## Normal matrix

In mathematics, a [complex](https://en.wikipedia.org/wiki/Complex_number) [square matrix](https://en.wikipedia.org/wiki/Square_matrix) *A* is **normal** if it [commutes](https://en.wikipedia.org/wiki/Commute_(mathematics)) with its [conjugate transpose](https://en.wikipedia.org/wiki/Conjugate_transpose) *A*\*:

*A*\**A = AA*\*

For real matrices, the conjugate transpose is just the transpose, A\* = AT

## Positive Definite Matrices

A square matrix is called positive definite if it is symmetric and all its eigenvalues λ are positive,

that is λ > 0.

If A is positive definite, then it is invertible and det A > 0.

## Condition number

In [numerical analysis](https://en.wikipedia.org/wiki/Numerical_analysis), the **condition number** of a [function](https://en.wikipedia.org/wiki/Function_(mathematics)) measures how much the output value of the function can change for a small change in the input argument. This is used to measure how [sensitive](https://en.wikipedia.org/wiki/Sensitivity_analysis) a function is to changes or errors in the input, and how much error in the output results from an error in the input.

In case of  [matrix norm induced by the (vector) Euclidean norm](https://en.wikipedia.org/wiki/Matrix_norm#Matrix_norms_induced_by_vector_norms)

κ2(A) = σmax(A) / σmin(A)

where σmax(A)σ and σmin(A) denote the largest and smallest singular value of A respectively.

## QR decomposition

In linear algebra, a QR decomposition, also known as a QR factorization or QU factorization, is a decomposition of a matrix A into a product A = QR of an orthonormal matrix Q (*QTQ*=*E*) and an upper triangular matrix R. QR decomposition is often used to solve the linear least squares (LLS) problem and is the basis for a particular eigenvalue algorithm, the QR algorithm.

## Eigen Decomposition

In [linear algebra](https://en.wikipedia.org/wiki/Linear_algebra), **eigendecomposition** is the [factorization](https://en.wikipedia.org/wiki/Matrix_factorization) of a [matrix](https://en.wikipedia.org/wiki/Matrix_(mathematics)) into a [canonical form](https://en.wikipedia.org/wiki/Canonical_form), whereby the matrix is represented in terms of its [eigenvalues and eigenvectors](https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors). Only [diagonalizable matrices](https://en.wikipedia.org/wiki/Diagonalizable_matrix) can be factorized in this way. When the matrix being factorized is a [normal](https://en.wikipedia.org/wiki/Normal_matrix) or real [symmetric matrix](https://en.wikipedia.org/wiki/Symmetric_matrix), the decomposition is called "spectral decomposition", derived from the [spectral theorem](https://en.wikipedia.org/wiki/Spectral_theorem).

Matrix A is *diagonalizable* if it's [similar](http://mlwiki.org/index.php/Similar_Matrices) to a diagonal matrix (a matrix A is similar to B if there exists an invertible M s.t. B=M^(−1)AM)

A is diagonalizable if there exists [invertible matrix](http://mlwiki.org/index.php/Inverse_Matrices) PP s.t. P^−1AP is diagonal .

Let A be a square n × n matrix with n linearly independent eigenvectors qi (where i = 1, ..., n). Then A can be factorized as

A = *Q ΛQ^-1*

where **Q** is the square *n* × *n* matrix whose *i*th column is the eigenvector *qi* of **A**, and **Λ** is the [diagonal matrix](https://en.wikipedia.org/wiki/Diagonal_matrix) whose diagonal elements are the corresponding eigenvalues, *Λii* = *λi*. Note that only [diagonalizable matrices](https://en.wikipedia.org/wiki/Diagonalizable_matrix) can be factorized in this way.

One nice thing about eigendecompositions is that we can write many operations we usually encounter cleanly in terms of the eigendecomposition:

A^n = *Q Λ^nQ^-1*

## Singular value decomposition

The Singular Value Decomposition (SVD) of a matrix is a factorization of that matrix into three matrices.

