Chapter 6

differentiation

In this chapter we define differentiation of mappings. I follow Edwards fairly closely, his approach is efficient and his langauge clarifies concepts which are easily confused. Susan Colley's text *Vector Calculus* is another good introductory text which describes much of the mathematics in this chapter. When I teach calculus III I do touch on the main thrust of this chapter but I shy away from proofs and real use of linear algebra. That is not the case here.

6.1 derivatives and differentials

In this section we motivate the general definition of the derivative for mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$. Naturally this definition must somehow encompass the differentiation concepts we've already discussed in the calculus sequence: let's recall a few examples to set the stage,

1. derivatives of functions of $\mathbb{R}$, for example $f(x) = x^2$ has $f'(x) = 2x$
2. derivatives of mappings of $\mathbb{R}$, for example $f(t) = (t, t^2, t^3)$ has $f'(t) =< 1, 2t, 3t^2 >$.
3. $f : dom(f) \subseteq \mathbb{R}^2 \to \mathbb{R}$ has directional derivative $(D_u f)(p) = (\nabla f)(p) \cdot u$ where $\nabla f = grad(f) =< \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} >$
4. $X : U \subset \mathbb{R}^2_{uv} \to \mathbb{R}^3_{xyz}$ parametrizes a surface $X(U)$ and $N(u, v) = \frac{\partial X}{\partial u} \times \frac{\partial X}{\partial v}$ gives the normal vector field to the surface.

We'd like to understand how these derivatives may be connected in some larger context. If we could find such a setting then that gives us a way to state theorems about derivatives in an efficient and general manner. We also should hope to gain a deeper insight into the geometry of differentiation.

6.1.1 derivatives of functions of a real variable

Let's revisit the start of Calculus I. We begin by defining the change in a function f between the point a and $a + h$:

$$\Delta f = f(a + h) - f(a).$$

We can approximate this change for small values of h by replacing the function with a line. Recall that the line closest to the function at that point is the **tangent line** which has slope $f'(a)$ which we define below.

Definition 6.1.1.

Suppose $f : U \subseteq \mathbb{R} \to \mathbb{R}$ then we say that f has **derivative** $f'(a)$ defined by the limit below (if the limit exists, otherwise we say f is not differentiable at a)

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

If f has a derivative at a then it also has a **differential** $df_a : \mathbb{R} \to \mathbb{R}$ at a which is a function defined by $df_a(h) = hf'(a)$. Finally, if f has derivative $f'(a)$ at a then the tangent line to the curve has equation $y = f(a) + f'(a)(x - a)$.

Notice that the derivative at a point is a number whereas the differential at a point is a linear map[1]. Also, the tangent line is a "*paralell translate*" of the line through the origin with slope $f'(a)$.

Example 6.1.2. . .

Definition 6.1.3.

Suppose $f : U \subseteq \mathbb{R} \to \mathbb{R}$ and suppose $f'(v)$ exists for each $v \in V \subset U$. We say that f has **derivative** $f' : V \subseteq \mathbb{R} \to \mathbb{R}$ defined by

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

for each $x \in V$.

In words, the derivative function is defined pointwise by the derivative at a point.

[1]We will maintain a similar distinction in the higher dimensional cases so I want to draw your attention to the distinction in terminology from the outset.

Proposition 6.1.4.

Suppose $a \in dom(f)$ where $f : dom(f) \subseteq \mathbb{R} \to \mathbb{R}$ and $a \in dom(f')$ then df_a is a linear transformation from $\mathbb{R}$ to $\mathbb{R}$.

Proof: Let $c, h, k \in \mathbb{R}$ and $a \in dom(f')$ which simply means $f'(a)$ is well-defined. Note that:

$$df_a(ch + k) = (ch + k)f'(a) = chf'(a) + kf'(a) = cdf_a(h) + df_a(k)$$

for all c, h, k thus df_a is linear transformation. □

The **differential** is likewise defined to be the **differential form** $df : dom(f) \to L(\mathbb{R}, \mathbb{R}) = \mathbb{R}^*$ where $df(a) = df_a$ and df_a is a linear function from $\mathbb{R}$ to $\mathbb{R}$. We'll study differential forms in more depth in a later section.

6.1.2 derivatives of vector-valued functions of a real variable

A **vector-valued function of a real variable** is a mapping from a subset of $\mathbb{R}$ to some subset $\mathbb{R}^n$. In this section we discuss how to differentiate such functions as well as a few interesting theorems which are known for the various vector products.

We can revisit the start of Calculus III. We begin by defining the change in a vector-valued function f between the inputs a and $a + h$:

$$\Delta f = f(a + h) - f(a).$$

This is a vector. We can approximate this change for small values of h by replacing the space curve $a \mapsto f(a)$ with a line $t \mapsto f(a) + tf'(a)$ in $\mathbb{R}^n$. The direction vector of the tangent line is $f'(a)$ which we define below.

Definition 6.1.5.

Suppose $f : U \subseteq \mathbb{R} \to \mathbb{R}$ then we say that f has **derivative** $f'(a)$ defined by the limit below (if the limit exists, otherwise we say f is not differentiable at a)

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

We define f' to be the function defined pointwise by the limit above for all such values as the limit converges. If f has a derivative at a then it also has a **differential** $df_a : \mathbb{R} \to \mathbb{R}^n$ at a which is a mapping defined by $df_a(h) = hf'(a)$. The vector-valued-differential form df is defined pointwise by $df(a) = df_a$ for all $a \in dom(f')$.

The tangent line is a "*paralell translate*" of the line through the origin with direction-vector $f'(a)$. In particular, if f has a derivative of $f'(a)$ at a then the tangent line to the curve has parametric equation $\vec{r}(t) = f(a) + tf'(a)$.

Proposition 6.1.6.

Suppose $a \in dom(f)$ where $f : dom(f) \subseteq \mathbb{R} \to \mathbb{R}^n$ and $a \in dom(f')$ then the differential df_a is a linear transformation from $\mathbb{R}$ to $\mathbb{R}^n$.

The proof is almost identical to the proof for real-valued functions of a real variable. Note:

$$df_a(ch + k) = (ch + k)f'(a) = chf'(a) + kf'(a) = cdf_a(h) + df_a(k)$$

for all $h, k, c \in \mathbb{R}$ hence df_a is a linear transformation.

6.1.3 directional derivatives

Let $m \geq n$, the image of a injective continuous mapping $F : dom(F) \subseteq \mathbb{R}^n \to \mathbb{R}^m$ gives an n-dimensional continuous surface in $\mathbb{R}^m$ provided the mapping F satisfy the topological requirement $dom(F) \approx \mathbb{R}^n$. This topological fine print is just a way to avoid certain pathological cases like space filling curves. We proved in Example 3.4.7 that the unit-sphere is a continuous surface. The proof that the sphere of radius 2 is a continuous surface is similar. In the example that follows we'll see how curves on the surface provide a definition for the tangent plane.

Example 6.1.7. *The sphere of radius 2 centered at the origin has equation $x^2 + y^2 + z^2 = 4$. We can view the top-half of the sphere as the image of the mapping $F : \mathbb{R}^2 \to \mathbb{R}^3$ where*

$$F(x, y) = \left(x,\ y,\ \sqrt{4 - x^2 - y^2}\ \right).$$

The tangent plane to the sphere at some point on the sphere can be defined as the set of all tangent vectors to curves on the sphere which pass through the point: let S be the sphere and $p \in S$ then the tangent space to p is intuitively defines as follows:

$$T_pS = \{\gamma'(0) \mid \gamma : \mathbb{R} \to S,\ \text{a smooth curve with } \gamma(0) = p\}$$

A line in the direction of $< a, b >$ through $(1, 1)$ in $\mathbb{R}^2$ has parametric representation $\vec{r}(t) = (1 + at, 1 + bt)$. We can construct curves on the sphere that pass through $F(1, 1) = (1, 1, \sqrt{2})$ by simply mapping the lines in the plane to curves on the sphere; $\gamma(t) = F(\vec{r}(t))$ which gives

$$\gamma(t) = \left(\ 1 + at,\ 1 + bt,\ \sqrt{4 - (1 + at)^2 - (1 + bt)^2}\ \right)$$

Now, not all curves through p have the same form as $\gamma(t)$ above but it is fairly clear that if we allow (a, b) to trace out all possible directions in $\mathbb{R}^2$ then we should cover T_pS. A short calculation reveals that

$$\gamma'(0) =< a, b, \tfrac{-1}{\sqrt{2}}(a + b) >$$

These are vectors we should envision as attached to the point $(1, 1, \sqrt{2})$. A generic point in the tangent plane to the point should have the form $p + \gamma'(0)$. This gives equations:

$$x = 1 + a, \qquad y = 1 + b, \qquad z = \sqrt{2} - \tfrac{1}{\sqrt{2}}(a + b)$$

we can find the Cartesian equation for the plane by eliminating a, b

$$a = x - 1, \qquad b = y - 1 \;\Rightarrow\; z = \sqrt{2} - \tfrac{1}{\sqrt{2}}(x + y - 2) \;\Rightarrow\; x + y + \sqrt{2}z = 4.$$

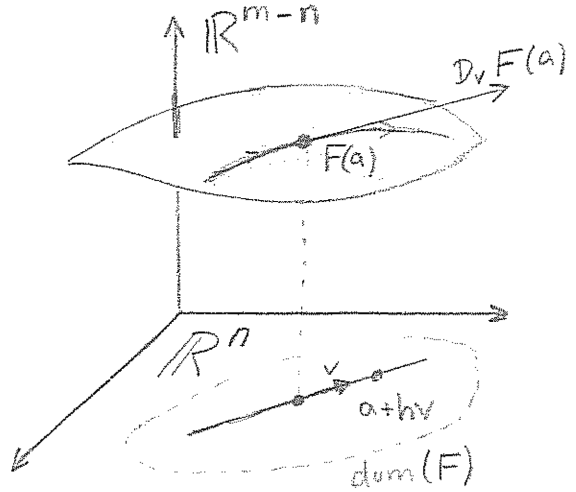
We find the tangent plane to the sphere $x^2 + y^2 + z^2 = 4$ has normal $< 1, 1, \sqrt{2} >$ at the point $(1, 1, \sqrt{2})$.

