(2.5)

Properties of the Derivative

Sum and difference rules

Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}^n$, then $f \pm g$ is differentiable at a with derivative $D(f \pm g)(a) : \mathbb{R}^n \to \mathbb{R}^m$ equal to $Df(a) \pm Dg(a)$

Proof: Since f, g are differentiable at a, we know that

$$\lim_{x \to a} \frac{\|f(x) - f(a) - Df(a)(x-a)\|}{\|x-a\|} = \lim_{x \to a} \frac{\|g(x) - g(a) - Dg(a)(x-a)\|}{\|x-a\|} = 0$$

We wish to show that $\displaystyle \lim_{x \to a} \frac{\|(f+g)(x) - (f+g)(a) - D(f+g)(a)(x-a)\|}{\|x-a\|} =$

$$= \lim_{x \to a} \frac{\|f(x) + g(x) - f(a) - g(a) - Df(a)(x-a) - Dg(a)(x-a)\|}{\|x-a\|} = 0$$

This is easy:

$$0 \leq \frac{\|f(x) + g(x) - f(a) - g(a) - Df(a)(x-a) - Dg(a)(x-a)\|}{\|x-a\|} \leq$$

$$\frac{\|f(x) - f(a) - Df(a)(x-a)\| + \|g(x) - g(a) - Dg(a)(x-a)\|}{\|x-a\|} =$$

$$= \frac{\|f(x)-f(a)-Df(a)(x-a)\|}{\|x-a\|} + \frac{\|g(x)-g(a)-Dg(a)(x-a)\|}{\|x-a\|}$$

Since $\lim_{x\to a} \frac{\|f(x)-f(a)-Df(a)(x-a)\|}{\|x-a\|} + \frac{\|g(x)-g(a)-Dg(a)(x-a)\|}{\|x-a\|} = 0$

and $\lim_{x\to a} 0 = 0$

We see that

$$0=\lim_{x\to a} 0 \leq \lim_{x\to a} \frac{\|(f+g)(x)-(f+g)(a)-Df(a)(x-a)-Dg(a)(x-a)\|}{\|x-a\|} \leq$$

$$\leq \lim_{x\to a} \frac{\|f(x)-f(a)-Df(a)(x-a)\|}{\|x-a\|} + \frac{\|g(x)-g(a)-Dg(a)(x-a)\|}{\|x-a\|} = 0$$

which proves that $f+g$ is differentiable at a with derivative

$D(f+g)(a)(x) = Df(a)(x) + Dg(a)(x)$.

<u>The Chain rule</u>

Suppose that $f:\mathbb{R}^p \to \mathbb{R}^n$ is differentiable at $a\in\mathbb{R}^p$ and $g:\mathbb{R}^n\to\mathbb{R}^m$ is differentiable at $f(a)\in\mathbb{R}^n$. What might be the derivative of $g\circ f$ at a? In Calc. I we have seen that if $f:\mathbb{R}\to\mathbb{R}$ is differentiable at $a\in\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}$ is differentiable at $f(a)$ then $\frac{d}{dx}(g\circ f)\Big|_{x=a} = g'(f(a))f'(a)$. A similar result holds for functions of several variables:

Thm: Let $f: \mathcal{U} \subset \mathbb{R}^p \to \mathbb{R}^n$ and $g: f(\mathcal{U}) \to \mathbb{R}^m$. Suppose that f is differentiable at $a \in \mathcal{U}$ with derivative $Df(a)$ and g is differentiable at $f(a)$ with derivative $Dg(f(a))$. Then the composite function $g \circ f: \mathbb{R}^p \to \mathbb{R}^m$ is differentiable at a and the derivative of $g \circ f$ at a is the linear transformation $D(g \circ f)(a)$ given by

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

Proof:
Let $S = Dg(f(a))$ and $T = Df(a)$. We wish to show that ST is the derivative of $g \circ f$ at a. In other words, we would like to prove that

$$\lim_{x \to a} \frac{\|g(f(x)) - g(f(a)) - ST(x-a)\|}{\|x-a\|} = 0$$

Observe that

$$0 \leq \frac{\|g(f(x)) - g(f(a)) - ST(x-a)\|}{\|x-a\|} =$$

$$= \frac{\|g(f(x)) - g(f(a)) - S\big(f(x) - f(a) + T(x-a) - f(x) + f(a)\big)\|}{\|x-a\|} =$$

$$= \frac{\|g(f(x)) - g(f(a)) - S(f(x) - f(a)) + S\big(f(x) - f(a) - T(x-a)\big)\|}{\|x-a\|} \leq$$

$$\leq \frac{\|g(f(x)) - g(f(a)) - S(f(x) - f(a))\|}{\|x-a\|} + \frac{\|S(f(x) - f(a) - T(x-a))\|}{\|x-a\|} \quad (1)$$

The left-hand side of (1) can be written as

$$\frac{\|g(f(x))-g(f(a))-S(f(x)-f(a))\|}{\|f(x)-f(a)\|}\cdot\frac{\|f(x)-f(a)\|}{\|x-a\|}\le$$

$$\le\frac{\|g(f(x))-g(f(a))-S(f(x)-f(a))\|}{\|f(x)-f(a)\|}\cdot\left(\frac{\|f(x)-f(a)-T(x-a)\|}{\|x-a\|}+\frac{\|T(x-a)\|}{\|x-a\|}\right)\ (2$$

By exercise 7 of H.W#6, $\|T(x-a)\|\le\|T\|\,\|x-a\|$. Therefore

$$(2)\le\frac{\|g(f(x))-g(f(a))-S(f(x)-f(a))\|}{\|f(x)-f(a)\|}\cdot\left(\frac{\|f(x)-f(a)-T(x-a)\|}{\|x-a\|}+\|T\|\right)\ (3)$$

