WIKIPEDIA
The Free Encyclopedia

Matrix calculus

In mathematics, **matrix calculus** is a specialized notation for doing multivariable calculus, especially over spaces of matrices. It collects the various partial derivatives of a single function with respect to many variables, and/or of a multivariate function with respect to a single variable, into vectors and matrices that can be treated as single entities. This greatly simplifies operations such as finding the maximum or minimum of a multivariate function and solving systems of differential equations. The notation used here is commonly used in statistics and engineering, while the tensor index notation is preferred in physics.

Two competing notational conventions split the field of matrix calculus into two separate groups. The two groups can be distinguished by whether they write the derivative of a scalar with respect to a vector as a column vector or a row vector. Both of these conventions are possible even when the common assumption is made that vectors should be treated as column vectors when combined with matrices (rather than row vectors). A single convention can be somewhat standard throughout a single field that commonly uses matrix calculus (e.g. econometrics, statistics, estimation theory and machine learning). However, even within a given field different authors can be found using competing conventions. Authors of both groups often write as though their specific conventions were standard. Serious mistakes can result when combining results from different authors without carefully verifying that compatible notations have been used. Definitions of these two conventions and comparisons between them are collected in the layout conventions section.

Scope

Matrix calculus refers to a number of different notations that use matrices and vectors to collect the derivative of each component of the dependent variable with respect to each component of the independent variable. In general, the independent variable can be a scalar, a vector, or a matrix while the dependent variable can be any of these as well. Each different situation will lead to a different set of rules, or a separate calculus, using the broader sense of the term. Matrix notation serves as a convenient way to collect the many derivatives in an organized way.

As a first example, consider the gradient from vector calculus. For a scalar function of three independent variables, $f(x_1, x_2, x_3)$, the gradient is given by the vector equation

$$\nabla f = \frac{\partial f}{\partial x_1}\hat{x}_1 + \frac{\partial f}{\partial x_2}\hat{x}_2 + \frac{\partial f}{\partial x_3}\hat{x}_3,$$

where $\hat{x}_i$ represents a unit vector in the x_i direction for $1 \le i \le 3$. This type of generalized derivative can be seen as the derivative of a scalar, f, with respect to a vector, **x**, and its result can be easily collected in vector form.

$$\nabla f = \left(\frac{\partial f}{\partial \mathbf{x}}\right)^{\mathsf{T}} = \left[\begin{array}{ccc} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \frac{\partial f}{\partial x_3} \end{array}\right]^{\mathsf{T}}.$$

More complicated examples include the derivative of a scalar function with respect to a matrix, known as the gradient matrix, which collects the derivative with respect to each matrix element in the corresponding position in the resulting matrix. In that case the scalar must be a function of each of the independent variables in the matrix. As another example, if we have an *n*-vector of dependent variables, or functions, of *m* independent variables we might consider the derivative of the dependent vector with respect to the independent vector. The result could be collected in an *m*×*n* matrix consisting of all of the possible derivative combinations.

There are a total of nine possibilities using scalars, vectors, and matrices. Notice that as we consider higher numbers of components in each of the independent and dependent variables we can be left with a very large number of possibilities. The six kinds of derivatives that can be most neatly organized in matrix form are collected in the following table.[1]

Types of matrix derivative

Types	Scalar	Vector	Matrix
Scalar	$\dfrac{\partial y}{\partial x}$	$\dfrac{\partial \mathbf{y}}{\partial x}$	$\dfrac{\partial \mathbf{Y}}{\partial x}$
Vector	$\dfrac{\partial y}{\partial \mathbf{x}}$	$\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$	
Matrix	$\dfrac{\partial y}{\partial \mathbf{X}}$		

Here, we have used the term "matrix" in its most general sense, recognizing that vectors are simply matrices with one column (and scalars are simply vectors with one row). Moreover, we have used bold letters to indicate vectors and bold capital letters for matrices. This notation is used throughout.

Notice that we could also talk about the derivative of a vector with respect to a matrix, or any of the other unfilled cells in our table. However, these derivatives are most naturally organized in a tensor of rank higher than 2, so that they do not fit neatly into a matrix. In the following three sections we will define each one of these derivatives and relate them to other branches of mathematics. See the layout conventions section for a more detailed table.

Relation to other derivatives

The matrix derivative is a convenient notation for keeping track of partial derivatives for doing calculations. The Fréchet derivative is the standard way in the setting of functional analysis to take derivatives with respect to vectors. In the case that a matrix function of a matrix is Fréchet differentiable, the two derivatives will agree up to translation of notations. As is the case in general for partial derivatives, some formulae may extend under weaker analytic conditions than the existence of the derivative as approximating linear mapping.

Usages

Matrix calculus is used for deriving optimal stochastic estimators, often involving the use of Lagrange multipliers. This includes the derivation of:

- Kalman filter
- Wiener filter
- Expectation-maximization algorithm for Gaussian mixture
- Gradient descent

Notation

The vector and matrix derivatives presented in the sections to follow take full advantage of matrix notation, using a single variable to represent a large number of variables. In what follows we will distinguish scalars, vectors and matrices by their typeface. We will let $M(n,m)$ denote the space of real $n{\times}m$ matrices with n rows and m columns. Such matrices will be denoted using bold capital letters: $\mathbf{A}$, $\mathbf{X}$, $\mathbf{Y}$, etc. An element of $M(n,1)$, that is, a column vector, is denoted with a boldface lowercase letter: $\mathbf{a}$, $\mathbf{x}$, $\mathbf{y}$, etc. An element of $M(1,1)$ is a scalar, denoted with lowercase italic typeface: a, t, x, etc. $\mathbf{X}^{\mathrm{T}}$ denotes matrix transpose, $\mathrm{tr}(\mathbf{X})$ is the trace, and $\det(\mathbf{X})$ or $|\mathbf{X}|$ is the determinant. All functions are assumed to be of differentiability class C^1 unless otherwise noted. Generally letters from the first half of the alphabet (a, b, c, ...) will be used to denote constants, and from the second half (t, x, y, ...) to denote variables.

NOTE: As mentioned above, there are competing notations for laying out systems of partial derivatives in vectors and matrices, and no standard appears to be emerging yet. The next two introductory sections use the numerator layout convention simply for the purposes of convenience, to avoid overly complicating the discussion. The section after them discusses layout conventions in more detail. It is important to realize the following:

1. Despite the use of the terms "numerator layout" and "denominator layout", there are actually more than two possible notational choices involved. The reason is that the choice of numerator vs. denominator (or in some situations, numerator vs. mixed) can be made independently for scalar-by-

vector, vector-by-scalar, vector-by-vector, and scalar-by-matrix derivatives, and a number of authors mix and match their layout choices in various ways.

2. The choice of numerator layout in the introductory sections below does not imply that this is the "correct" or "superior" choice. There are advantages and disadvantages to the various layout types. Serious mistakes can result from carelessly combining formulas written in different layouts, and converting from one layout to another requires care to avoid errors. As a result, when working with existing formulas the best policy is probably to identify whichever layout is used and maintain consistency with it, rather than attempting to use the same layout in all situations.

Alternatives

The tensor index notation with its Einstein summation convention is very similar to the matrix calculus, except one writes only a single component at a time. It has the advantage that one can easily manipulate arbitrarily high rank tensors, whereas tensors of rank higher than two are quite unwieldy with matrix notation. All of the work here can be done in this notation without use of the single-variable matrix notation. However, many problems in estimation theory and other areas of applied mathematics would result in too many indices to properly keep track of, pointing in favor of matrix calculus in those areas. Also, Einstein notation can be very useful in proving the identities presented here (see section on differentiation) as an alternative to typical element notation, which can become cumbersome when the explicit sums are carried around. Note that a matrix can be considered a tensor of rank two.

Derivatives with vectors

Because vectors are matrices with only one column, the simplest matrix derivatives are vector derivatives.

The notations developed here can accommodate the usual operations of vector calculus by identifying the space $M(n,1)$ of n-vectors with the Euclidean space $\mathbf{R}^n$, and the scalar $M(1,1)$ is identified with $\mathbf{R}$. The corresponding concept from vector calculus is indicated at the end of each subsection.

NOTE: The discussion in this section assumes the numerator layout convention for pedagogical purposes. Some authors use different conventions. The section on layout conventions discusses this issue in greater detail. The identities given further down are presented in forms that can be used in conjunction with all common layout conventions.

Vector-by-scalar

The derivative of a vector $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}^\mathsf{T}$, by a scalar x is written (in numerator layout notation) as

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

In vector calculus the derivative of a vector $\mathbf{y}$ with respect to a scalar x is known as the **tangent vector** of the vector $\mathbf{y}$, $\dfrac{\partial \mathbf{y}}{\partial x}$. Notice here that $\mathbf{y}: \mathbf{R}^1 \to \mathbf{R}^m$.

Example Simple examples of this include the velocity vector in Euclidean space, which is the tangent vector of the position vector (considered as a function of time). Also, the acceleration is the tangent vector of the velocity.