Of course there are easier ways to calculate the equation for a tangent plane. The directional derivative of a mapping F at a point $a \in dom(F)$ along v is defined to be the derivative of the curve $\gamma(t) = F(a + tv)$. In other words, the directional derivative gives you the instantaneous vector-rate of change in the mapping F at the point a along v. In the case that $m = 1$ then $F : dom(F) \subseteq \mathbb{R}^n \to \mathbb{R}$ and the directional derivative gives the instantaneous rate of change of the function F at the point a along v. You probably insisted that $||v|| = 1$ in calculus III but we make no such demand here. We define the directional derivative for mappings and vectors of non-unit length.

Definition 6.1.8.

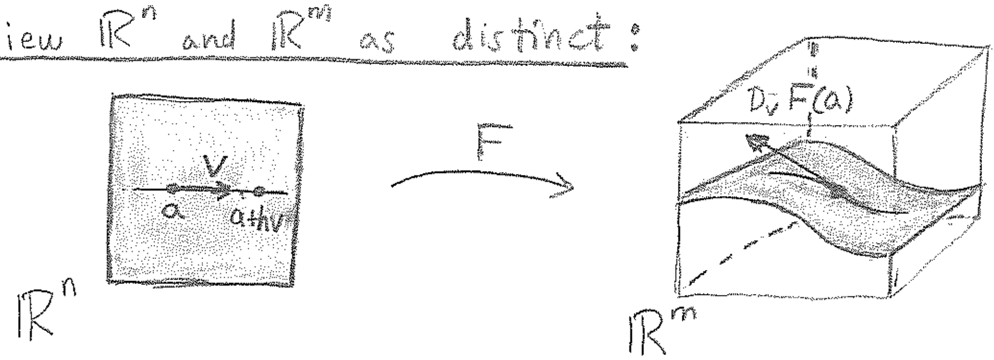
Let $F : dom(F) \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and suppose the limit below exists for $a \in dom(F)$ and $v \in \mathbb{R}^n$ then we define the **directional derivative of F at a along v** to be $D_vF(a) \in \mathbb{R}^m$ where

$$D_vF(a) = \lim_{h \to 0} \frac{F(a + hv) - F(a)}{h}$$



you can picture $\mathbb{R}^n \subset \mathbb{R}^m$ when $m > n$. Typically we identify $\mathbb{R}^n$ with $\mathbb{R}^n \times \{(0,0,\dots,0)\} \subset \mathbb{R}^m$ ($m - n$ zeros.)

Or, view $\mathbb{R}^n$ and $\mathbb{R}^m$ as distinct:



The directional derivative $D_vF(a)$ is homogenous in v.

Proposition 6.1.9.

Let $F : dom(F) \subseteq \mathbb{R}^n \to \mathbb{R}^m$ then if $D_vF(a)$ exists in $\mathbb{R}^m$ then $D_{cv}F(a) = cD_vF(a)$

See Edwards pg. 66 the proof is not hard. Let $F : U \to \mathbb{R}^m$ define a continuous surface S with dimension n. The tangent space of S at $p \in S$ should be the paralell translate of a n-dimensional subspace of $\mathbb{R}^m$. Moreover, we would like for the tangent space at a point $p \in S$ to be very close to the surface near that point. The change of F near $p = F(a)$ along the curve $\gamma(t) = F(a+tv)$ is given by

$$\Delta F = F(a+hv) - F(a).$$

It follows that $F(a+hv) \approxeq F(a) + hD_vF(a)$ for $h \approxeq 0$. We'd like for the the set of all directional derivatives at p to form a subspace of $\mathbb{R}^m$. Recall(or learn) that in linear algebra we learn that every subspaces of $\mathbb{R}^m$ is the range of some linear operator[2] This means that if $D_vF(a)$ was a linear operator with respect to v then we would know the set of all directional derivatives formed a subspace of $\mathbb{R}^m$. Note that directional derivative almost gives us linearity since its homogeneous but we also need the condition of additivity:

$$D_{v+w}F(a) = D_vF(a) + D_wF(a) \qquad \text{additivity of directional derivative}$$

This condition is familar. Recall that Propositions 6.1.4 and 6.1.6 showed the *differential* df_a was linear for $f : dom(f) \subseteq \mathbb{R} \to \mathbb{R}^m$. In fact the differential is the directional derivative in these special cases if we let $v = 1$; $D_1F(a) = dF_a(1)$ for $F : dom(F) \subseteq \mathbb{R} \to \mathbb{R}^m$ where $a \in dom(F')$. So we have already proved the directional derivative is linear in those special cases. Fortunately it's not so simple for a general mapping. We have to make an additional assumption if we wish for the tangent space to be well-defined.

Definition 6.1.10.

Suppose that U is open and $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a mapping the we say that F is **differentiable** at $a \in U$ iff there exists a linear mapping $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

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

In such a case we call the linear mapping L the **differential at** a and we denote $L = dF_a$. The matrix of the differential is called the **derivative of** F **at** a and we denote $[dF_a] = F'(a) \in \mathbb{R}^{\,m \times n}$ which means that $dF_a(v) = F'(a)v$ for all $v \in \mathbb{R}^n$.

[2]don't believe it? Let $W \leq \mathbb{R}^m$ and choose a basis $\beta = \{f_1, \ldots, f_n\}$ for W. You can verify that $L(v) = [f_1|f_2|\cdots|f_n|f_n|\cdots|f_n]v$ defines a linear transformation with $range(L) = Col[\beta] = W$.

The preceding definition goes hand in hand with the definition of the tangent space given below.

Definition 6.1.11.

Suppose that $U \approx \mathbb{R}^n$ is open and $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a mapping which is differentiable on U. If $rank(F'(a)) = n$ at each $a \in U$ then we say that $F(U)$ is a **differentiable surface of dimension** n. Also, a map such as F is said to be **regular**. Moreover, we define the tangent space to $S = F(U)$ at $p \in S$ to be the paralell translate of the subspace $Col(F'(a)) \leq \mathbb{R}^m$. A typical point in the tangent space at $p \in S$ has the form $p + F'(a)v$ for some $v \in \mathbb{R}^n$.

The condition that $rank(F'(a)) = n$ is the higher-dimensional analogue of the condition that the direction vector of a line must be nonzero for a line. If we want a genuine n-dimensional surface then there must be n-linearly independent vectors in the columns in the derivative matrix. If there were two columns which were linearly dependent then the subspace $W = \{F'(a)v \mid v \in \mathbb{R}^n\}$ would not be n-dimensional.

Remark 6.1.12.

If this all seems a little abstract, relax, the examples are in the next section. I want to wrap up the mostly theoretical aspects in this section then turn to more calculational ideas such as partial derivatives and the Jacobian matrix in the next section. We'll see that partial differentiation gives us an easy straight-forward method to calculate all the theoretical constructs of this section. Edwards has the calculations mixed with the theory, I've ripped them apart for better or worse. Also, we will discuss surfaces and manifolds independently in the next chapter. I wouldn't expect you to entirely understand them from the discussion in this chapter.

Example 6.1.13. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $F(v) = p + Av$ for all $v \in \mathbb{R}^n$ where the matrix $A \in \mathbb{R}^{\ m \times n}$ such that $rank(A) = n$ and $p \in \mathbb{R}^m$. We can calculate that $[dF_a] = A$. Observe that for $x \in \mathbb{R}^n$,*

$$\lim_{h \to 0} \frac{F(x+h) - F(x) - A(h)}{||h||} = \lim_{h \to 0} \frac{Ax + Ah - Ax - Ah}{||h||} = 0.$$

Therefore, $dF_x(h) = Ah$ for each $x \in \mathbb{R}^n$ and we find $F(\mathbb{R}^n)$ is an differentiable surface of dimensional n. Moreover, we find that $F(\mathbb{R}^n)$ is its own tangent space, the tangent space is the paralell translate of $Col(A)$ to the point $p \in \mathbb{R}^m$. This is the higher dimensional analogue of finding the tangent line to a line, it's just the line again.

The directional derivative helped us connect the definition of the derivative of mapping with the derivative of a function of $\mathbb{R}$. We now turn it around. If we're given the derivative of a mapping then the directional derivative exists. The converse is not true, see Example 4 on page 69 of Edwards.

