C

Linear Algebra: Matrix Calculus

TABLE OF CONTENTS

PDF-XChange Editor Click to BUY NOW! www.tracker-software.com

§C.1. Introduction

Matrix Calculus is the extension of ordinary calculus to matrices and vectors whose entries are functions of one or more independent variables. This Appendix collect formulas of matrix calculus that often appear in finite element derivations.

§C.2. The Derivatives of Vector Functions

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors of orders n and m respectively:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \tag{C.1}$$

where each component y_i may be a function of all the x_j, a fact represented by saying that $\mathbf{y}$ is a function of $\mathbf{x}$, or

$$\mathbf{y} = \mathbf{y}(\mathbf{x}). \tag{C.2}$$

If $n = 1$, $\mathbf{x}$ reduces to a scalar, which we call x. If $m = 1$, $\mathbf{y}$ reduces to a scalar, which we call y. Various applications are studied in the following subsections.

§C.2.1. Derivative of Vector with Respect to Vector

The derivative of the vector $\mathbf{y}$ with respect to vector $\mathbf{x}$ is the $n \times m$ matrix

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{C.3}$$

§C.2.2. Derivative of a Scalar with Respect to Vector

If y is a scalar,

$$\frac{\partial y}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}. \tag{C.4}$$

§C.2.3. Derivative of Vector with Respect to Scalar

If x is a scalar,

$$\frac{\partial \mathbf{y}}{\partial x} \stackrel{\text{def}}{=} \left[\frac{\partial y_1}{\partial x} \quad \frac{\partial y_2}{\partial x} \quad \ldots \quad \frac{\partial y_m}{\partial x} \right] \tag{C.5}$$

Remark C.1. Many authors, notably in statistics and economics, define the derivatives as the transposes of those given above.[1] This has the advantage of better agreement of matrix products with composition schemes such as the chain rule. Evidently the notation is not yet stable.

Example C.1. Given

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{C.6}$$

and

$$\begin{aligned} y_1 &= x_1^2 - x_2 \\ y_2 &= x_3^2 + 3x_2 \end{aligned} \tag{C.7}$$

the partial derivative matrix $\partial \mathbf{y}/\partial \mathbf{x}$ is computed as follows:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} \\ \frac{\partial y_1}{\partial x_3} & \frac{\partial y_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_1 & 0 \\ -1 & 3 \\ 0 & 2x_3 \end{bmatrix} \tag{C.8}$$

§C.2.4. Jacobian of a Variable Transformation

In multivariate analysis, if $\mathbf{x}$ and $\mathbf{y}$ are of the same order, the determinant of the square matrix $\partial \mathbf{x}/\partial \mathbf{y}$, that is

$$J = \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| \tag{C.9}$$

is called the *Jacobian* of the transformation determined by $\mathbf{y} = \mathbf{y}(\mathbf{x})$. The inverse determinant is

$$J^{-1} = \left| \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right|. \tag{C.10}$$

Example C.2. The transformation from spherical to Cartesian coordinates is defined by

$$x = r \sin\theta \cos\psi, \qquad y = r \sin\theta \sin\psi, \qquad z = r\cos\theta \tag{C.11}$$

where $r > 0$, $0 < \theta < \pi$ and $0 \le \psi < 2\pi$. To obtain the Jacobian of the transformation, let

$$\begin{aligned} x &\equiv x_1, & y &\equiv x_2, & z &\equiv x_3 \\ r &\equiv y_1, & \theta &\equiv y_2, & \psi &\equiv y_3 \end{aligned} \tag{C.12}$$

Then

$$\begin{aligned} J = \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| &= \begin{vmatrix} \sin y_2 \cos y_3 & \sin y_2 \sin y_3 & \cos y_2 \\ y_1 \cos y_2 \cos y_3 & y_1 \cos y_2 \sin y_3 & -y_1 \sin y_2 \\ -y_1 \sin y_2 \sin y_3 & y_1 \sin y_2 \cos y_3 & 0 \end{vmatrix} \\ &= y_1^2 \sin y_2 = r^2 \sin\theta. \end{aligned} \tag{C.13}$$

The foregoing definitions can be used to obtain derivatives to many frequently used expressions, including quadratic and bilinear forms.

[1] One author puts it this way: “When one does matrix calculus, one quickly finds that there are two kinds of people in this world: those who think the gradient is a row vector, and those who think it is a column vector.”

PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

Example C.3. Consider the quadratic form

$$y = \mathbf{x}^T \mathbf{A} \mathbf{x} \tag{C.14}$$

where $\mathbf{A}$ is a square matrix of order n. Using the definition (D.3) one obtains

$$\frac{\partial y}{\partial \mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{A}^T \mathbf{x} \tag{C.15}$$

and if $\mathbf{A}$ is symmetric,

$$\frac{\partial y}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}. \tag{C.16}$$

We can of course continue the differentiation process:

$$\frac{\partial^2 y}{\partial \mathbf{x}^2} = \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial y}{\partial \mathbf{x}} \right) = \mathbf{A} + \mathbf{A}^T, \tag{C.17}$$

and if $\mathbf{A}$ is symmetric,

$$\frac{\partial^2 y}{\partial \mathbf{x}^2} = 2\mathbf{A}. \tag{C.18}$$

The following table collects several useful vector derivative formulas.

