matrix inverses . . . 28
Invertible matrices . . . 28
Determinant . . . 28
Transpose . . . 29
Rank of the transpose . . . 29
Matrix rank . . . 29
Fundamental subspaces . . . 30
Special matrices . . . 31
Diagonal matrices . . . 31
Eigenvalues . . . 32
Triangular matrices . . . 32
Permutation matrices . . . 33
Block matrices . . . 33
Rank 1 . . . 34
SVD . . . 34
LU decomposition . . . 34
Symmetric matrices . . . 35
Normal matrices . . . 35

Linear maps **36**

Linear maps (homomorphisms) . . . 36
Geometric examples $\mathbb{R}^2 \to \mathbb{R}^2$ . . . 36
Differentiation and integration . . . 38
Matrix representation $f : x \mapsto Ax$ . . . 38
Change-of-basis matrices . . . 40
Similarity . . . 40
Composition of maps . . . 41
Inverse maps . . . 42

Endomorphisms, isomorphisms, automorphisms . . . 42
Finding preimages . . . 43
Isomorphism theorem . . . 44
Homomorphism theorem . . . 44
Adjoint map . . . 45

Inner product **47**

Inner product, norm, and orthogonality . . . 47
Inner product . . . 47
Cauchy–Schwarz inequality . . . 48
Norm . . . 49
Nonstandard norms on $\mathbb{R}^2$ . . . 50
Nonstandard inner products . . . 51
Positive definite matrices . . . 51
Quadratic forms and definiteness . . . 52
Orthogonality . . . 52
Orthogonal complement . . . 53
Orthonormal bases . . . 53
Fourier transform . . . 54
Orthogonal matrices . . . 54
Gram–Schmidt orthogonalization . . . 54
QR decomposition . . . 55
Orthogonal projection . . . 55
Method of least squares . . . 56
Fundamental subspaces . . . 56

Linear Algebra OrgPage

Author: Pavel Klavík

Last updated: 13th October 2026

Acknowledgements: Zdeněk Hedrlín, Peter Zeman, and the students in the advanced exercise classes.

This is English translation of Czech Lineární algebra was generated by GPT 6, including translated images. The original Czech text was written around 2016, and was used for teaching linear algebra at Charles University in Prague.

The following main topics are currently covered:

- Systems of linear equations A**x**=**b**.
- Matrices.
- Vector spaces.
- Linear maps (homomorphisms).

The following topics are in progress:

- Inner products, norms, and orthogonality.
- Determinants.
- Eigenvalues.
- Quadratic forms and definiteness.

Sources

The content is based on the teaching in the Advanced Exercise Classes in Linear Algebra and the books A Discussion of Linear Algebra and Linear Algebra II: Advanced Exercises (from which almost all the illustrations are taken).

Filtering

- All nodes
- Key nodes only
- Individual topics:
  - Systems of equations
  - Vectors
  - Matrices
  - Linear maps
  - Inner product

Systems of linear equations

Systems of linear equations: Ax=b

Let us begin with an example of a system of three linear equations in three unknowns:

$$\begin{aligned} 2x &- y &&= -1, \\ -x &+ 2y &- z &= 3, \\ &- y &+ 2z &= 1. \end{aligned}$$

We seek the **set of all solutions**: the triples (x, y, z) that satisfy all three equations simultaneously when substituted into them. In this case, **the only solution is the triple** $(1, 3, 2)$.

A general system of *m* linear equations in *n* unknowns:

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & a_{1,3}x_3 & + & \cdots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & a_{2,3}x_3 & + & \cdots & + & a_{2,n}x_n & = & b_2 \\ a_{3,1}x_1 & + & a_{3,2}x_2 & + & a_{3,3}x_3 & + & \cdots & + & a_{3,n}x_n & = & b_3 \\ \vdots & & \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & a_{m,3}x_3 & + & \cdots & + & a_{m,n}x_n & = & b_m \end{array}$$

The numbers $a_{i,j}$ and b_j are fixed and are called **coefficients**. The numbers x_j are called **unknowns**, and we want to find their values.

We seek the **solution set** consisting of the n-tuples $(x_1, \ldots, x_n)$ that, when substituted into the equations, **simultaneously satisfy** all of them.

Matrix notation

We can simplify the notation for a system by omitting unnecessary information:

- The variable names are the same within each column, so we omit them.
- We remove the symbols + and =.

Only the **coefficients** matter when solving the system. We arrange them in a table called a **matrix**.

Original system:

$$\begin{aligned} 2x &- y &&= -1, \\ -x &+ 2y &- z &= 3, \\ &- y &+ 2z &= 1. \end{aligned}$$

Matrix notation:

$$\left(\begin{array}{ccc|c} 2 & -1 & & -1 \\ -1 & 2 & -1 & 3 \\ & -1 & 2 & 1 \end{array}\right)$$

In the language of matrices and vectors, we can express the system in the elegant form $A\boldsymbol{x} = \boldsymbol{b}$. Here A is the coefficient matrix, $\boldsymbol{x}$ is the vector of unknowns, and $\boldsymbol{b}$ is the vector of right-hand sides.

$$A = \begin{pmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{pmatrix}, \qquad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix}.$$

Expanding the **definition of matrix multiplication** gives us a system of linear equations.

Interpreting **matrix multiplication in terms of the columns of** A leads to the column interpretation of the system.

History

Solving systems of linear equations is an age-old problem with many applications. Specific systems of linear equations appeared in mathematics more than five thousand years ago. The Chinese book **The Nine Chapters**, written around two hundred years before our era, gives a procedure for solving **a particular system of three linear equations in three unknowns**. There is no formal justification, but generalizing the procedure gives Gaussian elimination.

In Western mathematics, Gaussian elimination was first described by **Newton in his book on algebra**, although the procedure was already widely known among mathematicians.

Gauss studied the **method of least squares** (in connection with geodesy), for which he described an algorithm similar to Gaussian elimination. The name Gaussian elimination only came into use in the 1950s, with the implementation of least squares on computers. **The history of many concepts in linear algebra is quite complicated!**

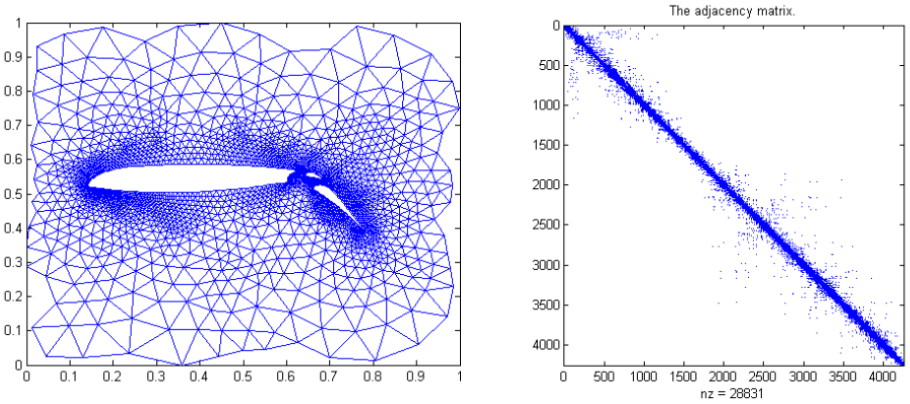
Analysis of dynamical systems

Much of mathematics is concerned with studying models of the real world. These are often **dynamical systems**, whose state (for example, a vector of real numbers) changes over time. Changes in state are described by mathematical relationships, for example locally by differential equations. In analyzing such systems, we want to address questions such as predicting their evolution from an initial state and understanding their qualitative behavior (convergence to a steady state, oscillations, or chaotic behavior).

The problem is that a typical dynamical system is too complicated for its evolution to be described exactly. In most situations, we settle for an approximate solution obtained by **discretization** (simulation, grid methods, or the finite element method), which in many cases reduces the problem to solving a **large system of linear equations**

$$A\boldsymbol{x} = \boldsymbol{b}.$$

Here the matrix A describes the discretized relationships in the system, while the vector $\boldsymbol{b}$ corresponds to the state or its evolution. The solution process often takes advantage of the particular structure of the coefficients of the matrix A.



Operations

We can derive **new information by combining equations**. We will consider two elementary operations and their combinations:

- **Multiplication by a real number:** from $A = B$ we obtain $\alpha A = \alpha B$.
- **Addition of two equations:** from $A = B$ and $C = D$ we obtain $A + C = B + D$.

Since we do not want the number of equations to **grow**, we always use the new equation to **replace** one of the previous equations, obtaining a **transformed system**. We call the operations above **elementary operations**.

We also consider **derived elementary operations**:

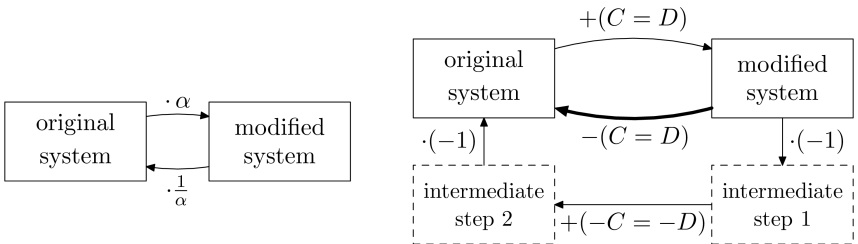
- Adding α times one equation to another.
- Interchanging two equations.

If an n-tuple $(x_1, \dots, x_n)$ solves the system, it also solves the transformed system. The converse need not hold, however: an operation may **enlarge** the solution set.

An operation is called **invertible** (or equivalence-preserving) if it leaves the solution set unchanged for every system.

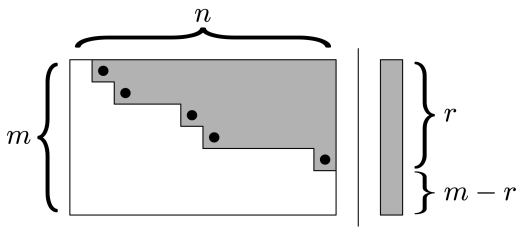
Proposition: Multiplication by a nonzero real number, addition of one equation to another, and combinations of these operations are invertible.

Proof: The operations above can be reversed: there are operations that take the transformed system back to its original form. Since the solution set can only grow, it cannot have changed.



Gaussian elimination

An algorithm for solving a system of linear equations. It is a **strategy** for applying elementary operations to simplify a matrix to echelon form.

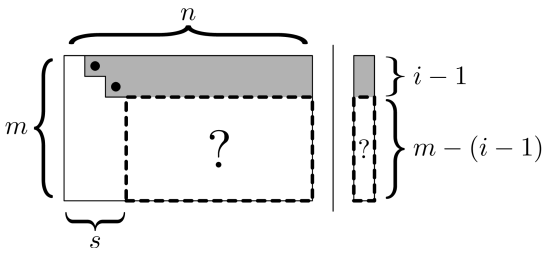


The coefficients marked in black are called **pivots**: they are the first nonzero entries in their respective rows. Gaussian elimination consists of two phases:

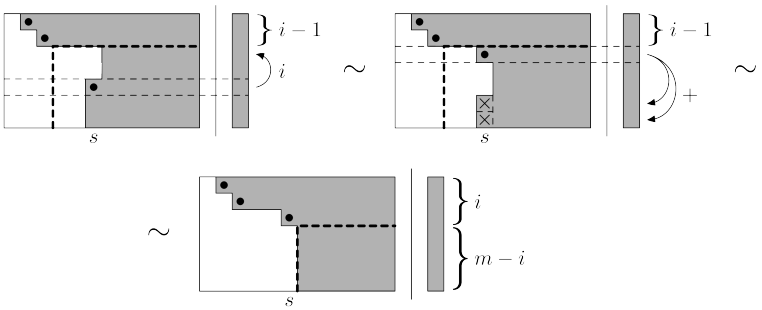
- **forward elimination**, which computes an echelon form,
- **back substitution**, which finds all solutions from the echelon form.

The echelon form is constructed in n **steps**, requiring a total of $n^3/3$ elementary arithmetic operations (multiplications and additions).

State at the beginning of step *i*:



In step *i*, we apply the following elementary operations:



Examples of elimination

Forward elimination:

$$\left(\begin{array}{ccc|c} 2 & -1 & & -1 \\ -1 & 2 & -1 & 3 \\ & -1 & 2 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ -1 & 2 & -1 & 3 \\ & -1 & 2 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & \frac{3}{2} & -1 & \frac{5}{2} \\ & -1 & 2 & 1 \end{array}\right) \sim$$

$$\left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & 1 & -\frac{2}{3} & \frac{5}{3} \\ & -1 & 2 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & 1 & -\frac{2}{3} & \frac{5}{3} \\ & & \frac{4}{3} & \frac{8}{3} \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & -\frac{1}{2} & & -\frac{1}{2} \\ & 1 & -\frac{2}{3} & \frac{5}{3} \\ & & 1 & 2 \end{array}\right).$$

Back substitution:

- First we find $z = 2$.
- We then substitute for z in the second equation and find $y = 3$.
- Finally, we substitute for y and z in the first equation and obtain $x = 1$.

Thus the system has the unique solution $(1, 3, 2)$.

Infinitely many solutions.

$$\left(\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & 6 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & 0 \end{array}\right).$$

There are infinitely many solutions of the form $(3 - 2y, y)$, where y is an arbitrary real number. They can also be written as $(3, 0) + y(-2, 1)$.

No solution.

$$\left(\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & 8 \end{array}\right) \sim \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & 2 \end{array}\right).$$

There is no solution: the operations have shown that $0x + 0y = 2$ must hold, but this is not satisfied by any pair (x, y) of real numbers.

Geometry of elementary operations

Row operations from the column perspective:

Row operations transform the space $\mathbb{R}^m$ together with the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ and the right-hand side $\boldsymbol{b}$. The key property is that, as long as the operation does not flatten the space in any direction (that is, it is invertible), the relationship between $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ and the right-hand side $\boldsymbol{b}$ is preserved.

Elementary row operations correspond to the following geometric transformations of the space $\mathbb{R}^m$:

- **Multiplying row i by a scalar α** corresponds to stretching the space by a factor of α along coordinate axis i.
- **Adding row i to row j** corresponds to shearing coordinate axis i toward coordinate axis j.

In matrix notation, let R be the invertible matrix of row operations representing the geometric transformation. The system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ becomes the transformed system $RA\boldsymbol{x} = R\boldsymbol{b}$.

Column operations from the column perspective:

A column operation transforms the individual column vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$. Their relationship to the right-hand side $\boldsymbol{b}$ may change, so the solution set changes. We can use column operations, but we must keep track of them and adjust the solution accordingly at the end.

Elementary column operations correspond to the following transformations:

- **Multiplying column** i **by a scalar** α corresponds to stretching the vector $\boldsymbol{u}_i$ by a factor of α.
- **Adding column** i **to column** j corresponds to a shear that adds $\boldsymbol{u}_i$ to $\boldsymbol{u}_j$.

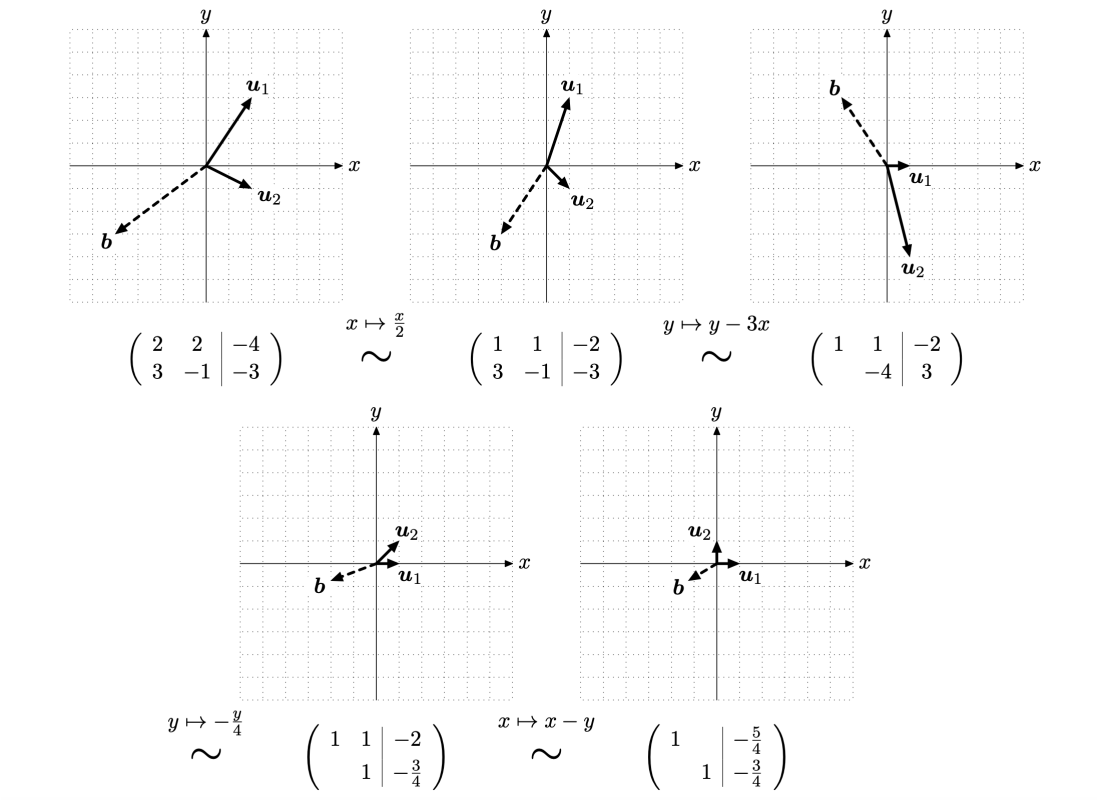
In matrix notation, let S be the invertible matrix of a column operation representing the transformation of the column vectors. The system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ becomes the transformed system $ASS^{-1}\boldsymbol{x} = \boldsymbol{b}$, leading to the system $AS\boldsymbol{y} = \boldsymbol{b}$. Once we have solved it and found $\boldsymbol{y}$, we must recover $\boldsymbol{x} = S\boldsymbol{y}$.

Geometry of Gaussian elimination

Gaussian elimination geometrically transforms the column vectors and the right-hand side by elementary row operations. It puts the system into echelon form, with column vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ and right-hand side $\boldsymbol{b}$, in which it is easy to solve.

If $\boldsymbol{b}$ points in a direction beyond those of the column vectors, we can detect this from a nonzero component along a coordinate axis on which all the column vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ have zero components. Thus $\boldsymbol{b}$ cannot be expressed in terms of the column vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$, and the system has no solution.

If all these coefficients of $\boldsymbol{b}$ are zero, a solution exists. We find it by choosing arbitrary values for the free variables, which determines the scaling of the corresponding column vectors. We can subtract these from the right-hand side to obtain a modified right-hand side $\boldsymbol{b}'$. We then compute the solution values for the pivot columns: at each stage, only one of them has a component in a particular direction of $\boldsymbol{b}'$, so its scaling factor is uniquely determined, and the corresponding scaled column vector can be subtracted.



Iterative methods

In mathematics, there are generally two kinds of computational methods. **Direct methods** (such as Gaussian elimination) carry out a long sequence of steps, at the end of which we obtain a solution. By contrast, **iterative methods** proceed in shorter steps, constructing better and better approximations to the solution.

