Multivariate Analysis

Iannis Petridis

Autum 2011

Contents

1 Mulivarible Calculus

1.1 Notation

$X \in \mathbb{R}^n$, $X = \{x_1, x_2, \ldots, x_n\}$ where $x_i \in \mathbb{R}$ $\mathbb{R}^n$ is a vector space
length norm$|x| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$
If $Y, X \in \mathbb{R}^n$ and $Y = \{y_1, y_2, \ldots, y_n\}$ then $X \cdot Y = x_1y_1 + x_2y_2 + \cdots + x_ny_n$
Standard Basis:

$$e_j = (0, \ldots, 0, 1, 0, \ldots)$$

j-1, j, j+1

Properties of norm

$$|x| \geq 0$$

$$|x| = 0 iff x = \vec{0}$$

1.2 linear Transformation

$$T : \mathbb{R}^n \to \mathbb{R}^n$$

(i) $T(x + y) = T(x) + T(y)$

(ii) $T(\lambda x) = \lambda T(x)$

Matrix Representation of T with respect to the standard basis:

$$T(e_i) = \sum_{j=1}^{m} a_{i,j} e_j \text{ where } [T]_\epsilon^\epsilon = A = (a_i, j)_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$$

Given: $T : \mathbb{R}^n \to \mathbb{R}^m, S : \mathbb{R}^n \to \mathbb{R}^m$ and $U : \mathbb{R}^m \to \mathbb{R}^k$

(i) $[UT]_{kxm} = [U]_{kxm}[T]_{mxn}$

(ii) $[T + S] = [T] + [S]$

(iii) $\lambda[T] = [\lambda T]$

$T : \mathbb{R}^n \to \mathbb{R}^m, X \in \mathbb{R}^n, Y \in \mathbb{R}^m, X = (x^1, \ldots, x^n), Y = (y^1, \ldots, y^m)$

$$\begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{pmatrix} = [T] \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix}$$

1.3 Functions & Continuity

$$f : \mathbb{R}^n \to \mathbb{R}^m$$

vector valued function

$$f : A \to \mathbb{R}^m$$

where $A \subset \mathbb{R}^n$
f has components which are scalar fields
$f^i : A \to \mathbb{R}$

$$f(x) = (f^1(x), \ldots, f^m(x))$$

$\Pi^i : \mathbb{R}^m \to \mathbb{R}$

$$\Pi^i((x)^1, \ldots, (x)^m)$$

Π^i is a linear transformation for i=1,...,m

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "f"] \arrow[dr, "f^i"'] & \mathbb{R}^m \arrow[d, "\Pi^i"] \\
 & \mathbb{R}
\end{tikzcd}$$

Definition 1.1. *$f : \mathbb{R}^n \to \mathbb{R}^m$ then $\lim_{x\to a}(f(x)) = b$ means:*

$$\forall\epsilon > 0, \exists\delta > 0 \ st,\ 0 < |x - a| < \delta \implies |f(x) - b| < \epsilon$$

Definition 1.2. *f is called continuious at a if:*

$$\lim_{x\to a}(f(x)) = f(a)$$

f is called continuious at the set of A if it is continuious at a $\forall a \in Al$

Theorem 1.1 (Combination Theorm)**.** *Assume*

$$\lim_{x\to a}(f(x)) = b, lim_{x\to a}(g(x)) = c$$

then:

(i) $\lim_{x\to a}(f(x) + g(x)) = b + c$

(ii) $\lim_{x\to a}(\lambda f(x)) = \lambda b$

(iii) $\lim_{x\to a}(f(x) \cdot g(x)) = b \cdot c$

(iv) $\lim_{x\to a}|f(x)| = |b|$

Proof. of (iii)

$$\begin{aligned}
f(x)\cdot g(x)-b\cdot c &= f(x)\cdot g(x)-b\cdot g(x)+b\cdot g(x)-b\cdot c\\
&= g(x)\dot{(}f(x)-b)+b\cdot(g(x)-c)\\
|f(x)\cdot g(x)-b\cdot c| &= |g(x)\dot{(}f(x)-b)+b\cdot(g(x)-c)|\\
&\leq |g(x)\dot{(}f(x)-b)|+|b\cdot(g(x)-c)|
\end{aligned}$$