$\mathbf{y}$	$\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{A}\mathbf{x}$	$\mathbf{A}^T$
$\mathbf{x}^T \mathbf{A}$	$\mathbf{A}$
$\mathbf{x}^T \mathbf{x}$	$2\mathbf{x}$
$\mathbf{x}^T \mathbf{A}\mathbf{x}$	$\mathbf{A}\mathbf{x} + \mathbf{A}^T \mathbf{x}$

§C.3. The Chain Rule for Vector Functions

Let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_r \end{bmatrix} \quad \text{and} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix} \tag{C.19}$$

where $\mathbf{z}$ is a function of $\mathbf{y}$, which is in turn a function of $\mathbf{x}$. Using the definition (D.2), we can write

$$\left(\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right)^T = \begin{bmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_1}{\partial x_2} & \ldots & \frac{\partial z_1}{\partial x_n} \\ \frac{\partial z_2}{\partial x_1} & \frac{\partial z_2}{\partial x_2} & \ldots & \frac{\partial z_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial z_m}{\partial x_1} & \frac{\partial z_m}{\partial x_2} & \ldots & \frac{\partial z_m}{\partial x_n} \end{bmatrix} \tag{C.20}$$

Each entry of this matrix may be expanded as

$$\frac{\partial z_i}{\partial x_j} = \sum_{q=1}^{r} \frac{\partial z_i}{\partial y_q} \frac{\partial y_q}{\partial x_j} \qquad \begin{cases} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n. \end{cases} \tag{C.21}$$

PDF-XChange Editor Click to BUY NOW! www.tracker-software.com

Then

$$\left(\frac{\partial \mathbf{z}}{\partial \mathbf{x}}\right)^T = \begin{bmatrix} \sum \frac{\partial z_1}{y_q}\frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_1}{\partial y_q}\frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_2}{\partial y_q}\frac{\partial y_q}{\partial x_n} \\ \sum \frac{\partial z_2}{\partial y_q}\frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_2}{\partial y_q}\frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_2}{\partial y_q}\frac{\partial y_q}{\partial x_n} \\ \vdots & & & \\ \sum \frac{\partial z_m}{\partial y_q}\frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_m}{\partial y_q}\frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_m}{\partial y_q}\frac{\partial y_q}{\partial x_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_r} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_2}{\partial y_r} \\ \vdots & & & \\ \frac{\partial z_m}{\partial y_1} & \frac{\partial z_m}{\partial y_2} & \cdots & \frac{\partial z_m}{\partial y_r} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial y_r}{\partial x_1} & \frac{\partial y_r}{\partial x_2} & \cdots & \frac{\partial y_r}{\partial x_n} \end{bmatrix}$$

$$= \left(\frac{\partial \mathbf{z}}{\partial \mathbf{y}}\right)^T \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right)^T = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial \mathbf{z}}{\partial \mathbf{y}}\right)^T. \tag{C.22}$$

On transposing both sides, we finally obtain

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial \mathbf{z}}{\partial \mathbf{y}}, \tag{C.23}$$

which is the *chain rule* for vectors. If all vectors reduce to scalars,

$$\frac{\partial z}{\partial x} = \frac{\partial y}{\partial x}\frac{\partial z}{\partial y} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}, \tag{C.24}$$

which is the conventional chain rule of calculus. Note, however, that when we are dealing with vectors, the chain of matrices builds "toward the left." For example, if $\mathbf{w}$ is a function of $\mathbf{z}$, which is a function of $\mathbf{y}$, which is a function of $\mathbf{x}$,

$$\frac{\partial \mathbf{w}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial \mathbf{z}}{\partial \mathbf{y}}\frac{\partial \mathbf{w}}{\partial \mathbf{z}}. \tag{C.25}$$

On the other hand, in the ordinary chain rule one can indistictly build the product to the right or to the left because scalar multiplication is commutative.

§C.4. The Derivative of Scalar Functions of a Matrix

Let $\mathbf{X} = (x_{ij})$ be a matrix of order $(m \times n)$ and let

$$y = f(\mathbf{X}), \tag{C.26}$$

be a scalar function of $\mathbf{X}$. The derivative of y with respect to $\mathbf{X}$, denoted by

$$\frac{\partial y}{\partial \mathbf{X}}, \tag{C.27}$$

PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

is defined as the following matrix of order $(m \times n)$:

$$\mathbf{G} = \frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} = \left[\frac{\partial y}{\partial x_{ij}} \right] = \sum_{i,j} \mathbf{E}_{ij} \frac{\partial y}{\partial x_{ij}}, \tag{C.28}$$

where $\mathbf{E}_{ij}$ denotes the elementary matrix* of order $(m \times n)$. This matrix $\mathbf{G}$ is also known as a *gradient matrix*.