For a system $A\boldsymbol{x} = \boldsymbol{b}$, iterative methods compute a sequence of approximate solutions $\boldsymbol{x}_0, \dots, \boldsymbol{x}_k$. We stop at step k when the error $\boldsymbol{e}_k = \boldsymbol{x} - \boldsymbol{x}_k$ is sufficiently small. Since we typically do not know the error vector $\boldsymbol{e}_k$, we test whether the residual is sufficiently small:

$$\boldsymbol{r}_k = A\boldsymbol{e}_k = A(\boldsymbol{x} - \boldsymbol{x}_k) = \boldsymbol{b} - A\boldsymbol{x}_k.$$

One crucial difference is that, if we stop an iterative method halfway through, we have at least some estimate of the solution, unlike with a direct method. (What do we know about the solution of a system if Gaussian elimination has computed only half of the echelon form?)

Splitting methods. Suppose that $A = S - T$ and that we can easily solve a system with the matrix S (say, we know S^{-1}). Then $A\boldsymbol{x} = \boldsymbol{b}$ leads to $S\boldsymbol{x} = T\boldsymbol{x} + \boldsymbol{b}$, or equivalently $\boldsymbol{x} = S^{-1}T\boldsymbol{x} + S^{-1}\boldsymbol{b}$. We can use this formula to iterate:

$$\boldsymbol{x}_{k+1} = S^{-1}T\boldsymbol{x}_k + S^{-1}\boldsymbol{b}$$

The method converges only for certain choices of the matrices S and T, and very slowly. A typical choice of S is the diagonal part of A or the upper triangular part of A. Another way to view these methods is that S approximates the matrix A.

The conjugate gradient method (CG). This constructs successive approximations by computing the Krylov subspace

$$K_k = \langle \boldsymbol{b}, A\boldsymbol{b}, A^2\boldsymbol{b}, \dots, A^{k-1}\boldsymbol{b} \rangle$$

We choose $\boldsymbol{x}_k \in K_k$ to minimize the solution error over all vectors in K_k, which can be done very efficiently. Extending the Krylov subspace requires multiplying a vector by the matrix, which takes n^2 operations. Since $n^{1/2}$ iterations typically suffice, the method is typically faster than Gaussian elimination: $n^{5/2}$ operations.

Geometric structure of the solution set

Theorem: The set of all solutions of the system $A\boldsymbol{x} = 0$ is a vector subspace, denoted by $\mathrm{Ker}(A)$.

Proof: It suffices to show that the solution set is **closed under addition and scalar multiplication**. If $\boldsymbol{x}$ is a solution of the system $A\boldsymbol{x} = 0$, then equation i gives:

$$a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = 0. \qquad (\star)$$

The solution set is closed under addition because equation i remains satisfied:

$$\begin{aligned} a_{i,1}(x_1 + y_1) + a_{i,2}(x_2 + y_2) + \cdots + a_{i,n}(x_n + y_n) &= \underbrace{a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n}_{=0 \text{ by } (\star)} \\ &\quad + \underbrace{a_{i,1}y_1 + a_{i,2}y_2 + \cdots + a_{i,n}y_n}_{=0 \text{ by } (\star) \text{ with } \boldsymbol{y} \text{ in place of } \boldsymbol{x}} \\ &= 0 + 0 = 0. \end{aligned}$$

The proof of closure under scalar multiplication is similar.

In **matrix notation**, this is even simpler. If $A\boldsymbol{x} = 0$ and $A\boldsymbol{y} = 0$, then

$$A(c\boldsymbol{x}) = c(A\boldsymbol{x}) = c \cdot 0 = 0,$$

$$A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y} = 0 + 0 = 0.$$

Theorem: The set of all solutions of the system $A\boldsymbol{x} = \boldsymbol{b}$ is the affine subspace $\mathrm{Ker}(A) + \boldsymbol{p}$, where $\boldsymbol{p}$ is any solution of the system.

Proof: It suffices to observe that, whenever we have two solutions $A\boldsymbol{x} = \boldsymbol{b}$ and $A\boldsymbol{y} = \boldsymbol{b}$, their difference $\boldsymbol{x} - \boldsymbol{y}$ belongs to $\mathrm{Ker}(A)$:

$$A(\boldsymbol{x} - \boldsymbol{y}) = A\boldsymbol{x} - A\boldsymbol{y} = \boldsymbol{b} - \boldsymbol{b} = 0.$$

The rest follows from the properties of affine subspaces.

Key consequence: Consider different right-hand sides $\boldsymbol{b}$. For some, there is no solution, which means that no vector $\boldsymbol{p}$ exists. For the others, the solution sets are affine subspaces obtained by different translations of W. Thus the solution sets of $A\boldsymbol{x} = \boldsymbol{b}$ and $A\boldsymbol{x} = \boldsymbol{c}$ have the same structure up to translation.

Let $\mathrm{Im}(A)$ denote the set of all right-hand sides for which a solution exists. This is a vector subspace of $\mathbb{R}^m$; we need only check closure under addition and scalar multiplication:

$$A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y} = \boldsymbol{b} + \boldsymbol{c}.$$

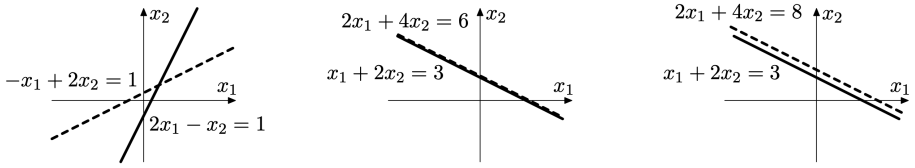
The **larger** $\operatorname{Ker}(A)$ is, the fewer distinct translates it has, and hence the **fewer** right-hand sides admit a solution. **This relationship can be described precisely in terms of dimensions:**

$$\dim \operatorname{Ker}(A) + \dim \operatorname{Im}(A) = n.$$

Row interpretation

Consider a system of m equations in n unknowns. We can identify assignments of values to the unknowns with vectors in $\mathbb{R}^n$. What does the **solution set look like geometrically?** We can proceed by finding the solution set R_i for each row $i = 1, \dots, m$. The solution set of the whole system is the intersection $R_1 \cap \dots \cap R_m$.

Examples of systems of 2 equations in 2 unknowns:



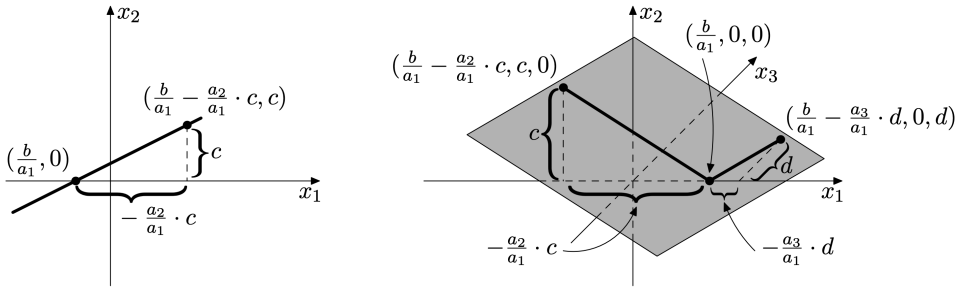
On the left, the intersection of the two lines is a single point. In the middle, the intersection is an entire line. On the right, the intersection is empty because the lines are parallel.

What does the set R_i look like in general?

Observe that we can **freely choose the values of any $n-1$ variables**, after which the value of the remaining variable is uniquely determined. Thus R_i has $n-1$ "degrees of freedom."

Geometrically, R_i is a **hyperplane translated away from the origin**: an affine subspace of dimension $n-1$, generalizing planes in $\mathbb{R}^3$. The coefficients $a_{i,1}, \dots, a_{i,n}$ determine the **direction of a normal vector**, which is perpendicular to the hyperplane. The coefficient b_i determines the **displacement of the hyperplane from the origin**.

To prove this, we would need another geometric definition of a hyperplane. Try to convince yourself, however, that it works for small values of n. In particular, for $n = 2$ (left) we obtain a line translated away from the origin, and for $n = 3$ (right) we obtain a plane translated away from the origin.



Column interpretation

Although this interpretation is often overlooked, I consider it **much more useful** than the row interpretation. The idea is to consider the **column vectors** of the matrix and the right-hand side, which belong to $\mathbb{R}^m$:

$$\boldsymbol{u}_i = \begin{pmatrix} a_{1,i} \\ a_{2,i} \\ \vdots \\ a_{m,i} \end{pmatrix}, \qquad i = 1, 2, \dots, n \qquad \text{and} \qquad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

We now make the key observation that the unknown x_i occurs precisely with the coefficients of a single column vector $\boldsymbol{u}_i$. Thus the system of linear equations is **equivalent to the vector equation**

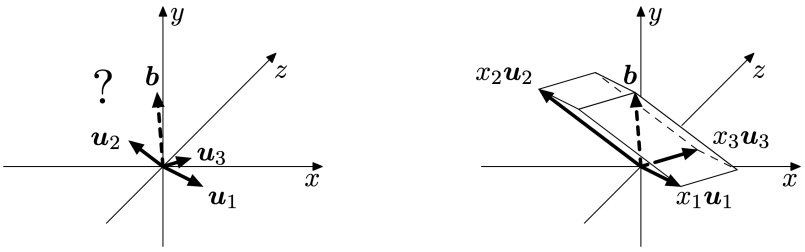
$$x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + \cdots + x_n\boldsymbol{u}_n = \boldsymbol{b},$$

where equality must hold in every component.

We therefore have the equation

$$x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + \cdots + x_n\boldsymbol{u}_n = \boldsymbol{b},$$

which we want to interpret geometrically. We seek scaling coefficients $x_1, \dots, x_n$ for the vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ so that they combine to give the right-hand side $\boldsymbol{b}$.



There may be many choices of coefficients, leading to infinitely many solutions. If the vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ point in "different" directions from $\boldsymbol{b}$, the system has no solution.

Vector spaces

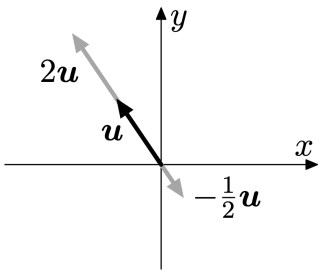
Geometry of vectors

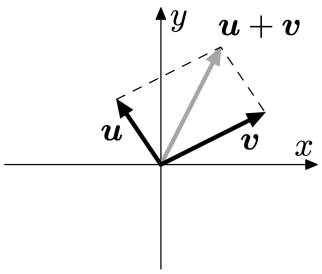
The vector space $\mathbb{R}^n$ consists of vectors that **geometrically correspond to points**. It generalizes $\mathbb{R}^2$ and $\mathbb{R}^3$. Descartes's idea was that, although we cannot see space directly, we can describe a position in it **using a triple of coordinates** (x, y, z).

Note that vectors are sometimes described as **directed line segments**. This distinction between points and vectors is historical and is no longer necessary here. Sometimes we will work with a vector as a point; at other times, it will be useful to think of it as an arrow from the origin to that point.

Scalar multiplication corresponds to **stretching a vector in a given direction**. If the coefficient is negative, the resulting vector points in the opposite direction.

The motivation for adding two vectors comes from physics, specifically Newtonian mechanics. If vectors represent forces (which have both magnitude and direction), then **addition represents combining these forces** into a single resultant force.





Properties of vectors in $\mathbb{R}^n$

If we work with the abstract definition of a vector space, several **properties of vectors are guaranteed**: vector addition is commutative and associative, the distributive laws hold for scalar multiplication, there is a zero vector, and so on.

If we use the constructive definition of $\mathbb{R}^n$, we can easily prove that the vectors **also satisfy all these properties**. For example, the zero vector is the origin $0 = (0, \dots, 0)$.

Let us illustrate the idea with at least one example.

Proposition: Vector addition in $\mathbb{R}^n$ is commutative, that is, $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.

Proof: First, recall that the vectors $\boldsymbol{u} + \boldsymbol{v}$ and $\boldsymbol{v} + \boldsymbol{u}$ are equal if and only if they agree in every component. Since these components are real numbers, their commutativity carries over to vectors:

$$(\boldsymbol{u} + \boldsymbol{v})_i = \boldsymbol{u}_i + \boldsymbol{v}_i \overset{\text{comm. } \mathbb{R}}{=} \boldsymbol{v}_i + \boldsymbol{u}_i = (\boldsymbol{v} + \boldsymbol{u})_i\,, \qquad \forall i = 1, \dots, n.$$

This proof illustrates a **general principle of algebra**. Mathematical structures are defined by combining **simpler structures with additional operations and relations**. Of course, this chain must end somewhere, so we must describe some basic structure directly. The situation resembles Russian nesting dolls, or matryoshkas.



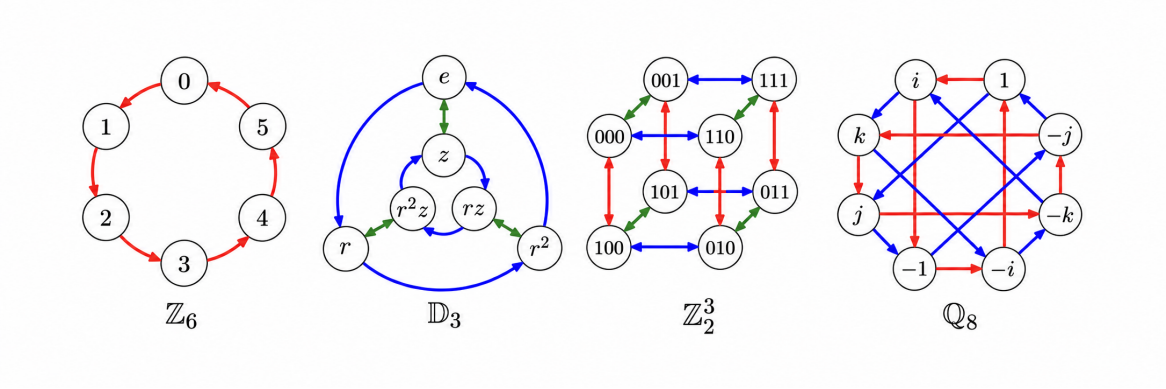
The important point is that the properties of simpler structures **carry over to the next level**, to the structure containing them. For example, vectors consist of real numbers, which are constructed from subsets of rational numbers, which are constructed from pairs of integers, which are constructed as two copies of the natural numbers. Thus the commutativity of vector addition is a consequence of the commutativity of addition of natural numbers.

Groups

A mathematical structure $\mathbb{G}$ with a **binary operation** $\circ : \mathbb{G} \times \mathbb{G} \to \mathbb{G}$ is called a **group** if it satisfies the following properties:

- Associativity holds: $a \circ (b \circ c) = (a \circ b) \circ c$.
- There is an identity element e such that $a \circ e = e \circ a = a$.
- There are inverse elements a^{-1} such that $a \circ a^{-1} = a^{-1} \circ a = e$.

Number systems often form groups; for example, $\mathbb{Z}$ is a group under addition. The structure $\mathbb{Z}_n$ consisting of the elements $\{0, 1, \dots, n-1\}$, with addition **modulo** n, is a group. These groups are quite special, however: a typical group is noncommutative.



In 1878, Cayley devised a way to visualize groups using diagrams. The vertices are the **individual elements** of the group $\mathbb{G}$. The edges describe a **set of generators** $g_1, \ldots, g_k$ for which

$$\langle g_1, \ldots, g_k \rangle = \mathbb{G}.$$

Assign a color to each generator and use color i to **draw arrows from x to $g_i \circ x$**. This captures the structure of the entire group and is much easier to follow than a table describing the binary operation $\circ$.

The groups on the left are **commutative**: the group of addition modulo 6 and the group of binary vectors. The groups on the right are **not commutative**: for example, in $\mathbb{D}_3$, following a blue arrow and then a green one is different from doing so in the opposite order. The group $\mathbb{D}_3$ consists of the **symmetries of an equilateral triangle**, with composition as the operation $\circ$. The group $\mathbb{Q}_8$ is the group of quaternion units (quaternions generalize complex numbers), and the operation describes their multiplication.

Fields

We have described almost all of linear algebra over the real numbers $\mathbb{R}$. Yet we have never used many of the special properties of $\mathbb{R}$. For example, to solve a system of linear equations by Gaussian elimination, we do not need to know that square roots or limits of Cauchy sequences exist. We can therefore replace $\mathbb{R}$ with any mathematical structure of numbers called a **field**, satisfying a few rules.

Definition of a field. A mathematical structure $\mathbb{T}$ with two **binary operations**, addition $+ : \mathbb{T} \times \mathbb{T} \to \mathbb{T}$ and multiplication $* : \mathbb{T} \times \mathbb{T} \to \mathbb{T}$, is called a **field** if it satisfies the following properties:

- $\mathbb{T}$ is a **group** under addition; denote its identity element by 0.
- $\mathbb{T} \setminus \{0\}$ is a **group** under multiplication; denote its identity element by 1.
- Distributivity holds: $a * (b + c) = a * b + a * c$.

Thus a field consists of two groups linked by distributivity.

Examples of fields. Besides $\mathbb{R}$, standard examples of infinite fields are the complex numbers $\mathbb{C}$ and the rational numbers $\mathbb{Q}$. Finite fields also exist and are completely classified.

Generalizing vector spaces. Most of this map considers vector spaces over the field of real numbers $\mathbb{R}$, that is, spaces $\mathbb{R}^n$. We can, however, consider spaces over any field $\mathbb{T}$. In the concrete definition, we obtain spaces $\mathbb{T}^n$; in the abstract definition, the scalars come from $\mathbb{T}$ instead of $\mathbb{R}$. Matrices over a field $\mathbb{T}$ are similarly considered as elements of $\mathbb{T}^{m \times n}$. For example, the space of binary vectors $\mathbb{Z}_2^n$ is useful in computer science.

Much of the linear algebra described in this map **works without any changes** for vector spaces over general fields $\mathbb{T}$. Differences arise in the following areas:

- The fundamental subspaces need not be independent; proving their independence requires orthogonality. For example, the matrix $A = (1, 1)$ over $\mathbb{Z}_2$ satisfies $R(A) = \text{Ker}(A)$.
- We can consider the standard inner product, but it need not be positive definite.
- A norm therefore does not make sense, and orthogonality behaves very differently: a nonzero vector can be perpendicular to itself, for example $(1, 1)$ over $\mathbb{Z}_2$.
- Eigenvalues need not exist, because a polynomial over a field other than $\mathbb{C}$ need not have roots.
- Definiteness of matrices does not make sense, because the field need not admit an ordering.

Finite fields

A finite field has **finitely many elements** and is called a Galois field, denoted by GF(n) when it has n elements. Finite fields are convenient for computation and have many applications, for example in cryptography and coding theory.

The simplest examples are fields of prime order:

Proposition: The structure $\mathbb{Z}_n$ consists of the elements $\{0, 1, \dots, n-1\}$, with addition and multiplication defined **modulo** n. The structure $\mathbb{Z}_n$ is a **field** if and only if n **is a prime** p.