Cauchy-Schwartz: $|x^1y^1+\cdots+x^ny^n| \leq \sqrt{(x^1)^2+\cdots+(x^n)^2}\cdot\sqrt{(y^1)^2+\cdots+(y^n)^2}$

$$|f(x)\cdot g(x)-b\cdot c| \leq |g(x)\dot{(}f(x)-b)|+|b\cdot(g(x)-c)| \leq |g(x)|\cdot|f(x)-b|+|b|\cdot|g(x)-c|$$

Since $lim_{x\to a}(g(x))=c$, g is a bounded neighbourhood of a, i.e:

$$\forall M\leq 0, \exists\delta>0 \; st, \; |g(x)|\leq M for|x-a|<\delta$$

$\square$

Remark. *We have:*

(i) $f:\mathbb{R}^n\to\mathbb{R}^m$ *is continuious iff:* $f^i:\mathbb{R}^n\to\mathbb{R}$ *is continuious for* $i=1,\ldots,m$

(ii) Polynomial functions in n *variables,* $f(x^1,\ldots,x^n)$*, are continuious*

(iii) Rational functions, $R(x)=\frac{P(x)}{Q(x)}$*, are continuious where defined, ie:* $Q(x)\neq 0$ *and* P*,* Q *are polynomials in* n *variables.*

Theorem 1.2. *Linear transformations are continuious.*

Proof. $T:\mathbb{R}^n\to\mathbb{R}^m$ let $a\in\mathbb{R}^n$ to show: $\lim_{x\to a}T(a+h)=T(a)$

$$\begin{aligned}
|T(a+h)-T(a)| = |T(h)| = |T(h^1e_1+\cdots+h^ne_n)| &= |h^1T(e_1)+\cdots+h^nT(e_n)|\\
\leq |h^1||T(e_1)|+\ldots|h^n||T(e_n)| &\leq |h|(T(e_1)+\ldots T(e_n))
\end{aligned}$$

$$So: \quad |T(a+h)-T(a)|\leq M|h| \quad where \quad M=\sum_{i=1}^{n}|T(e_i)|$$

$$So\,given \quad \epsilon>0, \quad choose \quad \delta=\frac{\epsilon}{M} \quad such\,that \quad |h|<\delta \implies |T(a+h)-T(a)|<\epsilon$$

$\square$

Example 1.1. $f(x,y) = \frac{x^2-y^2}{x^2+y^2}, \quad (x,y) = (0,0)$ *assume* $\quad \lim_{(x,y)\to(0,0)} f(x,y) = L$

$$\forall \epsilon > 0, \quad \exists \delta > 0 \quad \textit{such that} \quad 0 < |(x,y)| < \delta \implies |f(x,y) - L| < \epsilon$$

$$\begin{aligned}
\textit{Plug (x,0) into f:} f(x,0) &= \frac{x^2-0}{x^2-0} = 1 \\
\textit{Plug (0,y) into f:} f(0,y) &= \frac{0-y^2}{0+y^2} = -1 \\
If\ |x| < \delta \quad |f(x,0)| < \delta &\implies |f(x,0) - L| < \epsilon \quad ie \quad |1-L| < \epsilon \\
If\ |y| < \delta \quad |f(0,y)| < \delta &\implies |f(0,y) - L| < \epsilon \quad ie \quad |-1-L| < \epsilon \\
&\implies \epsilon = \frac{1}{2} \quad contradiction!
\end{aligned}$$

Now consider $\quad y = mx, m \in \mathbb{R}$

$$\begin{aligned}
&f(x,mx) = \frac{x^2-(mx)^2}{x^2+(mx)^2} = \frac{1-m^2}{1+m^2} \\
&\lim_{x\to 0}(\lim_{y\to 0} f(x,y)) = \lim_{x\to 0} 1 = 1 \\
&\lim_{y\to 0}(\lim_{x\to 0} f(x,y)) = \lim_{y\to 0} -1 = -1
\end{aligned}$$

However checking along straight lines is not enough to prove continuity.

Example 1.2.