Example C.4. Find the gradient matrix if y is the trace of a square matrix $\mathbf{X}$ of order n, that is

$$y = \mathrm{tr}(\mathbf{X}) = \sum_{i=1}^{n} x_{ii}. \tag{C.29}$$

Obviously all non-diagonal partials vanish whereas the diagonal partials equal one, thus

$$\mathbf{G} = \frac{\partial y}{\partial \mathbf{X}} = \mathbf{I}, \tag{C.30}$$

where $\mathbf{I}$ denotes the identity matrix of order n.

§C.4.1. Functions of a Matrix Determinant

An important family of derivatives with respect to a matrix involves functions of the determinant of a matrix, for example $y = |\mathbf{X}|$ or $y = |\mathbf{AX}|$. Suppose that we have a matrix $\mathbf{Y} = [y_{ij}]$ whose components are functions of a matrix $\mathbf{X} = [x_{rs}]$, that is $y_{ij} = f_{ij}(x_{rs})$, and set out to build the matrix

$$\frac{\partial |\mathbf{Y}|}{\partial \mathbf{X}}. \tag{C.31}$$

Using the chain rule we can write

$$\frac{\partial |\mathbf{Y}|}{\partial x_{rs}} = \sum_i \sum_j \mathbf{Y}_{ij} \frac{\partial |\mathbf{Y}|}{\partial y_{ij}} \frac{\partial y_{ij}}{\partial x_{rs}}. \tag{C.32}$$

But

$$|\mathbf{Y}| = \sum_j y_{ij} \mathbf{Y}_{ij}, \tag{C.33}$$

where $\mathbf{Y}_{ij}$ is the *cofactor* of the element y_{ij} in $|\mathbf{Y}|$. Since the cofactors $\mathbf{Y}_{i1}, \mathbf{Y}_{i2}, \ldots$ are independent of the element y_{ij}, we have

$$\frac{\partial |\mathbf{Y}|}{\partial y_{ij}} = \mathbf{Y}_{ij}. \tag{C.34}$$

It follows that

$$\frac{\partial |\mathbf{Y}|}{\partial x_{rs}} = \sum_i \sum_j \mathbf{Y}_{ij} \frac{\partial y_{ij}}{\partial x_{rs}}. \tag{C.35}$$

---

* The elementary matrix $\mathbf{E}_{ij}$ of order $m \times n$ has all zero entries except for the (i, j) entry, which is one.

PDF-XChange Editor Click to BUY NOW! www.tracker-software.com

There is an alternative form of this result which is ocassionally useful. Define

$$a_{ij} = \mathbf{Y}_{ij}, \quad \mathbf{A} = [a_{ij}], \qquad b_{ij} = \frac{\partial y_{ij}}{\partial x_{rs}}, \quad \mathbf{B} = [b_{ij}]. \tag{C.36}$$

Then it can be shown that

$$\frac{\partial |\mathbf{Y}|}{\partial x_{rs}} = \mathrm{tr}(\mathbf{A}\mathbf{B}^T) = \mathrm{tr}(\mathbf{B}^T \mathbf{A}). \tag{C.37}$$

Example C.5. If $\mathbf{X}$ is a nonsingular square matrix and $\mathbf{Z} = |\mathbf{X}|\mathbf{X}^{-1}$ its cofactor matrix,

$$\mathbf{G} = \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \mathbf{Z}^T. \tag{C.38}$$

If $\mathbf{X}$ is also symmetric,

$$\mathbf{G} = \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = 2\mathbf{Z}^T - \mathrm{diag}(\mathbf{Z}^T). \tag{C.39}$$

§C.5. The Matrix Differential

For a scalar function $f(\mathbf{x})$, where $\mathbf{x}$ is an n-vector, the ordinary differential of multivariate calculus is defined as

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i. \tag{C.40}$$

In harmony with this formula, we define the differential of an $m \times n$ matrix $\mathbf{X} = [x_{ij}]$ to be

$$d\mathbf{X} \stackrel{\text{def}}{=} \begin{bmatrix} dx_{11} & dx_{12} & \ldots & dx_{1n} \\ dx_{21} & dx_{22} & \ldots & dx_{2n} \\ \vdots & \vdots & & \vdots \\ dx_{m1} & dx_{m2} & \ldots & dx_{mn} \end{bmatrix}. \tag{C.41}$$

This definition complies with the multiplicative and associative rules

$$d(\alpha\mathbf{X}) = \alpha\, d\mathbf{X}, \qquad d(\mathbf{X} + \mathbf{Y}) = d\mathbf{X} + d\mathbf{Y}. \tag{C.42}$$

If $\mathbf{X}$ and $\mathbf{Y}$ are product-conforming matrices, it can be verified that the differential of their product is

$$d(\mathbf{X}\mathbf{Y}) = (d\mathbf{X})\mathbf{Y} + \mathbf{X}(d\mathbf{Y}). \tag{C.43}$$

which is an extension of the well known rule $d(xy) = y\,dx + x\,dy$ for scalar functions.