Proof: If n is not prime, it is easy to show that multiplicative inverses do not always exist. There are $1 < a, b < n$ such that $a * b = n$. Thus $a * b = 0$ in $\mathbb{Z}_n$, which is impossible in a field.

It is harder to prove the existence of multiplicative inverses in $\mathbb{Z}_p$. It suffices to show that multiplication by k is a bijection on $\mathbb{Z}_p$.

More complicated finite fields also exist with other sizes, for example GF(4).

$+$	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

$\cdot$	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

$\mathbb{GF}(4)$

Theorem: Every finite field has order p^k, where p is a prime and k is a natural number. Moreover, for every p^k, such a field exists and is unique up to isomorphism, so the notation GF(p^k) makes sense.

Proof sketch: Let $\mathbb{T}$ be a finite field; we want to prove that its size is p^k. The **characteristic** of $\mathbb{T}$ is the smallest number of ones that add up to zero, and it is always a prime p. These sums of ones form a subfield isomorphic to $\mathbb{Z}_p$. The main trick is that $\mathbb{T}$ forms a **vector space** over the field $\mathbb{Z}_p$. It has some dimension k, and therefore contains p^k elements.

The uniqueness of a field of size p^k follows from the **isomorphism theorem**. Constructing such a field is more involved and uses polynomial rings and splitting fields; we omit the details.

Vector spaces

One way to view linear algebra is as the study of vector spaces and their transformations (linear maps). Vector spaces are therefore among the most fundamental structures in mathematics. They consist of points called **vectors**, on which we can perform two operations: addition and scalar multiplication.

There are **two possible approaches** to defining a vector space. (These reflect two different approaches to describing mathematical objects in general.) The first is by construction: we **specify exactly** what the vectors are and how the operations act on them. The second, abstract approach is to **describe their properties**. We say that a vector space consists of a set of vectors together with operations satisfying a given list of properties.

There are two possible ways to define a vector space.

Concrete construction: The vector space $\mathbb{R}^n$ consists of n-tuples $\boldsymbol{x} = (x_1, \dots, x_n)$, called **vectors**, with addition and scalar multiplication defined componentwise:

- $\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \dots, x_n + y_n)$.
- $c \cdot \boldsymbol{x} = (c \cdot x_1, \dots, c \cdot x_n)$.

Abstract definition: A vector space consists of a set of elements called **vectors**, together with addition $+$ and scalar multiplication $\cdot$, satisfying the following properties:

- Commutativity of $+$: $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
- Associativity of $+$: $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$.
- Identity element for $+$: There is a vector 0 such that $\boldsymbol{u} + 0 = \boldsymbol{u}$ for every vector $\boldsymbol{u}$.
- Inverses for $+$: For every vector $\boldsymbol{u}$, there is a vector $-\boldsymbol{u}$ such that $\boldsymbol{u} + (-\boldsymbol{u}) = 0$.
- Associativity of $\cdot$: $(a \cdot b) \cdot \boldsymbol{u} = a \cdot (b \cdot \boldsymbol{u})$.
- Identity element for $\cdot$: $1 \cdot \boldsymbol{u} = \boldsymbol{u}$.
- First distributive law: $(a + b) \cdot \boldsymbol{u} = a \cdot \boldsymbol{u} + b \cdot \boldsymbol{u}$.
- Second distributive law: $a \cdot (\boldsymbol{u} + \boldsymbol{v}) = a \cdot \boldsymbol{u} + a \cdot \boldsymbol{v}$.

Vector subspaces

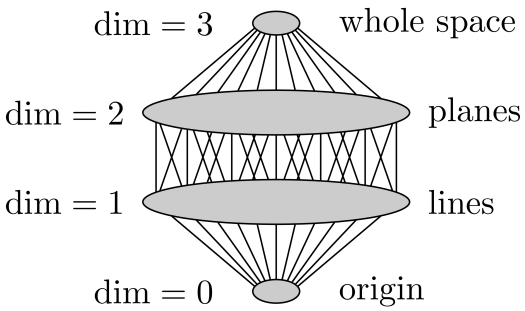
A vector subspace W is a **nonempty set of vectors closed under the operations**. More precisely, whenever **u** and **v** belong to W, so do $\boldsymbol{u} + \boldsymbol{v}$ and $c \cdot \boldsymbol{u}$ for every c. In particular, the origin 0 always belongs to W.

Studying vector subspaces is useful for two reasons:

- They have **special structural properties**; for example, they are vector spaces in their own right.
- We encounter them **at every turn** in linear algebra.

Geometrically, vector subspaces of $\mathbb{R}^n$ are lines, planes, and their higher-dimensional generalizations passing through the origin.

Consider **all vector subspaces ordered by inclusion**. For example, for $\mathbb{R}^3$ we obtain the Hasse diagram below:



We see that the subspaces are divided **into layers according to dimension**.

Proposition: The vector subspaces P form a **complete lattice**: every collection of vector subspaces S has an **infimum** and a **supremum**. The infimum is the intersection of all the subspaces in S, where the intersection of an empty collection is the whole space. The supremum is defined as the infimum of all subspaces containing every subspace in S, that is, as their intersection.

$$\inf(S) = \bigcap_{X \in S} X \qquad \text{and} \qquad \sup(S) = \bigcap_{W \in P, S \subseteq W} W.$$

If defining the supremum in terms of an infimum seems strange, think about why, in general, **it is enough to establish the existence of all infima or all suprema**: the existence of the other automatically follows.

Affine subspaces

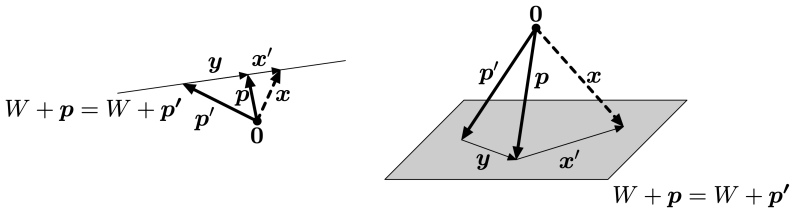
Affine subspaces **generalize vector subspaces**. An affine subspace $W + \boldsymbol{p}$ is obtained from a vector subspace W by translating it away from the origin by a vector $\boldsymbol{p}$:

$$W + \boldsymbol{p} = \{\boldsymbol{x} + \boldsymbol{p} \mid \boldsymbol{x} \in W\}.$$

We also regard the **empty set** as an affine subspace. For example, the affine subspaces of $\mathbb{R}^3$ are the empty set, all points, all lines, all planes, and the whole space.

Proposition: We have $W + \boldsymbol{p} = W + \boldsymbol{q}$ if and only if $\boldsymbol{p} - \boldsymbol{q} \in W$.

Proof:



Moreover, $W + 0 = W$, so $W + \boldsymbol{p} = W$ if and only if $\boldsymbol{p} \in W$.

Affine subspaces have a **very interesting structure**; see quotient spaces for more details.

Linear combinations

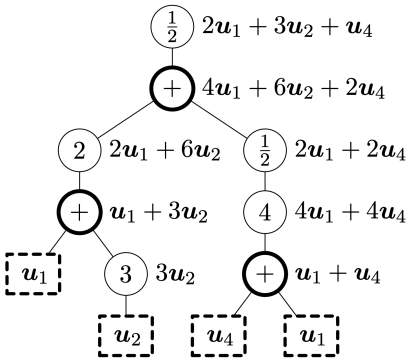
The expression $c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$ is called a **linear combination** of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

The resulting vector is also called a linear combination.

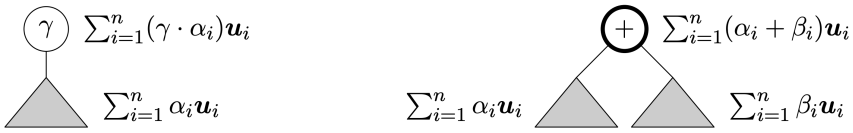
The motivation for studying linear combinations is that they are universal operations on vectors:

Proposition: Every finite sequence of operations is equivalent to some linear combination.

$$o'(\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4, \boldsymbol{u}_5) = \frac{1}{2}\left(2\left(\boldsymbol{u}_1 + 3\boldsymbol{u}_2\right) + \frac{1}{2}\left(4\left(\boldsymbol{u}_4 + \boldsymbol{u}_1\right)\right)\right)$$

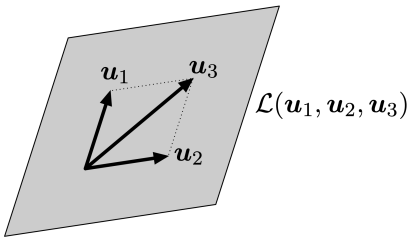


Proof: By induction on the tree representing the sequence of operations:



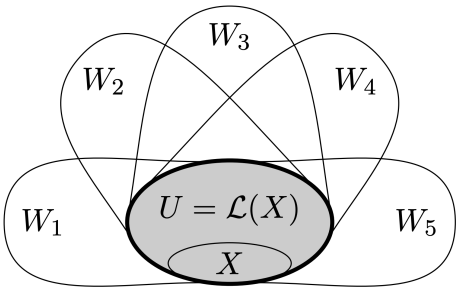
Linear span and spanning sets

The linear span $\langle X \rangle$ of a set of vectors X is the **set of all linear combinations** of vectors in X. By the lemma, these are all the vectors that can be obtained from X by finite sequences of operations. The linear span is sometimes also denoted by $L(X)$ or $\operatorname{span}(X)$. We say that X **spans** $\langle X \rangle$.



Observation: The linear span $\langle X \rangle$ is a **vector subspace**.

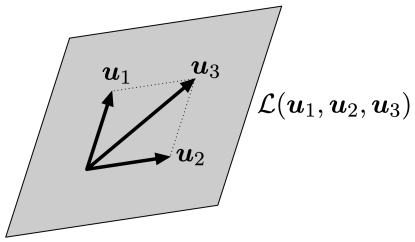
Proposition: The linear span $\langle X \rangle$ is the **smallest vector subspace** containing all vectors in X, an infimum in the lattice of subspaces. Thus $\langle X \rangle$ is the intersection of all vector subspaces containing X.



Linear dependence and independence

A vector $\boldsymbol{u}$ is **redundant** in a set X if $\langle X \setminus \{\boldsymbol{u}\} \rangle = \langle X \rangle$. We can remove redundant vectors from the set one by one without changing its linear span. The motivation is to describe the span $\langle X \rangle$ using **as few vectors as possible**.

Consider a plane spanned by three vectors. Any one of them is redundant: removing it leaves the span unchanged. After one is removed, the remaining two vectors are no longer redundant.



Notice that the origin 0 is always redundant.

A set of vectors X is **linearly dependent** if it contains a redundant vector, and **linearly independent** otherwise.

Proposition: The following definitions of linear dependence of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are equivalent:

1. The set contains a **redundant vector**.
2. There is a vector $\boldsymbol{x}_i$ that is a **linear combination of the others**.
3. There is a **nontrivial linear combination giving zero**: $0 = c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n$, with at least one coefficient c_i nonzero.
4. There is a vector in the span that can be expressed as **two different linear combinations**.

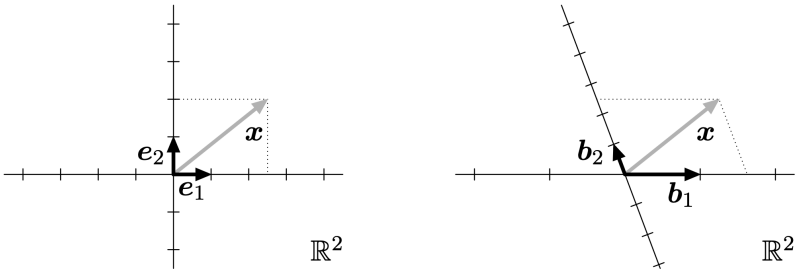
Bases

A set of vectors is called a **basis** if it is **linearly independent** and **spans the whole space**. We thus want as few vectors as possible that allow us to describe the vector space. Similarly, we can consider bases of any vector subspace.

Every basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ defines a **coordinate system on the vector space**:

If $\boldsymbol{x} = c_1\boldsymbol{b}_1 + \ldots + c_n\boldsymbol{b}_n$, its **coordinates** are $(c_1, \ldots, c_n)$.

Different bases define different coordinate systems. For example, with respect to the standard basis on the left, the vector $\boldsymbol{x}$ has coordinates $(5/2, 2)$, while with respect to the basis on the right, its coordinates are $(8/5, 5/2)$.



The following property follows from the linear independence of a basis:

Proposition: Every vector $\boldsymbol{x}$ has **uniquely determined coordinates**.

Constructing a basis

The whole theory of bases would be uninteresting if no basis existed. Under the following assumption, we will prove that **a basis always exists** (and, in fact, there are many ways to choose one).

Assumption: The vector space is **finite-dimensional**, so every basis has **finite size**. (By the Steinitz theorem, it would even suffice for one finite basis to exist.)

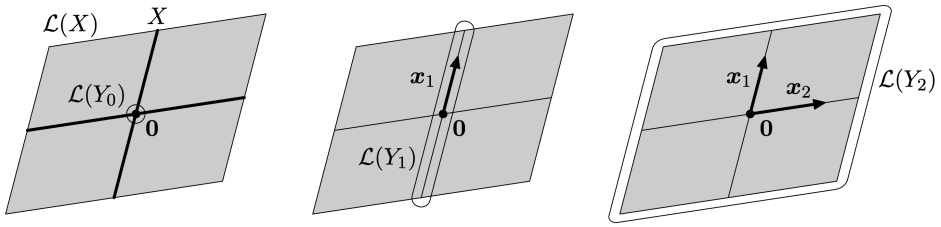
Without this assumption, we would also have to consider **infinite-dimensional vector spaces** (such as spaces of functions or sequences), which require **infinite bases**. Conceptually, the idea is the same, but there are various technical difficulties; for example, we would have to consider infinite linear combinations.

An idea that does not work: removing vectors from a spanning set. Start with any spanning set of the vector space (for example, the whole vector space U) and remove redundant vectors one by one. We end up with a linearly independent spanning set, that is, a basis.

The problem is that a spanning set is typically infinite, even in a finite-dimensional vector space. We are removing infinitely many vectors, and we may even end up with the empty set (a difficulty with infinity). For example, in $\mathbb{R}$, we could successively remove all vectors of the line, since each is redundant relative to those remaining.

An idea that works: adding vectors. We therefore proceed in the opposite direction: start with the empty set and add vectors one by one while preserving linear independence.

Let $Y_0 = \emptyset$, and let Y_k be the independent set of k vectors constructed so far. Stop if $\langle Y_k \rangle$ is the whole space U. Otherwise, choose any vector $\boldsymbol{y}_{k+1} \in U \setminus \langle Y_k \rangle$ and continue. By our assumption, the process stops after finitely many steps and constructs a basis of U.



Exchange lemma

A simpler version of the Steinitz theorem. It says that one vector in a set can be replaced by any nonzero vector in its span.

Lemma: Let X be **any set** of vectors, and let $\boldsymbol{y} \neq 0$ belong to $\langle X \rangle$. Then there exists $\boldsymbol{z} \in X$ such that

$$\big\langle X \setminus \{\boldsymbol{z}\} \cup \{\boldsymbol{y}\} \big\rangle = \langle X \rangle.$$

Proof: Since $\boldsymbol{y}$ belongs to $\langle X \rangle$, we have:

$$\boldsymbol{y} = \alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \cdots + \alpha_n \boldsymbol{x}_n.$$

Hence $\langle X \rangle = \big\langle X \cup \{\boldsymbol{y}\} \big\rangle \supseteq \big\langle X \setminus \{\boldsymbol{z}\} \cup \{\boldsymbol{y}\} \big\rangle$.

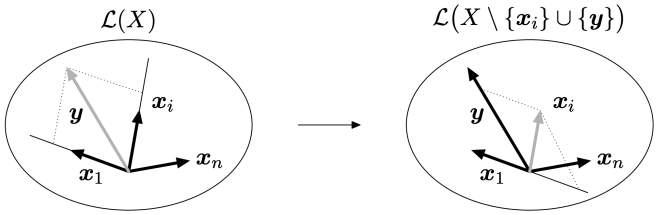
For the reverse inclusion, choose $\boldsymbol{z}$ to be any vector $\boldsymbol{x}_i$ with $\alpha_i \neq 0$. A simple rearrangement shows that $\boldsymbol{x}_i$ can be expressed in terms of the other $\boldsymbol{x}_j$ and $\boldsymbol{y}$:

$$\boldsymbol{x}_i = \frac{\boldsymbol{y} - \alpha_1 \boldsymbol{x}_1 - \cdots - \alpha_{i-1}\boldsymbol{x}_{i-1} - \alpha_{i+1}\boldsymbol{x}_{i+1} - \cdots - \alpha_n \boldsymbol{x}_n}{\alpha_i} = \\ = \frac{1}{\alpha_i}\boldsymbol{y} - \frac{\alpha_1}{\alpha_i}\boldsymbol{x}_1 - \cdots - \frac{\alpha_{i-1}}{\alpha_i}\boldsymbol{x}_{i-1} - \frac{\alpha_{i+1}}{\alpha_i}\boldsymbol{x}_{i+1} - \cdots - \frac{\alpha_n}{\alpha_i}\boldsymbol{x}_n.$$

Thus

$$\langle X \rangle \subseteq \big\langle X \setminus \{\boldsymbol{x}_i\} \cup \{\boldsymbol{y}\} \big\rangle$$

.



Steinitz exchange theorem

Theorem: Let X be **any set** of vectors, and let Y be a **linearly independent set** of vectors in $\langle X \rangle$. Then **there exists** $Z \subseteq X$ with $|Z| = |Y|$ such that

$$\langle X \setminus Z \cup Y \rangle = \langle X \rangle.$$

This theorem has several interesting consequences:

- A vector space **has many bases**, and there is nothing special about their individual vectors.
- Any linearly independent set **can be extended** to a basis.
- We always have $|X| \geq |Y|$. Thus every linearly independent set is no larger than any spanning set, and **all bases have the same size**.

Proof: Let $Y = \{y_1, \ldots, y_n\}$. Using the exchange lemma, insert the vectors of Y into X one by one. A potential problem is that, when inserting y_i, the lemma might remove a previously inserted vector y_j.

Recall, however, that **in the exchange lemma we can choose z** to be any x_k whose coefficient satisfies $\alpha_k \neq 0$. Thus, whenever we insert y_i, we remove one of the vectors of X. If this were impossible, y_i would be a linear combination of the other vectors in Y, contradicting the **linear independence of** Y.

Dimension

The **dimension of a vector space** or subspace is the size of any of its bases. The dimension of a vector space V is denoted by $\dim(V)$.

The fact that dimension is **well-defined** follows from the Steinitz exchange theorem. Take any two bases X and Y. We can exchange the vectors of X for those of Y and vice versa, so they must have the same number of elements.

Observation: If U is a subspace of V, then $\dim(U) \leq \dim(V)$.

Why use different bases?