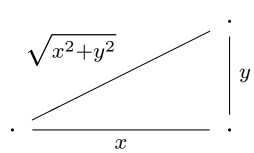
$$f(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & \textit{if } (x,y) \neq 90,0) \\ 0 & \textit{if } 9x,y) = (0,0) \end{cases}$$

Show f is continuious at (0,0)

$$\forall \epsilon > 0, \quad \exists \delta > 0$$

$$\left| \frac{xy}{\sqrt{x^2+y^2}} \right| \leq \frac{|x| \cdot |y|}{\sqrt{x^2+y^2}} \leq \frac{\sqrt{x^2+y^2} \cdot \sqrt{x^2+y^2}}{\sqrt{x^2+y^2}} = \sqrt{x^2+y^2} = |(x,y)|$$

Since:



Note. *if the total degree of the neumerator is higher than the denominator in a rational function. Then the limit should be 0.*

Theorem 1.3. *If f is continuious at a and g is continuious at $f(a)$ then $g \circ f$ is continuious at a.*

1.4 Partial Derivitives

Definition 1.3. *Let* $f : \mathbb{R}^n \to \mathbb{R},\ a \in \mathbb{R}$

$$Define: \quad D_i f(a) = \lim_{h\to 0} \frac{f(a^1, \ldots, a^{i-1}, a^i + h, a^{i+1}, \ldots, a^n)}{h}$$

Example 1.3. *if* $f : \mathbb{R}^n \to \mathbb{R}$

$$\left.\frac{df}{dx}\right|_{(a,b)} = D_1 f(a,b)$$

$$\left.\frac{df}{dy}\right|_{(a,b)} = D_2 f(a,b)$$

$$\textit{and in } \mathbb{R}^3 \textit{ we use } \frac{df}{dx}, \frac{df}{dy} \textit{ and } \frac{df}{dz} \textit{ etc.}$$

Example 1.4.

$$f(x,y) = \begin{cases} \frac{x^2-y^2}{\sqrt{x^2+y^2}} & \textit{if } (x,y) \neq (0,0) \\ 1 & \textit{if } (x,y) = (0,0) \end{cases}$$

$$D_1 f(0,0) = \left.\frac{df}{dx}\right|_{(0,0)} = \lim_{x\to 0} \frac{f(x,0) - f(0,0)}{x} = \lim_{x\to 0} \frac{\frac{x^2-0}{x^2-0} - 1}{x} = 0$$

$$D_2 f(0,0) = \left.\frac{df}{dy}\right|_{(0,0)} = \lim_{y\to 0} \frac{f(0,y) - f(0,0)}{y} = \lim_{y\to 0} \frac{\frac{0-y^2}{0+y^2} - 1}{y} = \frac{-2}{y} = \pm\infty$$

1.5 Total Derivitive

In 1 dimention we write the following for the derivitive of $f : \mathbb{R} \to \mathbb{R}$

$$f^{'}(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

we try to write it in higher dimentions $f : \mathbb{R}^n \to \mathbb{R}^m$ in this form

$$\lim_{h\to 0}\left[\frac{f(a+h) - f(a)}{h} - f^{'}(a)\right] = \lim_{h\to 0}\left[\frac{f(a+h) - f(a) - h\cdot f^{'}(a)}{h}\right]$$
$$= \lim_{h\to 0} \frac{|f(a+h) - f(a) - h\cdot f^{'}(a)|}{|h|} = 0$$

For $f : \mathbb{R}^n \to \mathbb{R}^m$ consider the tangent line at a: $y = f(a) + f^{'}(a)(x-a)$
call $x - a = h$ then we have: $y = f(a) + f^{'}(a)(h)$
this is an Affine transformation, not a linear map.
Look at the map:

$$\lambda : h \to hf^{'}(a), \quad h \in \mathbb{R}$$

This is a linear map.

$$\lambda(h_1 + h_2) = (h_1 + h_2)f^{'}(a) = h_1 f^{'}(a) + h_2 f^{'}(a) = \lambda(h_1) + \lambda(h_2)$$
$$\lambda(\alpha\cdot h) = (\alpha h)f^{'}(a) = \alpha(hf^{'}(a)) = \alpha\cdot\lambda(h)$$
$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0$$

Definition 1.4 (Total Derivitive). *$f : \mathbb{R}^n \to \mathbb{R}^m$ or ($f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$, A is open) is differentiable at a ($a \in A$) if we can rind a linear transformation $\lambda : \mathbb{R}^n \to \mathbb{R}^m$ st:*

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0$$

The linear transformation λ is called the total derivitive of f at a and denoted Df(a) st

$$Df(a) = \lambda(h)$$

Example 1.5. *$f : \mathbb{R}^n \to \mathbb{R}^m$, $f(x) = k$, $k \in \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$ with the 0 linear transformation $0 : f : \mathbb{R}^n \to \mathbb{R}^m$, $0(h) = 0$*

