2020/05/16

Topic: matrix derivative

主要内容: Matrix Calculus

1. 矩阵求导术，上
2. 矩阵求导术，下

Matrix Calculus 求解网站

The Matrix Cookbook

Other resouces

1. Matrix Calculus in The Matrix Reference Manual
2. Matrix Calculus by Johannes
3. Matrix Calculus by Barnes
4. Old and New Matrix Algebra Useful for Statistics by Minka

微积分

1. 微积分是什么
2. 微分和导数的关系
3. 微分是什么

4. 如何通俗易懂地理解全微分
5. 微积分的本质，系列视频 by 3Blue1Brown

Hessian

In mathematics, the **Hessian matrix** or **Hessian** is a square matrix of second-order partial derivatives of a scalar-valued function, or scalar field. It describes the local curvature of a function of many

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function taking as input a vector $\mathbf{x} \in \mathbb{R}^n$ and outputting a scalar $f(\mathbf{x}) \in \mathbb{R}$. If all second partial derivatives of f exist and are continuous over the domain of the function, then the Hessian matrix $\mathbf{H}$ of f is a square $n \times n$ matrix, usually defined and arranged as follows:

$$\mathbf{H} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\,\partial x_1} & \dfrac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix},$$

or, by stating an equation for the coefficients using indices i and j,

$$\mathbf{H}_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

The Hessian matrix is a symmetric matrix, since the hypothesis of continuity of the second derivatives implies that the order of differentiation does not matter (Schwarz's theorem)

The determinant of the Hessian matrix is called the *Hessian determinant*.[1]

The Hessian matrix of a function f is the Jacobian matrix of the gradient of the function: $\mathbf{H}(f(\mathbf{x})) = \mathbf{J}(\nabla f(\mathbf{x}))$.

Use in optimization [edit]

Hessian matrices are used in large-scale optimization problems within Newton-type methods because they are the coefficient of the quadratic term of a local Taylor expansion of a function. That is,

$$y = f(\mathbf{x} + \Delta\mathbf{x}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathrm{T}} \Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^{\mathrm{T}}\mathbf{H}(\mathbf{x})\Delta\mathbf{x}$$

where ∇f is the gradient $(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$. Computing and storing the full Hessian matrix takes $\Theta(n^2)$ memory, which is infeasible for high-dimensional functions such as the loss functions of neural nets, conditional random fields, and other statistical models with large numbers of parameters. For such situations, truncated-Newton and quasi-Newton algorithms have been developed. The latter family of algorithms use approximations to the Hessian; one of the most popular quasi-Newton algorithms is BFGS.[5]

Gradient

In vector calculus, the **gradient** of a scalar-valued differentiable function f of several variables, $f\colon \mathbf{R}^n \to \mathbf{R}$, is the vector field, or more simply a vector-valued function[a] $\nabla f\colon \mathbf{R}^n \to \mathbf{R}^n$, whose value at a point p is the vector[b] whose components are the partial derivatives of f at p:[1][2][3][4][5][6][7][8][9]

$$\nabla f(p) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \\ \vdots \\ \frac{\partial f}{\partial x_n}(p) \end{bmatrix}.$$

The gradient by Khan Academy 理解方向导数与梯度

Taylor expansion

In mathematics, the **Taylor series** of a function is an infinite sum of terms that are expressed in terms of the function's derivatives at a single point. For most common functions, the function and the sum of

The partial sum formed by the n first terms of a Taylor series is a polynomial of degree n that is called the nth **Taylor polynomial** of the function. Taylor polynomials are approximations of a function, which becomes generally better when n increases. Taylor's theorem gives quantitative estimates on the error introduced by the use of such approximations. If the Taylor series of a function is convergent, its sum is the limit of the infinite sequence of the Taylor polynomials. A

Definition [edit]

The Taylor series of a real or complex-valued function $f(x)$ that is infinitely differentiable at a real or complex number a is the power series

$$f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3+\cdots,$$

where $n!$ denotes the factorial of n. In the more compact sigma notation, this can be written as

$$\sum_{n=0}^{\infty}\frac{f^{(n)}(a)}{n!}(x-a)^n,$$

where $f^{(n)}(a)$ denotes the nth derivative of f evaluated at the point a. (The derivative of order zero of f is defined to be f itself and $(x-a)^0$ and $0!$ are both defined to be 1.)