Scalar-by-vector

The derivative of a scalar y by a vector $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\mathsf{T}$, is written (in numerator layout notation) as

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} & \dfrac{\partial y}{\partial x_2} & \cdots & \dfrac{\partial y}{\partial x_n} \end{bmatrix}.$$

In [vector calculus](#), the [gradient](#) of a scalar field $\mathbf{f} : \mathbf{R}^n \to \mathbf{R}$ (whose independent coordinates are the components of $\mathbf{x}$) is the transpose of the derivative of a scalar by a vector.

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \left(\frac{\partial f}{\partial \mathbf{x}} \right)^{\mathsf{T}}$$

By example, in physics, the [electric field](#) is the negative vector [gradient](#) of the [electric potential](#).

The [directional derivative](#) of a scalar function $f(\mathbf{x})$ of the space vector $\mathbf{x}$ in the direction of the unit vector $\mathbf{u}$ (represented in this case as a column vector) is defined using the gradient as follows.

$$\nabla_{\mathbf{u}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{u}$$

Using the notation just defined for the derivative of a scalar with respect to a vector we can re-write the directional derivative as $\nabla_{\mathbf{u}} f = \dfrac{\partial f}{\partial \mathbf{x}} \mathbf{u}.$ This type of notation will be nice when proving product rules and chain rules that come out looking similar to what we are familiar with for the scalar [derivative](#).

Vector-by-vector

Each of the previous two cases can be considered as an application of the derivative of a vector with respect to a vector, using a vector of size one appropriately. Similarly we will find that the derivatives involving matrices will reduce to derivatives involving vectors in a corresponding way.

The derivative of a [vector function](#) (a vector whose components are functions) $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}^{\mathsf{T}}$, with respect to an input vector, $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathsf{T}}$, is written (in [numerator layout notation](#)) as

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

In [vector calculus](#), the derivative of a vector function $\mathbf{y}$ with respect to a vector $\mathbf{x}$ whose components represent a space is known as the **pushforward (or differential)**, or the **Jacobian matrix**.

The pushforward along a vector function $\mathbf{f}$ with respect to vector $\mathbf{v}$ in $\mathbf{R}^n$ is given by $d\mathbf{f}(\mathbf{v}) = \dfrac{\partial \mathbf{f}}{\partial \mathbf{v}} d\mathbf{v}.$

Derivatives with matrices

There are two types of derivatives with matrices that can be organized into a matrix of the same size. These are the derivative of a matrix by a scalar and the derivative of a scalar by a matrix. These can be useful in minimization problems found in many areas of applied mathematics and have adopted the names **tangent matrix** and **gradient matrix** respectively after their analogs for vectors.

Note: The discussion in this section assumes the [numerator layout convention](#) for pedagogical purposes. Some authors use different conventions. The section on [layout conventions](#) discusses this issue in greater detail. The identities given further down are presented in forms that can be used in conjunction with all common layout conventions.

Matrix-by-scalar

The derivative of a matrix function $\mathbf{Y}$ by a scalar x is known as the **tangent matrix** and is given (in numerator layout notation) by

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}.$$

Scalar-by-matrix

The derivative of a scalar function y, with respect to a $p \times q$ matrix $\mathbf{X}$ of independent variables, is given (in numerator layout notation) by

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

Important examples of scalar functions of matrices include the trace of a matrix and the determinant.

In analog with vector calculus this derivative is often written as the following.

$$\nabla_{\mathbf{X}} y(\mathbf{X}) = \frac{\partial y(\mathbf{X})}{\partial \mathbf{X}}$$

Also in analog with vector calculus, the **directional derivative** of a scalar $f(\mathbf{X})$ of a matrix $\mathbf{X}$ in the direction of matrix $\mathbf{Y}$ is given by

$$\nabla_{\mathbf{Y}} f = \operatorname{tr}\left(\frac{\partial f}{\partial \mathbf{X}} \mathbf{Y} \right).$$

It is the gradient matrix, in particular, that finds many uses in minimization problems in estimation theory, particularly in the derivation of the Kalman filter algorithm, which is of great importance in the field.

Other matrix derivatives

The three types of derivatives that have not been considered are those involving vectors-by-matrices, matrices-by-vectors, and matrices-by-matrices. These are not as widely considered and a notation is not widely agreed upon.

Layout conventions

This section discusses the similarities and differences between notational conventions that are used in the various fields that take advantage of matrix calculus. Although there are largely two consistent conventions, some authors find it convenient to mix the two conventions in forms that are discussed below. After this section, equations will be listed in both competing forms separately.

The fundamental issue is that the derivative of a vector with respect to a vector, i.e. $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$, is often written in two competing ways. If the numerator $\mathbf{y}$ is of size m and the denominator $\mathbf{x}$ of size n, then the result can be laid out as either an $m \times n$ matrix or $n \times m$ matrix, i.e. the m elements of $\mathbf{y}$ laid out in rows and the n elements of $\mathbf{x}$ laid out in columns, or vice versa. This leads to the following possibilities:

1. *Numerator layout*, i.e. lay out according to $\mathbf{y}$ and $\mathbf{x}^{\mathrm{T}}$ (i.e. contrarily to $\mathbf{x}$). This is sometimes known as the *Jacobian formulation*. This corresponds to the $m{\times}n$ layout in the previous example, which means that the row number of $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ equals to the size of the numerator $\mathbf{y}$ and the column number of $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ equals to the size of $\mathbf{x}^{\mathrm{T}}$.

2. *Denominator layout*, i.e. lay out according to $\mathbf{y}^{\mathrm{T}}$ and $\mathbf{x}$ (i.e. contrarily to $\mathbf{y}$). This is sometimes known as the *Hessian formulation*. Some authors term this layout the *gradient*, in distinction to the *Jacobian* (numerator layout), which is its transpose. (However, <u>gradient</u> more commonly means the derivative $\frac{\partial y}{\partial \mathbf{x}}$, regardless of layout.). This corresponds to the $n{\times}m$ layout in the previous example, which means that the row number of $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ equals to the size of $\mathbf{x}$ (the denominator).

3. A third possibility sometimes seen is to insist on writing the derivative as $\frac{\partial \mathbf{y}}{\partial \mathbf{x}'}$, (i.e. the derivative is taken with respect to the transpose of $\mathbf{x}$) and follow the numerator layout. This makes it possible to claim that the matrix is laid out according to both numerator and denominator. In practice this produces results the same as the numerator layout.

When handling the <u>gradient</u> $\frac{\partial y}{\partial \mathbf{x}}$ and the opposite case $\frac{\partial \mathbf{y}}{\partial x}$, we have the same issues. To be consistent, we should do one of the following:

1. If we choose numerator layout for $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$, we should lay out the <u>gradient</u> $\frac{\partial y}{\partial \mathbf{x}}$ as a row vector, and $\frac{\partial \mathbf{y}}{\partial x}$ as a column vector.

2. If we choose denominator layout for $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$, we should lay out the <u>gradient</u> $\frac{\partial y}{\partial \mathbf{x}}$ as a column vector, and $\frac{\partial \mathbf{y}}{\partial x}$ as a row vector.

3. In the third possibility above, we write $\frac{\partial y}{\partial \mathbf{x}'}$ and $\frac{\partial \mathbf{y}}{\partial x}$, and use numerator layout.

Not all math textbooks and papers are consistent in this respect throughout. That is, sometimes different conventions are used in different contexts within the same book or paper. For example, some choose denominator layout for gradients (laying them out as column vectors), but numerator layout for the vector-by-vector derivative $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$.

Similarly, when it comes to scalar-by-matrix derivatives $\frac{\partial y}{\partial \mathbf{X}}$ and matrix-by-scalar derivatives $\frac{\partial \mathbf{Y}}{\partial x}$, then consistent numerator layout lays out according to $\mathbf{Y}$ and $\mathbf{X}^{\mathrm{T}}$, while consistent denominator layout lays out according to $\mathbf{Y}^{\mathrm{T}}$ and $\mathbf{X}$. In practice, however, following a denominator layout for $\frac{\partial \mathbf{Y}}{\partial x}$, and laying the result out according to $\mathbf{Y}^{\mathrm{T}}$, is rarely seen because it makes for ugly formulas that do not correspond to the scalar formulas. As a result, the following layouts can often be found:

1. *Consistent numerator layout*, which lays out $\frac{\partial \mathbf{Y}}{\partial x}$ according to $\mathbf{Y}$ and $\frac{\partial y}{\partial \mathbf{X}}$ according to $\mathbf{X}^{\mathrm{T}}$.

2. *Mixed layout*, which lays out $\frac{\partial \mathbf{Y}}{\partial x}$ according to $\mathbf{Y}$ and $\frac{\partial y}{\partial \mathbf{X}}$ according to $\mathbf{X}$.