Proposition 6.1.14.

If $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in U$ then the directional derivative $D_vF(a)$ exists for each $v \in \mathbb{R}^n$ and $D_vF(a) = dF_a(v)$.

Proof: Suppose $a \in U$ such that dF_a is well-defined then we are given that

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

This is a limit in $\mathbb{R}^n$, when it exists it follows that the limits that approach the origin along particular paths also exist and are zero. In particular we can consider the path $t \mapsto tv$ for $v \neq 0$ and $t > 0$, we find

$$\lim_{tv \to 0,\ t>0} \frac{F(a+tv) - F(a) - dF_a(tv)}{||tv||} = \frac{1}{||v||} \lim_{t \to 0^+} \frac{F(a+tv) - F(a) - tdF_a(v)}{|t|} = 0.$$

Hence, as $|t| = t$ for $t > 0$ we find

$$\lim_{t \to 0^+} \frac{F(a+tv) - F(a)}{t} = \lim_{t \to 0} \frac{tdF_a(v)}{t} = dF_a(v).$$

Likewise we can consider the path $t \mapsto tv$ for $v \neq 0$ and $t < 0$

$$\lim_{tv \to 0,\ t<0} \frac{F(a+tv) - F(a) - dF_a(tv)}{||tv||} = \frac{1}{||v||} \lim_{t \to 0^-} \frac{F(a+tv) - F(a) - tdF_a(v)}{|t|} = 0.$$

Note $|t| = -t$ thus the limit above yields

$$\lim_{t \to 0^-} \frac{F(a+tv) - F(a)}{-t} = \lim_{t \to 0^-} \frac{tdF_a(v)}{-t} \quad \Rightarrow \quad \lim_{t \to 0^-} \frac{F(a+tv) - F(a)}{t} = dF_a(v).$$

Therefore,

$$\lim_{t \to 0} \frac{F(a+tv) - F(a)}{t} = dF_a(v)$$

and we conclude that $D_vF(a) = dF_a(v)$ for all $v \in \mathbb{R}^n$ since the $v = 0$ case follows trivially. $\square$

6.2 partial derivatives and the existence of the derivative

Definition 6.2.1.

Suppose that $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is a mapping the we say that F is **has partial derivative** $\frac{\partial F}{\partial x_i}(a)$ at $a \in U$ iff the directional derivative in the e_i direction exists at a. In this case we denote,

$$\frac{\partial F}{\partial x_i}(a) = D_{e_i}F(a).$$

Also we may use the notation $D_{e_i}F(a) = D_iF(a)$ or $\partial_i F = \frac{\partial F}{\partial x_i}$ when convenient. We also construct the partial derivative mapping $\partial_i F : V \subseteq \mathbb{R}^n \to \mathbb{R}^m$ as the mapping defined pointwise for each $v \in V$ where $\partial_i F(v)$ exists.

Let's expand this definition a bit. Note that if $F = (F_1, F_2, \dots, F_m)$ then

$$D_{e_i}F(a) = \lim_{h \to 0} \frac{F(a + he_i) - F(a)}{h} \quad \Rightarrow \quad [D_{e_i}F(a)] \cdot e_j = \lim_{h \to 0} \frac{F_j(a + he_i) - F_j(a)}{h}$$

for each $j = 1, 2, \dots m$. But then the limit of the component function F_j is precisely the directional derivative at a along e_i hence we find the result

$$\frac{\partial F}{\partial x_i} \cdot e_j = \frac{\partial F_j}{\partial x_i} \qquad \text{in other words,} \qquad \boxed{\partial_i F = (\partial_i F_1, \partial_i F_2, \dots, \partial_i F_m).}$$

Proposition 6.2.2.

If $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in U$ then the directional derivative $D_vF(a)$ can be expressed as a sum of partial derivative maps for each $v =< v_1, v_2, \dots, v_n >\in \mathbb{R}^n$:

$$D_vF(a) = \sum_{j=1}^{n} v_j \partial_j F(a)$$

Proof: since F is differentiable at a the differential dF_a exists and $D_vF(a) = dF_a(v)$ for all $v \in \mathbb{R}^n$. Use linearity of the differential to calculate that

$$D_vF(a) = dF_a(v_1e_1 + \cdots + v_ne_n) = v_1dF_a(e_1) + \cdots + v_ndF_a(e_n).$$

Note $dF_a(e_j) = D_{e_j}F(a) = \partial_j F(a)$ and the prop. follows. $\square$

Proposition 6.2.3.

If $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in U$ then the differential dF_a has derivative matrix $F'(a)$ and it has components which are expressed in terms of partial derivatives of the component functions:

$$[dF_a]_{ij} = \partial_j F_i$$

for $1 \le i \le m$ and $1 \le j \le n$.

Perhaps it is helpful to expand the derivative matrix explicitly for future reference:

$$F'(a) = \begin{bmatrix} \partial_1 F_1(a) & \partial_2 F_1(a) & \cdots & \partial_n F_1(a) \\ \partial_1 F_2(a) & \partial_2 F_2(a) & \cdots & \partial_n F_2(a) \\ \vdots & \vdots & \vdots & \vdots \\ \partial_1 F_m(a) & \partial_2 F_m(a) & \cdots & \partial_n F_m(a) \end{bmatrix}$$

Let's write the operation of the differential for a differentiable mapping at some point $a \in \mathbb{R}$ in terms of the explicit matrix multiplication by $F'(a)$. Let $v = (v_1, v_2, \dots v_n) \in \mathbb{R}^n$,

$$dF_a(v) = F'(a)v = \begin{bmatrix} \partial_1 F_1(a) & \partial_2 F_1(a) & \cdots & \partial_n F_1(a) \\ \partial_1 F_2(a) & \partial_2 F_2(a) & \cdots & \partial_n F_2(a) \\ \vdots & \vdots & \vdots & \vdots \\ \partial_1 F_m(a) & \partial_2 F_m(a) & \cdots & \partial_n F_m(a) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

You may recall the notation from calculus III at this point, omitting the a-dependence,

$$\boxed{\nabla F_j = grad(F_j) = \left[\ \partial_1 F_j,\ \partial_2 F_j,\ \cdots,\ \partial_n F_j\ \right]^T}$$

So if the derivative exists we can write it in terms of a stack of gradient vectors of the component functions: (I used a transpose to write the stack side-ways),

$$F' = \left[\nabla F_1 | \nabla F_2 | \cdots | \nabla F_m\right]^T$$

Finally, just to collect everything together,

$$\boxed{F' = \begin{bmatrix} \partial_1 F_1 & \partial_2 F_1 & \cdots & \partial_n F_1 \\ \partial_1 F_2 & \partial_2 F_2 & \cdots & \partial_n F_2 \\ \vdots & \vdots & \vdots & \vdots \\ \partial_1 F_m & \partial_2 F_m & \cdots & \partial_n F_m \end{bmatrix} = \left[\partial_1 F \mid \partial_2 F \mid \cdots \mid \partial_n F\ \right] = \begin{bmatrix} (\nabla F_1)^T \\ \hline (\nabla F_2)^T \\ \hline \vdots \\ \hline (\nabla F_m)^T \end{bmatrix}}$$

Example 6.2.4. *Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ then $\nabla f = [\partial_x f, \partial_y f, \partial_z f]^T$ and we can write the directional derivative in terms of*

$$D_v f = [\partial_x f, \partial_y f, \partial_z f]^T v = \nabla f \cdot v$$

if we insist that $||v|| = 1$ then we recover the standard directional derivative we discuss in calculus III. Naturally the $||\nabla f(a)||$ yields the maximum value for the directional derivative at a if we limit the inputs to vectors of unit-length. If we did not limit the vectors to unit length then the directional derivative at a can become arbitrarily large as $D_v f(a)$ is proportional to the magnitude of v. Since our primary motivation in calculus III was describing rates of change along certain directions for some multivariate function it made sense to specialize the directional derivative to vectors of unit-length. The definition used in these notes better serves the theoretical discussion. If you read my calculus III notes you'll find a derivation of how the directional derivative in Stewart's calculus arises from the general definition of the derivative as a linear mapping. Look up page 305g. Incidentally, those notes may well be better than these in certain respects.

6.2.1 examples of derivatives

Our goal here is simply to exhbit the Jacobian matrix and partial derivatives for a few mappings. At the base of all these calculations is the observation that partial differentiation is just ordinary differentiation where we treat all the independent variable not being differentiated as constants. The criteria of indepedence is important. We'll study the case the variables are not independent in a later section.

Remark 6.2.5.