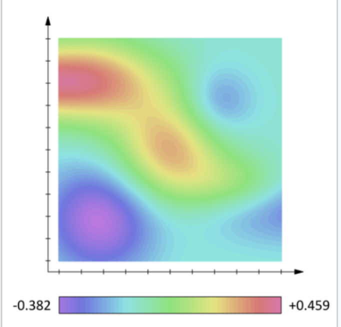
Field

In mathematics, a **field** is a set on which addition, subtraction, multiplication, and division are defined and behave as the corresponding operations on rational and real numbers do. A field is thus a fundamental algebraic structure which is widely used in algebra, number theory, and many other areas of mathematics.

The best known fields are the field of rational numbers, the field of real numbers and the field of complex numbers. Many other fields, such as

Scalar Field

In mathematics and physics, a **scalar field** associates a scalar value to every point in a space – possibly physical space. The scalar may either be a (dimensionless) mathematical number or a physical quantity. In a physical context, scalar fields are required to be independent of the choice of reference frame, meaning that any two observers using the same units will agree on the value of the scalar field at the same absolute point in space (or spacetime) regardless of their respective points of origin. Examples used in physics include the temperature distribution throughout space, the pressure distribution in a fluid, and spin-zero quantum fields, such as the Higgs field. These fields are the subject of scalar field theory.



A scalar field such as temperature or pressure, where intensity of the field is represented by different hues of colors.

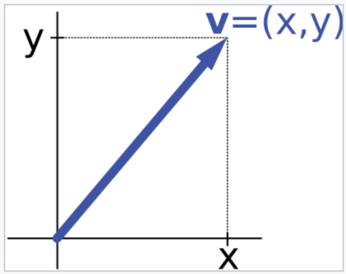
Contents [hide]

Scalar

A **scalar** is an element of a field which is used to define a vector space. A quantity described by multiple scalars, such as having both direction and magnitude, is called a vector.[1]

In linear algebra, real numbers or other elements of a field are called **scalars** and relate to vectors in a vector space through the operation of scalar multiplication, in which a vector can be multiplied by a number to produce another vector.[2][3][4] More generally, a vector space may be defined by using any field instead of real numbers, such as complex numbers. Then the scalars of that vector space will be the elements of the associated field.

A scalar product operation – not to be confused with scalar multiplication – may be defined on a vector space, allowing two vectors to be multiplied to produce a scalar. A vector space equipped with a scalar product is called an inner product space.



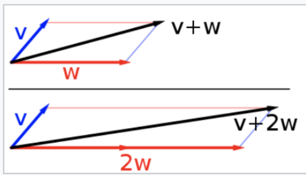
Scalars are real numbers used in linear algebra, as opposed to vectors. This image shows a Euclidean vector. Its coordinates *x* and *y* are scalars, as is its length, but **v** is not a scalar.

The real component of a quaternion is also called its **scalar part**.

The term is also sometimes used informally to mean a vector, matrix, tensor, or other usually "compound" value that is actually reduced to a single component. Thus, for example, the product of a 1×*n* matrix and an *n*×1 matrix, which is formally a 1×1 matrix, is often said to be a **scalar**.

Vector space

A **vector space** (also called a **linear space**) is a collection of objects called **vectors**, which may be added together and multiplied ("scaled") by numbers, called *scalars*. Scalars are often taken to be real numbers, but there are also vector spaces with scalar multiplication by complex numbers, rational numbers, or generally any field. The operations of vector addition and scalar multiplication must satisfy certain requirements, called *axioms*, listed below, in § Definition. For specifying that the scalars are real or complex numbers, the terms **real vector space** and **complex vector space** are often used.



