Section 4.2. Types of Differentiation

Note. In this section we define differentiation of various structures with respect to a scalar, a vector, and a matrix.

Definition. Let vector y be a function of scalar variable x so that $y = y(x) = [y_1, y_2, \ldots, y_n] = [y_1(x), y_2(x), \ldots, y_n(x)]$. The *derivative of vector y with respect to scalar x* is the vector

$$\frac{\partial y}{\partial x} = \left[\frac{\partial y_1}{\partial x}, \frac{\partial y_2}{\partial x}, \ldots, \frac{\partial y_n}{\partial x}\right] = [y_1', y_2', \ldots, y_n'].$$

Let matrix Y be a function of scalar variable x so that $Y = Y(x) = [y_{ij}(x)]$. The *derivative of matrix Y with respect to scalar x* is the matrix

$$\frac{\partial Y}{\partial x} = \left[\frac{\partial y_{ij}}{\partial x}\right] = [y_{ij}'].$$

Note. To define differentiation with respect to a vector x of an object Φ which is a function of a vector, we use the usual definition:

$$\lim_{t\to 0} \frac{\Phi(x + ty) - \Phi(x)}{t}$$

where y is any conformable vector with x (this will give the derivative of Φ "in the direction" of y). Notice that we can use the usual ε/δ definition of limit, provided we have a metric on the objects Φ. We start with scalar valued functions.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a scalar valued function of a vector. The *derivative of scalar valued function f with respect to vector* $x = [x_1, x_2, \ldots, x_n]$ is

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right].$$

This derivative is the *gradient* of f, denoted ∇f (sometimes read "dell f").

Note. Let A be a given (constant) $n \times n$ matrix and $x \in \mathbb{R}^n$. The quadratic form $x^T Ax$ is a scalar and the derivative of the quadratic form in the direction y is

$$\begin{aligned}
\lim_{t\to 0} \frac{(x+ty)^T A(x+ty) - x^T Ax}{t} &= \lim_{t\to 0} \frac{(x^T + ty^T)A(x+ty) - x^T Ax}{t} \\
&= \lim_{t\to 0} \frac{x^T Ax + ty^T Ax + tx^T Ay + t^2 y^T Ay - x^T Ax}{t} \\
&= \lim_{t\to 0} (y^T Ax + x^T AY + ty^T Ay) = y^T Ax + x^T Ay \\
&= y^T Ax + (x^T Ay)^T \text{ since } x^T Ay \text{ is a } 1 \times 1 \text{ matrix} \\
&\quad \text{and so is symmetric} \\
&= y^T Ax + y^T A^T x = y^T (A + A^T)x.
\end{aligned}$$

Notice that $\|x\|_2^2 = x^T x = x^T \mathcal{I} x$ so this example shows that the directional derivative of the Euclidean norm exists in all directions y.

Note. We now consider derivatives of vector valued functions of vectors, $f : \mathbb{R}^n \to \mathbb{R}^m$.

Definition. Let $f : S \to \mathbb{R}^m$ where $S \subset \mathbb{R}^n$. With

$f = f(x_1, x_2, \ldots, x_n) = (f_1(x_1, x_2, \ldots, x_n), f_2(x_1, x_2, \ldots, x_n), \ldots, f_m(x_1, x_2, \ldots, x_n))$,

define the *matrix gradient* $\partial f/\partial x$ as the $n \times m$ matrix with (i, j) entry $\partial f_i/\partial x_j$:

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

This is also sometimes denoted ∇f (Gentle uses the notation $\partial f^T/\partial x$). The $m \times n$ matrix $\partial f \partial x)^T = (\nabla f)^T$ is the *Jacobian* of f (denoted by Gentle as $\partial f/\partial x^T$).

Note. You have seen the Jacobian when dealing with substitution in a multiple integral (it sort of plays the role of "du" in one variable u-substitution). See my online Calculus 3 (MATH 2110) notes on 15.8. Substitutions in Multiple Integrals.

Definition. Let Y be a $p \times q$ matrix of functions $f_{ij} : \mathbb{R}^n \to \mathbb{R}$, $Y = [f_{ij}]$ where $f_{ij} = f_{ij}(x_1, x_2, \ldots, x_m)$. Define the *derivative* of Y as $\dfrac{\partial Y}{\partial x} = [\nabla f_{ij}]$. That is, $\partial Y/\partial x$ is a three dimensional object (or a matrix of vectors) of dimension $p \times q \times n$ where the (i, j, k) entry is $\partial f_{ij}/\partial x_k$. For fixed k^* we denote the $p \times q$ matrix with (i, j) entry as the (i, j, k^*) entry of $\partial Y/\partial x$ as $\partial Y/\partial x_{k^*}$.

Note. The linearity and chain rule properties of partial derivatives allow us to establish rules of differentiation in these new settings. For example, if $Y = [f_{ij}]$ where $f_{ij} : \mathbb{R}^n \to \mathbb{R}$ is a square nonsingular matrix then we show in Exercise 4.3 that

$$\frac{\partial Y^{-1}}{\partial x} = -Y^{-1}\left(\frac{\partial Y}{\partial x}\right)Y^{-1}.$$

In Exercise 4.2.A we show for $Y = [f_{ij}]$ that $\displaystyle\frac{\partial}{\partial x}[\mathrm{tr}(Y)] = \mathrm{tr}\left(\frac{\partial Y}{\partial x}\right)$.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a scalar valued function of a vector. The matrix H with (i,j) entry as $\displaystyle\frac{\partial^2 f}{\partial x_i \partial x_j}$ is the *Hessian* of f, denoted H:

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

Gentle denotes the Hessian as $\nabla^2 f$ and $\displaystyle\frac{\partial^2 f}{\partial x \partial x^T}$.