I have put remarks about the rank of the derivative in the examples below. Of course this has nothing to do with the process of calculating Jacobians. It's something to think about once we master the process of calculating the Jacobian. Ignore the red comments for now if you wish

Example 6.2.6. *Let* $f(t) = (t, t^2, t^3)$ *then* $f'(t) = (1, 2t, 3t^2)$. *In this case we have*

$$f'(t) = [df_t] = \begin{bmatrix} 1 \\ 2t \\ 3t^2 \end{bmatrix}$$

The Jacobian here is a single column vector. It has rank 1 provided the vector is nonzero. We see that $f'(t) \neq (0, 0, 0)$ *for all* $t \in \mathbb{R}$. *This corresponds to the fact that this space curve has a well-defined tangent line for each point on the path.*

Example 6.2.7. *Let* $f(\vec{x}, \vec{y}) = \vec{x} \cdot \vec{y}$ *be a mapping from* $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$. *I'll denote the coordinates in the domain by* $(x_1, x_2, x_3, y_1, y_2, y_3)$ *thus* $f(\vec{x}, \vec{y}) = x_1y_1 + x_2y_2 + x_3y_3$. *Calculate,*

$$[df_{(\vec{x},\vec{y})}] = \nabla f(\vec{x}, \vec{y})^T = [y_1, y_2, y_3, x_1, x_2, x_3]$$

The Jacobian here is a single row vector. It has rank 6 provided all entries of the input vectors are nonzero.

Example 6.2.8. *Let* $f(\vec{x}, \vec{y}) = \vec{x} \cdot \vec{y}$ *be a mapping from* $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. *I'll denote the coordinates in the domain by* $(x_1, \dots, x_n, y_1, \dots, y_n)$ *thus* $f(\vec{x}, \vec{y}) = \sum_{i=1}^n x_iy_i$. *Calculate,*

$$\frac{\partial}{x_j}\left[\sum_{i=1}^n x_iy_i\right] = \sum_{i=1}^n \frac{\partial x_i}{x_j} y_i = \sum_{i=1}^n \delta_{ij} y_i = y_j$$

Likewise,

$$\frac{\partial}{y_j}\left[\sum_{i=1}^n x_iy_i\right] = \sum_{i=1}^n x_i \frac{\partial y_i}{y_j} = \sum_{i=1}^n x_i \delta_{ij} = x_j$$

Therefore, noting that $\nabla f = (\partial_{x_1} f, \dots, \partial_{x_n} f, \partial_{y_1} f, \dots, \partial_{y_n} f)$,

$$[df_{(\vec{x},\vec{y})}]^T = (\nabla f)(\vec{x}, \vec{y}) = \vec{y} \times \vec{x} = (y_1, \dots, y_n, x_1, \dots, x_n)$$

The Jacobian here is a single row vector. It has rank 2n provided all entries of the input vectors are nonzero.

Example 6.2.9. *Suppose $F(x,y,z) = (xyz, y, z)$ we calculate,*

$$\frac{\partial F}{\partial x} = (yz, 0, 0) \qquad \frac{\partial F}{\partial y} = (xz, 1, 0) \qquad \frac{\partial F}{\partial z} = (xy, 0, 1)$$

Remember these are actually column vectors in my sneaky notation; $(v_1, \dots, v_n) = [v_1, \dots, v_n]^T$. This means the **derivative** *or* **Jacobian matrix** *of F at (x,y,z) is*

$$F'(x,y,z) = [dF_{(x,y,z)}] = \begin{bmatrix} yz & xz & xy \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note, $rank(F'(x,y,z)) = 3$ for all $(x,y,z) \in \mathbb{R}^3$ such that $y, z \neq 0$. There are a variety of ways to see that claim, one way is to observe $det[F'(x,y,z)] = yz$ and this determinant is nonzero so long as neither y nor z is zero. In linear algebra we learn that a square matrix is invertible iff it has nonzero determinant iff it has linearly indpendent column vectors.

Example 6.2.10. *Suppose $F(x,y,z) = (x^2 + z^2, yz)$ we calculate,*

$$\frac{\partial F}{\partial x} = (2x, 0) \qquad \frac{\partial F}{\partial y} = (0, z) \qquad \frac{\partial F}{\partial z} = (2z, y)$$

The derivative is a 2×3 matrix in this example,

$$F'(x,y,z) = [dF_{(x,y,z)}] = \begin{bmatrix} 2x & 0 & 2z \\ 0 & z & y \end{bmatrix}$$

The maximum rank for F' is 2 at a particular point (x,y,z) because there are at most two linearly independent vectors in $\mathbb{R}^2$. You can consider the three square submatrices to analyze the rank for a given point. If any one of these is nonzero then the rank (dimension of the column space) is two.

$$M_1 = \begin{bmatrix} 2x & 0 \\ 0 & z \end{bmatrix} \qquad M_2 = \begin{bmatrix} 2x & 2z \\ 0 & y \end{bmatrix} \qquad M_3 = \begin{bmatrix} 0 & 2z \\ z & y \end{bmatrix}$$

We'll need either $det(M_1) = 2xz \neq 0$ or $det(M_2) = 2xy \neq 0$ or $det(M_3) = -2z^2 \neq 0$. I believe the only point where all three of these fail to be true simulataneously is when $x = y = z = 0$. This mapping has maximal rank at all points except the origin.

Example 6.2.11. *Suppose $F(x,y) = (x^2 + y^2, xy, x + y)$ we calculate,*

$$\frac{\partial F}{\partial x} = (2x, y, 1) \qquad \frac{\partial F}{\partial y} = (2y, x, 1)$$

The derivative is a 3×2 matrix in this example,

$$F'(x,y) = [dF_{(x,y)}] = \begin{bmatrix} 2x & 2y \\ y & x \\ 1 & 1 \end{bmatrix}$$

The maximum rank is again 2, this time because we only have two columns. The rank will be two if the columns are not linearly dependent. We can analyze the question of rank a number of ways but I find determinants of submatrices a comforting tool in these sort of questions. If the columns are linearly dependent then all three sub-square-matrices of F' will be zero. Conversely, if even one of them is nonvanishing then it follows the columns must be linearly independent. The submatrices for this problem are:

$$M_1 = \begin{bmatrix} 2x & 2y \\ y & x \end{bmatrix} \qquad M_2 = \begin{bmatrix} 2x & 2y \\ 1 & 1 \end{bmatrix} \qquad M_3 = \begin{bmatrix} y & x \\ 1 & 1 \end{bmatrix}$$

You can see $det(M_1) = 2(x^2 - y^2)$, $det(M_2) = 2(x - y)$ and $det(M_3) = y - x$. Apparently we have $rank(F'(x,y,z)) = 2$ for all $(x,y) \in \mathbb{R}^2$ with $y \neq x$. In retrospect this is not surprising.

Example 6.2.12. *Suppose $P(x,v,m) = (P_o, P_1) = (\frac{1}{2}mv^2 + \frac{1}{2}kx^2, mv)$ for some constant k. Let's calculate the derivative via gradients this time,*

$$\nabla P_o = (\partial P_o/\partial x, \partial P_o/\partial v, \partial P_o/\partial m) = (kx, mv, \tfrac{1}{2}v^2)$$

$$\nabla P_1 = (\partial P_1/\partial x, \partial P_1/\partial v, \partial P_1/\partial m) = (0, m, v)$$

Therefore,

$$P'(x,v,m) = \begin{bmatrix} kx & mv & \frac{1}{2}v^2 \\ 0 & m & v \end{bmatrix}$$

Example 6.2.13. *Let $F(r,\theta) = (r\cos\theta, r\sin\theta)$. We calculate,*

$$\partial_r F = (\cos\theta, \sin\theta) \qquad \text{and} \qquad \partial_\theta F = (-r\sin\theta, r\cos\theta)$$

Hence,

$$F'(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

We calculate $det(F'(r,\theta)) = r$ thus this mapping has full rank everywhere except the origin.

Example 6.2.14. *Let $G(x,y) = (\sqrt{x^2+y^2}, \tan^{-1}(y/x))$. We calculate,*

$$\partial_x G = \left(\tfrac{x}{\sqrt{x^2+y^2}}, \tfrac{-y}{x^2+y^2}\right) \qquad \text{and} \qquad \partial_y G = \left(\tfrac{y}{\sqrt{x^2+y^2}}, \tfrac{x}{x^2+y^2}\right)$$

Hence,

$$G'(x,y) = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{bmatrix} = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \\ \frac{-y}{r^2} & \frac{x}{r^2} \end{bmatrix} \qquad \left(\ \text{using } r = \sqrt{x^2+y^2}\ \right)$$

We calculate $det(G'(x,y)) = 1/r$ thus this mapping has full rank everywhere except the origin.