Example C.6. If $\mathbf{X} = [x_{ij}]$ is a square nonsingular matrix of order n, and denote $\mathbf{Z} = |\mathbf{X}|\mathbf{X}^{-1}$. Find the differential of the determinant of $\mathbf{X}$:

$$d|\mathbf{X}| = \sum_{i,j} \frac{\partial |\mathbf{X}|}{\partial x_{ij}} dx_{ij} = \sum_{i,j} \mathbf{X}_{ij}\, dx_{ij} = \mathrm{tr}(|\mathbf{X}|\mathbf{X}^{-1})^T\, d\mathbf{X}) = \mathrm{tr}(\mathbf{Z}^T\, d\mathbf{X}), \tag{C.44}$$

where X_{ij} denotes the cofactor of x_{ij} in $\mathbf{X}$.

PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

Example C.7. With the same assumptions as above, find $d(\mathbf{X}^{-1})$. The quickest derivation follows by differentiating both sides of the identity $\mathbf{X}^{-1}\mathbf{X} = \mathbf{I}$:

$$d(\mathbf{X}^{-1})\mathbf{X} + \mathbf{X}^{-1}\,d\mathbf{X} = \mathbf{0}, \tag{C.45}$$

from which

$$d(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}\,d\mathbf{X}\,\mathbf{X}^{-1}. \tag{C.46}$$

If $\mathbf{X}$ reduces to the scalar x we have

$$d\left(\frac{1}{x}\right) = -\frac{dx}{x^2}. \tag{C.47}$$

2 Derivatives

This section is covering differentiation of a number of expressions with respect to a matrix $\mathbf{X}$. Note that it is always assumed that $\mathbf{X}$ has *no special structure*, i.e. that the elements of $\mathbf{X}$ are independent (e.g. not symmetric, Toeplitz, positive definite). See section 2.8 for differentiation of structured matrices. The basic assumptions can be written in a formula as

$$\frac{\partial X_{kl}}{\partial X_{ij}} = \delta_{ik}\delta_{lj} \tag{32}$$

that is for e.g. vector forms,

$$\left[\frac{\partial \mathbf{x}}{\partial y}\right]_i = \frac{\partial x_i}{\partial y} \qquad \left[\frac{\partial x}{\partial \mathbf{y}}\right]_i = \frac{\partial x}{\partial y_i} \qquad \left[\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right]_{ij} = \frac{\partial x_i}{\partial y_j}$$

The following rules are general and very useful when deriving the differential of an expression ([19]):

$$\begin{aligned}
\partial \mathbf{A} &= 0 \qquad (\mathbf{A} \text{ is a constant}) && (33)\\
\partial(\alpha \mathbf{X}) &= \alpha \partial \mathbf{X} && (34)\\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} && (35)\\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial \mathbf{X}) && (36)\\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) && (37)\\
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) && (38)\\
\partial(\mathbf{X} \otimes \mathbf{Y}) &= (\partial \mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (\partial \mathbf{Y}) && (39)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial \mathbf{X})\mathbf{X}^{-1} && (40)\\
\partial(\det(\mathbf{X})) &= \mathrm{Tr}(\mathrm{adj}(\mathbf{X})\partial \mathbf{X}) && (41)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) && (42)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) && (43)\\
\partial \mathbf{X}^T &= (\partial \mathbf{X})^T && (44)\\
\partial \mathbf{X}^H &= (\partial \mathbf{X})^H && (45)
\end{aligned}$$

2.1 Derivatives of a Determinant

2.1.1 General form

$$\frac{\partial \det(\mathbf{Y})}{\partial x} = \det(\mathbf{Y})\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right] \tag{46}$$

$$\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} = \delta_{ij}\det(\mathbf{X}) \tag{47}$$

$$\begin{aligned}
\frac{\partial^2 \det(\mathbf{Y})}{\partial x^2} = \det(\mathbf{Y}) \Bigg[&\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \frac{\partial \mathbf{Y}}{\partial x}}{\partial x}\right] \\
&+\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right]\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right] \\
&-\mathrm{Tr}\left[\left(\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right)\left(\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right)\right]\Bigg]
\end{aligned} \tag{48}$$

2.1.2 Linear forms

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(\mathbf{X}^{-1})^T \tag{49}$$

$$\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} = \delta_{ij} \det(\mathbf{X}) \tag{50}$$

$$\frac{\partial \det(\mathbf{AXB})}{\partial \mathbf{X}} = \det(\mathbf{AXB})(\mathbf{X}^{-1})^T = \det(\mathbf{AXB})(\mathbf{X}^T)^{-1} \tag{51}$$

2.1.3 Square forms

If $\mathbf{X}$ is square and invertible, then

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{X}^{-T} \tag{52}$$

If $\mathbf{X}$ is not square but $\mathbf{A}$ is symmetric, then

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^T\mathbf{AX})\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} \tag{53}$$

If $\mathbf{X}$ is not square and $\mathbf{A}$ is not symmetric, then

$$\frac{\partial \det(\mathbf{X}^T\mathbf{AX})}{\partial \mathbf{X}} = \det(\mathbf{X}^T\mathbf{AX})(\mathbf{AX}(\mathbf{X}^T\mathbf{AX})^{-1} + \mathbf{A}^T\mathbf{X}(\mathbf{X}^T\mathbf{A}^T\mathbf{X})^{-1}) \tag{54}$$