Note. In the case that $f = f(x,y)$ then

$$\det(H) = \det\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$$

(assuming the mixed partials are equals, which is really an assumption of continuity) and $\det(H)$ is related to the curvature of the surface $z = f(x,y)$; see online my Differential Geometry (MATH 5510) notes on 1.6. The Gauss Curvature in Detail (see the example on pages 3 and 4). Because of this, the Hessian (which is often defined as $\det(H)$ instead of H itself) can be used to find extrema and saddle points of $z = f(x, y)$ in a style similar to the use of the Second Derivative Test for functions $y = f(x)$. See my online Calculus 3 (MATH 2110) notes on 14.7. Extreme Values and Saddle Points.

Note. Gentle lists $\partial f/\partial x$ for several types of functions $f(x)$ where x is an n-vector. Let a be a constant scalar, b a constant conformable vector, and A a constant conformable matrix. Then we have (with corrections to Gentle as given in the Errata):

$f(x)$	$\partial f/\partial x$
ax	$a\mathcal{I}$
$b^T x$	b
$x^T b$	b
$x^T x$	$2x$
xx^T	$x \otimes \mathcal{I} + \mathcal{I} \otimes x$
$b^T Ax$	$A^T b$
$x^T Ab$	Ab
$x^T Ax$	$(A + A^T)x$
$\exp((-1/2)x^T Ax)$	$-\exp((-1/2)x^T Ax)Ax$ if $A = A^T$
$\|x\|_2^2$	$2x$
$V(x)$	$2x/(n-1)$

The proofs of some of these are to be given in Exercise 4.2.B.

Definition. Let X be a matrix of independent variables and f a function of X so that $f = f(X) = f([x_{ij}])$. The *derivative of function f with respect to matrix X* is the matrix $\dfrac{\partial f}{\partial X} = \left[\dfrac{\partial f}{\partial x_{ij}}\right]$.

Theorem 4.2.1. Differentiation of scalar valued function f satisfies the following.

(1) $\dfrac{\partial f}{\partial X^T} = \left(\dfrac{\partial f}{\partial X}\right)^T$.

(2) For X square and $f(X) = \mathrm{tr}(X)$, $\dfrac{\partial f}{\partial X} = \mathcal{I}$.

(3) For AX a square matrix where A is constant, $\dfrac{\partial[\mathrm{tr}(AX)]}{\partial X} = A^T$.

(4) $\dfrac{\partial[\mathrm{tr}(X^T X)]}{\partial X} = 2X$.

(5) With a and b constant vectors, $\dfrac{\partial[a^T X b]}{\partial X} = ab^T$.

(6) $\dfrac{\partial[\det(X)]}{\partial X} = (\mathrm{adj}(X))^T$.

Note. A quick search of the internet reveals that there are a number of definitions for the derivative of a matrix with respect to a matrix. Inspired by Gentle's general approach, we take the following definition.

Definition. Let X be a matrix of independent variables and let Y be a matrix of functions of X, $Y = [y_{ij}] = [y_{ij}(X)]$. The *derivative of matrix Y with respect to matrix X* is the matrix $\dfrac{\partial Y}{\partial X} = \left[\dfrac{\partial Y}{\partial x_{ij}}\right]$.

Note. Gentle claims that that the following hold. Given his cryptic statement of equation (4.15) on page 155, I am not totally confident that all of the following hold under our definition of a matrix with respect to a matrix.

Note. Gentle lists $\partial f/\partial X$ for several types of functions $f(X)$ where X is a matrix. Let a and b be a conformable vectors, and A a constant conformable matrix. Then we have:

$f(X)$	$\partial f/\partial X$	Justification
$a^T X b$	ab^T	Theorem 4.2.1(5)
$\mathrm{tr}(AX)$	A^T	Theorem 4.2.1(3)
$\mathrm{tr}(X^T X)$	$2X$	Theorem 4.2.1(4)
BX	$\mathcal{I} \otimes B$	
XC	$C^T \otimes \mathcal{I}$	
BXC	$C^T \otimes B$	

If X is square and invertible as required, then:

$f(X)$	$\partial f/\partial X$	Justification
$\mathrm{tr}(X)$	$\mathcal{I}$	Theorem 4.2.1(2)
$\mathrm{tr}(X^k)$	kX^{k-1}	Exercise 4.2.C(1)
$\mathrm{tr}(BX^{-1}C)$	$-(X^{-1}CBX^{-1})^T$	
$\det(X)$	$\det X(X^{-1})^T$	Exercise 4.2.C(2)
$\log(\det(X))$	$(X^{-1})^T$	Exercise 4.2.C(3)
$\det X^k$	$k\det(X)^k(X^{-1})^T$	Exercise 4.2.C(4)
$BX^{-1}C$	$-(X^{-1}C)^T \otimes BX^{-1}$	