3. Use the notation $\frac{\partial y}{\partial \mathbf{X}'}$, with results the same as consistent numerator layout.

In the following formulas, we handle the five possible combinations $\frac{\partial y}{\partial \mathbf{x}}, \frac{\partial \mathbf{y}}{\partial x}, \frac{\partial \mathbf{y}}{\partial \mathbf{x}}, \frac{\partial y}{\partial \mathbf{X}}$ and $\frac{\partial \mathbf{Y}}{\partial x}$ separately. We also handle cases of scalar-by-scalar derivatives that involve an intermediate vector or matrix. (This can arise, for example, if a multi-dimensional <u>parametric curve</u> is defined in terms of a scalar variable, and then a derivative of a

scalar function of the curve is taken with respect to the scalar that parameterizes the curve.) For each of the various combinations, we give numerator-layout and denominator-layout results, except in the cases above where denominator layout rarely occurs. In cases involving matrices where it makes sense, we give numerator-layout and mixed-layout results. As noted above, cases where vector and matrix denominators are written in transpose notation are equivalent to numerator layout with the denominators written without the transpose.

Keep in mind that various authors use different combinations of numerator and denominator layouts for different types of derivatives, and there is no guarantee that an author will consistently use either numerator or denominator layout for all types. Match up the formulas below with those quoted in the source to determine the layout used for that particular type of derivative, but be careful not to assume that derivatives of other types necessarily follow the same kind of layout.

When taking derivatives with an aggregate (vector or matrix) denominator in order to find a maximum or minimum of the aggregate, it should be kept in mind that using numerator layout will produce results that are transposed with respect to the aggregate. For example, in attempting to find the maximum likelihood estimate of a multivariate normal distribution using matrix calculus, if the domain is a $k{\times}1$ column vector, then the result using the numerator layout will be in the form of a $1{\times}k$ row vector. Thus, either the results should be transposed at the end or the denominator layout (or mixed layout) should be used.

<div align="center">Result of differentiating various kinds of aggregates with other kinds of aggregates</div>

		Scalar y		Column vector y (size $m{\times}1$)		Matrix Y (size $m{\times}n$)	
		Notation	**Type**	**Notation**	**Type**	**Notation**	**Type**
Scalar x	**Numerator**	$\dfrac{\partial y}{\partial x}$	Scalar	$\dfrac{\partial \mathbf{y}}{\partial x}$	Size-m column vector	$\dfrac{\partial \mathbf{Y}}{\partial x}$	$m{\times}n$ matrix
	Denominator				Size-m row vector		
Column vector x (size $n{\times}1$)	**Numerator**	$\dfrac{\partial y}{\partial \mathbf{x}}$	Size-n row vector	$\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$	$m{\times}n$ matrix	$\dfrac{\partial \mathbf{Y}}{\partial \mathbf{x}}$	
	Denominator		Size-n column vector		$n{\times}m$ matrix		
Matrix X (size $p{\times}q$)	**Numerator**	$\dfrac{\partial y}{\partial \mathbf{X}}$	$q{\times}p$ matrix	$\dfrac{\partial \mathbf{y}}{\partial \mathbf{X}}$		$\dfrac{\partial \mathbf{Y}}{\partial \mathbf{X}}$	
	Denominator		$p{\times}q$ matrix				

The results of operations will be transposed when switching between numerator-layout and denominator-layout notation.

Numerator-layout notation

Using numerator-layout notation, we have:[1]

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

The following definitions are only provided in numerator-layout notation:

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}.$$

$$d\mathbf{X} = \begin{bmatrix} dx_{11} & dx_{12} & \cdots & dx_{1n} \\ dx_{21} & dx_{22} & \cdots & dx_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ dx_{m1} & dx_{m2} & \cdots & dx_{mn} \end{bmatrix}.$$

Denominator-layout notation

Using denominator-layout notation, we have:[2]

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1q}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{p1}} & \frac{\partial y}{\partial x_{p2}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

Identities

As noted above, in general, the results of operations will be transposed when switching between numerator-layout and denominator-layout notation.

To help make sense of all the identities below, keep in mind the most important rules: the chain rule, product rule and sum rule. The sum rule applies universally, and the product rule applies in most of the cases below, provided that the order of matrix products is maintained, since matrix products are not commutative. The chain rule applies in some of the cases, but unfortunately does *not* apply in matrix-by-scalar derivatives or scalar-by-matrix derivatives (in the latter case, mostly involving the trace operator applied to matrices). In the latter case, the product rule can't quite be applied directly, either, but the equivalent can be done with a bit more work using the differential identities.

The following identities adopt the following conventions:

- the scalars, *a*, *b*, *c*, *d*, and *e* are constant in respect of, and the scalars, *u*, and *v* are functions of one of *x*, $\mathbf{x}$, or $\mathbf{X}$;
- the vectors, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$, and $\mathbf{e}$ are constant in respect of, and the vectors, $\mathbf{u}$, and $\mathbf{v}$ are functions of one of *x*, $\mathbf{x}$, or $\mathbf{X}$;
- the matrices, $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{E}$ are constant in respect of, and the matrices, $\mathbf{U}$ and $\mathbf{V}$ are functions of one of *x*, $\mathbf{x}$, or $\mathbf{X}$.

Vector-by-vector identities

This is presented first because all of the operations that apply to vector-by-vector differentiation apply directly to vector-by-scalar or scalar-by-vector differentiation simply by reducing the appropriate vector in the numerator or denominator to a scalar.

Identities: vector-by-vector $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by $\mathbf{y}$ and $\mathbf{x}^{\mathsf{T}}$	Denominator layout, i.e. by $\mathbf{y}^{\mathsf{T}}$ and $\mathbf{x}$
$\mathbf{a}$ is not a function of $\mathbf{x}$	$\dfrac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\dfrac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^{\mathsf{T}}$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\dfrac{\partial \mathbf{x}^{\mathsf{T}}\mathbf{A}}{\partial \mathbf{x}} =$	$\mathbf{A}^{\mathsf{T}}$	$\mathbf{A}$
a is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\dfrac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
$v = v(\mathbf{x})$, $\mathbf{a}$ is not a function of $\mathbf{x}$	$\dfrac{\partial v\mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{a}\dfrac{\partial v}{\partial \mathbf{x}}$	$\dfrac{\partial v}{\partial \mathbf{x}}\mathbf{a}^{\mathsf{T}}$
$v = v(\mathbf{x})$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\dfrac{\partial v\mathbf{u}}{\partial \mathbf{x}} =$	$v\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\dfrac{\partial v}{\partial \mathbf{x}}$	$v\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}} + \dfrac{\partial v}{\partial \mathbf{x}}\mathbf{u}^{\mathsf{T}}$
$\mathbf{A}$ is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\dfrac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^{\mathsf{T}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\dfrac{\partial(\mathbf{u} + \mathbf{v})}{\partial \mathbf{x}} =$	$\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}} + \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\dfrac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\dfrac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

Scalar-by-vector identities

The fundamental identities are placed above the thick black line.

Identities: scalar-by-vector $\dfrac{\partial y}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} y$

Condition	Expression	Numerator layout, i.e. by $\mathbf{x}^{\mathsf{T}}$; result is row vector	Denominator layout, i.e. by $\mathbf{x}$; result is column vector
a is not a function of $\mathbf{x}$	$\dfrac{\partial a}{\partial \mathbf{x}} =$	$\mathbf{0}^{\mathsf{T}}$ [nb 1]	$\mathbf{0}$ [nb 1]
a is not a function of $\mathbf{x}$, $u = u(\mathbf{x})$	$\dfrac{\partial au}{\partial \mathbf{x}} =$	$a\dfrac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$, $v = v(\mathbf{x})$	$\dfrac{\partial(u+v)}{\partial \mathbf{x}} =$	$\dfrac{\partial u}{\partial \mathbf{x}} + \dfrac{\partial v}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$, $v = v(\mathbf{x})$	$\dfrac{\partial uv}{\partial \mathbf{x}} =$	$u\dfrac{\partial v}{\partial \mathbf{x}} + v\dfrac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$	$\dfrac{\partial g(u)}{\partial \mathbf{x}} =$	$\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{x}}$	
$u = u(\mathbf{x})$	$\dfrac{\partial f(g(u))}{\partial \mathbf{x}} =$	$\dfrac{\partial f(g)}{\partial g}\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\dfrac{\partial(\mathbf{u}\cdot\mathbf{v})}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{u}^{\mathsf{T}}\mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^{\mathsf{T}}\dfrac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^{\mathsf{T}}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ in numerator layout	$\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{v} + \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{u}$ $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ in denominator layout
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$, $\mathbf{A}$ is not a function of $\mathbf{x}$	$\dfrac{\partial(\mathbf{u}\cdot\mathbf{A}\mathbf{v})}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{u}^{\mathsf{T}}\mathbf{A}\mathbf{v}}{\partial \mathbf{x}} =$	$\mathbf{u}^{\mathsf{T}}\mathbf{A}\dfrac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ in numerator layout	$\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}\mathbf{v} + \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{A}^{\mathsf{T}}\mathbf{u}$ $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ in denominator layout
	$\dfrac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^{\mathsf{T}}} =$	$\mathbf{H}^{\mathsf{T}}$	$\mathbf{H}$, the Hessian matrix [3]
$\mathbf{a}$ is not a function of $\mathbf{x}$	$\dfrac{\partial(\mathbf{a}\cdot\mathbf{x})}{\partial \mathbf{x}} = \dfrac{\partial(\mathbf{x}\cdot\mathbf{a})}{\partial \mathbf{x}} =$ $\dfrac{\partial \mathbf{a}^{\mathsf{T}}\mathbf{x}}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{x}^{\mathsf{T}}\mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{a}^{\mathsf{T}}$	$\mathbf{a}$
$\mathbf{A}$ is not a function of $\mathbf{x}$ $\mathbf{b}$ is not a function of $\mathbf{x}$	$\dfrac{\partial \mathbf{b}^{\mathsf{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{b}^{\mathsf{T}}\mathbf{A}$	$\mathbf{A}^{\mathsf{T}}\mathbf{b}$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\dfrac{\partial \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{x}^{\mathsf{T}}\left(\mathbf{A} + \mathbf{A}^{\mathsf{T}}\right)$	$\left(\mathbf{A} + \mathbf{A}^{\mathsf{T}}\right)\mathbf{x}$
$\mathbf{A}$ is not a function of $\mathbf{x}$ $\mathbf{A}$ is symmetric	$\dfrac{\partial \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$2\mathbf{x}^{\mathsf{T}}\mathbf{A}$	$2\mathbf{A}\mathbf{x}$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\dfrac{\partial^2 \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}\partial \mathbf{x}^{\mathsf{T}}} =$	$\mathbf{A} + \mathbf{A}^{\mathsf{T}}$	