Vector addition and scalar multiplication: a vector $\mathbf{v}$ (blue) is added to another vector $\mathbf{w}$ (red, upper illustration). Below, $\mathbf{w}$ is stretched by a factor of 2, yielding the sum $\mathbf{v} + 2\mathbf{w}$.

Euclidean vectors are an example of a vector space. They represent physical quantities such as forces: any two forces (of the same type) can be added to yield a third, and the multiplication of a force vector by a real multiplier is another force vector. In the same vein, but in a more geometric sense, vectors representing displacements in the plane or in three-dimensional space also form vector spaces. Vectors in vector spaces do not necessarily have to be arrow-like objects as they appear in the mentioned examples: vectors are regarded as abstract mathematical objects with particular properties, which in some cases can be visualized as arrows.

Jacobian

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial \mathbf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

Total derivative

In mathematics, the **total derivative** of a function f at a point is the best linear approximation near this point of the function with respec to its arguments. Unlike partial derivatives, the total derivative approximates the function with respect to all of its arguments, not just a single one. In many situations, this is the same as considerin all partial derivatives simultaneously. The term "total derivative" is primarily used when f is a function of several variables, because when f is a function of a single variable, the total derivative is the same as the derivative of the function.[1]:198–203

The total derivative as a differential form [edit]

When the function under consideration is real-valued, the total derivative can be recast using differential forms. For example, suppose that $f\colon \mathbf{R}^n \to \mathbf{R}$ is a differentiable function of variables $x_1, \ldots, x_n$. The total derivative of f at a may be written in terms of its Jacobian matrix, which in this instance is a row matrix (the transpose of the gradient):

$$df_a = \begin{pmatrix} \frac{\partial f}{\partial x_1}, & \cdots & , & \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

The linear approximation property of the total derivative implies that if

$$\Delta x = \begin{pmatrix} \Delta x_1, & \cdots & , & \Delta x_n \end{pmatrix}^T$$

is a small vector (where the T denotes transpose, so that this vector is a column vector), then

$$f(a + \Delta x) - f(a) \approx df_a \cdot \Delta x = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \Delta x_i.$$

Heuristically, this suggests that if $dx_1, \ldots, dx_n$ are infinitesimal increments in the coordinate directions, then

$$df_a = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) dx_i.$$

Some notes

梯度、散度、旋度、Jacobian、Hessian、Laplacian 的关系

by 王赟 Maigo

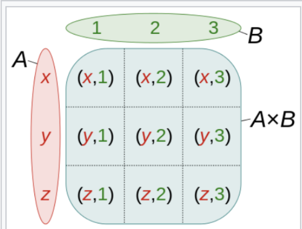
Products in Linear Algebra

Products between vectors

- dot product, also called as:
  - scalar product
  - inner product
  - projection product
- cross product (vector product)
- exterior product (wedge product)
- outer product
- Cartesian product (related to set theory)
  - In mathematics, specifically set theory, the **Cartesian product** of two sets *A* and *B*, denoted *A*×*B*, is the set of all ordered pairs (*a*, *b*) where *a* is in *A* and *b* is in *B*. In terms of set-builder notation, that is

    $$A \times B = \{\,(a,b) \mid a \in A \ \text{ and } \ b \in B\,\}.$$ [1]

    A table can be created by taking the Cartesian product of a set of rows and a set of columns. If the Cartesian product *rows* × *columns* is taken, the cells of the table contain ordered pairs of the form (row value, column value).

    Cartesian product $A \times B$ of the sets $A = \{x, y, z\}$ and $B = \{1, 2, 3\}$

    One can similarly define the Cartesian product of *n* sets, also known as an ***n*-fold Cartesian product**, which can be represented by an *n*-dimensional array, where each element is an *n*-tuple. An ordered pair is a 2-tuple or couple. More generally still, one can define the Cartesian product of an indexed family of sets.

Products between matrices

- matrix product
  - In mathematics, **matrix multiplication** is a binary operation that produces a matrix from two matrices. For matrix multiplication, the number of columns in the first matrix must be equal to the number of rows in the second matrix. The result matrix, known as the **matrix product**, has the number of rows of the first and the number of columns of the second matrix.