2.1.4 Other nonlinear forms

Some special cases are (See [9, 7])

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})|}{\partial \mathbf{X}} = 2(\mathbf{X}^+)^T \tag{55}$$

$$\frac{\partial \ln \det(\mathbf{X}^T\mathbf{X})}{\partial \mathbf{X}^+} = -2\mathbf{X}^T \tag{56}$$

$$\frac{\partial \ln |\det(\mathbf{X})|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^T = (\mathbf{X}^T)^{-1} \tag{57}$$

$$\frac{\partial \det(\mathbf{X}^k)}{\partial \mathbf{X}} = k\det(\mathbf{X}^k)\mathbf{X}^{-T} \tag{58}$$

2.2 Derivatives of an Inverse

From [27] we have the basic identity

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x} = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\mathbf{Y}^{-1} \tag{59}$$

from which it follows

$$\frac{\partial(\mathbf{X}^{-1})_{kl}}{\partial X_{ij}} = -(\mathbf{X}^{-1})_{ki}(\mathbf{X}^{-1})_{jl} \tag{60}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial \mathbf{X}} = -\mathbf{X}^{-T}\mathbf{a}\mathbf{b}^T\mathbf{X}^{-T} \tag{61}$$

$$\frac{\partial \det(\mathbf{X}^{-1})}{\partial \mathbf{X}} = -\det(\mathbf{X}^{-1})(\mathbf{X}^{-1})^T \tag{62}$$

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T \tag{63}$$

$$\frac{\partial \mathrm{Tr}((\mathbf{X}+\mathbf{A})^{-1})}{\partial \mathbf{X}} = -((\mathbf{X}+\mathbf{A})^{-1}(\mathbf{X}+\mathbf{A})^{-1})^T \tag{64}$$

From [32] we have the following result: Let $\mathbf{A}$ be an $n \times n$ invertible square matrix, $\mathbf{W}$ be the inverse of $\mathbf{A}$, and $J(\mathbf{A})$ is an $n \times n$ -variate and differentiable function with respect to $\mathbf{A}$, then the partial differentials of J with respect to $\mathbf{A}$ and $\mathbf{W}$ satisfy

$$\frac{\partial J}{\partial \mathbf{A}} = -\mathbf{A}^{-T}\frac{\partial J}{\partial \mathbf{W}}\mathbf{A}^{-T}$$

2.3 Derivatives of Eigenvalues

$$\frac{\partial}{\partial \mathbf{X}}\sum \mathrm{eig}(\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} \tag{65}$$

$$\frac{\partial}{\partial \mathbf{X}}\prod \mathrm{eig}(\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\det(\mathbf{X}) = \det(\mathbf{X})\mathbf{X}^{-T} \tag{66}$$

If $\mathbf{A}$ is real and symmetric, λ_i and $\mathbf{v}_i$ are distinct eigenvalues and eigenvectors of $\mathbf{A}$ (see (276)) with $\mathbf{v}_i^T\mathbf{v}_i = 1$, then [33]

$$\partial\lambda_i = \mathbf{v}_i^T\partial(\mathbf{A})\mathbf{v}_i \tag{67}$$

$$\partial\mathbf{v}_i = (\lambda_i\mathbf{I} - \mathbf{A})^+\partial(\mathbf{A})\mathbf{v}_i \tag{68}$$

2.4 Derivatives of Matrices, Vectors and Scalar Forms

2.4.1 First Order

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \tag{69}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{70}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T \tag{71}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T \tag{72}$$

$$\frac{\partial \mathbf{X}}{\partial X_{ij}} = \mathbf{J}^{ij} \tag{73}$$

$$\frac{\partial(\mathbf{X}\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{im}(\mathbf{A})_{nj} = (\mathbf{J}^{mn}\mathbf{A})_{ij} \tag{74}$$

$$\frac{\partial(\mathbf{X}^T\mathbf{A})_{ij}}{\partial X_{mn}} = \delta_{in}(\mathbf{A})_{mj} = (\mathbf{J}^{nm}\mathbf{A})_{ij} \tag{75}$$

2.4.2 Second Order

$$\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl}X_{mn} = 2\sum_{kl} X_{kl} \tag{76}$$

$$\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{b}\mathbf{c}^T + \mathbf{c}\mathbf{b}^T) \tag{77}$$

$$\frac{\partial (\mathbf{B}\mathbf{x}+\mathbf{b})^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d}) + \mathbf{D}^T\mathbf{C}^T(\mathbf{B}\mathbf{x}+\mathbf{b}) \tag{78}$$

$$\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})_{kl}}{\partial X_{ij}} = \delta_{lj}(\mathbf{X}^T\mathbf{B})_{ki} + \delta_{kj}(\mathbf{B}\mathbf{X})_{il} \tag{79}$$

$$\frac{\partial (\mathbf{X}^T\mathbf{B}\mathbf{X})}{\partial X_{ij}} = \mathbf{X}^T\mathbf{B}\mathbf{J}^{ij} + \mathbf{J}^{ji}\mathbf{B}\mathbf{X} \qquad (\mathbf{J}^{ij})_{kl} = \delta_{ik}\delta_{jl} \tag{80}$$