Since f is differentiable at a, f is continuous at a. Thus

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

By letting $y=f(x)$ and $b=f(a)$, we see that $\lim_{x\to a}\frac{\|g(f(x))-g(f(a))-S(f(x)-f(a))\|}{\|f(x)-f(a)\|}$

$$=\lim_{y\to b}\frac{\|g(y)-g(b)-S(y-b)\|}{\|y-b\|}=0$$

This shows that $\lim_{x\to a}(3)=0$

The right-hand side of (1), $\frac{\|S(f(x)-f(a)-T(x-a))\|}{\|x-a\|}\le\|S\|\frac{\|f(x)-f(a)-T(x-a)\|}{\|x-a\|}$

(why?)

Thus $\lim_{x\to a}\frac{\|S(f(x)-f(a)-T(x-a))\|}{\|x-a\|}\le\lim_{x\to a}\|S\|\frac{\|f(x)-f(a)-T(x-a)\|}{\|x-a\|}=0$

It follows that $\lim_{x\to a}(1)=0$ and therefore ST is the derivative

of $g\circ f$ at a as desired.

Since $D(g\circ f)(a) = Dg(f(a))(Df(a)(x))$, it follows that

$J(g\circ f)(a) = \mathcal{M}(D(g\circ f)(a)) = \mathcal{M}(Dg(f(a)))\,\mathcal{M}(Df(a)) =$

$= Jg(f(a))\,Jf(a)$

Ex. Let $f(x_1, x_2) = (x_1^2 + x_2^2,\ x_1x_2)$ and $g(x_1, x_2) =$
$= (3x_1x_2,\ x_1 - x_2,\ 7x_2^2)$. (*By Thomas Barr)

Then f is differentiable at $(-6,1)$ and g is differentiable at $f(-6,1) = (37,-6)$, so by the chain rule

$$J(g\circ f)(-6,1) = Jg(f(-6,1))\,Jf(-6,1) = Jg(37,-6)\,Jf(-6,1).$$

Now

$$Jg(37,-6) = \left.\begin{pmatrix} 3x_2 & 3x_1 \\ 1 & -1 \\ 0 & 14x_2 \end{pmatrix}\right|_{(37,-6)} = \begin{pmatrix} -18 & 111 \\ 1 & -1 \\ 0 & -84 \end{pmatrix}$$

and

$$Jf(-6,1) = \left.\begin{pmatrix} 2x_1 & 2x_2 \\ x_2 & x_1 \end{pmatrix}\right|_{(-6,1)} = \begin{pmatrix} -12 & 2 \\ 1 & -6 \end{pmatrix}$$

so

$$J(g\circ f)(-6,1) = \begin{pmatrix} -18 & 111 \\ 1 & -1 \\ 0 & -84 \end{pmatrix}\begin{pmatrix} -12 & 2 \\ 1 & -6 \end{pmatrix} = \begin{pmatrix} 327 & 702 \\ -13 & 8 \\ -84 & 504 \end{pmatrix}$$

Ex. (By Thomas Bair) Let $f: \mathbb{R} \to \mathbb{R}^3$ and $g: \mathbb{R}^3 \to \mathbb{R}$. With $u = g(f(t))$, find a formula for $\frac{du}{dt}$.

Solution: $f(t) = (f_1(t), f_2(t), f_3(t))$ so $Jf(t) = \begin{pmatrix} f_1'(t) \\ f_2'(t) \\ f_3'(t) \end{pmatrix}$

and $Jg(f(t)) = \left(\frac{\partial g}{\partial x_1}(f(t)) \ \frac{\partial g}{\partial x_2}(f(t)) \ \frac{\partial g}{\partial x_3}(f(t)) \right)$.

$$\frac{du}{dt} \equiv Ju(t) = \left(\frac{\partial g}{\partial x_1}(f(t)) \ \frac{\partial g}{\partial x_2}(f(t)) \ \frac{\partial g}{\partial x_3}(f(t)) \right) \begin{pmatrix} f_1'(t) \\ f_2'(t) \\ f_3'(t) \end{pmatrix} =$$

$$= \left(\frac{\partial g}{\partial x_1}(f(t)) f_1'(t) + \frac{\partial g}{\partial x_2}(f(t)) f_2'(t) + \frac{\partial g}{\partial x_3}(f(t)) f_3'(t) \right)$$

Hence $\frac{du}{dt} = \frac{\partial g}{\partial x_1}(f(t)) f_1'(t) + \frac{\partial g}{\partial x_2}(f(t)) f_2'(t) + \frac{\partial g}{\partial x_3}(f(t)) f_3'(t)$

Since $x_1 = f_1(t)$, $x_2 = f_2(t)$, and $x_3 = f_3(t)$, this can be written more simply as

$$\frac{du}{dt} = \frac{\partial u}{\partial x_1} \frac{\partial x_1}{\partial t} + \frac{\partial u}{\partial x_2} \frac{\partial x_2}{\partial t} + \frac{\partial u}{\partial x_3} \frac{\partial x_3}{\partial t}.$$

Ex. Suppose $g(x_1, x_2, x_3) = 3x_1^2 + x_2^2 x_3^2$ and $f(t) = \left(\frac{1}{t}, t^3, \ln t \right)$