Example 6.2.15. *Let $F(x,y) = (x, y, \sqrt{R^2 - x^2 - y^2})$ for a constant R. We calculate,*

$$\nabla\sqrt{R^2 - x^2 - y^2} = \left(\frac{-x}{\sqrt{R^2-x^2-y^2}}, \frac{-y}{\sqrt{R^2-x^2-y^2}} \right)$$

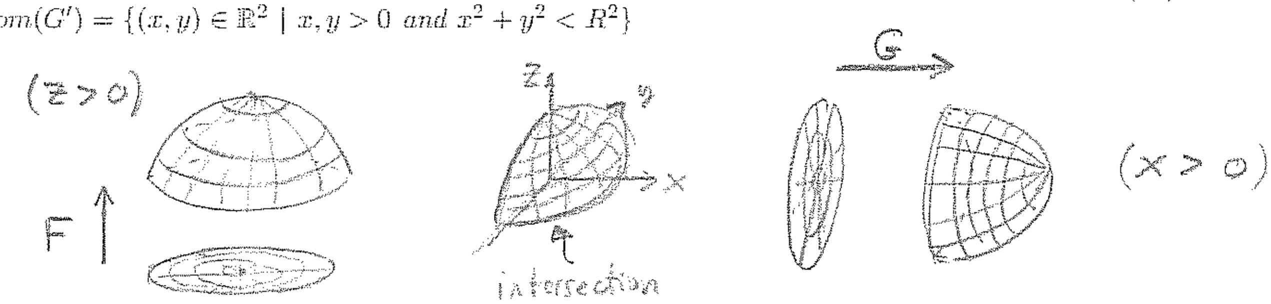
Also, $\nabla x = (1,0)$ and $\nabla y = (0,1)$ thus

$$F'(x,y) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{-x}{\sqrt{R^2-x^2-y^2}} & \frac{-y}{\sqrt{R^2-x^2-y^2}} \end{bmatrix}$$

This matrix clearly has rank 2 where is is well-defined. Note that we need $R^2 - x^2 - y^2 > 0$ for the derivative to exist. Moreover, we could define $G(y,z) = (\sqrt{R^2 - y^2 - z^2}, y, z)$ and calculate,

$$G'(y,z) = \begin{bmatrix} 1 & 0 \\ \frac{-y}{\sqrt{R^2-y^2-z^2}} & \frac{-z}{\sqrt{R^2-y^2-z^2}} \\ 0 & 1 \end{bmatrix}.$$

Observe that $G'(y,z)$ exists when $R^2 - y^2 - z^2 > 0$. Geometrically, F parametrizes the sphere above the equator at $z = 0$ whereas G parametrizes the right-half of the sphere with $x > 0$. These parametrizations overlap in the first octant where both x and z are positive. In particular, $dom(F') \cap dom(G') = \{(x,y) \in \mathbb{R}^2 \mid x, y > 0$ and $x^2 + y^2 < R^2\}$



Example 6.2.16. *Let $F(x,y,z) = (x, y, z, \sqrt{R^2 - x^2 - y^2 - z^2})$ for a constant R. We calculate,*

$$\nabla\sqrt{R^2 - x^2 - y^2 - z^2} = \left(\frac{-x}{\sqrt{R^2-x^2-y^2-z^2}}, \frac{-y}{\sqrt{R^2-x^2-y^2-z^2}}, \frac{-z}{\sqrt{R^2-x^2-y^2-z^2}} \right)$$

Also, $\nabla x = (1,0,0)$, $\nabla y = (0,1,0)$ and $\nabla z = (0,0,1)$ thus

$$F'(x,y,z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{-x}{\sqrt{R^2-x^2-y^2-z^2}} & \frac{-y}{\sqrt{R^2-x^2-y^2-z^2}} & \frac{-z}{\sqrt{R^2-x^2-y^2-z^2}} \end{bmatrix}$$

This matrix clearly has rank 3 where is is well-defined. Note that we need $R^2 - x^2 - y^2 - z^2 > 0$ for the derivative to exist. This mapping gives us a parametrization of the 3-sphere $x^2 + y^2 + z^2 + w^2 = R^2$ for $w > 0$. (drawing this is a little trickier)

Example 6.2.17. *Let $f(x,y,z) = (x+y, y+z, x+z, xyz)$. You can calculate,*

$$[df_{(x,y,z)}] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ yz & xz & xy \end{bmatrix}$$

This matrix clearly has rank 3 and is well-defined for all of $\mathbb{R}^3$.

Example 6.2.18. *Let $f(x,y,z) = xyz$. You can calculate,*

$$[df_{(x,y,z)}] = \begin{bmatrix} yz & xz & xy \end{bmatrix}$$

This matrix fails to have rank 3 if x, y or z are zero. In other words, $f'(x,y,z)$ has rank 3 in $\mathbb{R}^3$ provided we are at a point which is not on some coordinate plane. (the coordinate planes are $x = 0, y = 0$ and $z = 0$ for the yz, zx and xy coordinate planes respective)

Example 6.2.19. *Let $f(x,y,z) = (xyz, 1-x-y)$. You can calculate,*

$$[df_{(x,y,z)}] = \begin{bmatrix} yz & xz & xy \\ -1 & -1 & 0 \end{bmatrix}$$

This matrix has rank 3 if either $xy \neq 0$ or $(x-y)z \neq 0$. In contrast to the preceding example, the derivative does have rank 3 on certain points of the coordinate planes. For example, $f'(1,1,0)$ and $f'(0,1,1)$ both give $rank(f') = 3$.

Example 6.2.20. *Let $f : \mathbb{R}^3 \times \mathbb{R}^3$ be defined by $f(x) = x \times v$ for a fixed vector $v \neq 0$. We denote $x = (x_1, x_2, x_3)$ and calculate,*

$$\frac{\partial}{\partial x_a}(x \times v) = \frac{\partial}{\partial x_a}\Big(\sum_{i,j,k} \epsilon_{ijk} x_i v_j e_k \Big) = \sum_{i,j,k} \epsilon_{ijk} \frac{\partial x_i}{\partial x_a} v_j e_k = \sum_{i,j,k} \epsilon_{ijk} \delta_{ia} v_j e_k = \sum_{j,k} \epsilon_{ajk} v_j e_k$$

It follows,

$$\frac{\partial}{\partial x_1}(x \times v) = \sum_{j,k} \epsilon_{1jk} v_j e_k = v_2 e_3 - v_3 e_2 = (0, -v_3, v_2)$$

$$\frac{\partial}{\partial x_2}(x \times v) = \sum_{j,k} \epsilon_{2jk} v_j e_k = v_3 e_1 - v_1 e_3 = (v_3, 0, -v_1)$$

$$\frac{\partial}{\partial x_3}(x \times v) = \sum_{j,k} \epsilon_{3jk} v_j e_k = v_1 e_2 - v_2 e_1 = (-v_2, v_1, 0)$$

Thus the Jacobian is simply,

$$[df_{(x,y)}] = \begin{bmatrix} 0 & v_3 & -v_2 \\ -v_3 & 0 & -v_1 \\ v_2 & v_1 & 0 \end{bmatrix}$$

In fact, $df_p(h) = f(h) = h \times v$ for each $p \in \mathbb{R}^3$. The given mapping is linear so the differential of the mapping is precisely the mapping itself.

Example 6.2.21. *Let $f(x,y) = (x, y, 1-x-y)$. You can calculate,*

$$[df_{(x,y,z)}] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}$$

Example 6.2.22. *Let $X(u,v) = (x,y,z)$ where x,y,z denote functions of u,v and I prefer to omit the explicit depedendence to reduce clutter in the equations to follow.*

$$\frac{\partial X}{\partial u} = X_u = (x_u, y_u, z_u) \quad and \quad \frac{\partial X}{\partial v} = X_v = (x_v, y_v, z_v)$$

Then the Jacobian is the 3×2 matrix

$$\left[dX_{(u,v)}\right] = \begin{bmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{bmatrix}$$

The matrix $\left[dX_{(u,v)}\right]$ has rank 2 if at least one of the determinants below is nonzero,

$$det \begin{bmatrix} x_u & x_v \\ y_u & y_v \end{bmatrix} \quad det \begin{bmatrix} x_u & x_v \\ z_u & z_v \end{bmatrix} \quad det \begin{bmatrix} y_u & y_v \\ z_u & z_v \end{bmatrix}$$

Example 6.2.23. . .

Example 6.2.24. . .

6.2.2 sick examples and continuously differentiable mappings

We have noted that differentiablility on some set U implies all sorts of nice formulas in terms of the partial derivatives. Curiously the converse is not quite so simple. It is possible for the partial derivatives to exist on some set and yet the mapping may fail to be differentiable. We need an extra topological condition on the partial derivatives if we are to avoid certain pathological[3] examples.

Example 6.2.25. *I found this example in Hubbard's advanced calculus text(see Ex. 1.9.4, pg. 123). It is a source of endless odd examples, notation and bizarre quotes. Let* $f(x) = 0$ *and*

$$f(x) = \frac{x}{2} + x^2 \sin\frac{1}{x}$$

for all $x \neq 0$. *I can be shown that the derivative* $f'(0) = 1/2$. *Moreover, we can show that* $f'(x)$ *exists for all* $x \neq 0$, *we can calculate:*

$$f'(x) = \frac{1}{2} + 2x\sin\frac{1}{x} - \cos\frac{1}{x}$$

Notice that $dom(f') = \mathbb{R}$. *Note then that the tangent line at* $(0,0)$ *is* $y = x/2$. *You might be tempted to say then that this function is increasing at a rate of 1/2 for* x *near zero. But this claim would be false since you can see that* $f'(x)$ *oscillates wildly without end near zero. We have a tangent line at* $(0,0)$ *with positive slope for a function which is not increasing at* $(0,0)$ *(recall that increasing is a concept we must define in a open interval to be careful). This sort of thing cannot happen if the derivative is continuous near the point in question.*

The one-dimensional case is quite special, even though we had discontinuity of the derivative we still had a well-defined tangent line to the point. However, many interesting theorems in calculus of one-variable require the function to be continuously differentiable near the point of interest. For example, to apply the 2nd-derivative test we need to find a point where the first derivative is zero and the second derivative exists. We cannot hope to compute $f''(x_o)$ unless f' is continuous at x_o. The next example is *sick*.