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - 0(h)|}{|h|} = \lim_{h\to 0} \frac{|k - k - 0|}{|h|} = 0$$

Example 1.6. *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformaton, it is differentiable at $a \in \mathbb{R}^n$ with linear transformation $Df(a) = f$*

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - f(h)|}{|h|} = \lim_{h\to 0} \frac{|f(a+h-a-h)|}{|h|} = 0$$

Theorem 1.4 (Uniqueness of Total Derivitive). *If f is differentiable at a then there exists a unique linear transformation, $\lambda : \mathbb{R}^n \to \mathbb{R}^m$, such that*

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} = 0$$

Proof. suppose $\mu : \mathbb{R}^n \to \mathbb{R}^m$ is another linear transformation such that:

$$\lim_{h\to 0} \frac{|f(a+h) - f(a) - \mu(h)|}{|h|} = 0$$

deduce that $\lambda = \mu \, \forall h \in \mathbb{R}^n$ *ie* $\lambda(h) = \mu(h)$

$$\begin{aligned} \frac{|\lambda(h) - \mu(h)|}{|h|} &= \frac{|\lambda(h) + f(a) - f(a+h) + f(a+h) - f(a) - \mu(h)|}{|h|} \\ &\leq \frac{|f(a+h) - f(a) - \lambda(h)|}{|h|} + \frac{|f(a+h) - f(a) - \mu(h)|}{|h|} \end{aligned}$$

Conclude that:

$$lim_{h\to 0} \frac{|\lambda(h) - \mu(h)|}{|h|} \leq 0 + 0 = 0 \quad (*)$$

Let h=0 $\lambda = 0 = \mu$ since λ, μ are linear. Now fix $h \in \mathbb{R}^n$, $h \neq 0$ and let $t \in \mathbb{R}$ such that $th \in \mathbb{R}^n$ then replace h with th in (*):

$$\begin{aligned} \lim_{t\to 0} \frac{|\lambda(th) - \mu(th)|}{|th|} &= \lim_{t\to 0} \frac{|t\lambda(h) - t\mu(h)|}{|t||h|} \\ &= \lim_{t\to 0} \frac{|t|}{|t|} \frac{|\lambda(h) - \mu(h)|}{|h|} = \frac{|\lambda(h) - \mu(h)|}{|h|} = 0 \\ &So \quad \lambda(h) = \mu(h) \end{aligned}$$

$\square$

Definition 1.5 (Jacobian Matrix)**.** *$f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in \mathbb{R}^n$ and it is derivitive at a $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map. Then the matrix representation of $Df(a)$ is $f^{'}(a) \in |MM_{mxn}$ and is called the Jacobian Matrix of f at a.*

Example 1.7. $f : \mathbb{R}^2 \to \mathbb{R}^2, \quad f(x, y) = (x^2, x + 5) \quad x, y \in \mathbb{R}$
Show that $Df(1, 2)(h^1, h^2) = (4h^1 + h^2, h^1)$:

$$\begin{aligned}
&f((1,2) + (h^1, h^2)) - f(1,2) - Df(1,2)(h^1, h^2)\\
&= f(1 + h^1, 2 + h^2) - f(1,2) - (4h^1 + h^2, h^1)\\
&= ((1 + h^1)^2(2 + h^2), (1 + h^1 + 5)) - (2, 6) - (4h^1 + h^2, h^1)\\
&= (2 + h^2 + 2(h^1)^2 + (h^1)^2h^2 + 2h^1h^2 + 4h^1 - 2 - 4h^1 - h^2, 6 + h^1 - 6 - h^1)
\end{aligned}$$

Take length:

$$|(2(h^1)^2 + (h^1)^2h^2 + 2h^1h^2, 0)| \le 2|h|^2 + |h|^2|h| + 2|h||h| = 4|h|^2 + |h|^3$$

So:

$$\begin{aligned}
&\lim_{h\to 0} \frac{|f((1,2) + (h^1, h^2)) - f(1,2) - Df(1,2)(h^1, h^2)|}{|h|}\\
&\le \lim_{h\to 0} \frac{4|h|^2 + |h|^3}{|h|} = \lim_{h\to 0} 4|h| + |h|^2 = 0
\end{aligned}$$