A is not a function of **x** **A** is symmetric	$\dfrac{\partial^2 \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x} \partial \mathbf{x}^\top} =$	$2\mathbf{A}$	
	$\dfrac{\partial(\mathbf{x} \cdot \mathbf{x})}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = \dfrac{\partial \|\mathbf{x}\|^2}{\partial \mathbf{x}} =$	$2\mathbf{x}^\top$	$2\mathbf{x}$
a is not a function of **x**, **u** = **u**(**x**)	$\dfrac{\partial(\mathbf{a} \cdot \mathbf{u})}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{a}^\top \mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{a}^\top \dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ in numerator layout	$\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{a}$ $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ in denominator layout
a, **b** are not functions of **x**	$\dfrac{\partial\, \mathbf{a}^\top \mathbf{x} \mathbf{x}^\top \mathbf{b}}{\partial\, \mathbf{x}} =$	$\mathbf{x}^\top \left(\mathbf{a}\mathbf{b}^\top + \mathbf{b}\mathbf{a}^\top\right)$	$\left(\mathbf{a}\mathbf{b}^\top + \mathbf{b}\mathbf{a}^\top\right)\mathbf{x}$
A, **b**, **C**, **D**, **e** are not functions of **x**	$\dfrac{\partial\, (\mathbf{A}\mathbf{x} + \mathbf{b})^\top \mathbf{C}(\mathbf{D}\mathbf{x} + \mathbf{e})}{\partial\, \mathbf{x}} =$	$(\mathbf{D}\mathbf{x} + \mathbf{e})^\top \mathbf{C}^\top \mathbf{A} + (\mathbf{A}\mathbf{x} + \mathbf{b})^\top \mathbf{C}\mathbf{D}$	$\mathbf{D}^\top \mathbf{C}^\top (\mathbf{A}\mathbf{x} + \mathbf{b}) + \mathbf{A}^\top \mathbf{C}(\mathbf{D}\mathbf{x} + \mathbf{e})$
a is not a function of **x**	$\dfrac{\partial\, \|\mathbf{x} - \mathbf{a}\|}{\partial\, \mathbf{x}} =$	$\dfrac{(\mathbf{x} - \mathbf{a})^\top}{\|\mathbf{x} - \mathbf{a}\|}$	$\dfrac{\mathbf{x} - \mathbf{a}}{\|\mathbf{x} - \mathbf{a}\|}$

Vector-by-scalar identities

Identities: vector-by-scalar $\dfrac{\partial \mathbf{y}}{\partial x}$

Condition	**Expression**	**Numerator layout, i.e. by y, result is column vector**	**Denominator layout, i.e. by $\mathbf{y}^\mathrm{T}$, result is row vector**
a is not a function of x	$\dfrac{\partial \mathbf{a}}{\partial x} =$	$\mathbf{0}^{[nb\ 1]}$	
a is not a function of x, **u** = **u**(x)	$\dfrac{\partial a\mathbf{u}}{\partial x} =$	$a\dfrac{\partial \mathbf{u}}{\partial x}$	
A is not a function of x, **u** = **u**(x)	$\dfrac{\partial \mathbf{A}\mathbf{u}}{\partial x} =$	$\mathbf{A}\dfrac{\partial \mathbf{u}}{\partial x}$	$\dfrac{\partial \mathbf{u}}{\partial x}\mathbf{A}^\top$
u = **u**(x)	$\dfrac{\partial \mathbf{u}^\top}{\partial x} =$	$\left(\dfrac{\partial \mathbf{u}}{\partial x}\right)^\top$	
u = **u**(x), **v** = **v**(x)	$\dfrac{\partial(\mathbf{u} + \mathbf{v})}{\partial x} =$	$\dfrac{\partial \mathbf{u}}{\partial x} + \dfrac{\partial \mathbf{v}}{\partial x}$	
u = **u**(x), **v** = **v**(x)	$\dfrac{\partial(\mathbf{u}^\top \times \mathbf{v})}{\partial x} =$	$\left(\dfrac{\partial \mathbf{u}}{\partial x}\right)^\top \times \mathbf{v} + \mathbf{u}^\top \times \dfrac{\partial \mathbf{v}}{\partial x}$	$\dfrac{\partial \mathbf{u}}{\partial x} \times \mathbf{v} + \mathbf{u}^\top \times \left(\dfrac{\partial \mathbf{v}}{\partial x}\right)^\top$
u = **u**(x)	$\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial x} =$	$\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial x}$ Assumes consistent matrix layout; see below.	$\dfrac{\partial \mathbf{u}}{\partial x}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$ Assumes consistent matrix layout; see below.
u = **u**(x)	$\dfrac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial x} =$	$\dfrac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial x}$ Assumes consistent matrix layout; see below.	$\dfrac{\partial \mathbf{u}}{\partial x}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$ Assumes consistent matrix layout; see below.
U = **U**(x), **v** = **v**(x)	$\dfrac{\partial(\mathbf{U} \times \mathbf{v})}{\partial x} =$	$\dfrac{\partial \mathbf{U}}{\partial x} \times \mathbf{v} + \mathbf{U} \times \dfrac{\partial \mathbf{v}}{\partial x}$	$\mathbf{v}^\top \times \left(\dfrac{\partial \mathbf{U}}{\partial x}\right) + \dfrac{\partial \mathbf{v}}{\partial x} \times \mathbf{U}^\top$

NOTE: The formulas involving the vector-by-vector derivatives $\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$ and $\dfrac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$ (whose outputs are matrices) assume the matrices are laid out consistent with the vector layout, i.e. numerator-layout matrix when numerator-layout vector and vice versa; otherwise, transpose the vector-by-vector derivatives.

Scalar-by-matrix identities

Note that exact equivalents of the scalar product rule and chain rule do not exist when applied to matrix-valued functions of matrices. However, the product rule of this sort does apply to the differential form (see below), and this is the way to derive many of the identities below involving the trace function, combined with the fact that the trace function allows transposing and cyclic permutation, i.e.:

$$\mathrm{tr}(\mathbf{A}) = \mathrm{tr}\left(\mathbf{A}^\top\right)$$
$$\mathrm{tr}(\mathbf{ABCD}) = \mathrm{tr}(\mathbf{BCDA}) = \mathrm{tr}(\mathbf{CDAB}) = \mathrm{tr}(\mathbf{DABC})$$