Example 6.2.26. *Let us define* $f(0,0) = 0$ *and*

$$f(x,y) = \frac{x^2 y}{x^2 + y^2}$$

for all $(x,y) \neq (0,0)$ *in* $\mathbb{R}^2$. *It can be shown that* f *is continuous at* $(0,0)$. *Moreover, since* $f(x,0) = f(0,y) = 0$ *for all* x *and all* y *it follows that* f *vanishes identically along the coordinate axis. Thus the rate of change in the* e_1 *or* e_2 *directions is zero. We can calculate that*

$$\frac{\partial f}{\partial x} = \frac{2xy^3}{(x^2+y^2)^2} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = \frac{x^4 - x^2y^2}{(x^2+y^2)^2}$$

Consider the path to the origin $t \mapsto (t,t)$ *gives* $f_x(t,t) = 2t^4/(t^2+t^2)^2 = 1/2$ *hence* $f_x(x,y) \to 1/2$ *along the path* $t \mapsto (t,t)$, *but* $f_x(0,0) = 0$ *hence the partial derivative* f_x *is not continuous at* $(0,0)$. *In this example, the discontinuity of the partial derivatives makes the tangent plane fail to exist.*

[3]"pathological" as in, "your clothes are so pathological, where'd you get them?"

Definition 6.2.27.

A mapping $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is **continuously differentiable** at $a \in U$ iff the partial derivative mappings $D_j F$ exist on an open set containing a and are continuous at a.

The defintion above is interesting because of the proposition below. The import of the proposition is that we can build the tangent plane from the Jacobian matrix provided the partial derivatives are all continuous. This is a very nice result because the concept of the linear mapping is quite abstract but partial differentiation of a given mapping is easy.

Proposition 6.2.28.

If F is continuously differentiable at a then F is differentiable at a

We'll follow the proof in Edwards on pages 72-73.

Let $f = F \cdot e_p$ for some $p \in \{1, 2, \dots, m\}$. We seek to show continuous diff. of F at $a \Rightarrow f$ is diff. at a. Lemma 2.3 of Edwards then says $dF = (dF_1, \dots, dF_m)$ provided we have $dF_p = df$. (we need to show all components of F are differentiable at a)

Following Edwards (and other texts on this subject), let $h = (h_1, h_2, \dots, h_n)$ and note $h = \sum_{k=0}^{n} h_k e_k$ and we could break up h as follows,

$$h = \vec{h}_0 + h_1 e_1 + \cdots + h_n e_n = \vec{h}_1 + h_2 e_2 + \cdots + h_n e_n = \vec{h}_k + h_{k+1} e_{k+1} + \cdots + h_n e_n.$$

Anyway, since $\vec{h}_0 = 0$,

$$f(a+h) - f(a) = \sum_{k=1}^{n} \left[f(a + \vec{h}_k) - f(a + h_{k-1}) \right]$$
$$= f(a + \vec{h}_1) - f(a + \vec{h}_0) + f(a + \vec{h}_1) - f(a + \vec{h}_2) + \cdots + f(a + \vec{h}_n) - f(a + \vec{h}_{n-1})$$
$$= f(a + \vec{h}_n) - f(a).$$

Notice that,

$$f(a + \vec{h}_i) - f(a + \vec{h}_{i-1}) = f(a_1 + h_1, \dots, a_i + h_i, a_{i+1}, \dots, a_n) - f(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i, \dots, a_n)$$

entries match except at i^{th} position where we have $a_i + h_i$ verses a_i. Recognize this is $\Delta g(a_1 + h_1, \dots, a_{i-1} + h_{i-1}, a_i, \dots, a_n)$

Let $g(x) = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$ then $g'(x) = D_i f(a_1, \dots, x, \dots, a_n)$ which exists by assumption of continuous differentiability of F. Apply the mean value Th$^{\underline{m}}$ to conclude $\exists\, b_i$ such that $f(a + \vec{h}_i) - f(a + \vec{h}_{i-1}) = h_i D_i f(b_i)$

Thus,

$$\lim_{h \to 0} \frac{\| f(a+h) - f(a) - \sum_{i=1}^{n} D_i f(a) h_i \|}{\|h\|} = \lim_{h \to 0} \frac{\| \sum_{i=1}^{n} (D_i f(b_i) - D_i f(a)) h_i \|}{\|h\|}$$
$$\leq \lim_{h \to 0} \sum_{i=1}^{n} |D_i f(b_i) - D_i f(a)| \frac{|h_i|}{\|h\|}$$

Uses continuity of partials again $\longrightarrow$

$$\leq \lim_{h \to 0} \sum_{i=1}^{n} \| D_i f(b_i) - D_i f(a) \| = 0.$$

6.3 properties of the derivative

Of course much of what we discover in this section should be old news to you if you understood differentiation in calculus III. However, in our current context we have efficient methods of proof and the langauge of linear algebra allows us to summarize pages of calculations in a single line.

6.3.1 additivity and homogeneity of the derivative

Suppose $F_1 : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $F_2 : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Furthermore, suppose both of these are differentiable at $a \in \mathbb{R}^n$. It follows that $(dF_1)_a = L_1$ and $(dF_2)_a = L_2$ are linear operators from $\mathbb{R}^n$ to $\mathbb{R}^m$ which approximate the change in F_1 and F_2 near a, in particular,

$$\lim_{h\to 0} \frac{F_1(a+h) - F_1(a) - L_1(h)}{||h||} = 0 \qquad \lim_{h\to 0} \frac{F_2(a+h) - F_2(a) - L_2(h)}{||h||} = 0$$

To prove that $H = F_1 + F_2$ is differentiable at $a \in \mathbb{R}^n$ we need to find a differential at a for H. Naturally, we expect $dH_a = d(F_1 + F_2)_a = (dF_1)_a + (dF_2)_a$. Let $L = (dF_1)_a + (dF_2)_a$ and consider,

$$\begin{aligned}
\lim_{h\to 0} \tfrac{H(a+h)-H(a)-L(h)}{||h||} &= \lim_{h\to 0} \tfrac{F_1(a+h)+F_2(a+h)-F_1(a)-F_2(a)-L_1(h)-L_2(h)}{||h||} \\
&= \lim_{h\to 0} \tfrac{F_1(a+h)-F_1(a)-L_1(h)}{||h||} + \lim_{h\to 0} \tfrac{F_2(a+h)-F_2(a)-L_2(h)}{||h||} \\
&= 0 + 0 \\
&= 0
\end{aligned}$$

Note that breaking up the limit was legal because we knew the subsequent limits existed and were zero by the assumption of differentiability of F_1 and F_2 at a. Finally, since $L = L_1 + L_2$ we know L is a linear transformation since the sum of linear transformations is a linear transformation. Moreover, the matrix of L is the sum of the matrices for L_1 and L_2. Let $c \in \mathbb{R}$ and suppose $G = cF_1$ then we can also show that $dG_a = d(cF_1)_a = c(dF_1)_a$, the calculation is very similar except we just pull the constant c out of the limit. I'll let you write it out. Collecting our observations:

Proposition 6.3.1.

Suppose $F_1 : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $F_2 : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ are differentiable at $a \in U$ then $F_1 + F_2$ is differentiable at a and

$$\boxed{d(F_1 + F_2)_a = (dF_1)_a + (dF_2)_a} \quad \text{or} \quad \boxed{(F_1 + F_2)'(a) = F_1'(a) + F_2'(a)}$$

Likewise, if $c \in \mathbb{R}$ then

$$\boxed{d(cF_1)_a = c(dF_1)_a} \quad \text{or} \quad \boxed{(cF_1)'(a) = c(F_1'(a))}$$

6.3.2 product rules?

What sort of product can we expect to find among mappings? Remember two mappings have vector outputs and there is no way to multiply vectors in general. Of course, in the case we have two mappings that have equal-dimensional outputs we could take their dot-product. There is a product rule for that case: if $\vec{A}, \vec{B} : \mathbb{R}^n \to \mathbb{R}^m$ then

$$\boxed{\partial_j(\vec{A} \cdot \vec{B}) = (\partial_j \vec{A}) \cdot \vec{B}) + \vec{A} \cdot (\partial_j \vec{B})}$$

Or in the special case of $m = 3$ we could even take their cross-product and there is another product rule in that case:

$$\boxed{\partial_j(\vec{A} \times \vec{B}) = (\partial_j \vec{A}) \times \vec{B} + \vec{A} \times (\partial_j \vec{B})}$$

What other case can we "multiply" vectors? One very important case is $\mathbb{R}^2 = \mathbb{C}$ where is is customary to use the notation $(x, y) = x + iy$ and $f = u + iv$. If our range is complex numbers then we again have a product rule: if $f : \mathbb{R}^n \to \mathbb{C}$ and $g : \mathbb{R}^n \to \mathbb{C}$ then

$$\boxed{\partial_j(fg) = (\partial_j f)g + f(\partial_j g)}$$

I have relegated the proof of these product rules to the end of this chapter. One other object worth differentiating is a matrix-valued function of $\mathbb{R}^n$. If we **define** the partial derivative of a matrix to be the matrix of partial derivatives then partial differentiation will respect the sum and product of matrices (we may return to this in depth if need be later on):

$$\boxed{\partial_j(A + B) = \partial_j B + \partial_j B} \qquad \boxed{\partial_j(AB) = (\partial_j A)B + A(\partial_j B)}$$