Definition 1.6. *$f^{'}(a)$ is the matrix representation of $Df(a)$*

$$Df(a)(h)^t = \begin{pmatrix} y^1 \\ y^2 \\ \vdots \\ y^m \end{pmatrix} = f^{'}(a) \begin{pmatrix} h^1 \\ h^2 \\ \vdots \\ h^n \end{pmatrix}$$

$$f^{'}(a) = \begin{pmatrix} D_1f^1(a) & D_2f^1(a) & \cdots & D_nf^1(a) \\ D_1f^2(a) & D_2f^2(a) & \cdots & D_nf^2(a) \\ \vdots & \vdots & \ddots & \vdots \\ D_1f^m(a) & D_2f^m(a) & \cdots & D_nf^m(a) \end{pmatrix}$$

Example 1.8. *With this new information we can tackle example 1.7:*
$f : \mathbb{R}^2 \to \mathbb{R}^2, \quad f(x, y) = (x^2, x + 5) \quad x, y \in \mathbb{R}$
Show that $Df(1, 2)(h^1, h^2) = (4h^1 + h^2, h^1)$:

$$\frac{df^1}{dx} = 2xy, \quad \frac{df^1}{dy} = x^2, \quad \frac{df^2}{dx} = 1, \quad \frac{df^2}{dy} = 0$$

$$f^{'}(1, 2) = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}$$

$$f^{'}(1, 2) \begin{pmatrix} h^1 \\ h^2 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} h^1 \\ h^2 \end{pmatrix} = \begin{pmatrix} 4h^1 + h^2 \\ h^2 \end{pmatrix}$$

Remark. *Having directional derivitives in all directions $u \neq 0$ is not enough to guarantee $df(a)$ exists.*

Theorem 1.5. *If f is differentiable at a then f is continuious at a.*

Proof.

$$\begin{aligned}
\lim_{h\to 0}|f(a+h)-f(a)| &= \lim_{h\to 0}|f(a+h)-f(a)-Df(a)+Df(a)| \\
&\leq \lim_{h\to 0}\frac{|f(a+h)-f(a)-Df(a)(h)|}{|h|}\cdot|h| + \lim_{h\to 0}|Df(a)(h)| \\
&= 0
\end{aligned}$$

□

1.6 The Chain Rule

Theorem 1.6 (Chain Rule)**.** *if $f:\mathbb{R}^n\to\mathbb{R}^m$ is differentiable at a and $f:\mathbb{R}^m\to\mathbb{R}^k$ is differentiable at $f(a)$ then $g\circ f:\mathbb{R}^n\to\mathbb{R}^k$ is differentiable at a and*

$$D(g\circ f)(a) = Dg(f(a))\circ Df(a)$$

$$\begin{tikzcd}
\mathbb{R}^n \arrow[r, "Df(a)"] \arrow[dr, "D(g\circ f)(a)"'] & \mathbb{R}^m \arrow[d, "Dg(f(a))"] \\
& \mathbb{R}^k
\end{tikzcd}$$

$$(g\circ f)^{'}(a) = g^{'}(f(a))\cdot f^{'}(a), \quad \textit{where } \cdot \textit{ represents matrix multiplication}$$

Proof. if $b=f(a)$ and we let $Df(a)=\lambda$ and $Dg(f(a))=\mu$ then if we define:

$$\varphi(x) = f(x)-f(a)-\lambda(x-a) \tag{1}$$
$$\psi(y) = g(y)-g(b)-\mu(y-b) \tag{2}$$
$$\rho(x) = g\circ f(x)-g\circ f(a)-\mu\circ\lambda(x-a) \tag{3}$$

Then:

$$\lim_{h\to 0}\frac{|f(a+h)-f(a)-Df(a)(h)|}{|h|} = \lim_{x\to a}\frac{|\varphi(x)|}{|x-a|} = 0 \tag{4}$$
$$\lim_{h\to 0}\frac{|g(b+h)-g(b)-Dg(b)(h)|}{|h|} = \lim_{y\to b}\frac{|\psi(y)|}{|y-b|} = 0 \tag{5}$$

We must show:

$$\lim_{h\to 0}\frac{|g\circ f(x)-g\circ f(a)-\mu\circ\lambda(x-a)|}{|h|} = \lim_{x\to b}\frac{|\rho(x)|}{|x-b|} = 0$$