The SVD of  mxn matrix A is given by the formula  A=UΣVT

where:

* U:  *mxm* matrix of the orthonormal eigenvectors of AAT
* VT: transpose of a *nxn* matrix containing the orthonormal eigenvectors of ATA.
* Σ : diagonal matrix with r elements equal to the root of the positive eigenvalues of AAᵀ or Aᵀ A.

## Applications of SVD

* **Calculation of Pseudo-inverse:**Pseudo inverse or Moore-Penrose inverse is the generalization of the matrix inverse that may not be invertible (such as low-rank matrices). If the matrix is invertible then its inverse will be equal to Pseudo inverse but pseudo inverse exists for the matrix that is not invertible. It is denoted by A+.
* **Solving a set of Homogeneous Linear Equation (Mx =b):**if b=0,  calculate SVD and take any column of VT associated with a singular value (in W) equal to 0.
* **Curve Fitting Problem:**Singular value decomposition can be used to minimize the least square error. It uses the pseudo inverse to approximate it.
* Besides the above application, singular value decomposition and pseudo-inverse can also be used in Digital signal processing and image processing

## Computational complexity of matrix operations

* Matrix multiplication of n x m and m x p matrices: O(mnp)
* Matrix inversion: O(n^3)
* SVD of m x n matrix: O(m^2n)

## Gradient

In the case of **a univariate function**, it is simply the **first derivative at a selected point**. In the case of **a multivariate function**, it is a **vector of derivatives** in each main direction (along variable axes). Because we are interested only in a slope along one axis and we don’t care about others these derivatives are called **partial derivatives**.

A gradient for an n-dimensional function f(x) at a given point p is defined as follows:

Изображение выглядит как черный, снимок экрана, текст, черно-белый

Автоматически созданное описание

The upside-down triangle is a so-called *nabla* symbol and you read it “del”.

## Matrix differentiation

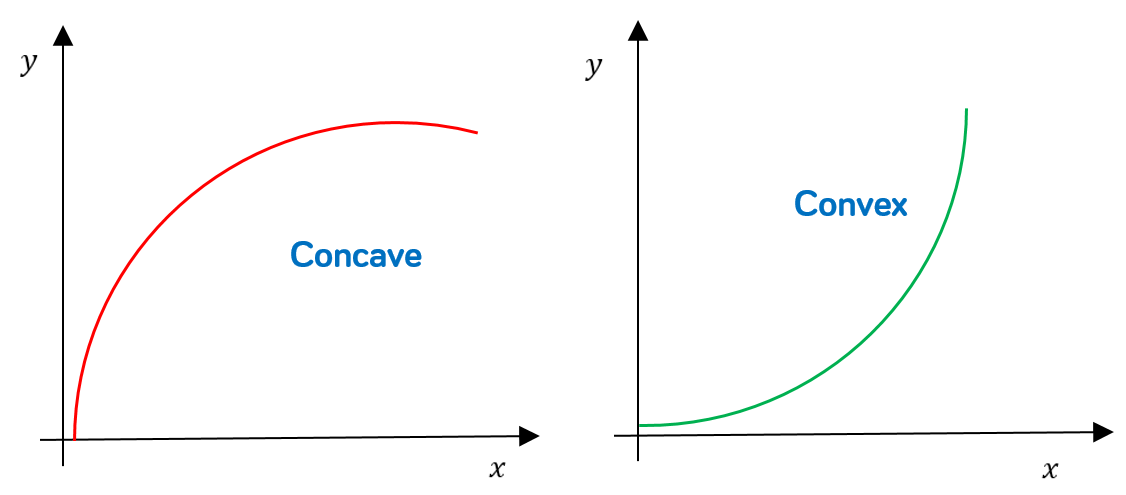
Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

## Convex and Concave functions

In [mathematics](https://en.wikipedia.org/wiki/Mathematics), a [real-valued function](https://en.wikipedia.org/wiki/Real-valued_function) is called **convex** if the [line segment](https://en.wikipedia.org/wiki/Line_segment) between any two distinct points on the [graph of the function](https://en.wikipedia.org/wiki/Graph_of_a_function) lies above the graph between the two points. Equivalently, a function is convex if its [*epigraph*](https://en.wikipedia.org/wiki/Epigraph_(mathematics)) (the set of points on or above the graph of the function) is a [convex set](https://en.wikipedia.org/wiki/Convex_set).