Moral of this story? If you have a pair mappings whose ranges allow some sort of product then it is entirely likely that there is a corresponding product rule [4]. There is one product rule which we can state for arbitrary mappings, note that we can always sensibly multiply a mapping by a function. Suppose then that $G : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ are differentiable at $a \in U$. It follows that there exist linear transformations $L_G : \mathbb{R}^n \to \mathbb{R}^m$ and $L_f : \mathbb{R}^n \to \mathbb{R}$ where

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

Since $G(a + h) \approx G(a) + L_G(h)$ and $f(a + h) \approx f(a) + L_f(h)$ we expect

$$\begin{aligned} fG(a+h) &\approx (f(a) + L_f(h))(G(a) + L_G(h)) \\ &\approx (fG)(a) + \underbrace{G(a)L_f(h) + f(a)L_G(h)}_{\text{linear in } h} + \underbrace{L_f(h)L_G(h)}_{2^{nd} \text{ order in } h} \end{aligned}$$

[4] In my research I consider functions on supernumbers, these also can be multiplied. Naturally there is a product rule for super functions, the catch is that super numbers z, w do not necessarily commute. However, if they're homogeneneous $zw = (-1)^{\epsilon_w \epsilon_z} wz$. Because of this the super product rule is $\partial_M(fg) = (\partial_M f)g + (-1)^{\epsilon_f \epsilon_M} f(\partial_M g)$

Thus we propose: $L(h) = G(a)L_f(h) + f(a)L_G(h)$ is the best linear approximation of fG.

$$\begin{aligned}
&\lim_{h\to 0} \frac{(fG)(a+h) - (fG)(a) - L(h)}{||h||} = \\
&= \lim_{h\to 0} \frac{f(a+h)G(a+h) - f(a)G(a) - G(a)L_f(h) - f(a)L_G(h)}{||h||} \\
&= \lim_{h\to 0} \frac{f(a+h)G(a+h) - f(a)G(a) - G(a)L_f(h) - f(a)L_G(h)}{||h||} + \\
&\quad + \lim_{h\to 0} \frac{f(a)G(a+h) - G(a+h)f(a)}{||h||} \\
&\quad + \lim_{h\to 0} \frac{f(a+h)G(a) - G(a)f(a+h)}{||h||} \\
&\quad + \lim_{h\to 0} \frac{f(a)G(a) - G(a)f(a)}{||h||} \\
&= \lim_{h\to 0} \Bigg[f(a)\frac{G(a+h) - G(a) - L_G(h)}{||h||} + \frac{f(a+h) - f(a) - L_f(h)}{||h||}G(a) + \\
&\quad + \Big(f(a+h) - f(a) \Big) \frac{G(a+h) - G(a)}{||h||} \Bigg] \\
&= f(a)\left[\lim_{h\to 0} \frac{G(a+h) - G(a) - L_G(h)}{||h||} \right] + \left[\lim_{h\to 0} \frac{f(a+h) - f(a) - L_f(h)}{||h||} \right] G(a) \\
&= 0
\end{aligned}$$

Where we have made use of the differentiability and the consequent continuity of both f and G at a. Furthermore, note

$$\begin{aligned}
L(h + ck) &= G(a)L_f(h + ck) + f(a)L_G(h + ck) \\
&= G(a)(L_f(h) + cL_f(k)) + f(a)(L_G(h) + cL_G(k)) \\
&= G(a)L_f(h) + f(a)(L_G(h) + c(G(a)L_f(k) + f(a)L_G(k)) \\
&= L(h) + cL(k)
\end{aligned}$$

for all $h, k \in \mathbb{R}^n$ and $c \in \mathbb{R}$ hence $L = G(a)L_f + f(a)L_G$ is a linear transformation. We have proved (most of) the following proposition:

Proposition 6.3.2.

If $G : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ are differentiable at $a \in U$ then fG is differentiable at a and

$$d(fG)_a = (df)_a G(a) + f(a)dG_a \qquad (fG)'(a) = ~~f'(a)G(a)~~ \; G(a)f'(a) + f(a)G'(a)$$

The argument above covers the ordinary product rule and a host of other less common rules. Note again that $G(a)$ and $G'(a)$ are vectors.

$\underbrace{G(a)}_{m\times 1}\underbrace{f'(a)}_{1\times n} = m\times n$

probably should write
$d(fG)_a = G(a)(df)_a + f(a)(dG)_a$
but, it's ok as is as well.

6.4 chain rule

The proof in Edwards is on 77-78. I'll give a heuristic proof here which captures the essence of the argument. The simplicity of this rule continues to amaze me.

Proposition 6.4.1.

If $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at a and $G : V \subseteq \mathbb{R}^p \to \mathbb{R}^m$ is differentiable at $F(a) \in V$ then $G \circ F$ is differentiable at a and

$$d(G \circ F)_a = (dG)_{F(a)} \circ dF_a \quad \text{or, in matrix notation,} \quad (G \circ F)'(a) = G'(F(a))F'(a)$$

Proof Sketch:

$$F(a+h) \approx F(a) + dF_a(h) = F(a) + F'(a)h$$
$$G(z+k) \approx G(z) + dG_z(k) = G(z) + G'(z)k$$

Consider then,

$$\begin{aligned}(G \circ F)(a+h) &= G(F(a+h)) \\ &\approx G(F(a) + F'(a)h) \quad : \text{let } z = F(a) \text{ and } k = F'(a)h, \text{ note } h \text{ small} \Rightarrow k \text{ small, hence this reasonable} \\ &= G(z+k) \\ &\approx G(z) + G'(z)k \\ &= G(F(a)) + G'(F(a))F'(a)h\end{aligned}$$

Thus $\Delta(G \circ F) \approx G'(F(a))F'(a)h \Rightarrow \underline{d(G \circ F) = dG \circ dF}$. ∎

In calculus III you may have learned how to calculate partial derivatives in terms of tree-diagrams and intermediate variable etc... We now have a way of understanding those rules and all the other chain rules in terms of one over-arching calculation: matrix multiplication of the constituent Jacobians in the composite function. Of course once we have this rule for the composite of two functions we can generalize to n-functions by a simple induction argument. For example, for three suitably defined mappings F, G, H,

$$(F \circ G \circ H)'(a) = F'(G(H(a)))G'(H(a))H'(a)$$

Example 6.4.2. . . Let $f(x,y) = x^2y^2$ and $g(t) = (t, t^2)$

We have $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}^2$ note,

$$f'(x,y) = [2xy^2, 2x^2y] \quad \text{and} \quad g'(t) = \begin{bmatrix} 1 \\ 2t \end{bmatrix}$$

Note $f \circ g : \mathbb{R} \to \mathbb{R}^2 \to \mathbb{R}$ has

$$(f \circ g)'(t) = f'(g(t))g'(t) = f'(t,t^2)g'(t) = [2t^5, 2t^4]\begin{bmatrix} 1 \\ 2t \end{bmatrix} = 6t^5.$$

Note that $(f \circ g)(t) = f(t,t^2) = t^2t^4 = t^6$ so this result is not surprising!

Remark: I omit some point dependence to reduce clutter

Example 6.4.3. . . $f : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}$

$$\frac{df}{dt} = \left[\frac{df_1}{dt}, \frac{df_2}{dt}, \dots, \frac{df_n}{dt}\right]^T \quad \text{and} \quad g'(\vec{x}) = (\nabla g)(\vec{x})^T = \left[\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_n}\right]$$

$$(g \circ f)'(t) = g'(f(t))f'(t) = \left[\frac{\partial g}{\partial x_1}, \dots, \frac{\partial g}{\partial x_n}\right]\begin{bmatrix} df_1/dt \\ \vdots \\ df_n/dt \end{bmatrix}$$

$$\frac{d}{dt}\Big[g(f(t))\Big] = \frac{\partial g}{\partial x_1}\frac{df_1}{dt} + \frac{\partial g}{\partial x_2}\frac{df_2}{dt} + \dots + \frac{\partial g}{\partial x_n}\frac{df_n}{dt}.$$

Example 6.4.4. . .