See Sec 9.7 for useful properties of the Single-entry matrix $\mathbf{J}^{ij}$

$$\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B}+\mathbf{B}^T)\mathbf{x} \tag{81}$$

$$\frac{\partial \mathbf{b}^T\mathbf{X}^T\mathbf{D}\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} = \mathbf{D}^T\mathbf{X}\mathbf{b}\mathbf{c}^T + \mathbf{D}\mathbf{X}\mathbf{c}\mathbf{b}^T \tag{82}$$

$$\frac{\partial}{\partial \mathbf{X}}(\mathbf{X}\mathbf{b}+\mathbf{c})^T\mathbf{D}(\mathbf{X}\mathbf{b}+\mathbf{c}) = (\mathbf{D}+\mathbf{D}^T)(\mathbf{X}\mathbf{b}+\mathbf{c})\mathbf{b}^T \tag{83}$$

Assume $\mathbf{W}$ is symmetric, then

$$\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) = -2\mathbf{A}^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) \tag{84}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s}) = 2\mathbf{W}(\mathbf{x}-\mathbf{s}) \tag{85}$$

$$\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s}) = -2\mathbf{W}(\mathbf{x}-\mathbf{s}) \tag{86}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) = 2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) \tag{87}$$

$$\frac{\partial}{\partial \mathbf{A}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) = -2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})\mathbf{s}^T \tag{88}$$

As a case with complex values the following holds

$$\frac{\partial (a-\mathbf{x}^H\mathbf{b})^2}{\partial \mathbf{x}} = -2\mathbf{b}(a-\mathbf{x}^H\mathbf{b})^* \tag{89}$$

This formula is also known from the LMS algorithm [14]

2.4.3 Higher-order and non-linear

$$\frac{\partial (\mathbf{X}^n)_{kl}}{\partial X_{ij}} = \sum_{r=0}^{n-1}(\mathbf{X}^r\mathbf{J}^{ij}\mathbf{X}^{n-1-r})_{kl} \tag{90}$$

For proof of the above, see B.1.3.

$$\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^T\mathbf{X}^n\mathbf{b} = \sum_{r=0}^{n-1}(\mathbf{X}^r)^T\mathbf{a}\mathbf{b}^T(\mathbf{X}^{n-1-r})^T \tag{91}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathbf{a}^T(\mathbf{X}^n)^T\mathbf{X}^n\mathbf{b} = \sum_{r=0}^{n-1}\Big[\mathbf{X}^{n-1-r}\mathbf{a}\mathbf{b}^T(\mathbf{X}^n)^T\mathbf{X}^r + (\mathbf{X}^r)^T\mathbf{X}^n\mathbf{a}\mathbf{b}^T(\mathbf{X}^{n-1-r})^T\Big] \tag{92}$$

See B.1.3 for a proof.

Assume $\mathbf{s}$ and $\mathbf{r}$ are functions of $\mathbf{x}$, i.e. $\mathbf{s} = \mathbf{s}(\mathbf{x}), \mathbf{r} = \mathbf{r}(\mathbf{x})$, and that $\mathbf{A}$ is a constant, then

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{s}^T\mathbf{A}\mathbf{r} = \left[\frac{\partial \mathbf{s}}{\partial \mathbf{x}}\right]^T\mathbf{A}\mathbf{r} + \left[\frac{\partial \mathbf{r}}{\partial \mathbf{x}}\right]^T\mathbf{A}^T\mathbf{s} \tag{93}$$

$$\frac{\partial}{\partial \mathbf{x}}\frac{(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x})}{(\mathbf{B}\mathbf{x})^T(\mathbf{B}\mathbf{x})} = \frac{\partial}{\partial \mathbf{x}}\frac{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x}} \tag{94}$$

$$= 2\frac{\mathbf{A}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{B}\mathbf{x}} - 2\frac{\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}\mathbf{B}^T\mathbf{B}\mathbf{x}}{(\mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x})^2} \tag{95}$$

2.4.4 Gradient and Hessian

Using the above we have for the gradient and the Hessian

$$f = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} \tag{96}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} + \mathbf{b} \tag{97}$$

$$\frac{\partial^2 f}{\partial \mathbf{x}\partial \mathbf{x}^T} = \mathbf{A} + \mathbf{A}^T \tag{98}$$

2.5 Derivatives of Traces

Assume $F(\mathbf{X})$ to be a differentiable function of each of the elements of X. It then holds that

$$\frac{\partial \mathrm{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$$

where $f(\cdot)$ is the scalar derivative of $F(\cdot)$.