Now:

$$\begin{aligned}
\rho(x) &= g(f(x))-g(b)-\mu(\lambda(x-a)) \\
&= g(f(x))-g(b)-\mu(f(x)-f(a)-\varphi(x)) && \text{by (1)} \\
&= [g(f(x))-g(b)-\mu(\lambda(f(x)-f(a)))] \\
&= \mu(\varphi(x)) = \psi(f(x))+\mu(f(x)) && \text{by (2)}
\end{aligned}$$

Thus we must Prove

$$\lim_{x\to a}\frac{|\psi(f(x))|}{|x-a|}=0 \tag{6}$$

$$\lim_{x\to a}\frac{|\mu\varphi(x)|}{|x-a|}=0 \tag{7}$$

It follows from (5) that for some $\delta>0$ we have

$$|\psi(f(x))|<\epsilon|f(x)-b| \quad if\ |f(x)-b|<\delta$$

which is true if $|x-a|<\delta_1$ for a suitable δ_1. We also have that if T is a linear transformation then $\exists M\geq 0\ such\ that\ |T(x)|<M|x|$. So then:

$$\begin{aligned}|\psi(f(x))|&<\epsilon|f(x)-b|\\&=\epsilon|\varphi(x)+\lambda(x-a)|\\&\leq\epsilon|\varphi(x)|+\epsilon M|x-a|\end{aligned}$$

So

$$\lim_{x\to a}\frac{|\psi(f(x))|}{|x-a|}\leq\lim_{x\to a}\frac{\epsilon|\varphi(x)|}{|x-a|}+\lim_{x\to a}\frac{\epsilon M|x-a|}{|x-a|}=\epsilon M\to 0$$

Also

$$\lim_{x\to a}\frac{|\mu\varphi(x)|}{|x-a|}\leq\lim_{x\to a}\frac{M|\varphi(x)|}{|x-a|}=0$$

□

Theorem 1.7. *Define $s:\mathbb{R}^2\to\mathbb{R}\quad s(x,y)=x+y$ then s is differentiable and $Ds=s$*

Proof. S is linear so

$$\begin{aligned}s((x,y)+(x',y'))=s(x+x',y+y')&=s(x,y)+s(x',y')\\s(\lambda(x,y))&=\lambda s(x,y)\end{aligned}$$

$$\lim_{h\to 0}\frac{|s(a+h)-s(a)-s(h)|}{|h|}=0$$

□

Theorem 1.8. *Define $p:\mathbb{R}^2\to\mathbb{R},\quad p(x,y)=xy$, then p is differentiable and: $Dp(a,b):\mathbb{R}^2\to\mathbb{R}$ is linear with $Dp(a,b)(h,k)=ak+bh$ and $p'=(b,a)$*

Proof. use of derivitive

$$\begin{aligned}p((a,b)+(h,k))-p(a,b)-Dp(a,b)(h,k)&=p(a+h,b+k)-p(a,b)-(ak+bh)\\&=(a+h)(b+k)-ab-(ak+bh)=hk\end{aligned}$$

$$\frac{|p((a,b)+(h,k))-p(a,b)-Dp(a,b)(h,k)|}{|(h,k)|}=\frac{|hk|}{\sqrt{h^2+k^2}}\leq\frac{\sqrt{h^2+k^2}\sqrt{h^2+k^2}}{\sqrt{h^2+k^2}}=\sqrt{h^2+k^2}\to 0$$

□

Remark. *To check some $T:\mathbb{R}^n\to\mathbb{R}^m$ is linear we listed two properties:*

$$T(x+y)=T(x)+T(y)$$
$$T(\lambda x)=\lambda T(x)$$

we can instead just check:

$$T(\lambda x+y)=\lambda T(x)+T(y)$$

1.7 Linear Functionals

Definition 1.7. *Let $g^i : \mathbb{R}^n \to \mathbb{R}$ be a linear map, such a map is called a linear functional. The set of all linear functionals from $\mathbb{R}^n \to \mathbb{R}$ is called the dual space of $\mathbb{R}^n$, denoted $(\mathbb{R}^n)*$ let $g^1, \dots, g^m$ be linear functionals $g^i : \mathbb{R}^n \to \mathbb{R}$, then I can combine them to get a map $g : \mathbb{R}^n \to \mathbb{R}^m$ by $g(x) = (g^1(x), \dots, g^m(x))$*
$g : \mathbb{R}^n \to \mathbb{R}^m$ is linear such for $x, y \in \mathbb{R}^n,\ \lambda \in \mathbb{R}$

$$\begin{aligned} g(\lambda x + y) &= \lambda g(x) + g(y) \\ \textit{this can be seen by} \qquad g(\lambda x + y) &= (g^1(\lambda x + y), \dots, g^m(\lambda x + y) \\ &= (\lambda g(x)^1 + g^1(y), \dots, \lambda g(x)^m + g^m(y)) \\ &= \lambda(g^1(x), \dots, g^m(x)) + (g^1, \dots, g^m) \end{aligned}$$