For example, to compute $\dfrac{\partial\,\mathrm{tr}(\mathbf{AXBX}^\top\mathbf{C})}{\partial\mathbf{X}}$:

$$
\begin{aligned}
d\,\mathrm{tr}(\mathbf{AXBX}^\top\mathbf{C}) &= d\,\mathrm{tr}\left(\mathbf{CAXBX}^\top\right) = \mathrm{tr}\left(d\left(\mathbf{CAXBX}^\top\right)\right) \\
&= \mathrm{tr}\left(\mathbf{CAX}d(\mathbf{BX}^\top) + d\left(\mathbf{CAX}\right)\mathbf{BX}^\top\right) \\
&= \mathrm{tr}\left(\mathbf{CAX}d\left(\mathbf{BX}^\top\right)\right) + \mathrm{tr}\left(d(\mathbf{CAX})\mathbf{BX}^\top\right) \\
&= \mathrm{tr}\left(\mathbf{CAXB}d\left(\mathbf{X}^\top\right)\right) + \mathrm{tr}\left(\mathbf{CA}(d\mathbf{X})\mathbf{BX}^\top\right) \\
&= \mathrm{tr}\left(\mathbf{CAXB}(d\mathbf{X})^\top\right) + \mathrm{tr}\left(\mathbf{CA}\left(d\mathbf{X}\right)\mathbf{BX}^\top\right) \\
&= \mathrm{tr}\left(\left(\mathbf{CAXB}(d\mathbf{X})^\top\right)^\top\right) + \mathrm{tr}\left(\mathbf{CA}(d\mathbf{X})\mathbf{BX}^\top\right) \\
&= \mathrm{tr}\left((d\mathbf{X})\mathbf{B}^\top\mathbf{X}^\top\mathbf{A}^\top\mathbf{C}^\top\right) + \mathrm{tr}\left(\mathbf{CA}(d\mathbf{X})\mathbf{BX}^\top\right) \\
&= \mathrm{tr}\left(\mathbf{B}^\top\mathbf{X}^\top\mathbf{A}^\top\mathbf{C}^\top(d\mathbf{X})\right) + \mathrm{tr}\left(\mathbf{BX}^\top\mathbf{CA}(d\mathbf{X})\right) \\
&= \mathrm{tr}\left(\left(\mathbf{B}^\top\mathbf{X}^\top\mathbf{A}^\top\mathbf{C}^\top + \mathbf{BX}^\top\mathbf{CA}\right)d\mathbf{X}\right) \\
&= \mathrm{tr}\left(\left(\mathbf{CAXB} + \mathbf{A}^\top\mathbf{C}^\top\mathbf{XB}^\top\right)^\top d\mathbf{X}\right)
\end{aligned}
$$

Therefore,

$$\frac{\partial\,\mathrm{tr}\left(\mathbf{AXBX}^\top\mathbf{C}\right)}{\partial\mathbf{X}} = \mathbf{B}^\top\mathbf{X}^\top\mathbf{A}^\top\mathbf{C}^\top + \mathbf{BX}^\top\mathbf{CA}.\text{ (numerator layout)}$$
$$\frac{\partial\,\mathrm{tr}\left(\mathbf{AXBX}^\top\mathbf{C}\right)}{\partial\mathbf{X}} = \mathbf{CAXB} + \mathbf{A}^\top\mathbf{C}^\top\mathbf{XB}^\top.\text{ (denominator layout)}$$

(For the last step, see the Conversion from differential to derivative form section.)

Identities: scalar-by-matrix $\dfrac{\partial y}{\partial \mathbf{X}}$

Condition	Expression	Numerator layout, i.e. by $\mathbf{X}^\mathsf{T}$	Denominator layout, i.e. by $\mathbf{X}$
a is not a function of $\mathbf{X}$	$\dfrac{\partial a}{\partial \mathbf{X}} =$	$\mathbf{0}^\mathsf{T}$ [nb 2]	$\mathbf{0}$ [nb 2]
a is not a function of $\mathbf{X}$, $u = u(\mathbf{X})$	$\dfrac{\partial au}{\partial \mathbf{X}} =$	$a\dfrac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$, $v = v(\mathbf{X})$	$\dfrac{\partial(u + v)}{\partial \mathbf{X}} =$	$\dfrac{\partial u}{\partial \mathbf{X}} + \dfrac{\partial v}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$, $v = v(\mathbf{X})$	$\dfrac{\partial uv}{\partial \mathbf{X}} =$	$u\dfrac{\partial v}{\partial \mathbf{X}} + v\dfrac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$	$\dfrac{\partial g(u)}{\partial \mathbf{X}} =$	$\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{X}}$	
$u = u(\mathbf{X})$	$\dfrac{\partial f(g(u))}{\partial \mathbf{X}} =$	$\dfrac{\partial f(g)}{\partial g}\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{X}}$	
$\mathbf{U} = \mathbf{U}(\mathbf{X})$	[3] $\dfrac{\partial g(\mathbf{U})}{\partial X_{ij}} =$	$\operatorname{tr}\left(\dfrac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\dfrac{\partial \mathbf{U}}{\partial X_{ij}}\right)$	$\operatorname{tr}\left(\left(\dfrac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\right)^\mathsf{T}\dfrac{\partial \mathbf{U}}{\partial X_{ij}}\right)$
		Both forms assume *numerator* layout for $\dfrac{\partial \mathbf{U}}{\partial X_{ij}}$, i.e. mixed layout if denominator layout for $\mathbf{X}$ is being used.	
$\mathbf{a}$ and $\mathbf{b}$ are not functions of $\mathbf{X}$	$\dfrac{\partial \mathbf{a}^\mathsf{T}\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} =$	$\mathbf{b}\mathbf{a}^\mathsf{T}$	$\mathbf{a}\mathbf{b}^\mathsf{T}$
$\mathbf{a}$ and $\mathbf{b}$ are not functions of $\mathbf{X}$	$\dfrac{\partial \mathbf{a}^\mathsf{T}\mathbf{X}^\mathsf{T}\mathbf{b}}{\partial \mathbf{X}} =$	$\mathbf{a}\mathbf{b}^\mathsf{T}$	$\mathbf{b}\mathbf{a}^\mathsf{T}$
$\mathbf{a}$, $\mathbf{b}$ and $\mathbf{C}$ are not functions of $\mathbf{X}$	$\dfrac{\partial(\mathbf{Xa} + \mathbf{b})^\mathsf{T}\mathbf{C}(\mathbf{Xa} + \mathbf{b})}{\partial \mathbf{X}} =$	$\left(\left(\mathbf{C} + \mathbf{C}^\mathsf{T}\right)(\mathbf{Xa} + \mathbf{b})\mathbf{a}^\mathsf{T}\right)^\mathsf{T}$	$\left(\mathbf{C} + \mathbf{C}^\mathsf{T}\right)(\mathbf{Xa} + \mathbf{b})\mathbf{a}^\mathsf{T}$
$\mathbf{a}$, $\mathbf{b}$ and $\mathbf{C}$ are not functions of $\mathbf{X}$	$\dfrac{\partial(\mathbf{Xa})^\mathsf{T}\mathbf{C}(\mathbf{Xb})}{\partial \mathbf{X}} =$	$\left(\mathbf{CXb}\mathbf{a}^\mathsf{T} + \mathbf{C}^\mathsf{T}\mathbf{Xa}\mathbf{b}^\mathsf{T}\right)^\mathsf{T}$	$\mathbf{CXb}\mathbf{a}^\mathsf{T} + \mathbf{C}^\mathsf{T}\mathbf{Xa}\mathbf{b}^\mathsf{T}$
	$\dfrac{\partial \operatorname{tr}(\mathbf{X})}{\partial \mathbf{X}} =$	$\mathbf{I}$	
$\mathbf{U} = \mathbf{U}(\mathbf{X})$, $\mathbf{V} = \mathbf{V}(\mathbf{X})$	$\dfrac{\partial \operatorname{tr}(\mathbf{U} + \mathbf{V})}{\partial \mathbf{X}} =$	$\dfrac{\partial \operatorname{tr}(\mathbf{U})}{\partial \mathbf{X}} + \dfrac{\partial \operatorname{tr}(\mathbf{V})}{\partial \mathbf{X}}$	
a is not a function of $\mathbf{X}$, $\mathbf{U} = \mathbf{U}(\mathbf{X})$	$\dfrac{\partial \operatorname{tr}(a\mathbf{U})}{\partial \mathbf{X}} =$	$a\dfrac{\partial \operatorname{tr}(\mathbf{U})}{\partial \mathbf{X}}$	
$g(\mathbf{X})$ is any polynomial with scalar coefficients, or any matrix function defined by an infinite polynomial series (e.g. $e^\mathbf{X}$, $\sin(\mathbf{X})$, $\cos(\mathbf{X})$, $\ln(\mathbf{X})$, etc. using a Taylor series); $g(x)$ is the	$\dfrac{\partial \operatorname{tr}(g(\mathbf{X}))}{\partial \mathbf{X}} =$	$\mathbf{g}'(\mathbf{X})$	$\left(\mathbf{g}'(\mathbf{X})\right)^\mathsf{T}$