Definition [edit]

If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times p$ matrix,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix}$$

the *matrix product* $\mathbf{C} = \mathbf{AB}$ (denoted without multiplication signs or dots) is defined to be the $m \times p$ matrix[4][5][6][7]

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{pmatrix}$$

such that

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj},$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, p$.

That is, the entry c_{ij} of the product is obtained by multiplying term-by-term the entries of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$, and summing these n products. In other words, c_{ij} is the dot product of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$.

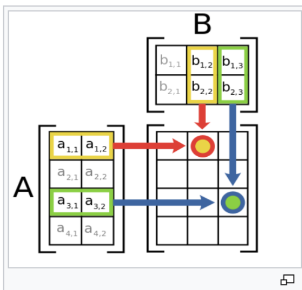
Therefore, $\mathbf{AB}$ can also be written as

$$\mathbf{C} = \begin{pmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + \cdots + a_{1n}b_{n2} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1p} + \cdots + a_{2n}b_{np} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + \cdots + a_{mn}b_{n2} & \cdots & a_{m1}b_{1p} + \cdots + a_{mn}b_{np} \end{pmatrix}$$

Thus the product $\mathbf{AB}$ is defined if and only if the number of columns in $\mathbf{A}$ equals the number of rows in $\mathbf{B}$, in this case n.

Usually the entries are numbers, but they may be any kind of mathematical objects for which an addition and a multiplication are defined, that are associative, and such that the addition is commutative, and the multiplication is distributive with respect to the addition. In particular, the entries may be matrices themselves (see block matrix).

Illustration [edit]



The figure to the right illustrates diagrammatically the product of two matrices $\mathbf{A}$ and $\mathbf{B}$, showing how each intersection in the product matrix corresponds to a row of $\mathbf{A}$ and a column of $\mathbf{B}$.

$$\overset{4\times 2 \text{ matrix}}{\begin{bmatrix} a_{11} & a_{12} \\ \cdot & \cdot \\ a_{31} & a_{32} \\ \cdot & \cdot \end{bmatrix}} \overset{2\times 3 \text{ matrix}}{\begin{bmatrix} \cdot & b_{12} & b_{13} \\ \cdot & b_{22} & b_{23} \end{bmatrix}} = \overset{4\times 3 \text{ matrix}}{\begin{bmatrix} \cdot & c_{12} & c_{13} \\ \cdot & \cdot & \cdot \\ \cdot & c_{32} & c_{33} \\ \cdot & \cdot & \cdot \end{bmatrix}}$$

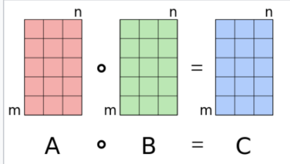
The values at the intersections marked with circles are:

$$\begin{aligned} c_{12} &= a_{11}b_{12} + a_{12}b_{22} \\ c_{33} &= a_{31}b_{13} + a_{32}b_{23} \end{aligned}$$

- Hadamard product

- In mathematics, the **Hadamard product** (also known as the **element-wise, entrywise**[1]:ch. 5 **or Schur**[2] **product**) is a binary operation that takes two matrices of the same dimensions and produces another matrix of the same dimension as the operands where each element i, j is the product of elements i, j of the original two matrices. It should not be confused with the more common matrix product. It is attributed to, and named after, either French mathematician Jacques Hadamard or German mathematician Issai Schur.

  The Hadamard product is associative and distributive. Unlike the matrix product, it is also commutative.

  The Hadamard product operates on identically shaped matrices and produces a third matrix of the same dimensions.

  ## Definition [edit]

  For two matrices A and B of the same dimension $m \times n$, the Hadamard product $A \circ B$ (or $A \odot B$[3][4][5]) is a matrix of the same dimension as the operands, with elements given by

  $$(A \circ B)_{ij} = (A \odot B)_{ij} = (A)_{ij}(B)_{ij}.$$

  For matrices of different dimensions ($m \times n$ and $p \times q$, where $m \neq p$ or $n \neq q$) the Hadamard product is undefined.