It is natural to ask why we want to consider different bases rather than simply using the natural standard basis:

- For $\mathbb{R}^n$, it is natural to use the **standard basis**, because it has a special role in terms of the components of vectors (which is why it is also called canonical). But an abstract vector space V has no natural standard basis. There are many bases, with no special one that we would always want to choose. It is therefore better to work with a **theory of general bases**.
- Many problems become **much simpler with a suitable choice of basis**. A linear map is represented by different matrices for different choices of bases. One of the central questions of linear algebra is which basis to choose to make the representing matrix as simple as possible. This leads to eigenvalues and, for example, the singular value decomposition (SVD). Another example is Lagrange interpolation.
- We often seek a **suitable basis** that describes a given situation as well as possible. For example, in statistics we may have data represented as vectors in a high-dimensional space. We seek a basis in which the data can be described as simply as possible: perhaps almost all of it lies in a plane or a subspace of low dimension. (Measurement errors typically prevent this from holding exactly, but linear algebra can handle that using SVD.) A useful metaphor, even outside linear algebra, is that of **processes with a hidden basis**. These are processes that seem terribly complicated until a suitable **hidden basis** is discovered, in which their description becomes very simple. For example, planetary motion is simple from the perspective of universal gravitation.

The Fibonacci numbers form the sequence defined by the recurrence

$$f_0 = 0, \quad f_1 = 1, \quad f_{n+2} = f_{n+1} + f_n.$$

This is an interesting sequence with many beautiful mathematical properties. By choosing a suitable basis, we will find the following formula for the nth Fibonacci number:

$$f_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n.$$

Consider the space of all Fibonacci-type sequences: sequences satisfying the recurrence with arbitrary values of the first two terms (a_0, a_1). These form a vector space of dimension two.

We seek a basis consisting of sequences with a simple formula for the nth term: $a_n = x^n$ for some real number x. Since such a sequence must satisfy the recurrence, we have $x^{n+2} = x^{n+1} + x^n$, or equivalently $x^2 = x + 1$. The solutions give two Fibonacci-type sequences forming a basis of the space:

$$a_n = \left(\frac{1+\sqrt{5}}{2}\right)^n, \qquad b_n = \left(\frac{1-\sqrt{5}}{2}\right)^n.$$

To obtain the formula, it remains to compute the coordinates of the vector $f_n = (0, 1) = \alpha a_n + \beta b_n$:

$$\begin{aligned} \alpha &+ \beta &&= 0, \\ \frac{1+\sqrt{5}}{2}\alpha &+ \frac{1-\sqrt{5}}{2}\beta &&= 1. \end{aligned}$$

The result is $\alpha = 1/\sqrt{5}$ and $\beta = -1/\sqrt{5}$.

Quotient spaces

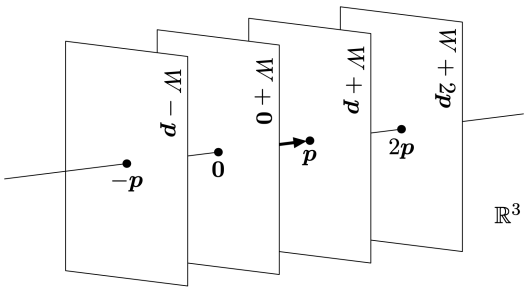
Affine subspaces have an interesting structure.

Proposition: The set of all affine subspaces obtained by translating W forms a vector space.

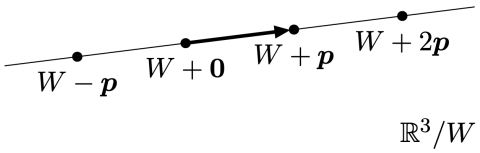
Proof: We need only define the vector operations:

- Addition: $(W + \boldsymbol{p}) + (W + \boldsymbol{q}) = W + (\boldsymbol{p} + \boldsymbol{q})$.
- Scalar multiplication: $c \cdot (W + \boldsymbol{p}) = W + c\boldsymbol{p}$.

Check that they satisfy the definition of a vector space.



These vector spaces are important and are called **quotient spaces**. Let V be a vector space and W a subspace. The vector space of all affine subspaces obtained by translating W is denoted by V/W and is called the **quotient space**, or simply the **quotient**.



We can picture the quotient space V/W by **"collapsing" each translate of W in V to a single vector**.

Proposition: We have $\dim(W) + \dim(V/W) = \dim(V)$.

Polynomial interpolation

Systems of linear equations can solve a problem that at first sight seems completely unrelated to them. Consider polynomials

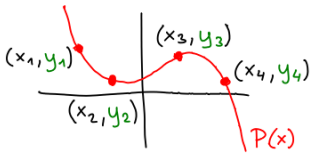
$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

of degree n. The following algebraic theorem holds, closely related to the fundamental theorem of algebra.

Theorem: There is **exactly one polynomial $P(x)$ of degree n** passing through any given $n+1$ points

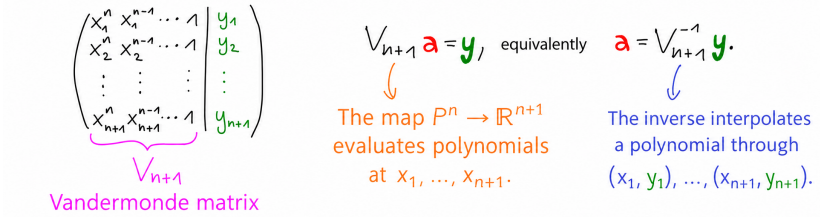
$$(x_1, y_1), ..., (x_{n+1}, y_{n+1}),$$

where $x_i \neq x_j$ for $i \neq j$, that is, $P(x_i) = y_i$.



Finding the coefficients of this polynomial is called **polynomial interpolation**, and can be done by solving a **system of linear equations**.

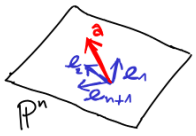
To find the coefficients, **it suffices to solve the following system**, where $\boldsymbol{y} = (y_1, \ldots, y_{n+1})$ and $\boldsymbol{a}$ is the vector of unknown coefficients of the polynomial $P(x)$.



The uniqueness theorem for interpolation can be proved using linear algebra. It suffices to show that the **Vandermonde matrix V_{n+1} is invertible** if and only if $x_i \neq x_j$ for $i \neq j$. Another proof follows from **Lagrange interpolation**.

Lagrange interpolation

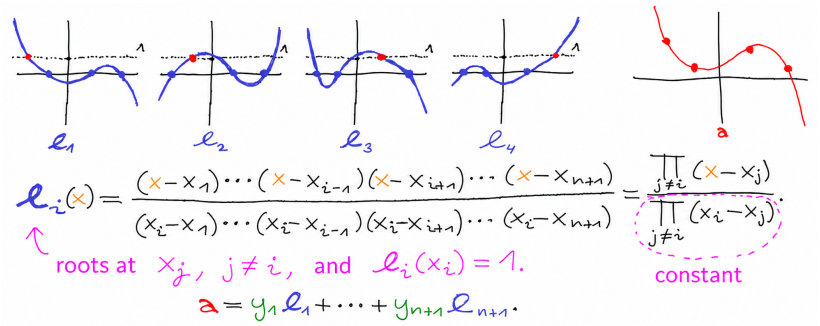
To solve the interpolation problem, we seek a **coefficient vector $\boldsymbol{a}$** in the space $\mathbb{P}_n$ of polynomials of degree at most n. Finding its **coordinates in the standard basis is difficult** and leads to solving the system



$$V_{n+1}\boldsymbol{a} = \boldsymbol{y}$$

Lagrange interpolation is based on the idea that **the problem becomes very simple with a suitable choice of basis** $\ell_1, \dots, \ell_{n+1}$.

We choose the basis $\ell_1, \dots, \ell_{n+1}$ so that each vector ℓ_i **interpolates one of the points** (x_i, y_i). A simple linear combination of these vectors gives the desired vector $\boldsymbol{a}$.



By the **Steinitz theorem**, we know that $\ell_1, \dots, \ell_{n+1}$ is a basis. Its existence **proves the uniqueness theorem for the polynomial** $P(x)$ of degree n passing through the points $(x_1, y_1), \dots, (x_{n+1}, y_{n+1})$.

Of course, other bases can be chosen. One example is **Newton interpolation**, with respect to the basis $\boldsymbol{n}_1, \dots, \boldsymbol{n}_{n+1}$. This works particularly well when adding further points, because the vector $\boldsymbol{n}_i$ depends only on the points $(x_1, y_1), \dots, (x_i, y_i)$.

Matrices

Matrices

A matrix is an $m \times n$ **array of real numbers:**

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{pmatrix}, \qquad \text{for example} \qquad A = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}.$$

An $m \times n$ matrix describes a **geometric transformation** from the vector space $\mathbb{R}^n$ to the vector space $\mathbb{R}^m$.

For a matrix A, let $(A)_{i,j} = a_{i,j}$ denote the entry in row i and column j. The rows of the matrix form **row vectors in** $\mathbb{R}^n$, and the columns form **column vectors in** $\mathbb{R}^m$.

The basic matrix operations are scalar multiplication and addition:

$$\alpha A = \begin{pmatrix} \alpha \cdot a_{1,1} & \alpha \cdot a_{1,2} & \alpha \cdot a_{1,3} & \cdots & \alpha \cdot a_{1,n} \\ \alpha \cdot a_{2,1} & \alpha \cdot a_{2,2} & \alpha \cdot a_{2,3} & \cdots & \alpha \cdot a_{2,n} \\ \alpha \cdot a_{3,1} & \alpha \cdot a_{3,2} & \alpha \cdot a_{3,3} & \cdots & \alpha \cdot a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha \cdot a_{m,1} & \alpha \cdot a_{m,2} & \alpha \cdot a_{m,3} & \cdots & \alpha \cdot a_{m,n} \end{pmatrix}.$$

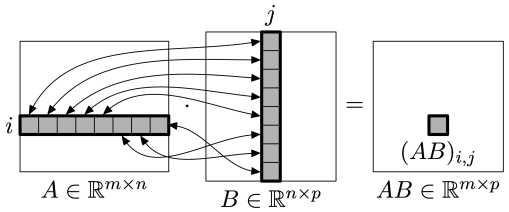
$$A + B = \begin{pmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} & a_{1,3} + b_{1,3} & \cdots & a_{1,n} + b_{1,n} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} & a_{2,3} + b_{2,3} & \cdots & a_{2,n} + b_{2,n} \\ a_{3,1} + b_{3,1} & a_{3,2} + b_{3,2} & a_{3,3} + b_{3,3} & \cdots & a_{3,n} + b_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} + b_{m,1} & a_{m,2} + b_{m,2} & a_{m,3} + b_{m,3} & \cdots & a_{m,n} + b_{m,n} \end{pmatrix}.$$

All $m \times n$ matrices therefore form a **vector space** $\mathbb{R}^{m \times n}$.

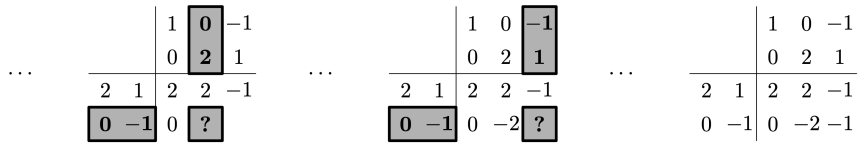
Matrix multiplication

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. Their **product** AB is an $m \times p$ matrix with the following entries:

$$(AB)_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}, \qquad \forall i = 1, \ldots, m,\ j = 1, \ldots, p.$$



Example of a matrix product: (A on the left, B above)



This is the **row-column view** of matrix multiplication. There are three other useful perspectives:

- **The column-column view:** First consider the product $A\boldsymbol{x}$. (Here we treat the vector as a matrix with one column.) By definition, the column vectors $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ of A are multiplied by the respective coefficients of $\boldsymbol{x}$ and added together. This gives the linear combination $A\boldsymbol{x} = x_1\boldsymbol{u}_1 + \cdots + x_n\boldsymbol{u}_n$. The columns of AB are **linear combinations of the columns of A, with coefficients taken from the columns of B.**
- **The row-row view:** This is the same as the previous view, but transposed: we obtain linear combinations of the rows of B using the coefficients of A.
- **The column-row view:** In the definition of the product, the entries in column i of A multiply only the entries in row i of B. Using distributivity, we can therefore pair the columns $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ of A with the rows $\boldsymbol{y}_1^\mathsf{T}, \dots, \boldsymbol{y}_n^\mathsf{T}$ of B. We obtain **the following sum of n matrices of size $m \times p$:** $AB = \boldsymbol{x}_1\boldsymbol{y}_1^\mathsf{T} + \cdots + \boldsymbol{x}_n\boldsymbol{y}_n^\mathsf{T}$.

Matrices of row operations

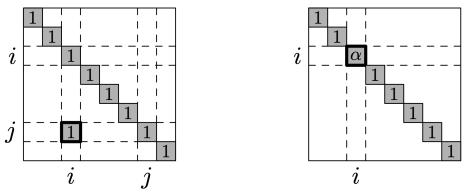
Invertible operations correspond to **invertible matrices**. The system $A\boldsymbol{x} = \boldsymbol{b}$ is transformed into the system $RA\boldsymbol{x} = R\boldsymbol{b}$, where R is an invertible matrix.

Proposition: If R is an invertible matrix, then $A\boldsymbol{x} = \boldsymbol{b}$ and $RA\boldsymbol{x} = R\boldsymbol{b}$ have the **same solution set**.

Proof: If **x** solves the original system, it certainly solves the transformed system. Thus each operation can only enlarge the solution set. Since R is invertible, there is an inverse operation R^{-1} that **takes the transformed system back to its original form:**

$R^{-1}RA\boldsymbol{x} = R^{-1}R\boldsymbol{b} \implies A\boldsymbol{x} = \boldsymbol{b}$. The **solution set must therefore be the same**; otherwise, the operation R^{-1} would make it smaller, which is impossible.

The invertible matrices R of elementary row operations are very simple. On the left is the operation of adding row i to row j. On the right is multiplication of row i by the coefficient $\alpha \neq 0$.



The reader can work out the matrices of the derived elementary operations: adding α times row i to row j, and interchanging two rows.

Matrix inverses

Let A be an $m \times n$ matrix. If there is an $n \times m$ matrix A^{-1} satisfying $AA^{-1} = I_m$, then A^{-1} is called a **right inverse** of A. It need not be unique. A **left inverse** A^{-1} is defined similarly, satisfying $A^{-1}A = I_n$.

Let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ denote the columns of a right inverse A^{-1}. **They must satisfy the systems**

$$A\boldsymbol{u}_i = \boldsymbol{e}_i, \qquad \forall i = 1, \dots, m.$$

They can be found by **a single Gaussian elimination** with multiple right-hand sides.

For example, for

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

we solve the system:

$$\left(\begin{array}{cc|cc} 2 & -1 & 1 & 0 \\ -1 & 2 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & -1/2 & 1/2 & 0 \\ -1 & 2 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{cc|cc} 1 & -1/2 & 1/2 & 0 \\ 0 & 3/2 & 1/2 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cc|cc} 1 & 0 & 2/3 & 1/3 \\ 0 & 1 & 1/3 & 2/3 \end{array}\right), \qquad A^{-1} = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

Here are several propositions about the existence of inverses; the reader can work out their proofs:

Proposition: A matrix A has a right inverse if and only if the system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every right-hand side $\boldsymbol{b}$. In other words, $\operatorname{Im}(A)$ is the whole space $\mathbb{R}^m$.

Proposition: Let A be an $m \times n$ matrix. If it has a right inverse, then $m \leq n$. If it has a left inverse, then $m \geq n$. Thus only square matrices can have a two-sided inverse.

Proposition: A matrix A has a right inverse if and only if A^{T} has a left inverse.

Proposition: An $m \times n$ matrix A has a right inverse if and only if $\operatorname{rank}(A) = m$, and a left inverse if and only if $\operatorname{rank}(A) = n$. (This follows from the nontrivial fact that $\operatorname{rank}(A) = \operatorname{rank}(A^{\mathsf{T}})$.)

Multiplying by an inverse matrix: If A has a left inverse, we can transform $A\boldsymbol{x} = \boldsymbol{b}$ into $A^{-1}A\boldsymbol{x} = A^{-1}\boldsymbol{b}$, or equivalently $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. This is not used in computations, because computing an inverse matrix is more demanding than solving the system; it is useful in theory, however.

Invertible matrices

A square matrix is called **nonsingular** if it is invertible. Other square matrices are called **singular**. The following theorem describes the inverses of nonsingular matrices:

Theorem: A square matrix has a left inverse if and only if it has a right inverse. Moreover, they are equal and uniquely determined.

The first part is the difficult one; it is a special case of the equality $\operatorname{rank}(A) = \operatorname{rank}(A^{\mathsf{T}})$. The second part follows easily from associativity. If X is a left inverse and Y is a right inverse, then:

$$X = XI_m = X(AY) = (XA)Y = I_nY = Y.$$

Determinant

Transpose

Let A be an $m \times n$ matrix. Its **transpose** A^T is the $n \times m$ matrix with entries

$$(A^\mathsf{T})_{i,j} = (A)_{j,i}.$$

Thus the entries of A^T are obtained by **reflecting** the entries of A across the main diagonal.

For matrices with **complex entries, we use a different version of the transpose**. It is denoted by A^H and also takes the complex conjugate of each entry (corresponding to a different inner product). The transpose is often also denoted by A^* (in both the real and complex cases).

Transposed matrices satisfy the **following identities**; work out their proofs from the definitions of the matrix operations:

- $(A^\mathsf{T})^\mathsf{T} = A$.
- $(cA)^\mathsf{T} = cA^\mathsf{T}$.
- $(A + B)^\mathsf{T} = A^\mathsf{T} + B^\mathsf{T}$.
- $(AB)^\mathsf{T} = B^\mathsf{T}A^\mathsf{T}$. Notice that the order of the transposed matrices **must be reversed** for the product to have **compatible dimensions**. What happens when we transpose a product of more than two matrices?
- If A is invertible, then $(A^{-1})^\mathsf{T} = (A^\mathsf{T})^{-1} = A^{-\mathsf{T}}$.

Rank of the transpose

The following theorem is one of the wonders of linear algebra. It says that, for any matrix A, we have $\dim \mathrm{Im}(A) = \dim R(A)$, even though these are entirely different spaces.

Theorem: For any matrix A, we have $\mathrm{rank}(A) = \mathrm{rank}(A^\mathsf{T})$.

Matrix rank

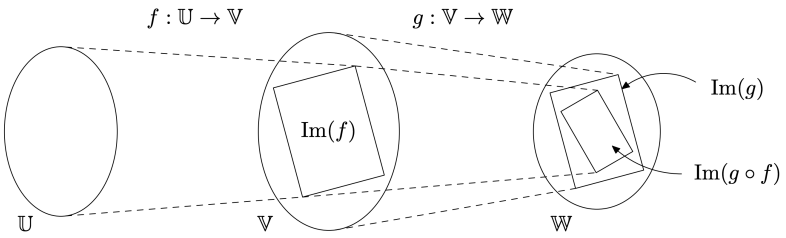
The rank of a matrix A, denoted by $\mathrm{rank}(A)$, measures how close A is to an invertible matrix. There are several equivalent definitions of $\mathrm{rank}(A)$:

- $\dim \mathrm{Im}(A)$, the number of linearly independent columns of the matrix, and also the dimension of the image of the corresponding linear map.
- $\dim R(A)$, the number of linearly independent rows of the matrix.
- The number of nonzero rows, or equivalently the number of pivots, in an echelon form of A.
- The smallest number of rank-one matrices whose sum is A.