equivalent scalar function, $g'(x)$ is its derivative, and $\mathbf{g}'(\mathbf{X})$ is the corresponding matrix function			
$\mathbf{A}$ is not a function of $\mathbf{X}$	$\dfrac{\partial \operatorname{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \dfrac{\partial \operatorname{tr}(\mathbf{XA})}{\partial \mathbf{X}} =$ [4]	$\mathbf{A}$	$\mathbf{A}^\top$
$\mathbf{A}$ is not a function of $\mathbf{X}$	$\dfrac{\partial \operatorname{tr}(\mathbf{AX}^\top)}{\partial \mathbf{X}} = \dfrac{\partial \operatorname{tr}(\mathbf{X}^\top \mathbf{A})}{\partial \mathbf{X}} =$ [3]	$\mathbf{A}^\top$	$\mathbf{A}$
$\mathbf{A}$ is not a function of $\mathbf{X}$	[3] $\dfrac{\partial \operatorname{tr}(\mathbf{X}^\top \mathbf{AX})}{\partial \mathbf{X}} =$	$\mathbf{X}^\top \left(\mathbf{A} + \mathbf{A}^\top\right)$	$\left(\mathbf{A} + \mathbf{A}^\top\right)\mathbf{X}$
$\mathbf{A}$ is not a function of $\mathbf{X}$	[3] $\dfrac{\partial \operatorname{tr}(\mathbf{X}^{-1}\mathbf{A})}{\partial \mathbf{X}} =$	$-\mathbf{X}^{-1}\mathbf{AX}^{-1}$	$-\left(\mathbf{X}^{-1}\right)^\top \mathbf{A}^\top \left(\mathbf{X}^{-1}\right)^\top$
$\mathbf{A}, \mathbf{B}$ are not functions of $\mathbf{X}$	$\dfrac{\partial \operatorname{tr}(\mathbf{AXB})}{\partial \mathbf{X}} = \dfrac{\partial \operatorname{tr}(\mathbf{BAX})}{\partial \mathbf{X}} =$	$\mathbf{BA}$	$\mathbf{A}^\top \mathbf{B}^\top$
$\mathbf{A}, \mathbf{B}, \mathbf{C}$ are not functions of $\mathbf{X}$	$\dfrac{\partial \operatorname{tr}\left(\mathbf{AXB}\mathbf{X}^\top \mathbf{C}\right)}{\partial \mathbf{X}} =$	$\mathbf{BX}^\top \mathbf{CA} + \mathbf{B}^\top \mathbf{X}^\top \mathbf{A}^\top \mathbf{C}^\top$	$\mathbf{A}^\top \mathbf{C}^\top \mathbf{XB}^\top + \mathbf{CAXB}$
n is a positive integer	[3] $\dfrac{\partial \operatorname{tr}(\mathbf{X}^n)}{\partial \mathbf{X}} =$	$n\mathbf{X}^{n-1}$	$n\left(\mathbf{X}^{n-1}\right)^\top$
$\mathbf{A}$ is not a function of $\mathbf{X}$, n is a positive integer	[3] $\dfrac{\partial \operatorname{tr}(\mathbf{AX}^n)}{\partial \mathbf{X}} =$	$\displaystyle\sum_{i=0}^{n-1} \mathbf{X}^i \mathbf{AX}^{n-i-1}$	$\displaystyle\sum_{i=0}^{n-1} \left(\mathbf{X}^i \mathbf{AX}^{n-i-1}\right)^\top$
	[3] $\dfrac{\partial \operatorname{tr}\left(e^{\mathbf{X}}\right)}{\partial \mathbf{X}} =$	$e^{\mathbf{X}}$	$\left(e^{\mathbf{X}}\right)^\top$
	[3] $\dfrac{\partial \operatorname{tr}(\sin(\mathbf{X}))}{\partial \mathbf{X}} =$	$\cos(\mathbf{X})$	$(\cos(\mathbf{X}))^\top$
	[5] $\dfrac{\partial |\mathbf{X}|}{\partial \mathbf{X}} =$	$\operatorname{cofactor}(X)^\top = |\mathbf{X}|\mathbf{X}^{-1}$	$\operatorname{cofactor}(X) = |\mathbf{X}|\left(\mathbf{X}^{-1}\right)^\top$
a is not a function of $\mathbf{X}$	[3] $\dfrac{\partial \ln|a\mathbf{X}|}{\partial \mathbf{X}} =$ [nb 3]	$\mathbf{X}^{-1}$	$\left(\mathbf{X}^{-1}\right)^\top$
$\mathbf{A}, \mathbf{B}$ are not functions of $\mathbf{X}$	[3] $\dfrac{\partial |\mathbf{AXB}|}{\partial \mathbf{X}} =$	$|\mathbf{AXB}|\mathbf{X}^{-1}$	$|\mathbf{AXB}|\left(\mathbf{X}^{-1}\right)^\top$
n is a positive integer	[3] $\dfrac{\partial |\mathbf{X}^n|}{\partial \mathbf{X}} =$	$n\,|\mathbf{X}^n|\,\mathbf{X}^{-1}$	$n\,|\mathbf{X}^n|\left(\mathbf{X}^{-1}\right)^\top$
(see pseudo-inverse)	[3] $\dfrac{\partial \ln\left|\mathbf{X}^\top \mathbf{X}\right|}{\partial \mathbf{X}} =$	$2\mathbf{X}^+$	$2\left(\mathbf{X}^+\right)^\top$
(see pseudo-inverse)	[3] $\dfrac{\partial \ln\left|\mathbf{X}^\top \mathbf{X}\right|}{\partial \mathbf{X}^+} =$	$-2\mathbf{X}$	$-2\mathbf{X}^\top$
$\mathbf{A}$ is not a function of $\mathbf{X}$, $\mathbf{X}$ is square and invertible	$\dfrac{\partial \left|\mathbf{X}^\top \mathbf{AX}\right|}{\partial \mathbf{X}} =$	$2\left|\mathbf{X}^\top \mathbf{AX}\right|\mathbf{X}^{-1} = 2\left|\mathbf{X}^\top\right||\mathbf{A}||\mathbf{X}|\mathbf{X}^{-1}$	$2\left|\mathbf{X}^\top \mathbf{AX}\right|\left(\mathbf{X}^{-1}\right)^\top$
$\mathbf{A}$ is not a function of $\mathbf{X}$, $\mathbf{X}$ is non-square,	$\dfrac{\partial \left|\mathbf{X}^\top \mathbf{AX}\right|}{\partial \mathbf{X}} =$	$2\left|\mathbf{X}^\top \mathbf{AX}\right|\left(\mathbf{X}^\top \mathbf{A}^\top \mathbf{X}\right)^{-1}\mathbf{X}^\top \mathbf{A}^\top$	$2\left|\mathbf{X}^\top \mathbf{AX}\right|\mathbf{AX}\left(\mathbf{X}^\top \mathbf{AX}\right)^{-1}$

A is symmetric		
A is not a function of **X**, **X** is non-square, **A** is non-symmetric	$\dfrac{\partial \lvert \mathbf{X}^\top \mathbf{AX} \rvert}{\partial \mathbf{X}} =$	$\lvert \mathbf{X}^\top \mathbf{AX} \rvert \left((\mathbf{X}^\top \mathbf{AX})^{-1} \mathbf{X}^\top \mathbf{A} + (\mathbf{X}^\top \mathbf{A}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{A}^\top \right)$ $\lvert \mathbf{X}^\top \mathbf{AX} \rvert \left(\mathbf{AX}(\mathbf{X}^\top \mathbf{AX})^{-1} + \mathbf{A}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{A}^\top \mathbf{X})^{-1} \right)$

Matrix-by-scalar identities

Identities: matrix-by-scalar $\dfrac{\partial \mathbf{Y}}{\partial x}$

Condition	Expression	Numerator layout, i.e. by $\mathbf{Y}$
$\mathbf{U} = \mathbf{U}(x)$	$\dfrac{\partial a\mathbf{U}}{\partial x} =$	$a\dfrac{\partial \mathbf{U}}{\partial x}$
$\mathbf{A}, \mathbf{B}$ are not functions of x, $\mathbf{U} = \mathbf{U}(x)$	$\dfrac{\partial \mathbf{AUB}}{\partial x} =$	$\mathbf{A}\dfrac{\partial \mathbf{U}}{\partial x}\mathbf{B}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\dfrac{\partial (\mathbf{U} + \mathbf{V})}{\partial x} =$	$\dfrac{\partial \mathbf{U}}{\partial x} + \dfrac{\partial \mathbf{V}}{\partial x}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\dfrac{\partial (\mathbf{UV})}{\partial x} =$	$\mathbf{U}\dfrac{\partial \mathbf{V}}{\partial x} + \dfrac{\partial \mathbf{U}}{\partial x}\mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\dfrac{\partial (\mathbf{U} \otimes \mathbf{V})}{\partial x} =$	$\mathbf{U} \otimes \dfrac{\partial \mathbf{V}}{\partial x} + \dfrac{\partial \mathbf{U}}{\partial x} \otimes \mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$, $\mathbf{V} = \mathbf{V}(x)$	$\dfrac{\partial (\mathbf{U} \circ \mathbf{V})}{\partial x} =$	$\mathbf{U} \circ \dfrac{\partial \mathbf{V}}{\partial x} + \dfrac{\partial \mathbf{U}}{\partial x} \circ \mathbf{V}$
$\mathbf{U} = \mathbf{U}(x)$	$\dfrac{\partial \mathbf{U}^{-1}}{\partial x} =$	$-\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}$
$\mathbf{U} = \mathbf{U}(x,y)$	$\dfrac{\partial^2 \mathbf{U}^{-1}}{\partial x \partial y} =$	$\mathbf{U}^{-1}\left(\dfrac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial y} - \dfrac{\partial^2 \mathbf{U}}{\partial x \partial y} + \dfrac{\partial \mathbf{U}}{\partial y}\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x} \right)\mathbf{U}^{-1}$
$\mathbf{A}$ is not a function of x, $\mathbf{g}(\mathbf{X})$ is any polynomial with scalar coefficients, or any matrix function defined by an infinite polynomial series (e.g. $e^{\mathbf{X}}$, $\sin(\mathbf{X})$, $\cos(\mathbf{X})$, $\ln(\mathbf{X})$, etc.); $g(x)$ is the equivalent scalar function, $g'(x)$ is its derivative, and $\mathbf{g}'(\mathbf{X})$ is the corresponding matrix function	$\dfrac{\partial \mathbf{g}(x\mathbf{A})}{\partial x} =$	$\mathbf{A}\mathbf{g}'(x\mathbf{A}) = \mathbf{g}'(x\mathbf{A})\mathbf{A}$
$\mathbf{A}$ is not a function of x	$\dfrac{\partial e^{x\mathbf{A}}}{\partial x} =$	$\mathbf{A}e^{x\mathbf{A}} = e^{x\mathbf{A}}\mathbf{A}$