If $u(t) = g(f(t))$, then $\frac{\partial u}{\partial x_1} = \frac{\partial g}{\partial x_1} = 6x_1$, $\frac{\partial u}{\partial x_2} = \frac{\partial g}{\partial x_2} = 2x_2 x_3^2$,

and $\frac{\partial u}{\partial x_3} = \frac{\partial g}{\partial x_3} = 2x_2^2 x_3$. $f'(t) = \left(-\frac{1}{t^2}, 3t^2, \frac{1}{t} \right)$

$$\frac{du}{dt} = 6x_1 \frac{-1}{t} + 2x_2 x_3^2 \, 3t^2 + 2x_2^2 x_3 \frac{1}{t} \Bigg|_{(x_1, x_2, x_3) = \left(\frac{1}{t}, t^3, \ln t \right)}$$

More generally, if $f: \mathbb{R}^p \to \mathbb{R}^n$ and $g: \mathbb{R}^n \to \mathbb{R}^m$ we may define $u(t_1, t_2, \ldots, t_p) = g(f(t_1, t_2, \ldots, t_p))$. Observe that $u = (u_1, u_2, \ldots, u_m)$, because u is a vector-valued function $u: \mathbb{R}^p \to \mathbb{R}^m$.

What is $\frac{du_i}{dt_j}$? Well, $\frac{du_i}{dt_j}$ is the entry in the i^{th} row and j^{th} column of $Ju(t_1, t_2, \ldots, t_p)$. This entry is the dot product of the i^{th} row of $Jg(x_1, x_2, \ldots, x_n)\Big|_{(x_1, x_2, \ldots, x_n) = f(t_1, t_2, \ldots, t_p)}$ and the j^{th} column of $Jf(t_1, t_2, \ldots, t_p)$:

$$\frac{du_i}{dt_j} = \left(\frac{\partial g_i}{\partial x_1}, \frac{\partial g_i}{\partial x_2}, \ldots, \frac{\partial g_i}{\partial x_n}\right) \cdot \left(\frac{\partial f_1}{\partial t_j}, \frac{\partial f_2}{\partial t_j}, \ldots, \frac{\partial f_n}{\partial t_j}\right) =$$

$$= \frac{\partial g_i}{\partial x_1}\frac{\partial f_1}{\partial t_j} + \frac{\partial g_i}{\partial x_2}\frac{\partial f_2}{\partial t_j} + \ldots + \frac{\partial g_i}{\partial x_n}\frac{\partial f_n}{\partial t_j} \quad \text{or}$$

$$\frac{\partial u_i}{\partial x_1}\frac{\partial x_1}{\partial t_j} + \frac{\partial u_i}{\partial x_2}\frac{\partial x_2}{\partial t_j} + \ldots + \frac{\partial u_i}{\partial x_n}\frac{\partial x_n}{\partial t_j}.$$

Ex. Let $g(x,y,z) = (x^2 - z^3, 2x + y, e^{x+y+z})$ and $f(s,t) = (6s + \ln t, t^7, s - t)$. Define $u(s,t) = g(f(s,t))$

Then $$\frac{du_3}{dt} = \frac{\partial u_3}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u_3}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial u_3}{\partial z}\frac{\partial z}{\partial t} =$$

$$= e^{x+y+z}\frac{1}{t} + e^{x+y+z} \cdot 7t^6 + e^{x+y+z}(-1) =$$

$$= e^{x+y+z}\left(\frac{1}{t}+7t^6-1\right) = e^{6s+\ln t+t^7+s-t}\left(\frac{1}{t}+7t^6-1\right)$$

$$= e^{7s+t^7-t+\ln t}\left(\frac{1}{t}+7t^6-1\right)$$

$$\frac{du_2}{ds} = \frac{\partial u_2}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u_2}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u_2}{\partial z}\frac{\partial z}{\partial s} =$$

$$= 2\cdot 6 + 1\cdot 0 + 0\cdot 1 = 12$$

$$\frac{du_1}{ds} = \frac{\partial u_1}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u_1}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial u_1}{\partial z}\frac{\partial z}{\partial s} = 2x\cdot 6 + 0\cdot 0 - 3z^2\cdot 1 =$$

$$= 12x - 3z^2 = 12(6s+\ln t) - 3(s-t)^2$$

Product and Quotient rules

Product Rule: let $f:\mathbb{R}^n \to \mathbb{R}$ and $g:\mathbb{R}^n \to \mathbb{R}$ be differentiable at $a \in \mathbb{R}^n$ with derivatives $Df(a):\mathbb{R}^n \to \mathbb{R}$ and $Dg(a):\mathbb{R}^n \to \mathbb{R}$. Then $fg:\mathbb{R}^n \to \mathbb{R}$ is differentiable at a with derivative $Dfg(a):\mathbb{R}^n \to \mathbb{R}$ defined by $Dfg(a)(x) = Df(a)(x)\,g(x) +$
$+ f(x)\,Dg(a)(x)$

Proof: let $P:\mathbb{R}^2 \to \mathbb{R}$ be given by $P(x,y) = xy$. Then P is differentiable at any point (α,β) with derivative $DP(\alpha,\beta)(x,y)$ $= \beta x + \alpha y$ as you should verify.

Consider $H:\mathbb{R}^2 \to \mathbb{R}$ defined by $H(x) = P(f(x), g(x)) = f(x)g(x)$

Then $H = P\circ u$, where $u:\mathbb{R}^n \to \mathbb{R}^2$ is given by $u(x) = (f(x), g(x))$.