Rank satisfies the following upper bounds:

- $\mathrm{rank}(A) \leq \min\{m, n\}$ for an $m \times n$ matrix A.
- $\mathrm{rank}(A + B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$.
- $\mathrm{rank}(AB) \leq \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}$.

The last inequality has the following proof in the language of linear maps:



Proposition: Multiplication by an invertible matrix on the left or right **does not change the rank**.

Proof: We prove this for multiplication on the left. By the inequality,

$$\mathrm{rank}(RA) \leq \mathrm{rank}(A).$$

On the other hand, since R is invertible, $A = R^{-1}RA$. Thus

$$\mathrm{rank}(A) = \mathrm{rank}(R^{-1}RA) \leq \mathrm{rank}(RA).$$

Fundamental subspaces

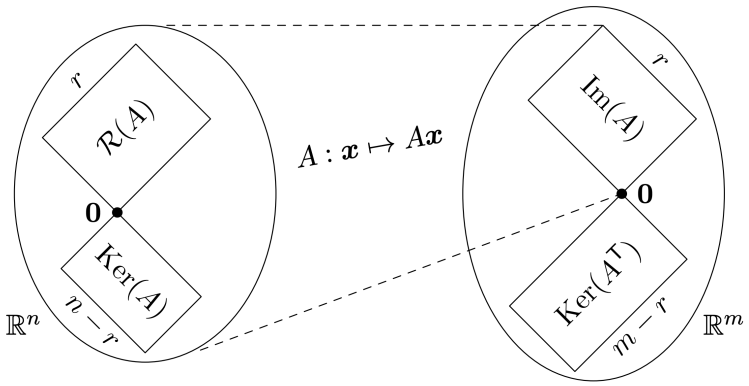
Let A be an $m \times n$ matrix with row vectors $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m$ and column vectors $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. The matrix A defines four **fundamental subspaces**:

- The **null space** $\mathrm{Ker}(A) = \{\boldsymbol{x} \,:\, A\boldsymbol{x} = 0\}$ (or **kernel**).
- The **row space** $R(A) = \mathrm{Im}(A^\mathsf{T}) = \langle \boldsymbol{r}_1, \ldots, \boldsymbol{r}_m \rangle = \{A^\mathsf{T}\boldsymbol{x} \,:\, \boldsymbol{x} \in \mathbb{R}^m\}$ (or **left image**).
- The **left null space** $\mathrm{Ker}(A^\mathsf{T}) = \{\boldsymbol{y} \,:\, \boldsymbol{y}^\mathsf{T}A = 0\}$ (or **left kernel**).
- The **column space** $S(A) = \mathrm{Im}(A) = \langle \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \rangle = \{A\boldsymbol{x} \,:\, \boldsymbol{x} \in \mathbb{R}^m\}$ (or **image**).

Notice that the kernel and row space are **subspaces of** $\mathbb{R}^n$, while the left kernel and column space are **subspaces of** $\mathbb{R}^m$.

Fundamental theorem of linear algebra:

1. Let $r = \mathrm{rank}(A)$. Then $\dim R(A) = \dim \mathrm{Im}(A) = r$, $\dim \mathrm{Ker}(A) = n - r$, and $\dim \mathrm{Ker}(A^\mathsf{T}) = m - r$.
2. $R(A)$ and $\mathrm{Ker}(A)$ are **linearly independent and together span** $\mathbb{R}^n$. Similarly, $\mathrm{Im}(A)$ and $\mathrm{Ker}(A^\mathsf{T})$ are **linearly independent and together span** $\mathbb{R}^m$.
3. Moreover, $R(A)$ and $\mathrm{Ker}(A)$ are **orthogonal complements**: every vector in $R(A)$ is **perpendicular** to every vector in $\mathrm{Ker}(A)$. Similarly for $\mathrm{Im}(A)$ and $\mathrm{Ker}(A^\mathsf{T})$.



Note that property 1 holds over any field. **Properties 2 and 3 require** $\mathbb{R}$ (and, with a slight change in the definitions, also hold over $\mathbb{C}$).

Special matrices

Some kinds of matrices occur so often in linear algebra that it is useful to give them **special names**.

They are interesting for two reasons:

1. They **occur frequently** in linear algebra and in various applications.
2. They have **special properties** that can be exploited, for example, in computations or structural analysis.

Diagonal matrices

A **diagonal matrix** D has zeros in every position off the main diagonal, so $(D)_{i,j} = 0$ for $i \neq j$. The definition also works for rectangular matrices.

$$D = \begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix}, \qquad D' = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}, \qquad D'' = \begin{pmatrix} 3 & & \\ & 0 & \\ & & -4 \end{pmatrix}.$$

The **identity matrix** I_n is the $n \times n$ diagonal matrix with ones on the diagonal.

$$I_3 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \qquad I_4 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}.$$

When diagonal matrices represent an endomorphism with respect to a single basis, they correspond to **scaling the coordinate axes**. The identity matrix represents the **identity map**.

Proposition: Every linear map f is represented, with respect to suitable bases X and Y, by a diagonal matrix with only 1 and 0 on the diagonal.

Proof: Let X_r be any **basis of the left image** $\mathrm{Im}(f^*)$ and X_k any **basis of the kernel** $\mathrm{Ker}(f)$. Choose $X = X_r \cup X_k$.

Since f defines an isomorphism between $\mathrm{Im}(f^*)$ and $\mathrm{Im}(f)$, it maps any basis of $\mathrm{Im}(f^*)$ to a basis of $\mathrm{Im}(f)$. Thus $f(X_r)$ is a basis of $\mathrm{Im}(f)$; we adjoin a basis of $\mathrm{Ker}(f^*)$ to obtain the second desired basis Y. Since $f(X_k) = \{0\}$, the map f is represented with respect to X and Y by the diagonal matrix described above.

Proposition: A square matrix A commutes with every square matrix of the same size if and only if $A = cI_n$.

Eigenvalues

Triangular matrices

We distinguish two kinds of triangular matrices. We usually consider square matrices, but the definition also works for rectangular ones.

A **lower triangular matrix** L has zeros in every position above the main diagonal. Thus $(L)_{i,j} = 0$ for $i < j$.

An **upper triangular matrix** U has zeros in every position below the main diagonal. Thus $(U)_{i,j} = 0$ for $i > j$.

$$L = \begin{pmatrix} 1 & & \\ 2 & 1 & \\ 3 & 2 & 1 \end{pmatrix}, \quad L' = \begin{pmatrix} 4 & & & \\ 0 & 0 & & \\ 0 & 1 & 0 & \\ 1 & 0 & 0 & \frac{1}{2} \end{pmatrix},$$

$$U = \begin{pmatrix} 1 & 2 & 3 \\ & 2 & 3 \\ & & 3 \end{pmatrix}, \quad U' = \begin{pmatrix} -1 & 1 & -1 & 1 \\ & -1 & 1 & -1 \\ & & -1 & 1 \\ & & & -1 \end{pmatrix}.$$

Diagonal matrices are both lower and upper triangular.

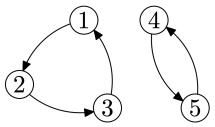
Let an upper triangular matrix U represent a map with respect to a single basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$. Since $U\boldsymbol{b}_i$ lies in $\langle \boldsymbol{b}_1, \ldots, \boldsymbol{b}_i \rangle$ (and similarly for lower triangular matrices L), many problems are considerably simpler for triangular matrices. For example, the systems $L\boldsymbol{x} = \boldsymbol{b}$ and $U\boldsymbol{x} = \boldsymbol{b}$ can be solved directly by substitution. It is also easy to find inverses and determine eigenvalues (the diagonal entries).

The reader can work out the following properties:

- **Triangular matrices are closed under multiplication**. Thus the product of lower triangular matrices is lower triangular, and similarly for upper triangular matrices.
- A triangular matrix is invertible if and only if its **diagonal entries are nonzero**.
- **Triangular matrices are closed under taking inverses**. Thus the inverse of a lower triangular matrix is lower triangular, and the inverse of an upper triangular matrix is upper triangular.
- Directly from the definition, we can see that **a square triangular matrix has an inverse on one side if and only if it has an inverse on the other side**. This is important for proving the general theorem on two-sided inverses using the LU decomposition.

Permutation matrices

Let π be a permutation of the set $\{1, \dots, n\}$. It is represented by a **permutation matrix** P: a square $n \times n$ matrix whose entries are zeros and ones, with $(P)_{i,j} = 1$ if and only if $\pi(j) = i$.



$$P = \begin{pmatrix} & & 1 & & \\ 1 & & & & \\ & 1 & & & \\ & & & & 1 \\ & & & 1 & \end{pmatrix}.$$

Let A be any matrix that we multiply by a permutation matrix P.

- Multiplication **on the left, PA, permutes the rows** of A according to π: row i moves to row $\pi(i)$.
- Multiplication **on the right, AP, permutes the columns** of A according to π^{-1}: column i moves to column $\pi^{-1}(i)$.

Let P_π denote the permutation matrix representing π. The reader can work out and prove the following properties.

The product of permutation matrices is a permutation matrix representing the composition of the corresponding permutations. Thus

$$P_\pi P_\sigma = P_{\pi \circ \sigma}.$$

Every permutation matrix P is **invertible**, and even **orthogonal**, so it has an **inverse matrix** that is also a permutation matrix. It satisfies

$$P_\pi^{-1} = P_{\pi^{-1}} = P_\pi^{\mathsf{T}}.$$

It is easy to check directly from the definition that this is a two-sided inverse.

Permutation matrices form a **group**. By Cayley's theorem, any group with n elements can be represented as a **group of permutations of** $\{1, \dots, n\}$. Hence any group with n elements can be represented by $n \times n$ **permutation matrices**. This naturally leads to the study of matrix groups.

Block matrices

Block matrices consist of smaller matrices called **blocks**. We can define operations on block matrices of compatible sizes. This lets us use the language of blocks to describe certain things efficiently.

Let A and B be two $2n \times 2n$ block matrices, each consisting of four $n \times n$ blocks.

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix},$$

These matrices can be **added and multiplied blockwise**:

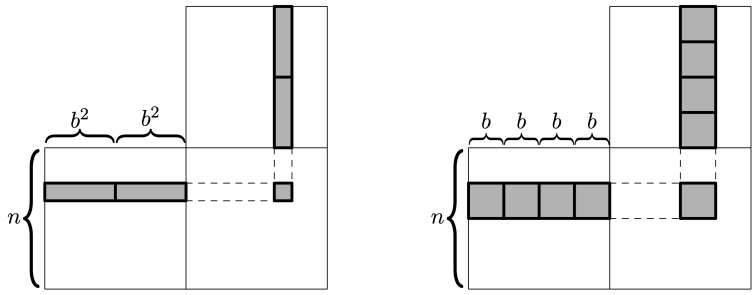
$$AB = \begin{pmatrix} A_{1,1}B_{1,1} + A_{1,2}B_{2,1} & A_{1,1}B_{1,2} + A_{1,2}B_{2,2} \\ A_{2,1}B_{1,1} + A_{2,2}B_{2,1} & A_{2,1}B_{1,2} + A_{2,2}B_{2,2} \end{pmatrix}.$$

The reader can work out the conditions under which these operations can be defined in general. We can also define **block diagonal matrices**, **block triangular matrices**, and so on.

Block multiplication has practical uses when **multiplying large matrices**. Their coefficients must be read from memory (say, RAM), which is **much slower than the computer's arithmetic.** Suppose we are multiplying $n \times n$ matrices and can load roughly $2b^2$ coefficients at a time.

If we follow the definition, we compute each entry of AB by multiplying the coefficients of a row of A and a column of B. This requires n/b^2 memory reads, giving $\boldsymbol{n^3/b^2}$ **products in total**.

Instead, we split both matrices into $b \times b$ blocks (for simplicity, suppose n is a multiple of b). To compute one block of AB, we must multiply one row of blocks of A by one column of blocks of B. This requires n/b reads. There are n^2/b^2 blocks in AB, so altogether we perform $\boldsymbol{n^3/b^3}$ **reads, which is $\boldsymbol{b}$ times more efficient**.



Rank 1

A matrix A has rank 1 if it is nonzero and every row or column is a multiple of the others. It has the elegant representation

$$A = \boldsymbol{x}\boldsymbol{y}^\mathsf{T}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are nonzero vectors.

SVD

LU decomposition

Theorem: Every $m \times n$ matrix A **has a decomposition** $PA = LU$, where

- P is an $m \times m$ **permutation matrix**,
- L is an $m \times m$ **lower triangular matrix** with ones on the diagonal,
- U is an $m \times n$ **upper triangular matrix**.

Sometimes we use the more symmetric decomposition $PA = LDU$, where both triangular matrices L and U have **ones on the diagonal**, and D is a **diagonal matrix of pivots**.

Example: Let A be the following matrix:

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

We obtain the following LU and LDU decompositions with $P = I_3$:

$$LU = \begin{pmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ & -\frac{2}{3} & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & \\ & \frac{3}{2} & -1 \\ & & \frac{4}{3} \end{pmatrix},$$

$$LDU = \begin{pmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ & -\frac{2}{3} & 1 \end{pmatrix} \begin{pmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & \\ & 1 & -\frac{2}{3} \\ & & 1 \end{pmatrix}.$$

The LU decomposition is **Gaussian elimination written in matrix form**, applied to A. The individual matrices in the decomposition have the following meanings:

- The permutation matrix P describes a **reordering of the rows of** A that allows **Gaussian elimination of** PA **without pivoting**.
- The matrix L describes the **elementary operations of forward elimination**. Since we have reordered the rows using P, we need only use **elementary operations that add** α **times** row i to row j, where $i < j$.
- The matrix U is the resulting **echelon form of** A, including the pivot values.

Let $R_1, \ldots, R_k$ be the matrices of these elementary operations, which transform PA into U. We have the matrix equality

$$R_k R_{k-1} \cdots R_2 R_1 PA = U.$$

The operation matrices R_i are **lower triangular matrices with ones on the diagonal**. Hence $L = (R_k R_{k-1} \cdots R_2 R_1)^{-1}$ is also lower triangular with ones on the diagonal. This gives the decomposition $PA = LU$.

Remark: If the matrix A is **symmetric and can be eliminated without interchanging rows**, its LDU decomposition has the special **symmetric form** $A = LDL^{\mathsf{T}}$.

Symmetric matrices

A square matrix A is called **symmetric** if $A = A^{\mathsf{T}}$. Symmetric matrices represent endomorphisms that are their own adjoints.

Symmetric matrices have many special properties; for example, they are closed under addition and taking inverses.

Normal matrices

A matrix A is called **normal** if it commutes with its adjoint: $AA^* = A^*A$. Normal matrices generalize symmetric matrices.

Linear maps

Linear maps (homomorphisms)

Let U and V be two vector spaces. A **map** $f\colon U \to V$ **is linear** if it satisfies:

$$\begin{aligned} f(\boldsymbol{x}+\boldsymbol{y}) &= f(\boldsymbol{x}) + f(\boldsymbol{y}) \quad && \forall \boldsymbol{x}, \boldsymbol{y} \in U, \\ f(\alpha \boldsymbol{x}) &= \alpha f(\boldsymbol{x}) && \forall \boldsymbol{x} \in U, \alpha \in \mathbb{R}. \end{aligned}$$

A linear map f **carries the structure of** U **into that of** V. For example, whenever an equality $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{z}$ holds, it also holds for the images: $f(\boldsymbol{x}) + f(\boldsymbol{y}) = f(\boldsymbol{z})$. Note that the images of these equalities may be very trivial.

Linear maps are called **homomorphisms**. Homomorphisms are considered between general mathematical structures, and studying them is very important. The name consists of **homo** (preserving structure) and **morphism** (a transformation from one structure to another).

First observe that every linear map f satisfies $f(0) = 0$. By linearity, the origin must map to the origin:

$$f(0) = f(0 \cdot 0) = 0 f(0) = 0.$$

Similarly, we have the following:

Proposition: The image of any linear combination is the corresponding linear combination of the images. Thus a **linear map is uniquely determined by the images of any basis**.

Proof: We use linearity:

$$f\left(\sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i\right) = \sum_{i=1}^{n} f(\alpha_i \boldsymbol{x}_i) = \sum_{i=1}^{n} \alpha_i f(\boldsymbol{x}_i).$$

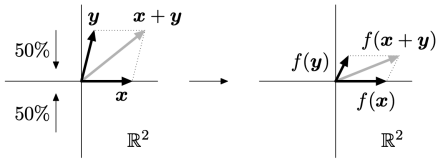
Geometric examples $\mathbb{R}^2 \to \mathbb{R}^2$

Many geometric maps $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ **are linear:**

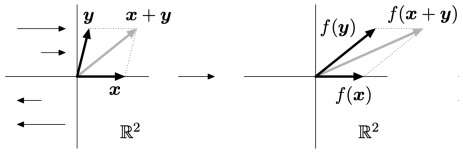
- scalings,
- rotations,
- shears,
- projections, and
- their combinations.

An example of a **nonlinear** geometric map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is translation of $\mathbb{R}^2$ by any nonzero vector $\boldsymbol{x}$.

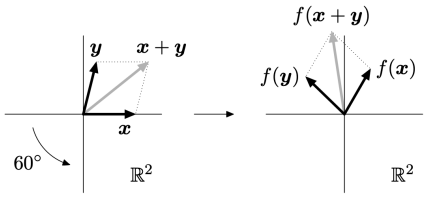
Scaling:



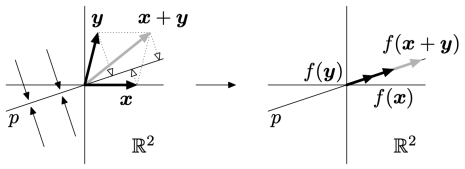
Shear:



Rotation:



Orthogonal projection:



Let us construct matrix representations of simple geometric maps in $\mathbb{R}^2$. We need only find the coordinates of the images $f(\boldsymbol{e}_1)$ and $f(\boldsymbol{e}_2)$ and write them as columns:

$$\begin{pmatrix} | & | \\ f(\boldsymbol{e}_1) & f(\boldsymbol{e}_2) \\ | & | \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}.$$

Stretching the y-axis by a factor of two

We have $f(\boldsymbol{e}_1) = \boldsymbol{e}_1 = (1, 0)$, but $f(\boldsymbol{e}_2) = 2\boldsymbol{e}_2 = (0, 2)$. The resulting matrix is diagonal:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Orthogonal projection onto the x-axis

We have $f(\boldsymbol{e}_1) = \boldsymbol{e}_1 = (1, 0)$ and $f(\boldsymbol{e}_2) = (0, 0)$. The resulting matrix is diagonal and has rank 1.