$[g^i]$ is the matrix representation of g^i
$[g^i] = (g^i_1, \dots, g^i_n)$

$$[g]_{mxn} = \begin{pmatrix} g^1_1 & \cdots & g^1_n \\ \vdots & & \vdots \\ g^m_1 & \cdots & g^m_n \end{pmatrix}$$

Theorem 1.9. *$f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a iff f^i are differentiable at a, $i = 1, \dots, m$ and $Df(a) = (Df^1, \dots, Df^m(a))$*

Proof. assume f is differentiable at a we take the linear function $\Pi^i(x^1, \dots, x^m) = x^i$ and compose it with f we get

$$f^i = \Pi^i \circ f$$

this is differentiable by chain rule since f and Π^i are differentiable $\forall i = 1, \dots, m$

$$\implies Df^i = D\Pi^i(a) \cdot Df(a)$$

$D\Pi^i = \Pi^i$

$$\implies Df^i = \Pi^i(a) \cdot Df(a)$$

Now assume the all f^i are differentiable at a $\forall i = 1, \dots, m$

$$\begin{aligned} &f(a+h) - f(a) - (Df^1(a)(h), \dots, Df^m(a)(h)) \\ &= (f^1(a+h), \dots, f^m(a+h)) - (f^1, \dots, f^m) - (Df^1 * (a)(h), \dots, Df^m(a)(h)) \\ &= (f^1(a+h) - f^1(a) - df^1(a), \dots, f^m(a+h) - f^m(a) - df^m(a)) \end{aligned}$$

So

$$\begin{aligned} &\frac{|f(a+h) - f(a) - (Df^1(a)(h), \dots, Df^m(a)(h))|}{|h|} \\ &\leq \frac{|f^1(a+h) - f^1(a) - df^1(a)|}{|h|}, \dots, \frac{|f^m(a+h) - f^m(a) - df^m(a)|}{|h|} \to 0 \end{aligned}$$

□

Remark. *If $T, S : \mathbb{R}^n \to \mathbb{R}^m$ are linear then $(T + S) : \mathbb{R}^n \to \mathbb{R}^m, (T + S)(x) = T(x) + S(x)$ is linear.*
If $\lambda \in \mathbb{R}$ then $(\lambda T) : \mathbb{R}^n \to \mathbb{R}^m, (\lambda T)(x) = \lambda \cdot T(x)$ is also linear.

Corollary 1.1. *$f, g : \mathbb{R}^n \to \mathbb{R}$ differentiable at $a \in \mathbb{R}^n$*

(i) $D(f+g)(a) = Df(a) + Dg(a)$

(ii) Product rule: $D(f \cdot g)(a) = g(a) \cdot Df(a) + f(a) \cdot Dg(a)$

(iii) Quotient rule: if $g(a) \neq 0,\ D(\frac{f}{g})(a) = \frac{1}{g(a)^2} \cdot (g(a) \cdot Df(a) - f(a) \cdot Dg(a))$

Proof. For (i):
We can consder the function s from theorem 1.7, $s : \mathbb{R}^2 \to \mathbb{R} \quad s(x,y) = x + y$, but acting on f and g ie $s(f,g) = f + g$ and $Ds = s$

$$D(f+g)(a) = Ds(f(a), g(a)) \circ D(f,g)(a) = s \circ (Df(a), Dg(a)) = Df(a) + Dg(a)$$

For (ii):
We can consder the function p from theorem 1.8, $p : \mathbb{R}^2 \to \mathbb{R} \quad p(x,y) = xy$, but acting on f and g ie $p(f,g) = fg$ with $Dp(f,g)(h,k) = fk + gh$

$$D(f \cdot g)(a) = Dp(f,g) \cdot D(f,g)(a) = Dp(f(a), g(a)) \cdot (Df(a), Dg(a)) = f(a) \cdot Dg(a) + g(a) \cdot Df(a)$$

(iii) follows from (ii) $\square$