By the chain rule, $Dfg(a) = DH(a) = D(P\circ u)(a) = DP(u(a))\circ Du(a)$

Since $DP(u(a))(Du(a)(x)) = DP(f(a), g(a))(Df(a)(x), Dg(a)(x)) =$
$= g(a)Df(a)(x) + f(a)Dg(a)(x)$

In terms of the Jacobian matrices, $Jfg(a) = g(a)Jf(a) + f(a)Jg(a)$

Quotient Rule: Let $f, g: \mathbb{R}^n \to \mathbb{R}$ be differentiable at a.
If $g(a) \neq 0$, $\frac{f}{g}$ is differentiable at a with derivative

$$D\frac{f}{g}(a)(x) = \frac{g(a)Df(a)(x) - f(a)Dg(a)(x)}{[g(a)]^2}$$

Proof: Let $Q: \mathbb{R}^2 \cap \{(x,y): y \neq 0\} \to \mathbb{R}$ be defined by $Q(x,y) = \frac{x}{y}$.

Then $DQ(\alpha,\beta)(x,y) = \frac{1}{\beta}x - \frac{\alpha}{\beta^2}y = \frac{\beta x - \alpha y}{\beta^2}$

Let $u(x,y) = (f(x), g(x))$. and

$H: \mathbb{R}^n \to \mathbb{R}$ given by $H(x_1, x_2, \ldots, x_n) = H(x) = Q(f(x), g(x))$

By the chain rule, $D\frac{f}{g}(a)(x) = DH(a)(x) = D(Q \circ u)(a)(x) =$
$= DQ(u(a))(Du(a)(x)) = DQ(f(a), g(a))(D(f(a), g(a))(x)) =$

$$= \left.\frac{g(a)x - f(a)y}{[g(a)]^2}\right|_{(x,y) = (Df(a)(x),\, Dg(a)(x))} = \frac{g(a)Df(a)(x) - f(a)Dg(a)(x)}{[g(a)]^2}$$

Ex. let $f(x,y,z) = x^2+y^2+z^2$ and $g(x,y,z) = x^2+1$.

Compute $Dfg(a)$ and $D\frac{f}{g}(a)$

Solution: let $a=(a_1,a_2,a_3)$ then

$$Jf(a) = \begin{pmatrix} 2x & 2y & 2z \end{pmatrix}\Big|_{(x,y,z)=(a_1,a_2,a_3)} = \begin{pmatrix} 2a_1 & 2a_2 & 2a_3 \end{pmatrix}$$

Hence $Df(a)(x,y,z) = 2a_1x + 2a_2y + 2a_3z$.

$$Jg(a) = \begin{pmatrix} 2x & 0 & 0 \end{pmatrix}\Big|_{(x,y,z)=(a_1,a_2,a_3)} = (2a_1, 0, 0)$$

Hence $Dg(a)(x,y,z) = 2a_1x$.

It follows that $Dfg(a) = g(a_1,a_2,a_3)\,Df(a)(x,y,z) + f(a_1,a_2,a_3)\,Dg(a)(x,y,z)$

$= (a_1^2+1)(2a_1x+2a_2y+2a_3z) + (a_1^2+a_2^2+a_3^2)(2a_1x) =$

$= 2a_1\left[(a_1^2+1)+a_1^2+a_2^2+a_3^2\right]x + 2a_2(a_1^2+1)y + 2a_3(a_1^2+1)z$

$= 2a_1\left[2a_1^2+a_2^2+a_3^2+1\right]x + 2a_2\left[a_1^2+1\right]y + 2a_3\left[a_1^2+1\right]z.$

Similarly, $D\frac{f}{g}(a) = \dfrac{g(a)\,Df(a)(x) - f(a)\,Dg(a)(x)}{[g(a)]^2} =$

$$= \frac{(a_1^2+1)(2a_1x+2a_2y+2a_3z) - (a_1^2+a_2^2+a_3^2)(2a_1x)}{(a_1^2+1)^2}$$

Derivatives of expressions of the form $f(x) \times g(x)$ and $f(x) \cdot g(x)$

To be able to evoke the chain rule in this case, we'll need to develop the theory of multi-linear maps and their derivatives.

Def: Let $\mathbb{R}^n$ and $\mathbb{R}^m$ be any two Euclidian sets. We define the Cartesian product of $\mathbb{R}^n$ and $\mathbb{R}^m$, $\mathbb{R}^n \times \mathbb{R}^m$ to be the set $\{(x, y) : x \in \mathbb{R}^n, y \in \mathbb{R}^m\}$.

Remark: Do not confuse with cross product! The Cartesian product is an operation on two sets, while the cross product is an operation on two vectors.

Ex. $((1,-4),(1,7,1))$ is an element of $\mathbb{R}^2 \times \mathbb{R}^3$. $((1,2,2,4),1)$ is an element of $\mathbb{R}^4 \times \mathbb{R}$. What is $((0,0,1),(1,1,4))$ an element of?

The Cartesian product can be extended to an arbitrary collection of sets:

Ex. $(0,(1,2),(-3,1,1,4),5,(-5,18))$ is an element of $\mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}^4 \times \mathbb{R} \times \mathbb{R}^2$. What is $((3,4,1),(1,1),6)$ an element of?

You should observe that $\mathbb{R}^n \times \mathbb{R}^m \equiv \mathbb{R}^{n+m}$.

Ex. $((1,-4),(1,7,1)) \in \mathbb{R}^2 \times \mathbb{R}^3$ may be naturally associated with the vector $(1,-4,1,7,1) \in \mathbb{R}^{2+3} = \mathbb{R}^5$

More generally, $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_k} \equiv \mathbb{R}^{\sum_{i=1}^{k} n_i}$

Ex. $(0,(1,2),(-3,1,1,4),5,(-5,18)) \in \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}^4 \times \mathbb{R} \times \mathbb{R}^2$ may be naturally associated with the vector $(0,1,2,-3,1,1,4,5,-5,18) \in$ $\in \mathbb{R}^{1+2+4+1+2} = \mathbb{R}^{10}$

We are now ready to define multi-linear functions.