2.5.1 First Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} \tag{99}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\mathbf{A}) = \mathbf{A}^T \tag{100}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) = \mathbf{A}^T\mathbf{B}^T \tag{101}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}) = \mathbf{B}\mathbf{A} \tag{102}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{A}) = \mathbf{A} \tag{103}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\mathbf{X}^T) = \mathbf{A} \tag{104}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{A}\otimes\mathbf{X}) = \mathrm{Tr}(\mathbf{A})\mathbf{I} \tag{105}$$

PDF-XChange Editor
Click to BUY NOW!
www.tracker-software.com

2.5.2 Second Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2) = 2\mathbf{X}^T \tag{106}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^2\mathbf{B}) = (\mathbf{XB}+\mathbf{BX})^T \tag{107}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{BX}) = \mathbf{BX}+\mathbf{B}^T\mathbf{X} \tag{108}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BXX}^T) = \mathbf{BX}+\mathbf{B}^T\mathbf{X} \tag{109}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T\mathbf{B}) = \mathbf{BX}+\mathbf{B}^T\mathbf{X} \tag{110}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XBX}^T) = \mathbf{XB}^T+\mathbf{XB} \tag{111}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{BX}^T\mathbf{X}) = \mathbf{XB}^T+\mathbf{XB} \tag{112}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{XB}) = \mathbf{XB}^T+\mathbf{XB} \tag{113}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}) = \mathbf{A}^T\mathbf{X}^T\mathbf{B}^T+\mathbf{B}^T\mathbf{X}^T\mathbf{A}^T \tag{114}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^T\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{XX}^T) = 2\mathbf{X} \tag{115}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{B}^T\mathbf{X}^T\mathbf{CXB}) = \mathbf{C}^T\mathbf{XBB}^T+\mathbf{CXBB}^T \tag{116}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{X}^T\mathbf{BXC}\right] = \mathbf{BXC}+\mathbf{B}^T\mathbf{XC}^T \tag{117}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXBX}^T\mathbf{C}) = \mathbf{A}^T\mathbf{C}^T\mathbf{XB}^T+\mathbf{CAXB} \tag{118}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\Big[(\mathbf{AXB}+\mathbf{C})(\mathbf{AXB}+\mathbf{C})^T\Big] = 2\mathbf{A}^T(\mathbf{AXB}+\mathbf{C})\mathbf{B}^T \tag{119}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}\otimes\mathbf{X}) = \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X})\mathrm{Tr}(\mathbf{X}) = 2\mathrm{Tr}(\mathbf{X})\mathbf{I} \tag{120}$$

See [7].

2.5.3 Higher Order

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{X}^k) = k(\mathbf{X}^{k-1})^T \tag{121}$$

$$\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^k) = \sum_{r=0}^{k-1}(\mathbf{X}^r\mathbf{AX}^{k-r-1})^T \tag{122}$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}\left[\mathbf{B}^T\mathbf{X}^T\mathbf{CXX}^T\mathbf{CXB}\right] = {} & \mathbf{CXX}^T\mathbf{CXBB}^T \\ & +\mathbf{C}^T\mathbf{XBB}^T\mathbf{X}^T\mathbf{C}^T\mathbf{X} \\ & +\mathbf{CXBB}^T\mathbf{X}^T\mathbf{CX} \\ & +\mathbf{C}^T\mathbf{XX}^T\mathbf{C}^T\mathbf{XBB}^T\end{aligned} \tag{123}$$

PDF-XChange Editor Click to BUY NOW! www.tracker-software.com

2.5.4 Other

$$\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}) = -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^T = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T} \tag{124}$$

Assume $\mathbf{B}$ and $\mathbf{C}$ to be symmetric, then

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{A}\Big] &= -(\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1})(\mathbf{A}+\mathbf{A}^T)(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} && (125)\\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\Big[(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] &= -2\mathbf{C}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \\
&\quad +2\mathbf{B}\mathbf{X}(\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} && (126)\\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}\Big[(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{B}\mathbf{X})\Big] &= -2\mathbf{C}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} \\
&\quad +2\mathbf{B}\mathbf{X}(\mathbf{A}+\mathbf{X}^T\mathbf{C}\mathbf{X})^{-1} && (127)
\end{aligned}$$

See [7].

$$\frac{\partial \mathrm{Tr}(\sin(\mathbf{X}))}{\partial \mathbf{X}} = \cos(\mathbf{X})^T \tag{128}$$

2.6 Derivatives of vector norms

2.6.1 Two-norm

$$\frac{\partial}{\partial \mathbf{x}} ||\mathbf{x}-\mathbf{a}||_2 = \frac{\mathbf{x}-\mathbf{a}}{||\mathbf{x}-\mathbf{a}||_2} \tag{129}$$

$$\frac{\partial}{\partial \mathbf{x}} \frac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|_2} = \frac{\mathbf{I}}{\|\mathbf{x}-\mathbf{a}\|_2} - \frac{(\mathbf{x}-\mathbf{a})(\mathbf{x}-\mathbf{a})^T}{\|\mathbf{x}-\mathbf{a}\|_2^3} \tag{130}$$

$$\frac{\partial ||\mathbf{x}||_2^2}{\partial \mathbf{x}} = \frac{\partial ||\mathbf{x}^T\mathbf{x}||_2}{\partial \mathbf{x}} = 2\mathbf{x} \tag{131}$$

2.7 Derivatives of matrix norms

For more on matrix norms, see Sec. 10.4.