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Rotation through an angle φ

We need only compute the rotations of $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ through the angle φ:

$$\begin{aligned} f(\boldsymbol{e}_1) &= (\cos\varphi, \sin\varphi), \\ f(\boldsymbol{e}_2) &= \big(\cos(\varphi + \pi/2), \sin(\varphi + \pi/2)\big) = (-\sin\varphi, \cos\varphi). \end{aligned}$$

The resulting matrix is:

$$C = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}.$$

Differentiation and integration

Consider the space C^1 of continuously differentiable functions on $\mathbb{R}$ and the space C of all continuous functions on $\mathbb{R}$. Differentiation $\frac{d}{dx} : C^1 \to C$ is a linear map:

$$\begin{aligned} \frac{d}{dx}(f+g) &= \frac{d}{dx}f + \frac{d}{dx}g \quad \forall f, g \in C^1, \\ \frac{d}{dx}(\alpha f) &= \alpha \frac{d}{dx} f \qquad \forall f \in C^1, \alpha \in \mathbb{R}. \end{aligned}$$

Similarly, integration $\int : C \to C^1$ is a linear map. Moreover, integration is a right inverse of differentiation, so

$$\frac{d}{dx} \circ \int = \mathrm{id}\,.$$

Many of the properties of linear algebra described in this text can be applied in analysis.

For example, consider finding an antiderivative of a given function $\boldsymbol{f}$, that is, solving $\int \boldsymbol{f} = \boldsymbol{g}$. This corresponds to solving a "system of linear equations" $\frac{d}{dx}\boldsymbol{g} = \boldsymbol{f}$, analogous to $A\boldsymbol{x} = \boldsymbol{b}$.

Theorem: An antiderivative $\boldsymbol{g}$ is uniquely determined up to an additive constant.

Proof: We know that, for $A\boldsymbol{x} = \boldsymbol{b}$, if a solution $\boldsymbol{p}$ exists, the solution set is the affine subspace $\mathrm{Ker}(A) + \boldsymbol{p}$. Thus we need only determine $\mathrm{Ker}\left(\frac{d}{dx}\right)$, consisting precisely of the functions $\boldsymbol{h}$ whose derivative is the zero function 0. These are the constant functions.

This is not how the result is usually proved in analysis. But if you examine the usual proof, it is exactly the same argument as the proof of the general theorem on the structure of the solution set of $A\boldsymbol{x} = \boldsymbol{b}$.

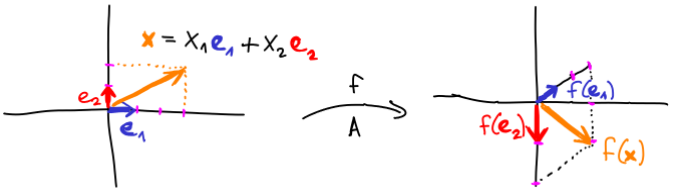
Let P_n be the space of polynomials of degree at most n. Try to work out the matrices representing differentiation $\frac{d}{dx} : P_{n+1} \to P_n$ and integration $\int : P_n \to P_{n+1}$.

Matrix representation f : x ↦ Ax

Theorem: Every linear map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is represented by a matrix A such that $f\colon \boldsymbol{x} \mapsto A\boldsymbol{x}$, that is, $f(\boldsymbol{x}) = A\boldsymbol{x}$.

Proof: The map $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is always linear by the distributive laws for matrix multiplication: $A(\boldsymbol{x} + \boldsymbol{y}) = A\boldsymbol{x} + A\boldsymbol{y}$ and $A(c\boldsymbol{x}) = c(A\boldsymbol{x})$.

Conversely, linearity uniquely determines the map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ from the images $f(\boldsymbol{e}_1), \dots, f(\boldsymbol{e}_n)$. For $\boldsymbol{x} = (x_1, \dots, x_n) = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n$, we have $f(\boldsymbol{x}) = x_1 f(\boldsymbol{e}_1) + \cdots + x_n f(\boldsymbol{e}_n)$.



Each image $f(\boldsymbol{e}_i)$ has m coordinates, so f is uniquely determined by a total of $m \times n$ coordinates. We write these as columns in an array A called a matrix.

$$A = \begin{pmatrix} | & | & & | \\ f(\boldsymbol{e}_1) & f(\boldsymbol{e}_2) & \cdots & f(\boldsymbol{e}_n) \\ | & | & & | \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}.$$

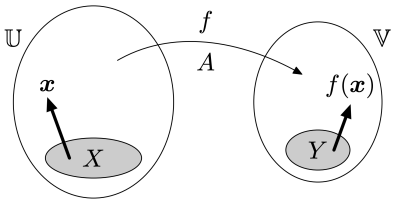
By the definition of matrix multiplication, $A\boldsymbol{x} = x_1 f(\boldsymbol{e}_1) + \cdots + x_n f(\boldsymbol{e}_n)$, so $f(\boldsymbol{x}) = A\boldsymbol{x}$. Thus this matrix represents f.

We want to generalize the whole procedure to arbitrary spaces U and V with arbitrary bases, since there is nothing special about standard bases. Take **any two bases**:

- $X = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ as a basis of U,
- $Y = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\}$ as a basis of V.

Theorem: Every linear map $f\colon U \to V$ is represented, with respect to bases X and Y, by a matrix A such that $f\colon \boldsymbol{x} \mapsto A\boldsymbol{x}$, that is, $f(\boldsymbol{x}) = A\boldsymbol{x}$.

Proof: We know that a linear map $f\colon U \to V$ is **uniquely determined by the images of any basis**. To describe it, it suffices to know $f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)$, each of which can be **described by m coordinates** with respect to the basis Y.



We choose to arrange these $m \times n$ coordinates in an array A, which we call a **matrix**. We write the coordinates of the individual vectors $f(\boldsymbol{x}_i)$ as columns:

$$A = \begin{pmatrix} | & | & & | \\ f(\boldsymbol{x}_1) & f(\boldsymbol{x}_2) & \cdots & f(\boldsymbol{x}_n) \\ | & | & & | \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{pmatrix}.$$

Thus $a_{i,j}$ is the coordinate of $f(\boldsymbol{x}_j)$ along $\boldsymbol{y}_i$.

Key point: **Different choices of bases X and Y can give different matrix representations** of the linear map f. Conversely, **one matrix can represent different linear maps** for different choices of bases.

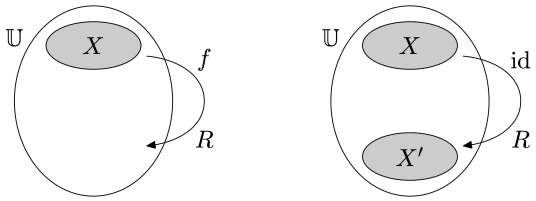
For example, consider rotation through an angle φ in the plane. It is represented by A when $X = Y$ is the standard basis consisting of $(1, 0)$ and $(0, 1)$. But if X is the standard basis and Y is X rotated through φ, the rotation is represented by B. On the other hand, if $X = Y$ and the map is an endomorphism, then B represents the identity map.

$$A = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Change-of-basis matrices

A square invertible matrix representing the identity map is called a **change-of-basis matrix**. It converts coordinates from one basis X to another basis X'.

A square invertible matrix can represent different linear maps. If we choose a single basis X, we obtain an endomorphism. We can also view it as representing the identity map with respect to two bases X and X'.



If $\boldsymbol{x}$ gives the coordinates of a vector in the basis X, then $A\boldsymbol{x}$ gives the coordinates of the same vector in the basis X'.

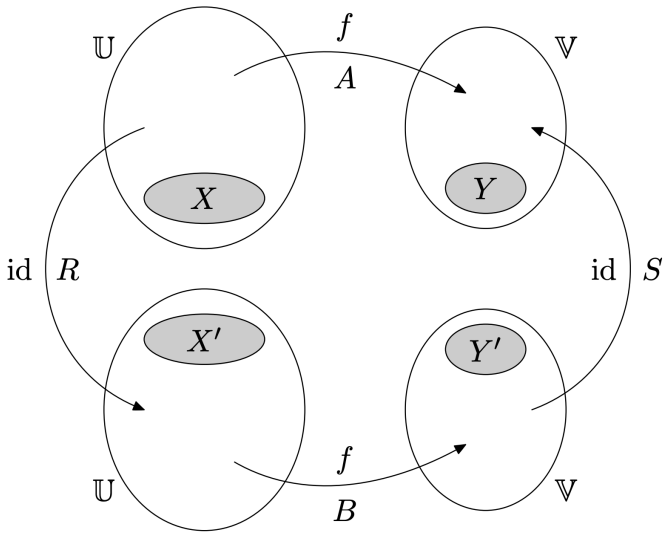
Similarity

Before defining similarity, let us first define the weaker notion of matrix equivalence.

Matrices A and B are **equivalent** if they represent the same linear map with different choices of bases. They are equivalent if and only if there are two invertible (change-of-basis) matrices R and S such that:

$$A = SBR, \quad \text{equivalently} \quad B = S^{-1}AR^{-1}.$$

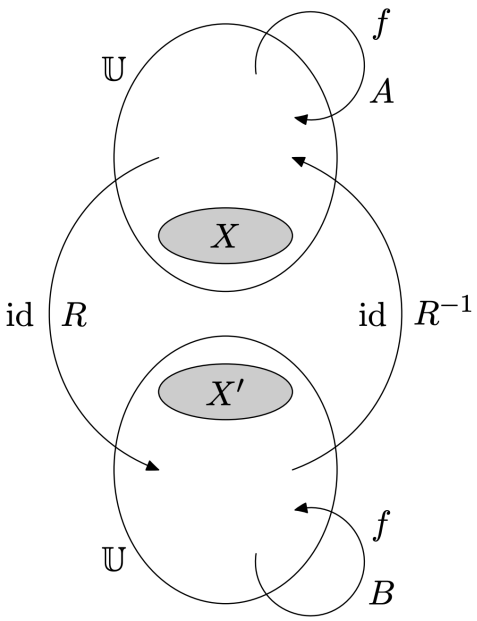
Proposition: Matrices A and B of the same size are equivalent if and only if $\operatorname{rank}(A) = \operatorname{rank}(B)$.



Similarity is matrix equivalence for square matrices representing endomorphisms with respect to a single basis. Thus two matrices A and B are **similar** if and only if they represent the **same endomorphism with different choices of basis**. This holds precisely when there is an invertible matrix R (a change-of-basis matrix) such that:

$$A = R^{-1}BR, \quad \text{equivalently} \quad B = RAR^{-1}.$$

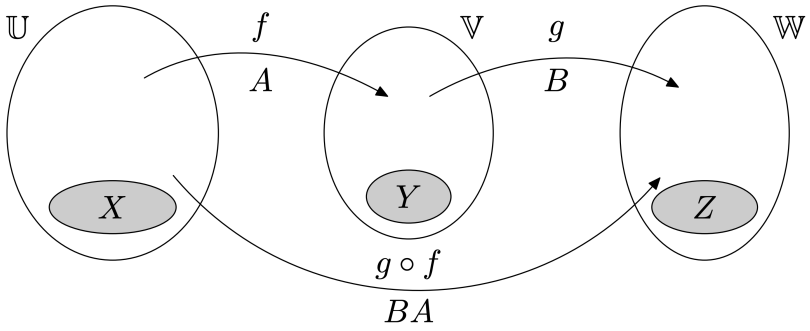
One of the central problems of linear algebra is to find the simplest possible matrix representation of a given endomorphism.



Composition of maps

Let $f\colon U \to V$ and $g\colon V \to W$ be linear maps. Their **composition** $g \circ f\colon U \to W$ is also **linear**. What is its matrix representation?

Theorem: Suppose A **represents** $f\colon U \to V$ with respect to bases X and Y, and B **represents** $g\colon V \to W$ with respect to bases Y and Z. Then BA **represents** $g \circ f\colon U \to W$ with respect to bases X and Z.



This theorem explains the definition of matrix multiplication. **We must multiply matrices in this unusual way precisely so that multiplication corresponds to composition of linear maps**. If we developed linear algebra starting from linear maps, we could derive the definition of matrix multiplication from that of matrix representation.

This is also why matrix multiplication **requires compatible dimensions**. Let A be an $n \times p$ matrix representing a map $\mathbb{R}^p \to \mathbb{R}^n$, and B an $m \times n$ matrix representing a map $\mathbb{R}^n \to \mathbb{R}^m$. The intermediate space is the same, and therefore has the same dimension, and the composite map $\mathbb{R}^p \to \mathbb{R}^m$ is represented by the $m \times p$ matrix BA.

Proof from the definition: The matrix representing $g \circ f$ with respect to bases X and Z must have, in column i, the coordinates of $g(f(\boldsymbol{x}_i))$ in the basis Z. Notice that column i of A contains the coordinates of $f(\boldsymbol{x}_i)$ in the basis Y, and the images of the individual vectors of Y are the columns of B. Thus we obtain the coordinates of $g(f(\boldsymbol{x}_i))$ in the basis Z by combining the columns of B using the coordinates in column i of A, which is exactly matrix multiplication.

Proof from associativity: By associativity of matrix multiplication, $(BA)\boldsymbol{x} = B(A\boldsymbol{x})$ for every vector $\boldsymbol{x}$. The right-hand side corresponds to the composite map $g \circ f \colon \boldsymbol{x} \mapsto B(A\boldsymbol{x})$. Since the equality holds for every vector $\boldsymbol{x}$, the map $g \circ f$ must be represented by BA.

Conversely, notice that the theorem implies associativity of matrix multiplication: it corresponds to composition of linear maps, and composition of arbitrary maps is associative.

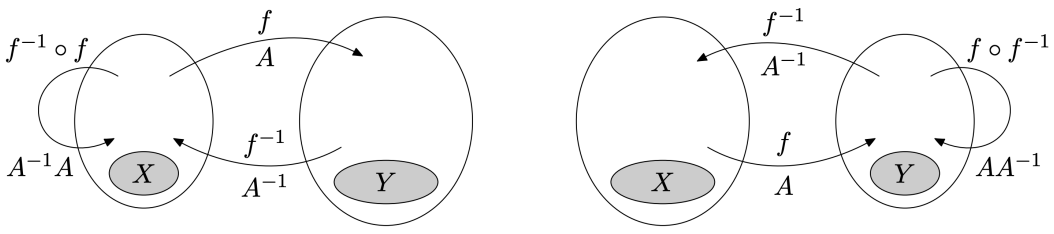
Inverse maps

A linear map $f \colon U \to V$ may or may not have an inverse map $f^{-1} : V \to U$. We distinguish the following kinds of inverses:

- A **left inverse** $f^{-1} \circ f = \mathrm{id}$ exists if and only if f is **injective**.
- A **right inverse** $f \circ f^{-1} = \mathrm{id}$ exists if and only if f is **surjective**.
- A **two-sided inverse** $f^{-1} \circ f = f \circ f^{-1} = \mathrm{id}$ exists if and only if f is **bijective**.

Since the identity map is represented by the identity matrix I_n, an inverse map is represented by an inverse matrix. More precisely, suppose f is represented by A with respect to bases X and Y.

- A **left inverse** $\boldsymbol{f^{-1}}$ is represented by a **left inverse** $\boldsymbol{A^{-1}}$ with respect to bases Y and X (left-hand diagram).
- A **right inverse** $\boldsymbol{f^{-1}}$ is represented by a **right inverse** $\boldsymbol{A^{-1}}$ with respect to bases Y and X (right-hand diagram).
- A **two-sided inverse** $\boldsymbol{f^{-1}}$ is represented by a **two-sided inverse** $\boldsymbol{A^{-1}}$ with respect to bases Y and X.

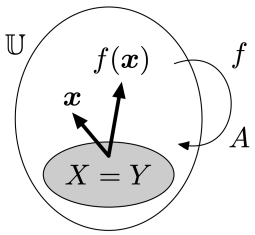


Endomorphisms, isomorphisms, automorphisms

Three prefixes distinguish special kinds of homomorphisms: **endomorphisms**, **isomorphisms**, and **automorphisms**. They distinguish whether f is an arbitrary map or a bijection, and whether f maps between two spaces U and V, or from a space U to the same space U.

	$f\colon U \to V$	$f\colon U \to U$
in general	*homo*morphism	*endo*morphism
***f* is a bijection**	*iso*morphism	*auto*morphism

Every endomorphism is a homomorphism, and every automorphism is an isomorphism, but the names specifically emphasize that we view them as maps from a space to itself. Compositions of endomorphisms are endomorphisms, and we can also take their powers:
$f^k = f \circ f \circ \cdots \circ f$.



Endomorphisms are represented by **square matrices**, and we usually choose just one basis, so $X = Y$. The advantage of such representations is that iterating the homomorphism corresponds to taking powers of its representing matrix:

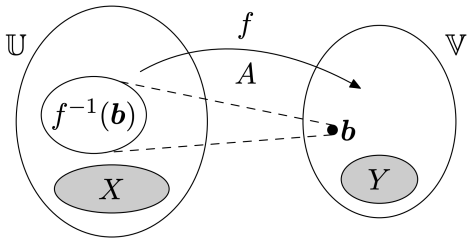
If f is represented by A, then f^k is represented by A^k.

Isomorphisms are represented by **invertible matrices**. If an isomorphism $f\colon U \to V$ exists, the spaces U and V have the same algebraic structure for the vector operations + and ·. Since f is a bijection, we can **pair the vectors of U with those of V**, and the operations act on them in exactly the same way. This does not mean the spaces are identical mathematical objects, only that their **algebraic structure is identical**, and in many situations we can **work in U just as we would in V**.

Finding preimages

Let $f\colon U \to V$ be a linear map, and let $\boldsymbol{b}$ be any vector in V. We want to find the preimage

$$f^{-1}(\boldsymbol{b}) = \{\boldsymbol{x} \,:\, f(\boldsymbol{x}) = \boldsymbol{b}\}.$$



Suppose the linear map f is **represented by a matrix $\boldsymbol{A}$.** We know that $f\colon \boldsymbol{x} \mapsto A\boldsymbol{x}$. We therefore seek all vectors $\boldsymbol{x}$ satisfying $A\boldsymbol{x} = \boldsymbol{b}$, which leads to a **system of linear equations**. Thus, even if we developed linear algebra abstractly, starting from vector spaces and their homomorphisms, solving systems of linear equations would still be **essential**.

We know that the preimage $f^{-1}(\boldsymbol{b})$ is nonempty if and only if $\boldsymbol{b} \in \mathrm{Im}(f)$. In that case, it is the affine subspace $\mathrm{Ker}(f) + \boldsymbol{p}$, where $\boldsymbol{p}$ is any preimage vector of $\boldsymbol{b}$.

Isomorphism theorem

Theorem: Every vector space V of dimension n is **isomorphic to the vector space** $\mathbb{R}^n$.

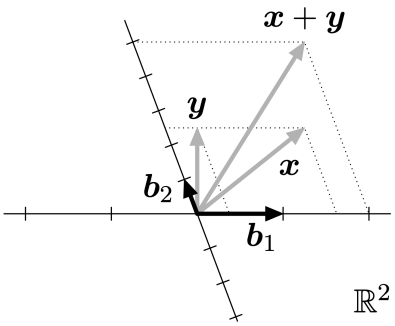
The word isomorphic means that there is an **isomorphism** $f\colon V \to \mathbb{R}^n$. This guarantees that the algebraic structure of V and $\mathbb{R}^n$—the structure of their operations—is **the same**. We have $f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$, and similarly for scalar multiplication, so we can identify each vector $\boldsymbol{x}$ in V with the vector $f(\boldsymbol{x})$ in $\mathbb{R}^n$.