Def: Let $f: \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \ldots \times \mathbb{R}^{n_k} \to \mathbb{R}^m$. Then f is multi-linear if f is linear in each $\mathbb{R}^{n_i}$. That is, if $(a_1, \ldots, a_i + b_i, \ldots, a_k) \in \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_k}$, then $f(a_1, \ldots, a_i + b_i, \ldots, a_k) = f(a_1, \ldots, a_i, \ldots, a_k) +$ $+ f(a_1, \ldots, b_i, \ldots, a_k)$ and if $(a_1, \ldots, ca_i, \ldots, a_k) \in \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_k}$ where $c \in \mathbb{R}$ $f(a_1, \ldots, ca_i, \ldots, a_k) = c f(a_1, \ldots, a_i, \ldots, a_k)$.

In these notes, we'll restrict our attention to multi-linear functions of the form $f: \mathbb{R}^n \times \mathbb{R}^n \times \ldots \times \mathbb{R}^n \to \mathbb{R}^m$.

Ex. Let $f: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be defined by $f(x,y) = x \cdot y$.
For example, if $n=4$, $f((1,0,0,1),(2,3,4,0)) = (1,0,0,1) \cdot (2,3,4,0) = 2$.
Then f is multi-linear (bi-linear in this case):

$f(x+z, y) = f(x,y) + f(z,y)$ because $f(x+z, y) = (x+z) \cdot y =$ $= x \cdot y + z \cdot y$. Similarly, $f(x, y+z) = f(x,y) + f(x,z)$ for $z \in \mathbb{R}^n$

you should verify that for $c \in \mathbb{R}$, $f(cx,y) = f(x,cy) = c f(x,y)$.

Ex. Let $f : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be given by $f(x,y,z) = \det\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. you should verify that f is tri-linear.

Ex. Let $f : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ be given by $f(x,y) = x \times y$.
then $f(x+z,y) = (x*z) \times y = x \times y + z \times y$ and $f(x,y+z) = f(x,y) + f(x,z)$ (You should finish the verification yourself).

Just as a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is completely determined by its action on the coordinate vectors e_i, a multi-linear map $f : \mathbb{R}^n_1 \times \dots \times \mathbb{R}^n_k \to \mathbb{R}^m$ is determined by its action on $(e_{i_1}, e_{i_2}, \dots, e_{i_k}) \in \mathbb{R}^n_1 \times \dots \times \mathbb{R}^n_k$ $i_1, i_2, \dots, i_k \in \{1, \dots, n\}$ Before we prove this, let's see what is meant with the following example:

Ex. let $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = \det\begin{pmatrix} x \\ y \end{pmatrix}$.

Then $f(x,y) = f((x_1,x_2),(y_1,y_2)) = \det\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} =$

$= \det\begin{pmatrix} x_1 e_1 + x_2 e_2 \\ y_1 e_1 + y_2 e_2 \end{pmatrix} = \det\begin{pmatrix} x_1 e_1 \\ y_1 e_1 + y_2 e_2 \end{pmatrix} + \det\begin{pmatrix} x_2 e_2 \\ y_1 e_1 + y_2 e_2 \end{pmatrix} =$

$= x_1 y_1 \det\begin{pmatrix} e_1 \\ e_1 \end{pmatrix} + x_1 y_2 \det\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} + x_2 y_1 \det\begin{pmatrix} e_2 \\ e_1 \end{pmatrix} + x_2 y_2 \det\begin{pmatrix} e_2 \\ e_2 \end{pmatrix}$

$= x_1 y_1 f(e_1,e_1) + x_1 y_2 f(e_1,e_2) + x_2 y_1 f(e_2,e_1) + x_2 y_2 f(e_2,e_2)$

Thm: Let $f: \mathbb{R}^n_1 \times \dots \times \mathbb{R}^n_k \to \mathbb{R}^m$ be multi-linear. Then

$$f(x_1, x_2, \dots, x_k) = f((x_{11}, x_{21}, \dots, x_{n1}), (x_{12}, x_{22}, \dots, x_{n2}), \dots, (x_{1k}, x_{2k}, \dots, x_{nk}))$$

$$= f\left(\left(\sum_{i_1=1}^{n} x_{i_1 1} e_{i_1}\right), \left(\sum_{i_2=1}^{n} x_{i_2 2} e_{i_2}\right), \dots \left(\sum_{i_k=1}^{n} x_{i_k k} e_{i_k}\right)\right) =$$

$$= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \dots \sum_{i_k=1}^{n} x_{i_1 1} x_{i_2 2} \dots x_{i_k k} f(e_{i_1}, e_{i_2}, \dots, e_{i_k})$$

$$= \sum_{i_1, i_2, \dots, i_k=1}^{n} x_{i_1 1} x_{i_2 2} \dots x_{i_k k} f(e_{i_1}, e_{i_2}, \dots, e_{i_k})$$

proof: Apply the linearity to each (vector) coordinate of

f: $$f(x_1, x_2, \dots, x_k) = f\left(\sum_{i_1=1}^{n} x_{i_1 1} e_{i_1}, \sum_{i_2=1}^{n} x_{i_2 2} e_{i_2}, \dots, \sum_{i_k=1}^{n} x_{i_k k} e_{i_k}\right)$$

$$= \sum_{i_1=1}^{n} f\left(x_{i_1 1} e_{i_1}, \sum_{i_2=1}^{n} x_{i_2 2} e_{i_2}, \dots, \sum_{i_k=1}^{n} x_{i_k k} e_{i_k}\right) =$$

