# D

# Matrix Calculus

In this Appendix we collect some useful formulas of matrix calculus that often appear in finite element derivations.

## §D.1 THE DERIVATIVES OF VECTOR FUNCTIONS

Let **x** and **y** be vectors of orders $n$ and $m$ respectively:

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

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

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

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

### §D.1.1 Derivative of Vector with Respect to Vector

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

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

### §D.1.2 Derivative of a Scalar with Respect to Vector

If $y$ is a scalar,

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

### §D.1.3 Derivative of Vector with Respect to Scalar

If $x$ is a scalar,

$$\frac{\partial \mathbf{y}}{\partial x} \overset{\text{def}}{=} \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix} \tag{D.5}$$

**REMARK D.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 D.1**

Given

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

and

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

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

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

### §D.1.4 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

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

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

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

[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.”

### EXAMPLE D.2

The transformation from spherical to Cartesian coordinates is defined by

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

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

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

Then

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

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

### EXAMPLE D.3

Consider the quadratic form

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

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

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

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

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

We can of course continue the differentiation process:

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

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

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

The following table collects several useful vector derivative formulas.

| $\mathbf{y}$ | $\frac{\partial\mathbf{y}}{\partial\mathbf{x}}$ |
|---|---|
| $\mathbf{A}\mathbf{x}$ | $\mathbf{A}^T$ |
| $\mathbf{x}^T\mathbf{A}$ | $\mathbf{A}$ |
| $\mathbf{x}^T\mathbf{x}$ | $2\mathbf{x}$ |
| $\mathbf{x}^T\mathbf{A}\mathbf{x}$ | $\mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x}$ |

## §D.2 THE CHAIN RULE FOR VECTOR FUNCTIONS

Let

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

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

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

Each entry of this matrix may be expanded as

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

Then

$$\begin{aligned} \left(\frac{\partial \mathbf{z}}{\partial \mathbf{x}}\right)^T &= \begin{bmatrix} \sum \frac{\partial z_1}{y_q} \frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_1}{\partial y_q} \frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_2}{\partial y_q} \frac{\partial y_q}{\partial x_n} \\ \sum \frac{\partial z_2}{\partial y_q} \frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_2}{\partial y_q} \frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_2}{\partial y_q} \frac{\partial y_q}{\partial x_n} \\ \vdots & & & \\ \sum \frac{\partial z_m}{\partial y_q} \frac{\partial y_q}{\partial x_1} & \sum \frac{\partial z_m}{\partial y_q} \frac{\partial y_q}{\partial x_2} & \cdots & \sum \frac{\partial z_m}{\partial y_q} \frac{\partial y_q}{\partial x_n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \frac{\partial z_1}{\partial y_2} & \cdots & \frac{\partial z_1}{\partial y_r} \\ \frac{\partial z_2}{\partial y_1} & \frac{\partial z_2}{\partial y_2} & \cdots & \frac{\partial z_2}{\partial y_r} \\ \vdots & & & \\ \frac{\partial z_m}{\partial y_1} & \frac{\partial z_m}{\partial y_2} & \cdots & \frac{\partial z_m}{\partial y_r} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial y_r}{\partial x_1} & \frac{\partial y_r}{\partial x_2} & \cdots & \frac{\partial y_r}{\partial x_n} \end{bmatrix} \\ &= \left(\frac{\partial \mathbf{z}}{\partial \mathbf{y}}\right)^T \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right)^T = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \frac{\partial \mathbf{z}}{\partial \mathbf{y}}\right)^T. \end{aligned} \tag{D.22}$$

On transposing both sides, we finally obtain

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

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

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

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 $\mathbf{x}$,

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

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.

## §D.3 THE DERIVATIVE OF SCALAR FUNCTIONS OF A MATRIX

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

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

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

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

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

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

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 D.4

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

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

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

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

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

---

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

### §D.3.1 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

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

Using the chain rule we can write

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

But

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

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

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

It follows that

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

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

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

Then it can be shown that

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

#### EXAMPLE D.5

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

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

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

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

## §D.4 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

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

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

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

This definition complies with the multiplicative and associative rules

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

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

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

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

#### EXAMPLE D.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 $\mathbf{X}$:

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

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

#### EXAMPLE D.7

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

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

from which

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

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

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