This does not mean that the spaces themselves are identical. For example, V may be the space of polynomials of degree at most $n-1$, whose elements are entirely different mathematical objects from n-tuples of real numbers. The isomorphism tells us, however, that **in many situations we can work with** V **exactly as we would with** $\mathbb{R}^n$; for example, we can solve systems in $\mathbb{R}^n$ instead of the abstract space V.

Proof: Choose any basis of V. It defines coordinates $(c_1, \dots, c_n)$ on V. This coordinate assignment is the desired isomorphism:

$$f\colon\ c_1\boldsymbol{b}_1 + \dots + c_n\boldsymbol{b}_n \mapsto (c_1, \dots, c_n).$$

Since the coordinates of every vector in V are uniquely defined, this map is a bijection. It remains to check that it is indeed linear. We need only observe that adding vectors in V corresponds to adding their coordinates, and similarly for scalar multiplication.



Homomorphism theorem

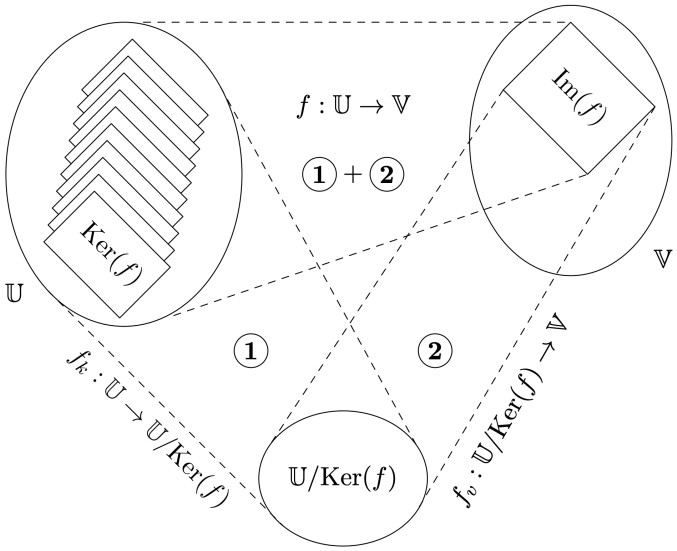
Theorem: For every linear map $f\colon U \to V$, we have

$$U/\operatorname{Ker}(f) \cong \operatorname{Im}(f).$$

We know that the preimages of individual vectors in $\operatorname{Im}(f)$ are affine subspaces of U obtained by translating $\operatorname{Ker}(f)$. Thus the preimages of individual vectors in $\operatorname{Im}(f)$ have the same structure up to translation.

The **larger** $\operatorname{Ker}(f)$ is, the fewer distinct translates it has, and hence the **fewer** vectors have nonempty preimages. **More precisely, in terms of dimensions:**

$$\dim\operatorname{Ker}(f) + \dim\operatorname{Im}(f) = \dim(U).$$



By the theorem, a homomorphism $f\colon U \to V$ can be written as a composition of two simpler homomorphisms:

- A **quotient homomorphism** $f_k : U \to U/\operatorname{Ker}(f)$, which is surjective and generally not injective.
- An **embedding homomorphism** $f_v : U/\operatorname{Ker}(f) \to V$, which is injective but generally not surjective.

Adjoint map

Associated with a linear map $f\colon U \to V$ is an **adjoint map** $f^* : V \to U$, which has special properties. Notice that the inverse map $f^{-1} : V \to U$, if it exists, maps between the same spaces, but the **adjoint is something quite different** (for example, it always exists).

Definition using the inner product: The adjoint map $f^* : V \to U$ is the unique map satisfying, for all vectors $\boldsymbol{x} \in U$ and $\boldsymbol{y} \in V$:

$$\langle f(\boldsymbol{x}) \mid \boldsymbol{y} \rangle = \langle \boldsymbol{x} \mid f^*(\boldsymbol{y}) \rangle.$$

Notice that the left-hand side is an inner product in V, while the right-hand side is an inner product in U.

Suppose A represents f with respect to bases X and Y. For the standard inner product, f^* **is represented by** A^T. Thus **transposition corresponds to taking the adjoint**.

With respect to which bases does A^T represent f^*? We can describe the coordinates of the bases X and Y relative to the standard bases by change-of-basis matrices X and Y (whose columns are the coordinates of the basis vectors).

Proposition: For the standard inner product, the map f^* is uniquely determined and is represented by A^T with respect to the **dual bases** Y^* **and** X^*, consisting of the columns of $Y^{-\mathsf{T}}$ and $X^{-\mathsf{T}}$.

Proof: First prove this for the standard bases. Let B be a matrix satisfying $\langle A\boldsymbol{x} \mid \boldsymbol{y} \rangle = \langle \boldsymbol{x} \mid B\boldsymbol{y} \rangle$. Clearly, this holds for $B = A^\mathsf{T}$. It is the only such matrix because

$$a_{ji} = (A\boldsymbol{e}_i)_j = \sum_{k=1}^{n} (A\boldsymbol{e}_i)_k \cdot (\boldsymbol{e}_j)_k = \sum_{k=1}^{n} (\boldsymbol{e}_i)_k \cdot (B\boldsymbol{e}_j)_k = b_{ij}.$$

nonzero only for $k = j$ — Thus $B = A^\mathsf{T}$. — nonzero only for $k = i$

For general bases, the result follows for the following reason. The matrix $Y^{-1}AX$ represents f with respect to the standard bases. Hence the adjoint map f^* is represented by $(Y^{-1}AX)^\mathsf{T} = X^\mathsf{T}A^\mathsf{T}Y^{-\mathsf{T}}$ with respect to the standard bases.

Inner product

Inner product, norm, and orthogonality

Our aim is to introduce further geometric notions on the vector space $\mathbb{R}^n$.

- The **inner product** $\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ describes how much they point in a common direction and is related to the angle between them.
- **Perpendicular vectors** (or orthogonal vectors) have no direction in common, which occurs precisely when $\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = 0$.
- The **norm** $\|\boldsymbol{x}\|$ is the geometric length of a vector.

Mathematics often studies spaces of elements (not necessarily vector spaces) equipped with geometric properties. This gives the following hierarchy of possible geometries:

inner product ⇒ norm ⇒ metric ⇒ topology.

For example, a space might be the surface of a sphere, on which we can consider geometric distances. Of these structures, we can consider a topology and a metric, but cannot define an inner product or a norm.

- **Inner product space.** We can consider angles and orthogonality in the space. An inner product induces a norm: $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle}$.
- **Normed space.** We can consider distances that scale proportionally when vectors are stretched: $\|c\boldsymbol{x}\| = |c| \cdot \|\boldsymbol{x}\|$. A norm induces a metric $\rho(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$.
- **Metric space.** A metric is a distance function $\rho(x, y)$ that is nonnegative (and positive for $x \neq y$), symmetric, and satisfies the triangle inequality. Unlike a normed space, the space need not have a vector space structure, so no linearity is required. A norm induces a topology through open and closed ε-neighborhoods, the sets $B_\varepsilon(x) = \{y \mid \rho(x, y) \leq \varepsilon\}$ (or with a strict inequality).
- **Topological space.** Distance is not defined directly, but the topology describes nearness. This is done through a system of closed and open neighborhoods. The advantage is that in many situations we do not need distance at all. Much of analysis can be developed on topological spaces. For example, we can define the limit of a sequence: $\lim x_n = x$ if, for every open neighborhood of x, there is an $N > 0$ such that, for $n > N$, every x_n lies in this neighborhood of x.

Inner product

An inner product can be defined concretely by a formula or abstractly by its properties.

The **standard inner product** is defined by the expression

$$\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \boldsymbol{x}^\mathsf{T} \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n.$$

The **abstract definition** describes an inner product as any map $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying the following properties:

1. **Linearity:** $\forall\, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n :\ \langle \boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{z} \rangle = \langle \boldsymbol{x} \mid \boldsymbol{z} \rangle + \langle \boldsymbol{y} \mid \boldsymbol{z} \rangle$ and $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \alpha \in \mathbb{R} :\ \langle \alpha \boldsymbol{x} \mid \boldsymbol{y} \rangle = \alpha \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$.
2. **Symmetry:** $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n :\ \langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \langle \boldsymbol{y} \mid \boldsymbol{x} \rangle$.
3. **Positive definiteness:** $\forall\, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq 0 :\ \langle \boldsymbol{x} \mid \boldsymbol{x} \rangle > 0$. (Note: For $\boldsymbol{x} = 0$, linearity gives $\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle = 0$.)

Any map satisfying the first property (in both arguments) is called a **bilinear form**: it assigns real numbers to pairs of vectors and is linear in both arguments. Thus an inner product is a **symmetric positive definite bilinear form**. Check that the standard inner product satisfies properties 1 to 3.

Over the field of complex numbers $\mathbb{C}$, we use complex conjugation: The standard inner product takes the complex conjugates of the coefficients of $\boldsymbol{x}$. The abstract definition of an inner product includes complex conjugation in property 2.

An inner product (such as the standard one) can be defined in the same way over general fields, but **positive definiteness cannot be satisfied** (over a general field, the inequality does not even make sense). Thus the notion of an inner product is considerably weaker over general fields.

Cauchy–Schwarz inequality

Theorem: (Cauchy–Schwarz) For any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\left|\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle\right| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|.$$

Remark: The theorem has the following application in **functional analysis**. On a space of functions, we can define an inner product and its induced norm, for example

$$\langle f \mid g \rangle = \int_0^{2\pi} f(x)\, g(x)\, dx.$$

$\int$ is like Σ_i; check that it satisfies properties (i)–(iii).

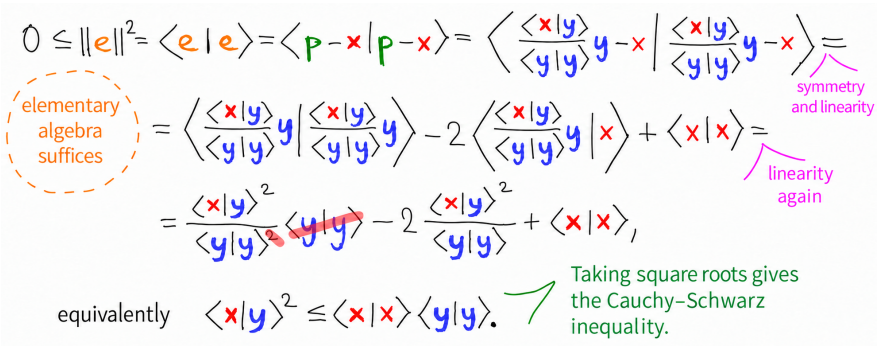
Since we want to avoid infinite values, we can consider the **Hilbert subspace of all functions with finite norm**. By the Cauchy–Schwarz inequality, their inner products are finite as well.

Cosine formula: For any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ making an angle φ, we have

$$\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos(\varphi)$$

The inner product is **positive** for an **acute angle** φ, **zero** for a **right angle** φ, and **negative** for an **obtuse angle** φ.

Proof: Consider the orthogonal projection $\boldsymbol{p}$ of the vector $\boldsymbol{x}$ onto the vector $\boldsymbol{y}$. The norm $\|\boldsymbol{e}\|$ of the difference between the projection and the original vector, $\boldsymbol{e} = \boldsymbol{p} - \boldsymbol{x}$, is nonnegative. Rearranging gives the Cauchy–Schwarz inequality:



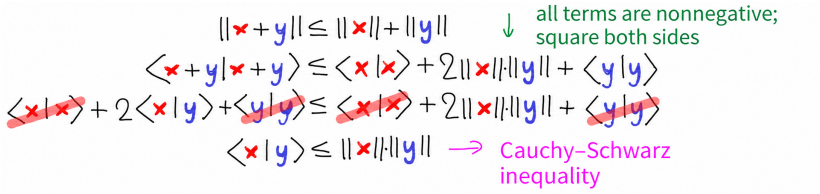
Norm

The norm $\|\boldsymbol{x}\|$ is the geometric length of a vector $\boldsymbol{x}$. Again, there are several definitions.

Norm induced by an inner product. Every inner product $\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ defines a norm $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle}$. For example, the standard inner product defines the **standard norm**

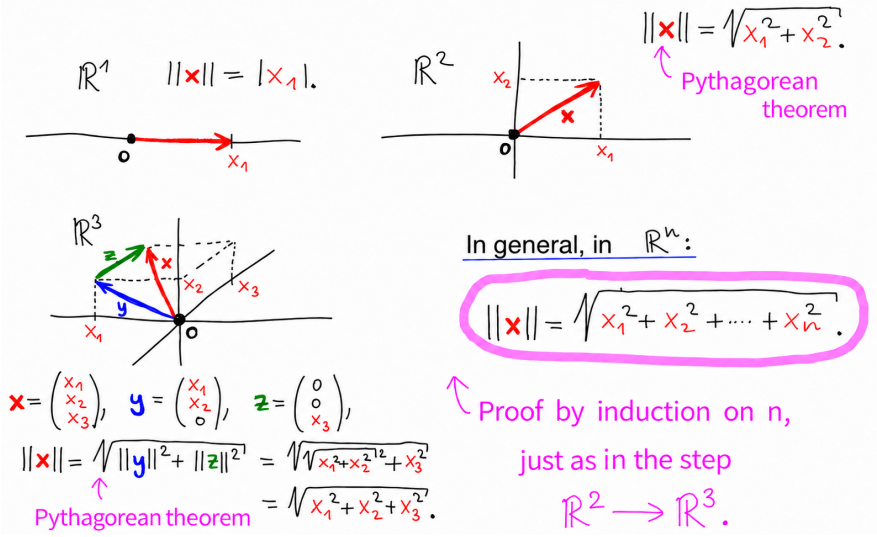


Properties 1 and 3 follow from the corresponding properties of the inner product. For the triangle inequality:



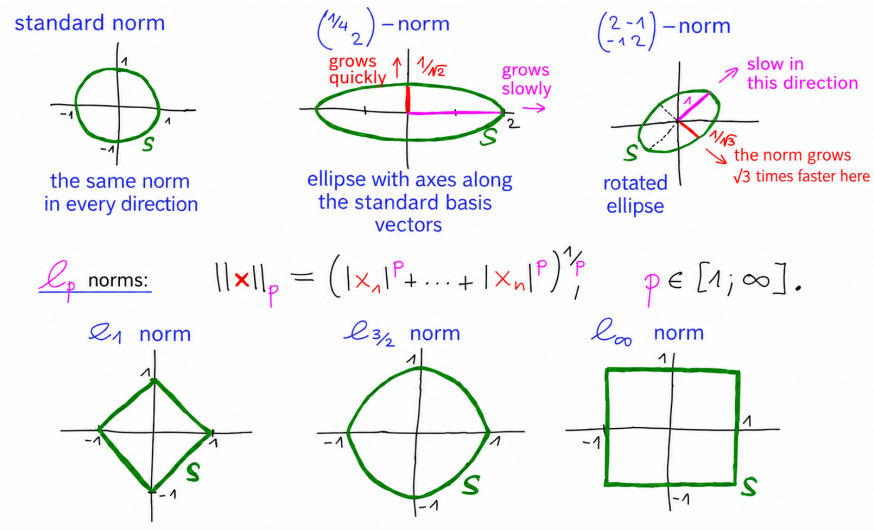
The **abstract definition** describes a norm as any map $\mathbb{R}^n \to \mathbb{R}$ satisfying the following properties:

1. **Absolute homogeneity:** $\forall\, \boldsymbol{x} \in \mathbb{R}^n, \alpha \in \mathbb{R} :\ \|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$.
2. **Triangle inequality:** $\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n :\ \|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.
3. **Nonnegativity:** $\forall\, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq 0 :\ \|\boldsymbol{x}\| > 0$. (Note: For $\boldsymbol{x} = 0$, homogeneity gives $\|\boldsymbol{x}\| = 0$.)



Nonstandard norms on $\mathbb{R}^2$

By the characterization, norms induced by inner products are the so-called A-**norms** $\|\boldsymbol{x}\|_A = \sqrt{\boldsymbol{x}^\mathsf{T} A \boldsymbol{x}}$, where A is a symmetric positive definite matrix. There are, however, **far more norms than inner products**.



For a given norm, consider the unit sphere S and the unit ball B:

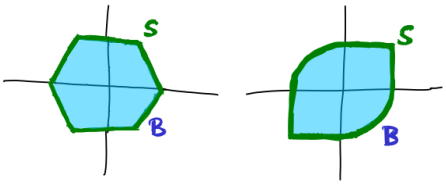
$$S = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| = 1\}, \quad B = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| \leq 1\}$$

.

Geometric properties of the sphere and ball. The norm axioms imply geometric properties of S and B.

- The set S intersects every line through the origin at **exactly two points** (by homogeneity and nonnegativity).
- The set S is **symmetric about the origin**: $S = -S$ (by homogeneity with a negative coefficient), and does not contain the origin: $0 \notin S$ (by nonnegativity).
- The set B is **convex** (the hardest part, following from the triangle inequality).

The key point is that a norm is **uniquely determined by its unit sphere** S. On every ray from the origin, exactly one vector $\boldsymbol{y}$ has norm one. Homogeneity determines the norms of all the other vectors $\boldsymbol{x}$ on the ray.

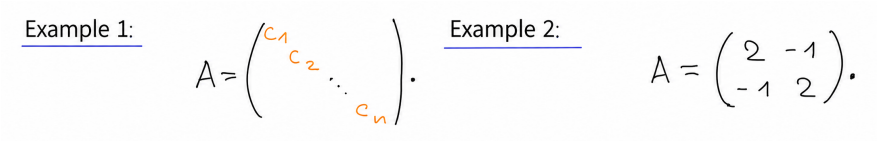


Proposition: Let B be a closed, bounded subset of $\mathbb{R}^n$, symmetric about the origin. Let S be the boundary of B, and suppose $0 \notin S$. Then the map $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ defined by $\|\boldsymbol{x}\| = |\alpha|$ for $\boldsymbol{x} = \alpha \boldsymbol{y}$ and $\boldsymbol{y} \in S$, is a norm.

Proposition: For a norm induced by an inner product, the **unit sphere** S **is a rotated ellipsoid** whose axes are eigenvectors of the matrix A.

Nonstandard inner products

The expression $\boldsymbol{x}^\mathsf{T}\boldsymbol{y}$ naturally generalizes to $\boldsymbol{x}^\mathsf{T}A\boldsymbol{y}$, where A is a matrix.



In Example 1, the third property requires every c_i to be positive. In Example 2, property 3 can be proved by rewriting the expression as a sum of squares:

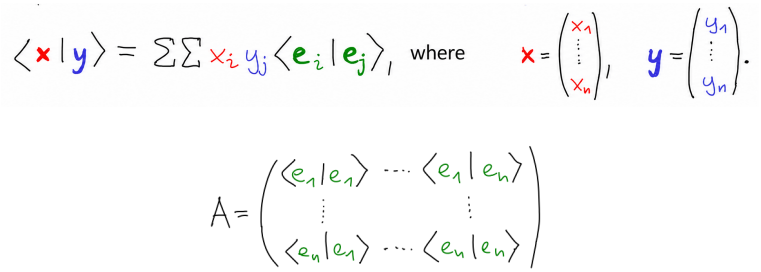
$$\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle = 2x_1^2 - 2x_1x_2 + 2x_2^2 = x_1^2 + x_2^2 + (x_1 - x_2)^2 > 0$$

In general, the expression $\boldsymbol{x}^\mathsf{T}A\boldsymbol{y}$ is a bilinear form.