Scalar-by-scalar identities

With vectors involved

Identities: scalar-by-scalar, with vectors involved

Condition	Expression	Any layout (assumes dot product ignores row vs. column layout)

$\mathbf{u} = \mathbf{u}(x)$	$\dfrac{\partial g(\mathbf{u})}{\partial x} =$	$\dfrac{\partial g(\mathbf{u})}{\partial \mathbf{u}} \cdot \dfrac{\partial \mathbf{u}}{\partial x}$
$\mathbf{u} = \mathbf{u}(x),\ \mathbf{v} = \mathbf{v}(x)$	$\dfrac{\partial(\mathbf{u} \cdot \mathbf{v})}{\partial x} =$	$\mathbf{u} \cdot \dfrac{\partial \mathbf{v}}{\partial x} + \dfrac{\partial \mathbf{u}}{\partial x} \cdot \mathbf{v}$

With matrices involved

Identities: scalar-by-scalar, with matrices involved[3]

Condition	Expression	Consistent numerator layout, i.e. by $\mathbf{Y}$ and $\mathbf{X}^{\mathbf{T}}$	Mixed layout, i.e. by $\mathbf{Y}$ and $\mathbf{X}$
$\mathbf{U} = \mathbf{U}(x)$	$\dfrac{\partial \lvert\mathbf{U}\rvert}{\partial x} =$	$\lvert\mathbf{U}\rvert \operatorname{tr}\left(\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x}\right)$	
$\mathbf{U} = \mathbf{U}(x)$	$\dfrac{\partial \ln\lvert\mathbf{U}\rvert}{\partial x} =$	$\operatorname{tr}\left(\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x}\right)$	
$\mathbf{U} = \mathbf{U}(x)$	$\dfrac{\partial^2 \lvert\mathbf{U}\rvert}{\partial x^2} =$	$\lvert\mathbf{U}\rvert\left[\operatorname{tr}\left(\mathbf{U}^{-1}\dfrac{\partial^2 \mathbf{U}}{\partial x^2}\right) + \operatorname{tr}^2\left(\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x}\right) - \operatorname{tr}\left(\left(\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x}\right)^2\right)\right]$	
$\mathbf{U} = \mathbf{U}(x)$	$\dfrac{\partial g(\mathbf{U})}{\partial x} =$	$\operatorname{tr}\left(\dfrac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\dfrac{\partial \mathbf{U}}{\partial x}\right)$	$\operatorname{tr}\left(\left(\dfrac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\right)^{\top}\dfrac{\partial \mathbf{U}}{\partial x}\right)$
$\mathbf{A}$ is not a function of x, $\mathbf{g}(\mathbf{X})$ is any polynomial with scalar coefficients, or any matrix function defined by an infinite polynomial series (e.g. $e^{\mathbf{X}}$, $\sin(\mathbf{X})$, $\cos(\mathbf{X})$, $\ln(\mathbf{X})$, etc.); $g(x)$ is the equivalent scalar function, $g'(x)$ is its derivative, and $\mathbf{g}'(\mathbf{X})$ is the corresponding matrix function.	$\dfrac{\partial \operatorname{tr}(\mathbf{g}(x\mathbf{A}))}{\partial x} =$	$\operatorname{tr}(\mathbf{A}\mathbf{g}'(x\mathbf{A}))$	
$\mathbf{A}$ is not a function of x	$\dfrac{\partial \operatorname{tr}(e^{x\mathbf{A}})}{\partial x} =$	$\operatorname{tr}(\mathbf{A}e^{x\mathbf{A}})$	

Identities in differential form

It is often easier to work in differential form and then convert back to normal derivatives. This only works well using the numerator layout. In these rules, a is a scalar.

Differential identities: scalar involving matrix[1][3]

Expression	Result (numerator layout)
$d(\operatorname{tr}(\mathbf{X})) =$	$\operatorname{tr}(d\mathbf{X})$
$d(\lvert\mathbf{X}\rvert) =$	$\lvert\mathbf{X}\rvert\operatorname{tr}(\mathbf{X}^{-1}d\mathbf{X}) = \operatorname{tr}(\operatorname{adj}(\mathbf{X})d\mathbf{X})$
$d(\ln\lvert\mathbf{X}\rvert) =$	$\operatorname{tr}(\mathbf{X}^{-1}d\mathbf{X})$

Differential identities: matrix[1][3][6] [7]

Condition	Expression	Result (numerator layout)
A is not a function of **X**	$d(\mathbf{A}) =$	**0**
a is not a function of **X**	$d(a\mathbf{X}) =$	$a\,d\mathbf{X}$
	$d(\mathbf{X} + \mathbf{Y}) =$	$d\mathbf{X} + d\mathbf{Y}$
	$d(\mathbf{XY}) =$	$(d\mathbf{X})\mathbf{Y} + \mathbf{X}(d\mathbf{Y})$
(Kronecker product)	$d(\mathbf{X} \otimes \mathbf{Y}) =$	$(d\mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (d\mathbf{Y})$
(Hadamard product)	$d(\mathbf{X} \circ \mathbf{Y}) =$	$(d\mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (d\mathbf{Y})$
	$d\left(\mathbf{X}^\top\right) =$	$(d\mathbf{X})^\top$
	$d\left(\mathbf{X}^{-1}\right) =$	$-\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}$
(conjugate transpose)	$d\left(\mathbf{X}^{\mathrm{H}}\right) =$	$(d\mathbf{X})^{\mathrm{H}}$
n is a positive integer	$d\left(\mathbf{X}^n\right) =$	$\displaystyle\sum_{i=0}^{n-1} \mathbf{X}^i (d\mathbf{X})\mathbf{X}^{n-i-1}$
	$d\left(e^{\mathbf{X}}\right) =$	$\displaystyle\int_0^1 e^{a\mathbf{X}}(d\mathbf{X})e^{(1-a)\mathbf{X}}\,da$
	$d\left(\log X\right) =$	$\displaystyle\int_0^\infty (\mathbf{X} + z\,\mathbf{I})^{-1}(d\mathbf{X})(\mathbf{X} + z\,\mathbf{I})^{-1}\,dz$
$\mathbf{X} = \sum_i \lambda_i \mathbf{P}_i$ is diagonalizable $\mathbf{P}_i\mathbf{P}_j = \delta_{ij}\mathbf{P}_i$ *f* is differentiable at every eigenvalue λ_i	$d(f(\mathbf{X})) =$	$\displaystyle\sum_{ij} \mathbf{P}_i(d\mathbf{X})\mathbf{P}_j \begin{cases} f'(\lambda_i) & \lambda_i = \lambda_j \\ \frac{f(\lambda_i)-f(\lambda_j)}{\lambda_i-\lambda_j} & \lambda_i \neq \lambda_j \end{cases}$

In the last row, δ_{ij} is the Kronecker delta and $(\mathbf{P}_k)_{ij} = (\mathbf{Q})_{ik}(\mathbf{Q}^{-1})_{kj}$ is the set of orthogonal projection operators that project onto the *k*-th eigenvector of **X**. **Q** is the matrix of eigenvectors of $\mathbf{X} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$, and $(\mathbf{\Lambda})_{ii} = \lambda_i$ are the eigenvalues. The matrix function $f(\mathbf{X})$ is defined in terms of the scalar function $f(x)$ for diagonalizable matrices by $f(\mathbf{X}) = \sum_i f(\lambda_i)\mathbf{P}_i$ where $\mathbf{X} = \sum_i \lambda_i \mathbf{P}_i$ with $\mathbf{P}_i\mathbf{P}_j = \delta_{ij}\mathbf{P}_i$.