$$= \sum_{i_1=1}^{n} x_{i_1 1} f\left(e_{i_1}, \sum_{i_2=1}^{n} x_{i_2 2} e_{i_2}, \dots, \sum_{i_k=1}^{n} x_{i_k k} e_{i_k}\right) =$$

$$= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} x_{i_1 1} x_{i_2 2} f\left(e_{i_1}, e_{i_2}, \dots, \sum_{i_k=1}^{n} x_{i_k k} e_{i_k}\right) = \dots =$$

$$= \sum_{i_1=1}^{n} \sum_{i_2, \dots, i_k=1}^{n} x_{i_1 1} x_{i_2 2} \dots x_{i_k k} f(e_{i_1}, e_{i_2}, \dots, e_{i_k})$$

which proves the desired result.

Recall that $\mathbb{R}_1^n \times \ldots \times \mathbb{R}_k^n \equiv \mathbb{R}^{kn}$. This allows us to define a norm on $\mathbb{R}_1^n \times \ldots \times \mathbb{R}_k^n$ in a natural way:

Def: Let $(x_1, x_2, \ldots, x_k) \in \mathbb{R}_1^n \times \ldots \times \mathbb{R}_k^n$. Then $\|(x_1, x_2, \ldots, x_k)\| = \sqrt{\|x_1\|^2 + \|x_2\|^2 + \ldots + \|x_k\|^2}$

Ex. $\|((1,2),(3,3),(0,1))\| = \sqrt{\|(1,2)\|^2 + \|(3,3)\|^2 + \|(0,1)\|^2} =$

$= \sqrt{1^2+2^2+3^2+3^2+0^2+1^2} = \|(1,2,3,3,0,1)\|$

We can now prove a very useful lemma:

Lemma: Let $f : \mathbb{R}_1^n \times \ldots \times \mathbb{R}_k^n \to \mathbb{R}^m$ be multi-linear. Then there exists a scalar $C \in \mathbb{R}$ s.t. $\|f(x)\| \leq C\|x\|^k$ where x is any element of $\mathbb{R}_1^n \times \ldots \times \mathbb{R}_k^n$.

Proof: $\|f(x)\| = \|f(x_1, x_2, \ldots, x_k)\| = \left\| \sum_{i_1, i_2, \ldots, i_k = 1}^{n} x_{i_1 1} x_{i_2 2} \ldots x_{i_k k} f(e_{i_1}, \ldots, e_{i_k}) \right\|$

$$\leq \sum_{i_1, i_2, \ldots, i_k = 1}^{n} |x_{i_1 1} x_{i_2 2} \ldots x_{i_k k}| \, \|f(e_{i_1}, \ldots, e_{i_k})\| \qquad (1)$$

Each $|x_{i_j j}| \leq \|x_j\| \leq \|x\|$ so

$$(1) \leq \sum_{i_1, i_2, \ldots, i_k = 1}^{n} \|x\|^k \|f(e_{i_1}, \ldots, e_{i_k})\| = \left(\sum_{i_1, i_2, \ldots, i_k = 1}^{n} \|f(e_{i_1}, \ldots, e_{i_k})\| \right) \|x\|^k$$

$= C\|x\|^k$ where $C = \sum_{i_1, i_2, \ldots, i_k = 1}^{n} \|f(e_{i_1}, \ldots, e_{i_k})\|$

The lemma above is an important step in the proof and caracterization of derivatives of multilinear maps. Let's try to see why with the help of a simple example:

Ex. Let $f: \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be defined by $f(x,y) = x \cdot y$.

What might be $Df(a,b)(u)$? (We can think of $Df(a,b)$ as a linear map $T: \mathbb{R}^{3+3} \to \mathbb{R}$)

Solution: $Df(a,b) = T$ must satisfy the limit

$$\lim_{u \to 0} \frac{|f(\alpha+u) - f(\alpha) - T(u)|}{\|u\|} = 0$$ where α is the natural vector in $\mathbb{R}^{3+3} = \mathbb{R}^6$ corresponding to $(a,b) \in \mathbb{R}^3 \times \mathbb{R}^3$.

We may write the above limit as

$$\lim_{(u_1,u_2) \to (0,0)} \frac{|f(a+u_1, b+u_2) - f(a,b) - T(u_1,u_2)|}{\|(u_1,u_2)\|}$$ where

$(u_1,u_2) \in \mathbb{R}^3 \times \mathbb{R}^3$ is the natural element that corresponds to $u \in \mathbb{R}^6$.

Because f is bi-linear, $f(a+u_1, b+u_2) = f(a,b) + f(u_1,b) + f(a,u_2) + f(u_1,u_2)$ and $f(a+u_1,b+u_2) - f(a,b) = f(u_1,b) + f(a,u_2) + f(u_1,u_2)$

If we set $T(u_1,u_2) = f(u_1,b) + f(a,u_2)$, we will have

$$\lim_{(u_1,u_2) \to (0,0)} \frac{|f(a+u_1, b+u_2) - f(a,b) - T(u_1,u_2)|}{\|(u_1,u_2)\|} = \lim_{(u_1,u_2) \to (0,0)} \frac{|f(u_1,u_2)|}{\|(u_1,u_2)\|} \leq$$

$$\leq \lim_{(u_1,u_2) \to (0,0)} \frac{C\|(u_1,u_2)\|^2}{\|(u_1,u_2)\|} = \lim_{(u_1,u_2) \to (0,0)} \|(u_1,u_2)\| = 0$$

Note that $T(u_1,u_2)$ is linear: $T((u_1,u_2)+(v_1,v_2)) =$

$= T(u_1+v_1, u_2+v_2) = f(u_1+v_1, b) + f(a, u_2+v_2) =$

$= f(u_1,b) + f(v_1,b) + f(a,u_2) + f(a,v_2) = T(u_1,u_2) + T(v_1,v_2)$

Similarly, $T(c(u_1,u_2)) = c\,T(u_1,u_2)$ as you should verify.