- If the matrix A is **symmetric**, it is a **symmetric bilinear form**.
- If A is also **positive definite** ($\forall\, \boldsymbol{x} \neq 0 :\ \boldsymbol{x}^\mathsf{T}A\boldsymbol{x} > 0$), it is a **symmetric positive definite bilinear form**, that is, an inner product.

Theorem: Every inner product $\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle$ equals $\boldsymbol{x}^\mathsf{T}A\boldsymbol{y}$ for some **symmetric positive definite matrix** A.

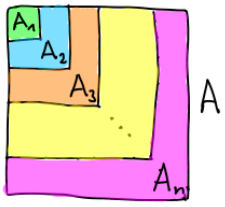
Proof: Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ be the standard basis. By linearity, an inner product is uniquely determined by its values on any basis:



We choose A to be the **Gram matrix** of inner products $\langle \boldsymbol{e}_i \mid \boldsymbol{e}_j \rangle$. Then $\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \boldsymbol{x}^\mathsf{T}A\boldsymbol{y}$. Properties 2 and 3 of the inner product imply that A is symmetric and positive definite.

Positive definite matrices

A square symmetric matrix A is called **positive definite** if and only if it satisfies any of the following six equivalent properties:



1. For every $\boldsymbol{x} \neq 0$, we have $\boldsymbol{x}^\mathsf{T}A\boldsymbol{x} > 0$.
2. **All eigenvalues satisfy** $\lambda_i > 0$. (The eigenvalues are real because A is symmetric.)

3. The **leading principal minors satisfy** $\det(A_1), \dots, \det(A_n) > 0$. Here A_k is the square $k \times k$ submatrix in the upper-left corner of A.
4. **Gaussian elimination can be performed without interchanging or scaling rows**, and all pivots satisfy $p_i > 0$.
5. There is an invertible matrix R such that $A = R^\mathsf{T} R$, giving a Cholesky decomposition.
6. The expression $\boldsymbol{x}^\mathsf{T} A \boldsymbol{y}$ is an inner product, and $\sqrt{\boldsymbol{x}^\mathsf{T} A \boldsymbol{x}}$ is a norm.

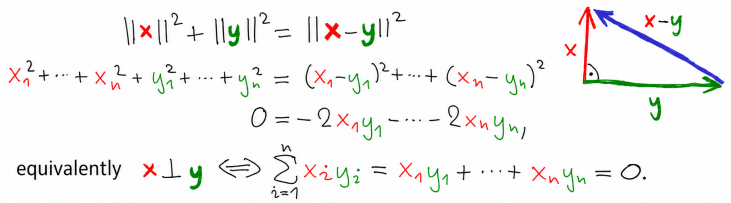
Proposition: For symmetric matrices A, properties 1 to 6 are equivalent.

Quadratic forms and definiteness

Orthogonality

Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal**, or **perpendicular**, if and only if $\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = 0$. We denote orthogonality by $\boldsymbol{x} \perp \boldsymbol{y}$.

Derivation using the Pythagorean theorem. Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are perpendicular if and only if they satisfy the Pythagorean theorem.



By definition, the origin 0 is perpendicular to every vector.

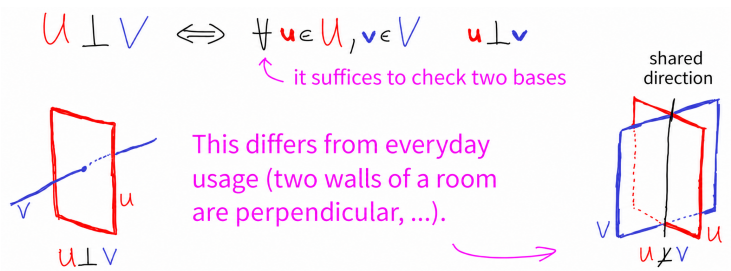
Lemma: Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be nonzero, **pairwise orthogonal vectors**, that is, $\boldsymbol{x}_i \perp \boldsymbol{x}_j$ for $i \neq j$. Then the vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are **linearly independent**.

Proof: Suppose $\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_n \boldsymbol{x}_n = 0$. We want to prove that $\alpha_i = 0$ for every i. Multiplying the entire expression on the left by $\boldsymbol{x}_i^\mathsf{T}$, **orthogonality gives**

$$0 = \alpha_1 \boldsymbol{x}_i^\mathsf{T} \boldsymbol{x}_1 + \cdots + \alpha_i \boldsymbol{x}_i^\mathsf{T} \boldsymbol{x}_i + \cdots + \alpha_n \boldsymbol{x}_i^\mathsf{T} \boldsymbol{x}_n = \alpha_i \boldsymbol{x}_i^\mathsf{T} \boldsymbol{x}_i = \alpha_i \|\boldsymbol{x}_i\|^2.$$

Since $\boldsymbol{x}_i \neq 0$, we know that $\alpha_i = 0$. This holds when we multiply by any $\boldsymbol{x}_i^\mathsf{T}$, so the linear combination $\alpha_1 \boldsymbol{x}_1 + \cdots + \alpha_n \boldsymbol{x}_n$ is **trivial**.

Orthogonality of vector subspaces.



Orthogonal complement

Let U be any vector subspace. The **orthogonal complement of** U is the set

$$U^\perp = \{\boldsymbol{x} \,:\, \boldsymbol{x} \perp \boldsymbol{y}\, \forall \boldsymbol{y} \in U\}.$$

Proposition: The orthogonal complement $U^\perp$ has the following properties:

1. $U^\perp$ is a vector subspace.
2. $U \cap U^\perp = \{0\}$ and $\langle U \cup U^\perp \rangle = \mathbb{R}^n$.
3. $\dim U + \dim U^\perp = n$.
4. $(U^\perp)^\perp = U$.

Proof: Property 1 and the first part of property 2 are easy. For the rest, we need the existence of orthogonal bases. Choose an orthogonal basis $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ of U and extend it by vectors $\boldsymbol{b}_{k+1}, \dots, \boldsymbol{b}_n$ to an orthogonal basis of $\mathbb{R}^n$. The vectors $\boldsymbol{b}_{k+1}, \dots, \boldsymbol{b}_n$ form an orthogonal basis of $U^\perp$. Indeed, a vector with a nonzero coordinate along any of $\boldsymbol{b}_1, \dots, \boldsymbol{b}_k$ cannot lie in $U^\perp$.

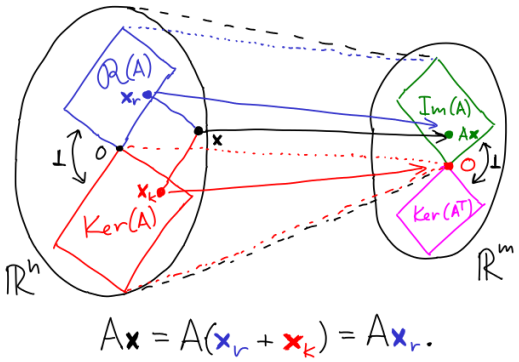
Theorem: The fundamental subspaces of a matrix A are paired as **orthogonal complements**:

$$\mathrm{Ker}(A) = \mathrm{R}(A)^\perp, \qquad \mathrm{Ker}(A^\mathsf{T}) = \mathrm{Im}(A)^\perp.$$

Proof: Let $\boldsymbol{x} \in \mathrm{Ker}(A)$, and let $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m$ be the rows of A. By definition,

$$\boldsymbol{u}_1^\mathsf{T}\boldsymbol{x} = 0, \quad \boldsymbol{u}_2^\mathsf{T}\boldsymbol{x} = 0, \quad \dots, \quad \boldsymbol{u}_m^\mathsf{T}\boldsymbol{x} = 0.$$

Thus $\boldsymbol{x}$ is orthogonal to $\mathrm{R}(A)$, and conversely every vector orthogonal to $\mathrm{R}(A)$ belongs to $\mathrm{Ker}(A)$. Hence $\mathrm{Ker}(A) = \mathrm{R}(A)^\perp$. Applying this to A^T gives the second equality.



Orthonormal bases

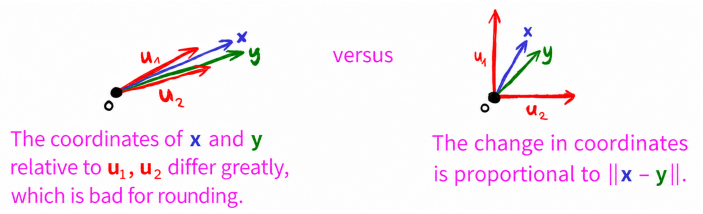
Since orthogonal nonzero vectors are linearly independent, the notion of an **orthogonal basis makes sense**. We usually also take the vectors to be **orthonormal**, with norm one: this is not essential, but it simplifies the formulas.

Proposition: Any orthogonal or orthonormal set of nonzero vectors **can be extended** to an orthogonal or orthonormal basis, respectively.

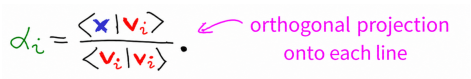
Proof: Extend the orthonormal set to any basis and apply Gram–Schmidt orthogonalization. An alternative proof uses the fundamental theorem of linear algebra.

There are therefore **plenty** of orthogonal bases, and they are **nice bases** with many useful properties.

1. Orthonormal bases have favorable numerical properties.



2. Coordinates are easy to compute using orthogonal projections. For a general basis, we must solve the system of linear equations $A\boldsymbol{\alpha} = \boldsymbol{x}$. For an orthogonal basis $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$, we have



The proof follows from orthogonality: $\langle \boldsymbol{x} \mid \boldsymbol{v}_i \rangle = \langle \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_n \boldsymbol{v}_n \mid \boldsymbol{v}_i \rangle = \alpha_i \langle \boldsymbol{v}_i \mid \boldsymbol{v}_i \rangle$.

Another perspective uses orthogonal matrices: $\boldsymbol{\alpha} = Q^{\mathsf{T}} Q \boldsymbol{\alpha} = Q^{\mathsf{T}} \boldsymbol{x}$, provided the columns of Q are the orthonormal vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$.

Fourier transform

Orthogonal matrices

A matrix Q is called **orthogonal** if its **columns are orthonormal** (orthogonal and of norm one). In matrix notation, this gives $Q^{\mathsf{T}} Q = I_n$.

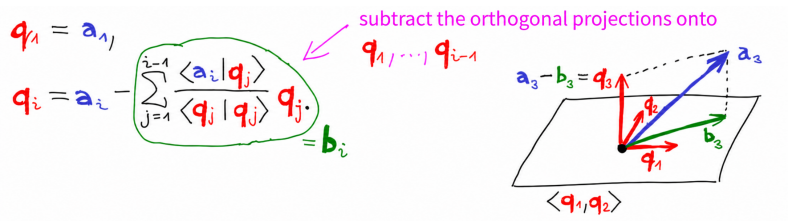
Thus an orthogonal matrix satisfies $Q^{-1} = Q^{\mathsf{T}}$. Since the inverse of any square matrix is two-sided, we also obtain $QQ^{\mathsf{T}} = I_n$. Hence, if a square matrix has orthonormal columns, it also has orthonormal rows, and Q^{T} is orthogonal as well.

Gram–Schmidt orthogonalization

An algorithm for orthogonalizing a set of input vectors.

Given input vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$, it produces orthogonal vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_k$ such that

$$\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_k \rangle = \langle \boldsymbol{q}_1, \dots, \boldsymbol{q}_k \rangle.$$



Proposition: The vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_k$ are orthogonal.

$$\langle q_i \mid q_t \rangle = \left\langle a_i - \sum_{j=1}^{i-1} \frac{\langle a_i | q_j \rangle}{\langle q_j \mid q_j \rangle} q_j \;\middle|\; q_t \right\rangle \overset{\text{linearity}}{=} \langle a_i \mid q_t \rangle - \sum_{j=1}^{i-1} \frac{\langle a_i | q_j \rangle}{\langle q_j \mid q_j \rangle} \langle q_j \mid q_t \rangle = \langle a_i \mid q_t \rangle - \langle a_i \mid q_t \rangle = 0.$$

($\forall t < i$; nonzero only for $j = t$)

Proposition: For every $i \leq k$ we have

$$\langle a_1, \ldots, a_i \rangle = \langle q_1, \ldots, q_i \rangle.$$

QR decomposition

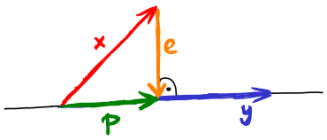
Proposition: Every square matrix A has a decomposition

$$A = QR,$$

where Q is an **orthogonal matrix** and R is an **upper triangular matrix**.

Proof: This is **Gram–Schmidt orthogonalization written in matrix form**, just as LU decomposition is Gaussian elimination written in matrix form.

Orthogonal projection



Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two vectors. We want to find the orthogonal projection of $\boldsymbol{x}$ onto the line $\langle \boldsymbol{y} \rangle$. Equivalently, the orthogonal projection is the vector on the line closest to $\boldsymbol{x}$.

We know that $\boldsymbol{p} = c\boldsymbol{y}$. We seek the value of c for which the vector $\boldsymbol{e} = \boldsymbol{p} - \boldsymbol{x}$ is perpendicular to $\boldsymbol{y}$:

$$0 = \langle p - x \mid y \rangle = \langle cy - x \mid y \rangle = c\langle y | y \rangle - \langle x | y \rangle, \text{ equivalently } c = \frac{\langle x | y \rangle}{\langle y | y \rangle}.$$

(just a normalization factor)

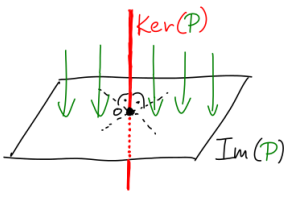
Proposition: Orthogonal projection onto the line $\langle \boldsymbol{y} \rangle$ is the linear map

$$x \mapsto \frac{\langle x | y \rangle}{\langle y | y \rangle} y = \frac{x^T y}{y^T y} y = \frac{y y^T}{y^T y} x$$

(matrix P of the orthogonal projection)

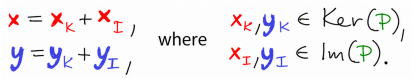
Let P be the matrix of orthogonal projection onto the subspace $\mathrm{Im}(P)$. It satisfies the following two conditions:

1. Characterizing any projection: $P^2 = P$. A projection is the identity on $\mathrm{Im}(P)$, so $P\boldsymbol{y} = \boldsymbol{y}$ for every $\boldsymbol{y} \in \mathrm{Im}(P)$. Since $\boldsymbol{y} = P\boldsymbol{x}$ for each $\boldsymbol{x}$, the equality follows.

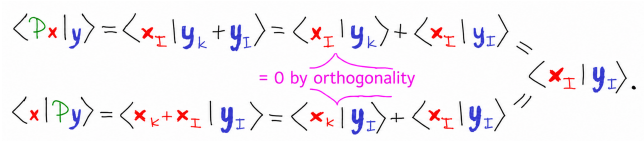


2. Characterizing orthogonality of the projection: $P^\mathsf{T} = P$. $\mathrm{Ker}(P)$ consists of the vectors projected onto 0, which are exactly the vectors perpendicular to $\mathrm{Im}(P)$. Thus $\mathrm{Ker}(P) = \mathrm{Im}(P)^\perp$, and the fundamental theorem of linear algebra gives $\mathrm{Ker}(P) = \mathrm{Ker}(P^\mathsf{T})$ and $\mathrm{R}(P) = \mathrm{Im}(P)$.

The rest follows from the adjoint relation if $\langle P\boldsymbol{x} \mid \boldsymbol{y}\rangle = \langle \boldsymbol{x} \mid P\boldsymbol{y}\rangle$. Choose



Since $P\boldsymbol{x} = \boldsymbol{x}_I$ and $P\boldsymbol{y} = \boldsymbol{y}_I$,



Proposition: If the columns of A form a basis of a subspace W, then orthogonal projection onto W is described by the matrix

$$P = A(A^\mathsf{T}A)^{-1}A^\mathsf{T}.$$

Method of least squares

Fundamental subspaces

A linear map $f\colon U \to V$ defines two **fundamental subspaces**:

- The **null space** $\mathrm{Ker}(f) = \{\boldsymbol{x} \in U \mid f(\boldsymbol{x}) = 0\}$ (or **kernel**).
- The **image** $\mathrm{Im}(f) = \{f(\boldsymbol{x}) \mid \boldsymbol{x} \in U\}$.

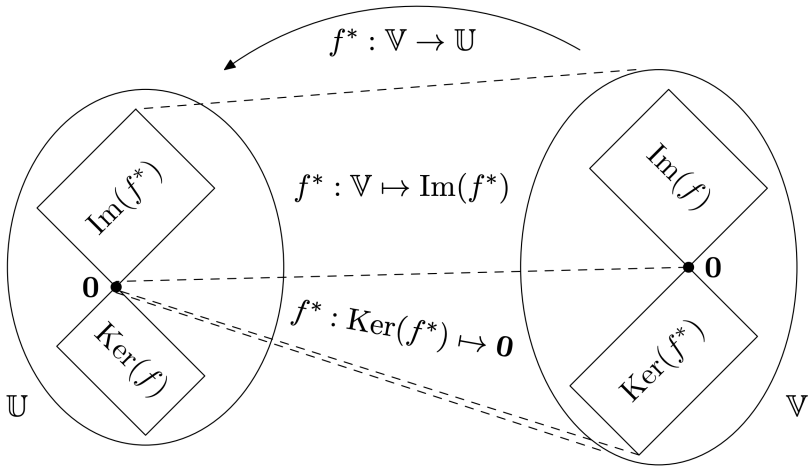
The kernel is a subspace of U, while the image is a subspace of V. If a matrix A represents f, these subspaces correspond to $\mathrm{Ker}(A)$ and $\mathrm{Im}(A)$.

But what do the remaining two fundamental subspaces associated with A^T correspond to?

They are the kernel and image of the adjoint map $f^*\colon V \to U$.

- The **left null space** $\mathrm{Ker}(f^*) = \{\boldsymbol{x} \in V \mid f^*(\boldsymbol{x}) = 0\}$ (or **left kernel**).
- The **left image** $\mathrm{Im}(f^*) = \{f^*(\boldsymbol{x}) \mid \boldsymbol{x} \in V\}$.

The diagram shows the adjoint map together with its fundamental subspaces.


$f^* : \mathbb{V} \to \mathbb{U}$
$\mathrm{Im}(f^*)$
$f^* : \mathbb{V} \mapsto \mathrm{Im}(f^*)$
$\mathrm{Im}(f)$
$\mathbf{0}$
$\mathbf{0}$
$f^* : \mathrm{Ker}(f^*) \mapsto \boldsymbol{0}$
$\mathrm{Ker}(f)$
$\mathrm{Ker}(f^*)$
$\mathbb{U}$
$\mathbb{V}$