To convert to normal derivative form, first convert it to one of the following canonical forms, and then use these identities:

Conversion from differential to derivative form[1]

Canonical differential form	Equivalent derivative form (numerator layout)
$dy = a\,dx$	$\dfrac{dy}{dx} = a$
$dy = \mathbf{a}^\top d\mathbf{x}$	$\dfrac{dy}{d\mathbf{x}} = \mathbf{a}^\top$
$dy = \mathrm{tr}(\mathbf{A}\,d\mathbf{X})$	$\dfrac{dy}{d\mathbf{X}} = \mathbf{A}$
$d\mathbf{y} = \mathbf{a}\,dx$	$\dfrac{d\mathbf{y}}{dx} = \mathbf{a}$
$d\mathbf{y} = \mathbf{A}\,d\mathbf{x}$	$\dfrac{d\mathbf{y}}{d\mathbf{x}} = \mathbf{A}$
$d\mathbf{Y} = \mathbf{A}\,dx$	$\dfrac{d\mathbf{Y}}{dx} = \mathbf{A}$

Applications

Matrix differential calculus is used in statistics and econometrics, particularly for the statistical analysis of multivariate distributions, especially the multivariate normal distribution and other elliptical distributions.[8][9][10]

It is used in regression analysis to compute, for example, the ordinary least squares regression formula for the case of multiple explanatory variables.[11] It is also used in random matrices, statistical moments, local sensitivity and statistical diagnostics.[12] [13]

See also

√x **Mathematics portal**

- Derivative (generalizations)
- Product integral
- Ricci calculus
- Tensor derivative

Notes

1. Here, **0** refers to a column vector of all 0's, of size n, where n is the length of $\mathbf{x}$.
2. Here, **0** refers to a matrix of all 0's, of the same shape as $\mathbf{X}$.
3. The constant a disappears in the result. This is intentional. In general,

$$\frac{d\ln au}{dx} = \frac{1}{au}\frac{d(au)}{dx} = \frac{1}{au}a\frac{du}{dx} = \frac{1}{u}\frac{du}{dx} = \frac{d\ln u}{dx}.$$

or, also

$$\frac{d\ln au}{dx} = \frac{d(\ln a + \ln u)}{dx} = \frac{d\ln a}{dx} + \frac{d\ln u}{dx} = \frac{d\ln u}{dx}.$$

References

1. Thomas P., Minka (December 28, 2000). "Old and New Matrix Algebra Useful for Statistics" (http://research.microsoft.com/en-us/um/people/minka/papers/matrix/). MIT Media Lab note (1997; revised 12/00). Retrieved 5 February 2016.
2. Felippa, Carlos A. "Appendix D, Linear Algebra: Determinants, Inverses, Rank" (http://www.colorado.edu/engineering/cas/courses.d/IFEM.d/IFEM.AppD.d/IFEM.AppD.pdf) (PDF). *ASEN 5007: Introduction To Finite Element Methods*. Boulder, Colorado: University of Colorado. Retrieved 5 February 2016. Uses the Hessian (transpose to Jacobian) definition of vector and matrix derivatives.
3. Petersen, Kaare Brandt; Pedersen, Michael Syskind. *The Matrix Cookbook* (https://web.archive.org/web/20100302210536/http://www.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf) (PDF). Archived from the original (http://matrixcookbook.com) on 2 March 2010. Retrieved 5 February 2016. This book uses a mixed layout, i.e. by $\mathbf{Y}$ in $\frac{\partial \mathbf{Y}}{\partial x}$, by $\mathbf{X}$ in $\frac{\partial y}{\partial \mathbf{X}}$.
4. Duchi, John C. "Properties of the Trace and Matrix Derivatives" (https://web.stanford.edu/~jduchi/projects/matrix_prop.pdf) (PDF). Stanford University. Retrieved 5 February 2016.
5. See Determinant § Derivative for the derivation.
6. Giles, Michael B. (2008). "An extended collection of matrix derivative results for forward and reverse mode algorithmic differentiation" (https://web.archive.org/web/20200227075201/https://pdfs.semanticscholar.org/c74c/5e11ed05246c12165ce7e4b6222bd32d68dc.pdf) (PDF). S2CID 17431500 (https://api.semanticscholar.org/CorpusID:17431500). Archived from the original (https://pdfs.semanticscholar.org/c74c/5e11ed05246c12165ce7e4b6222bd32d68dc.pdf) (PDF) on 2020-02-27. `{{cite journal}}`: Cite journal requires `|journal=` (help)
7. Unpublished memo (https://www.ias.edu/sites/default/files/sns/files/1-matrixlog_tex(1).pdf) by S Adler (IAS)
8. Fang, Kai-Tai; Zhang, Yao-Ting (1990). *Generalized multivariate analysis*. Science Press (Beijing) and Springer-Verlag (Berlin). ISBN 3540176519. 9783540176510.

9. Pan, Jianxin; Fang, Kaitai (2007). *Growth curve models and statistical diagnostics*. Beijing: Science Press. ISBN 9780387950532.

10. Kollo, Tõnu; von Rosen, Dietrich (2005). *Advanced multivariate statistics with matrices*. Dordrecht: Springer. ISBN 978-1-4020-3418-3.

11. Magnus, Jan; Neudecker, Heinz (2019). *Matrix differential calculus with applications in statistics and econometrics*. New York: John Wiley. ISBN 9781119541202.

12. Liu, Shuangzhe; Leiva, Victor; Zhuang, Dan; Ma, Tiefeng; Figueroa-Zúñiga, Jorge I. (2022). "Matrix differential calculus with applications in the multivariate linear model and its diagnostics" (https://doi.org/10.1016%2Fj.jmva.2021.104849). *Journal of Multivariate Analysis*. **188**: 104849. doi:10.1016/j.jmva.2021.104849 (https://doi.org/10.1016%2Fj.jmva.2021.104849).

13. Liu, Shuangzhe; Trenkler, Götz; Kollo, Tõnu; von Rosen, Dietrich; Baksalary, Oskar Maria (2023). "Professor Heinz Neudecker and matrix differential calculus". *Statistical Papers*. doi:10.1007/s00362-023-01499-w (https://doi.org/10.1007%2Fs00362-023-01499-w). S2CID 263661094 (https://api.semanticscholar.org/CorpusID:263661094).

Further reading

- Abadir, Karim M.; Magnus, Jan R. (2005). *Matrix algebra*. Econometric Exercises. Cambridge: Cambridge University Press. ISBN 978-0-511-64796-3. OCLC 569411497 (https://www.worldcat.org/oclc/569411497).
- Lax, Peter D. (2007). "9. Calculus of Vector- and Matrix-Valued Functions". *Linear algebra and its applications* (2nd ed.). Hoboken, N.J.: Wiley-Interscience. ISBN 978-0-471-75156-4.
- Magnus, Jan R. (October 2010). "On the concept of matrix derivative". *Journal of Multivariate Analysis*. **101** (9): 2200–2206. doi:10.1016/j.jmva.2010.05.005 (https://doi.org/10.1016%2Fj.jmva.2010.05.005).. Note that this Wikipedia article has been nearly completely revised from the version criticized in this article.

External links

Software

- MatrixCalculus.org (http://www.matrixcalculus.org/), a website for evaluating matrix calculus expressions symbolically
- NCAlgebra (https://math.ucsd.edu/~ncalg/), an open-source Mathematica package that has some matrix calculus functionality
- SymPy supports symbolic matrix derivatives in its matrix expression module (https://docs.sympy.org/latest/modules/matrices/expressions.html), as well as symbolic tensor derivatives in its array expression module (https://docs.sympy.org/latest/modules/tensor/array_expressions.html).

Information

- Matrix Reference Manual (https://web.archive.org/web/20120630192238/http://www.psi.toronto.edu/matrix/calculus.html), Mike Brookes, Imperial College London.
- Matrix Differentiation (and some other stuff) (http://www.atmos.washington.edu/~dennis/MatrixCalculus.pdf), Randal J. Barnes, Department of Civil Engineering, University of Minnesota.
- Notes on Matrix Calculus (http://www4.ncsu.edu/~pfackler/MatCalc.pdf), Paul L. Fackler, North Carolina State University.
- Matrix Differential Calculus (https://wiki.inf.ed.ac.uk/twiki/pub/CSTR/ListenSemester1_2006_7/slide.pdf) Archived (https://web.archive.org/web/20120916044332/https://wiki.inf.ed.ac.uk/twiki/pub/CSTR/ListenSemester1_2006_7/slide.pdf) 2012-09-16 at the Wayback Machine (slide presentation), Zhang Le, University of Edinburgh.
- Introduction to Vector and Matrix Differentiation (https://web.archive.org/web/20120526142207/http://www.econ.ku.dk/metrics/Econometrics2_05_II/LectureNotes/matrixdiff.pdf) (notes on matrix differentiation, in the context of Econometrics), Heino Bohn Nielsen.
- A note on differentiating matrices (http://mpra.ub.uni-muenchen.de/1239/1/MPRA_paper_1239.pdf) (notes on matrix differentiation), Pawel Koval, from Munich Personal RePEc Archive.

- Vector/Matrix Calculus (http://www.personal.rdg.ac.uk/~sis01xh/teaching/CY4C9/ANN3.pdf) More notes on matrix differentiation.
- Matrix Identities (http://www.cs.nyu.edu/~roweis/notes/matrixid.pdf) (notes on matrix differentiation), Sam Roweis.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Matrix_calculus&oldid=1217202866"

-