Thm: Let $f: \mathbb{R}^n_1 \times \dots \times \mathbb{R}^n_k \to \mathbb{R}^m$ be multi-linear. Then f is differentiable at any point $a = (a_1, a_2, \dots, a_k) \in \mathbb{R}^n_1 \times \dots \times \mathbb{R}^n_k$ with derivative $Df(a)(x) = Df(a_1,\dots,a_k)(x_1,\dots,x_k) =$

$= \sum_{i=1}^{k} f(a_1,\dots,x_i,\dots,a_k) = T(x) = \sum_{i=1}^{k} f_i$

Proof: We wish to show that $\lim_{(u_1,\dots,u_k)\to(0,\dots,0)} \frac{\| f(a_1+u_1,\dots,a_k+u_k) - f(a_1,\dots,a_k) - \sum_{i=1}^{k} f_i}{\|(u_1,\dots,u_k)\|}$

$= 0$

Observe that $f(a_1+u_1,\dots,a_k+u_k) - f(a_1,\dots,a_k) - \sum_{i=1}^{k} f(a_1,\dots,x_i,\dots,a_k) =$

$= \sum_{i_1<i_2} f(a_1,\dots,u_{i_1},\dots,u_{i_2},\dots a_k) + \sum_{i_1<i_2<i_3} f(a_1,\dots,u_{i_1},\dots,u_{i_2},\dots,u_{i_3},\dots,a_k) +$

$+ \dots + f(u_1,u_2,\dots,u_k)$.

That is, the sum above consists of sums of bilinear maps of the form $f(a_1,\dots,u_{i_1},\dots,u_{i_2},\dots,a_k) = g(u_{i_1},u_{i_2})$, sums of tri-linear maps, etc.

It follows, from triangle inequality and the lemma, that

$$\|f(a_1+u_1,\dots,a_k+u_k)-f(a_1,\dots,a_k)-\sum_{i=1}^{n} f(a_1,\dots,u_i,\dots,a_k)\| \leq$$

$$\leq \sum_{i_1<i_2}\|f(a_1,\dots,u_{i_1},\dots,u_{i_2},\dots,a_k)\| + \sum_{i_1<i_2<i_3}\|f(a_1,\dots,u_{i_1},\dots,u_{i_2},\dots,u_{i_3},\dots,a_k)\| +$$

$$+\dots+\|f(u_1,u_2,\dots,u_k)\| \leq \sum_{i_1<i_2} C_{i_1i_2}\|(u_1,\dots,u_k)\|^2 + \sum_{i_1<i_2<i_3} C_{i_1i_2i_3}\|(u_1,\dots,u_k)\|^3 +$$

$$+\dots+C_k\|(u_1,u_2,\dots,u_k)\|^k$$

So

$$\lim_{(u_1,u_2,\dots,u_k)\to(0,0,0,\dots,0)} \frac{\|f(a_1+u_1,\dots,a_k+u_k)-f(a_1,\dots,a_k)-\sum_{i=1}^{n} f(a_1,\dots,u_i,\dots,a_k)\|}{\|(u_1,\dots,u_k)\|} \leq$$

$$\leq \lim_{(u_1,\dots,u_k)\to(0,\dots,0)} \frac{\sum_{i_1<i_2} C_{i_1i_2}\|(u_1,\dots,u_k)\|^2 + \sum_{i_1<i_2<i_3} C_{i_1i_2i_3}\|(u_1,\dots,u_k)\|^3 + \dots + C_k\|(u_1,\dots,u_k)\|^k}{\|(u_1,\dots,u_k)\|}$$

$$= \lim_{(u_1,\dots,u_k)\to(0,0,\dots,0)} \sum_{i_1<i_2} C_{i_1i_2}\|(u_1,\dots,u_k)\| + \sum_{i_1<i_2<i_3} C_{i_1i_2i_3}\|u_1,\dots,u_k\|^2 +$$

$$+\dots+C_k\|(u_1,\dots,u_k)\|^{k-1} = 0$$ which proves the desired result.

Ex. Let $f:\mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}^3$ be given by $f(x,y)=x\times y$.

Then $Df(a,b)(x,y) = x\times b + a\times y$

Ex. Let $g:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ be given by $g(x,y)=x\cdot y$, then

$Dg(a,b)(x,y) = b\cdot x + a\cdot y$

Ex. Let $h: \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ be given by $h(x,y,z) = \det\begin{pmatrix} x \\ y \\ z \end{pmatrix}$.

Then $Dh(a,b,c)(x,y,z) = D(\det)(a,b,c)(x,y,z) =$

$$= \det\begin{pmatrix} x \\ b \\ c \end{pmatrix} + \det\begin{pmatrix} a \\ y \\ c \end{pmatrix} + \det\begin{pmatrix} a \\ b \\ z \end{pmatrix}$$

Now we can apply the chain rule to figure the derivatives of expressions of the form $f(x) \cdot g(x)$ and $f(x) \times g(x)$ for suitable functions f and g.

Thm: Let $f, g: \mathbb{R}^n \to \mathbb{R}^3$. Then $D(f \times g)(a)(x) = Df(a)(x) \times g(a) + f(a) \times Dg(a)(x)$.