C**oncave function** is one for which the value at any convex combination of elements in the domain is greater than or equal to the convex combination of the values at the endpoints.



Let function f : RD → R be a function whose domain is a convex set. The function f is a convex function if for all x, y in the domain convex function of f, and for any scalar θ with 0 ⩽ θ ⩽ 1, we have

f(θx + (1 − θ)y) ⩽ θf(x) + (1 − θ)f(y).

A concave function is the negative of a convex function

A function f(x) is convex if and only if for any two points x, y it holds that

f(y) ⩾ f(x) + ∇xf(x) ⊤(y − x).

If we further know that a function f(x) is twice differentiable, that is, the Hessian exists for all values in the domain of x, then the function f(x) is convex if and only if ∇2 x f(x) is positive semidefinite

## Supporting hyperplane

Изображение выглядит как Красочность

Автоматически созданное описание

In [geometry](https://en.wikipedia.org/wiki/Geometry), a **supporting hyperplane** of a [set](https://en.wikipedia.org/wiki/Set_(mathematics)) S in [Euclidean space](https://en.wikipedia.org/wiki/Euclidean_space) Rn is a [hyperplane](https://en.wikipedia.org/wiki/Hyperplane) that has both of the following two properties:

* Sis entirely contained in one of the two [closed](https://en.wikipedia.org/wiki/Closed_set) [half-spaces](https://en.wikipedia.org/wiki/Half-space_(geometry)) bounded by the hyperplane,
* S has at least one boundary-point on the hyperplane.

## Legendre–Fenchel Transform and Convex Conjugate

The Legendre-Fenchel transform is a transformation from a convex differentiable function f(x) to a function that depends on the tangents s(x) = ∇f(x). It is worth stressing that this is a transformation of the function f(·) and not the variable x or the function evaluated at x. The Legendre-Fenchel transform is also known as the convex conjugate and is closely related to duality.

The convex conjugate of a function f : RD → R is a function f ∗ defined by

f ∗ (s) = supremum x∈RD (⟨s, x⟩ − f(x)) .

We can reconstruct any function f(x), with some restriction, by **just knowing its tangent line at each point on its graph**.

Describing the tangent line of a function, on the other hand, requires two pieces of information; the slope of the line at each point, given by the value of df/dx, and the y-interception of the line at each point, b.

Therefore, we can encode *all* the information of a function f(x) into just these two values and this is indeed *exactly* what the Legendre transformation does; **generates a new function from df/dx and b**.

The important thing about this is that the Legendre transformation of a function then contains exactly the same information as the original function, just “presented” in a different way. This is why it’s useful in the first place.

Изображение выглядит как текст, снимок экрана, Шрифт, дизайн

Автоматически созданное описание

## Dirac delta function

Изображение выглядит как снимок экрана, пространство, линия

Автоматически созданное описание

The Dirac delta function δ(x) can be loosely thought of as a function on the real line which is zero everywhere except at the origin, where it is infinite, аnd its integral is 1.

## Taylor series

In [mathematics](https://en.wikipedia.org/wiki/Mathematics), the **Taylor series** or **Taylor expansion** of a [function](https://en.wikipedia.org/wiki/Function_(mathematics)) is an [infinite sum](https://en.wikipedia.org/wiki/Series_(mathematics)) of terms that are expressed in terms of the function's [derivatives](https://en.wikipedia.org/wiki/Derivative) at a single point. For most common functions, the function and the sum of its Taylor series are equal near this point.∫−∞∞δ(x)dx=1.

The [partial sum](https://en.wikipedia.org/wiki/Partial_sum) formed by the first *n* + 1 terms of a Taylor series is a [polynomial](https://en.wikipedia.org/wiki/Polynomial) of degree *n* that is called the *n*th **Taylor polynomial** of the function. Taylor polynomials are approximations of a function, which become generally more accurate as *n* increases. [Taylor's theorem](https://en.wikipedia.org/wiki/Taylor%27s_theorem) gives quantitative estimates on the error introduced by the use of such approximations.

Изображение выглядит как Шрифт, белый, типография, дизайн

Автоматически созданное описание