  ## Example [edit]

  For example, the Hadamard product for a 3 × 3 matrix $\mathbf{A}$ with a 3 × 3 matrix $\mathbf{B}$ is

  $$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \circ \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}\, b_{11} & a_{12}\, b_{12} & a_{13}\, b_{13} \\ a_{21}\, b_{21} & a_{22}\, b_{22} & a_{23}\, b_{23} \\ a_{31}\, b_{31} & a_{32}\, b_{32} & a_{33}\, b_{33} \end{bmatrix}.$$

  ## Properties [edit]

  - The Hadamard product is commutative (when working with a commutative ring), associative and distributive over addition. That is, if **A**, **B**, and **C** are matrices of the same size, and *k* is a scalar:

    $$\begin{aligned} &\mathbf{A} \circ \mathbf{B} = \mathbf{B} \circ \mathbf{A}, \\ &\mathbf{A} \circ (\mathbf{B} \circ \mathbf{C}) = (\mathbf{A} \circ \mathbf{B}) \circ \mathbf{C}, \\ &\mathbf{A} \circ (\mathbf{B} + \mathbf{C}) = \mathbf{A} \circ \mathbf{B} + \mathbf{A} \circ \mathbf{C}, \\ &(k\mathbf{A}) \circ \mathbf{B} = \mathbf{A} \circ (k\mathbf{B}) = k\,(\mathbf{A} \circ \mathbf{B})\,, \\ &\mathbf{A} \circ \mathbf{0} = \mathbf{0} \circ \mathbf{A} = \mathbf{0}. \end{aligned}$$

- Kronecker product (generalization of cross product from vectors to matrices)
  - In mathematics, the **Kronecker product**, sometimes denoted by ⊗, is an operation on two matrices of arbitrary size resulting in a block matrix. It is a generalization of the outer product (which is denoted by the same symbol) from vectors to matrices, and gives the matrix of the tensor product with respect to a standard choice of basis. The Kronecker product should not be confused with the usual matrix multiplication, which is an entirely different operation.

Definition [edit]

If **A** is an $m \times n$ matrix and **B** is a $p \times q$ matrix, then the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $pm \times qn$ block matrix:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix},$$

more explicitly:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix}
a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\
a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\
\vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\
a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\
\vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\
\vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\
a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\
a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\
\vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\
a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq}
\end{bmatrix}.$$

- Frobenius inner product (generalization of regular inner product from vectors to matrices)

    In mathematics, the **Frobenius inner product** is a binary operation that takes two matrices and returns a number. It is often denoted $\langle \mathbf{A}, \mathbf{B} \rangle_\mathrm{F}$. The operation is a component-wise inner product of two matrices as though they are vectors. The two matrices must have the same dimension—same number of rows and columns—but are not restricted to be square matrices.

Definition [edit]

Given two complex number-valued $n \times m$ matrices **A** and **B**, written explicitly as

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ B_{21} & B_{22} & \cdots & B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nm} \end{pmatrix}$$

the Frobenius inner product is defined by the following summation Σ of matrix elements,

$$\langle \mathbf{A}, \mathbf{B} \rangle_\mathrm{F} = \sum_{i,j} \overline{A_{ij}} B_{ij} = \mathrm{tr}\left(\overline{\mathbf{A}^T} \mathbf{B}\right)$$

where the overline denotes the complex conjugate. Explicitly this sum is

$$\begin{aligned} \langle \mathbf{A}, \mathbf{B} \rangle_\mathrm{F} = & \overline{A}_{11} B_{11} + \overline{A}_{12} B_{12} + \cdots + \overline{A}_{1m} B_{1m} \\ & + \overline{A}_{21} B_{21} + \overline{A}_{22} B_{22} + \cdots + \overline{A}_{2m} B_{2m} \\ & \vdots \\ & + \overline{A}_{n1} B_{n1} + \overline{A}_{n2} B_{n2} + \cdots + \overline{A}_{nm} B_{nm} \end{aligned}$$

The calculation is very similar to the dot product, which in turn is an example of an inner product.