Proof: Let $h: \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $h(x,y) = x \times y$.

Then $D(f \times g)(a)(x) = Dh(f(x), g(x))\Big|_{x=a} = Dh(f(a), g(a))(Df(a)(x), Dg(a)(x))$

$= Df(a)(x) \times g(a) + f(a) \times Dg(a)(x)$.

Thm: Let $f, g: \mathbb{R}^n \to \mathbb{R}^m$. Then $D(f \cdot g)(a)(x) = Df(a)(x) \cdot g(a) + f(a) \cdot Dg(a)(x)$

Proof: The proof is analogous to the one given in the previous theorem.

Implicit differentiation

Consider the equation $e^x y + 2e^y z - e^z = 0$. This equation is of the form $f(x,y,z)=0$ where f is a differentiable function of x, y, and z. Suppose that the equation defines $z = z(x,y)$ as a differentiable function of x and y. Then, to emphasize the dependence, we write $f(x,y,z(x,y)) = 0$.

Using the chain rule, we can obtain the derivative $\frac{\partial z}{\partial x}$ as follows:

$\frac{\partial}{\partial x}\left(f(x,y,z(x,y))\right) = \frac{\partial}{\partial x}(0) = 0$ Hence

$$\frac{\partial f}{\partial x}(x,y,z(x,y)) \frac{\partial x}{\partial x} + \frac{\partial f}{\partial y}(x,y,z(x,y))\frac{\partial y}{\partial x} + \frac{\partial f}{\partial z}(x,y,z(x,y))\frac{\partial z}{\partial x} = 0$$

or, since $\frac{\partial x}{\partial x} = 1$ and $\frac{\partial y}{\partial x} = 0$,

$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 0$ which implies that $\frac{\partial z}{\partial x} = \dfrac{-\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial z}}$

In our particular case, this means that $\frac{\partial z}{\partial x} = \dfrac{-e^x y}{2e^y - e^z}$

Similarly, you could show that $\frac{\partial z}{\partial y} = \dfrac{-\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial z}} = \dfrac{-e^x - 2e^y z}{2e^y - e^z}$

Ex. Consider the function $u = xy^2z$, where (x,y,z) is constrained to lie on the sphere S with equation $x^2+y^2+z^2=1$. Find $\left.\frac{\partial u}{\partial x}\right|_S$ and $\left.\frac{\partial u}{\partial y}\right|_S$

Solution: The expression $x^2+y^2+z^2=1$ defines z as an implicit function $z(x,y)$. Therefore $u=xy^2z(x,y)$.

$$\frac{\partial u}{\partial x}\Big|_S = y^2\frac{\partial}{\partial x}\left(x\,z(x,y)\right) = y^2\left(z(x,y)+\frac{\partial z}{\partial x}\right) = y^2\left(z+\frac{\partial z}{\partial x}\right)$$

Now to solve for $\frac{\partial z}{\partial x}$, we differentiate implicitly the expression $x^2+y^2+z^2=1$ with respect to x:

$$\frac{\partial}{\partial x}\left(x^2+y^2+z^2\right) = \frac{\partial}{\partial x}(1) = 0$$

$$2x\frac{\partial x}{\partial x}+2y\frac{\partial y}{\partial x}+2z\frac{\partial z}{\partial x} = 0 \quad \text{or} \quad 2x+2z\frac{\partial z}{\partial x} = 0 \text{ which implies}$$

that $\frac{\partial z}{\partial x} = \frac{-x}{z}$

Hence $\frac{\partial u}{\partial x}\Big|_S = y^2\left(z+\frac{-x}{z}\right) = y^2\,\frac{z^2-x}{z} = \frac{y^2z^2-xy^2}{z}$

To find $\frac{\partial u}{\partial y}\Big|_S$, we follow the same procedure:

$$\frac{\partial u}{\partial y}\Big|_S = \frac{\partial}{\partial y}\left(xy^2z(x,y)\right) = x\frac{\partial}{\partial y}\left(y^2z(x,y)\right) = x\left(2yz+y^2\frac{\partial z}{\partial y}\right)$$

$$\frac{\partial}{\partial y}\left(x^2+y^2+z^2\right) = 0$$

$$2x\frac{\partial x}{\partial y}+2y\frac{\partial y}{\partial y}+2z\frac{\partial z}{\partial y} = 2y+2z\frac{\partial z}{\partial y} = 0 \quad \text{or} \quad \frac{\partial z}{\partial y} = \frac{-y}{z}$$

Hence $\frac{\partial u}{\partial y}\Big|_S = x\left(2yz+y^2\left(\frac{-y}{z}\right)\right) = x\left(\frac{2yz^2-y^3}{z}\right) = \frac{2xyz^2-xy^3}{z}$

The example above illustrates the need for some assurance that a given equation actually does implicitly define one of the variables as a function of the others near a solution to the equation. That assurance comes from the implicit function theorem of advanced calculus. One form of that theorem is stated below:

Thm: (Implicit Function theorem): Suppose that $f: U \subset \mathbb{R}^n \to \mathbb{R}$ has continuous partial derivatives in an open ball centered at a point $a \in U$. For any i in the range 1 to n such that $\frac{\partial f}{\partial x_i}(a) \neq 0$, the equation

$$f(x_1, x_2, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) = 0$$

implicitly defines x_i as a differentiable function of the other variables that is defined on some open ball surrounding a. Moreover,

$$\frac{\partial x_i}{\partial x_j} = -\frac{\partial f/\partial x_j}{\partial f/\partial x_i} \quad \text{for } i \neq j.$$