2.7.1 Frobenius norm

$$\frac{\partial}{\partial \mathbf{X}} ||\mathbf{X}||_\mathrm{F}^2 = \frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = 2\mathbf{X} \tag{132}$$

See (248). Note that this is also a special case of the result in equation 119.

2.8 Derivatives of Structured Matrices

Assume that the matrix $\mathbf{A}$ has some structure, i.e. symmetric, toeplitz, etc. In that case the derivatives of the previous section does not apply in general. Instead, consider the following general rule for differentiating a scalar function $f(\mathbf{A})$

$$\frac{df}{dA_{ij}} = \sum_{kl} \frac{\partial f}{\partial A_{kl}} \frac{\partial A_{kl}}{\partial A_{ij}} = \mathrm{Tr}\left[\left[\frac{\partial f}{\partial \mathbf{A}}\right]^T \frac{\partial \mathbf{A}}{\partial A_{ij}}\right] \tag{133}$$

Click to BUY NOW! PDF-XChange Editor www.tracker-software.com

The matrix differentiated with respect to itself is in this document referred to as the *structure matrix* of $\mathbf{A}$ and is defined simply by

$$\frac{\partial \mathbf{A}}{\partial A_{ij}} = \mathbf{S}^{ij} \tag{134}$$

If $\mathbf{A}$ has no special structure we have simply $\mathbf{S}^{ij} = \mathbf{J}^{ij}$, that is, the structure matrix is simply the single-entry matrix. Many structures have a representation in singleentry matrices, see Sec. 9.7.6 for more examples of structure matrices.

2.8.1 The Chain Rule

Sometimes the objective is to find the derivative of a matrix which is a function of another matrix. Let $\mathbf{U} = f(\mathbf{X})$, the goal is to find the derivative of the function g($\mathbf{U}$) with respect to $\mathbf{X}$:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}} \tag{135}$$

Then the Chain Rule can then be written the following way:

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M}\sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}} \tag{136}$$

Using matrix notation, this can be written as:

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \mathrm{Tr}\Big[(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \frac{\partial \mathbf{U}}{\partial X_{ij}}\Big]. \tag{137}$$

2.8.2 Symmetric

If $\mathbf{A}$ is symmetric, then $\mathbf{S}^{ij} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij}\mathbf{J}^{ij}$ and therefore

$$\frac{df}{d\mathbf{A}} = \left[\frac{\partial f}{\partial \mathbf{A}}\right] + \left[\frac{\partial f}{\partial \mathbf{A}}\right]^T - \mathrm{diag}\left[\frac{\partial f}{\partial \mathbf{A}}\right] \tag{138}$$

That is, e.g., ([5]):

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A} \circ \mathbf{I}), \text{ see (142)} \tag{139}$$

$$\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \det(\mathbf{X})(2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})) \tag{140}$$

$$\frac{\partial \ln \det(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I}) \tag{141}$$

2.8.3 Diagonal

If $\mathbf{X}$ is diagonal, then ([19]):

$$\frac{\partial \mathrm{Tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A} \circ \mathbf{I} \tag{142}$$

PDF-XChange Editor Click to BUY NOW! www.tracker-software.com

2.8.4 Toeplitz

Like symmetric matrices and diagonal matrices also Toeplitz matrices has a special structure which should be taken into account when the derivative with respect to a matrix with Toeplitz structure.

$$
\begin{aligned}
& \frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} \\
= \; & \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} \\
= \; & \begin{bmatrix}
\mathrm{Tr}(\mathbf{A}) & \mathrm{Tr}([\mathbf{A}^T]_{n1}) & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) & \cdots & A_{n1} \\
\mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A}) & \ddots & \ddots & \vdots \\
\mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \ddots & \ddots & \ddots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{n-1,2}) \\
\vdots & \ddots & \ddots & \ddots & \mathrm{Tr}([\mathbf{A}^T]_{n1}) \\
A_{1n} & \cdots & \mathrm{Tr}([[\mathbf{A}^T]_{1n}]_{2,n-1}) & \mathrm{Tr}([\mathbf{A}^T]_{1n})) & \mathrm{Tr}(\mathbf{A})
\end{bmatrix} \\
\equiv \; & \boldsymbol{\alpha}(\mathbf{A})
\end{aligned}
\tag{143}
$$

As it can be seen, the derivative $\boldsymbol{\alpha}(\mathbf{A})$ also has a Toeplitz structure. Each value in the diagonal is the sum of all the diagonal valued in $\mathbf{A}$, the values in the diagonals next to the main diagonal equal the sum of the diagonal next to the main diagonal in $\mathbf{A}^T$. This result is only valid for the unconstrained Toeplitz matrix. If the Toeplitz matrix also is symmetric, the same derivative yields

$$
\frac{\partial \mathrm{Tr}(\mathbf{AT})}{\partial \mathbf{T}} = \frac{\partial \mathrm{Tr}(\mathbf{TA})}{\partial \mathbf{T}} = \boldsymbol{\alpha}(\mathbf{A}) + \boldsymbol{\alpha}(\mathbf{A})^T - \boldsymbol{\alpha}(\mathbf{A}) \circ \mathbf{I} \tag{144}
$$