Recall, $x = x(u,v,w)$, $y = y(u,v,w)$, $f = f(x,y)$ then,

$$\frac{\partial f}{\partial u} = \frac{\partial x}{\partial u}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial u}\frac{\partial f}{\partial y}$$

$$\frac{\partial f}{\partial v} = \frac{\partial x}{\partial v}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial f}{\partial y}$$

$$\frac{\partial f}{\partial w} = \frac{\partial x}{\partial w}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial w}\frac{\partial f}{\partial y}$$

note we derive these via matrix multiplication in example below ↴

Example 6.4.5. . .

$$f = f(x,y) \longrightarrow f' = (\nabla f)^T = \underbrace{\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]}_{1\times 2 \text{ Jacobian}}$$

$f : \mathbb{R}^2 \to \mathbb{R}$

$$\vec{r}(u,v,w) = [x(u,v,w), y(u,v,w)]^T \longrightarrow \vec{r}\,' = \left[\frac{\partial \vec{r}}{\partial u} \Big| \frac{\partial \vec{r}}{\partial v} \Big| \frac{\partial \vec{r}}{\partial w}\right]$$

$\vec{r} : \mathbb{R}^3 \to \mathbb{R}^2$

$$\vec{r}\,' = \underbrace{\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{bmatrix}}_{(2\times 3) \text{ Jacobian}}$$

$f \circ \vec{r} : \mathbb{R}^3 \to \mathbb{R}^2 \to \mathbb{R}$

$$(f \circ \vec{r})' = f'\vec{r}\,'$$

$$= \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{bmatrix}$$

$$= \Bigg[\underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}}_{\frac{\partial f}{\partial u}}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}}_{\frac{\partial f}{\partial v}}, \underbrace{\frac{\partial f}{\partial x}\frac{\partial x}{\partial w} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial w}}_{\frac{\partial f}{\partial w}}\Bigg]$$

Example 6.4.6. . .

$$T(r,\theta) = (r\cos\theta,\ r\sin\theta) = (x,y)$$

$$T' = \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

If $w = f(x,y)$ then $w = \underbrace{g(r,\theta) = f(T(r,\theta))}_{f \text{ rewritten in polars.}}$

$$\left[\frac{\partial g}{\partial r}, \frac{\partial g}{\partial \theta}\right] = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} = \left[\cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y},\ -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}\right]$$

With the proper understanding we have derived,

$$\frac{\partial}{\partial r} = \cos\theta\frac{\partial}{\partial x} + \sin\theta\frac{\partial}{\partial y}$$

$$\frac{\partial}{\partial \theta} = -r\sin\theta\frac{\partial}{\partial x} + r\cos\theta\frac{\partial}{\partial y}$$

You can invert these, $r = \sqrt{x^2+y^2}$ & $\theta = \tan^{-1}(y/x)$

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial x}\frac{\partial}{\partial \theta} = \frac{x}{r}\frac{\partial}{\partial r} - \frac{y}{r^2}\frac{\partial}{\partial \theta} = \cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}$$

$$\frac{\partial}{\partial y} = \frac{\partial r}{\partial y}\frac{\partial}{\partial r} + \frac{\partial \theta}{\partial y}\frac{\partial}{\partial \theta} = \frac{y}{r}\frac{\partial}{\partial r} + \frac{x}{r^2}\frac{\partial}{\partial \theta} = \sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}$$

You can use these to change coordinates. For example

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \left(\cos\theta\frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\right)\left(\cos\theta\frac{\partial f}{\partial r} - \frac{\sin\theta}{r}\frac{\partial f}{\partial \theta}\right) + \left(\sin\theta\frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}\right)\left(\sin\theta\frac{\partial f}{\partial r} + \frac{\cos\theta}{r}\frac{\partial f}{\partial \theta}\right)$$

$$= \cos^2\theta\frac{\partial^2 f}{\partial r^2} - \cos\theta\sin\theta\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial f}{\partial \theta}\right] - \frac{\sin\theta}{r}\frac{\partial}{\partial \theta}\left[\cos\theta\frac{\partial f}{\partial r}\right] + \frac{\sin\theta}{r^2}\frac{\partial}{\partial \theta}\left[\sin\theta\frac{\partial f}{\partial \theta}\right]$$
$$+ \sin^2\theta\frac{\partial^2 f}{\partial r^2} + \sin\theta\cos\theta\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial f}{\partial \theta}\right] + \frac{\cos\theta}{r}\frac{\partial}{\partial \theta}\left[\sin\theta\frac{\partial f}{\partial r}\right] + \frac{\cos\theta}{r^2}\frac{\partial}{\partial \theta}\left[\cos\theta\frac{\partial f}{\partial \theta}\right]$$

$$= \frac{\partial^2 f}{\partial r^2} + \frac{\sin^2\theta}{r}\frac{\partial f}{\partial r} + \frac{\cos^2\theta}{r}\frac{\partial f}{\partial r} - \frac{\sin\theta\cos\theta}{r}\frac{\partial^2 f}{\partial \theta\partial r} + \frac{\cos\theta\sin\theta}{r}\frac{\partial^2 f}{\partial \theta\partial r} +$$
$$+ \frac{\sin\theta\cos\theta}{r^2}\frac{\partial f}{\partial \theta} - \frac{\cos\theta\sin\theta}{r^2}\frac{\partial f}{\partial \theta} + \frac{\sin^2\theta}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\cos^2\theta}{r^2}\frac{\partial^2 f}{\partial \theta^2}$$

$$\therefore \boxed{\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2}}$$

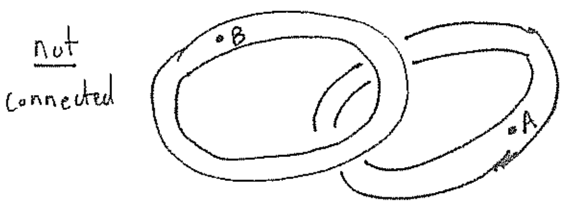
6.4.1 theorems

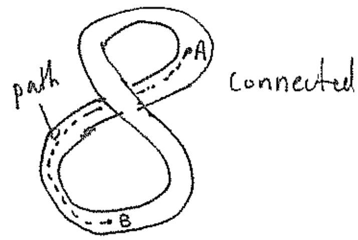
The goal of this section is to prove the partial derivatives commute for nice functions. Of course some of the results we discuss on the way to that goal are interesting in their own right as well.

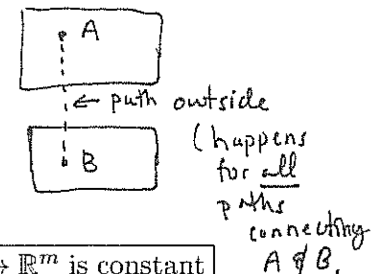
Definition 6.4.8.

> We say $U \subseteq \mathbb{R}^n$ is **path connected** iff any two points in U can be connected by a path which is contained within the set.

For example, $\mathbb{R}^n$ is connected since given any two points $a, b \in \mathbb{R}^n$ we can construct the path $\phi(t) = a + t(b-a)$ from a to b and naturally the path is within the set. You can easily verify that open and closed balls are also path connected. Even a donut is path connected. However, a pair donuts is not path connected unless it's one of those artsy figure-8 deals.







Proposition 6.4.9.

> If U is a connected open subset of $\mathbb{R}^n$ then a differenitable mapping $F : U \to \mathbb{R}^m$ is constant iff $F'(u) = 0$ for all $u \in U$. (?) ☹

Let $A, B \in U$ and connect these points with a smooth path $\gamma : [a,b] \longrightarrow U$.
(perhaps we should insist path connected $\Rightarrow$ smooth paths exist inside U...)
Consider, $F(\gamma(t)) = g(t)$. If F is constant then $g(t)$ is constant hence $F'(\gamma(t))\gamma'(t) = 0$ but γ smooth $\Rightarrow \gamma'(t) \neq 0$ hence $F'(\gamma(a)) = F'(a) = 0 \Rightarrow F'(a) = 0 \ \forall a \in U$. Conversely
if $F'(u) = 0 \ \forall u \in U \Rightarrow \frac{d}{dt}[g(t)] = \underbrace{F'(\gamma(t))}_{\text{zero}}\gamma'(t) = 0$
Thus $g(a) = g(b) \Rightarrow F(\gamma(a)) = F(\gamma(b))$
$\Rightarrow F(A) = F(B) \ \ \forall A, B \in U \therefore F$ constant on U.

Proposition 6.4.10.

☹

> If U is a connected open subset of $\mathbb{R}^n$ and the differenitable mappings $F, G : U \to \mathbb{R}$ such that $F'(x) = G'(x)$ for all $x \in U$ then there exists a constant vector $c \in \mathbb{R}^m$ such that $F(x) = G(x) + c$ for all $x \in U$.

Construct $H(x) = F(x) - G(x)$ and note $H'(x) = F'(x) - G'(x) = 0 \ \forall x \in U$.
By the previous prop. $H(x) = c \ \ \forall x \in U$ hence

$$F(x) = G(x) + c \quad \forall x \in U.$$

Remark: if U was not connected we could only conclude that the F was constant on connected subsets of U.
In topology one discusses breaking down a space into its path components, these are maximal connected subsets.

There is no mean value theorem for mappings since counter-examples exist. For example, Exercise 1.12 on page 63 shows the mean value theorem fails for the helix. In particular, you can find average velocity vector over a particular time interval such that the velocity vector never matches the average velocity over that time period. Fortunately, if we restrict our attention to mappings with one-dimensional codomains we still have a nice theorem:

Proposition 6.4.11. *(Mean Value Theorem)*

Suppose that $f : U \to \mathbb{R}$ is a differentiable function and U is an open set. Furthermore, suppose U contains the line segment from a to b in U;

$$L_{a,b} = \{a + t(b-a) \mid t \in [0,1]\} \subset U.$$

It follows that there exists some point $c \in L_{a,b}$ such that

$$f(b) - f(a) = f'(c)(b-a).$$

The proof follows from the construction of a function on $\mathbb{R}$ to which the elementary mean value Th$^{\underline{m}}$ is applied. Let

$$g(t) = f(a + t(b-a)) \quad \text{for } 0 \le t \le 1.$$

Or if you prefer, construct $\varphi(t) = a + t(b-a)$ which parametrizes the line segment from a to b. Clearly $\varphi'(t) = b - a$ and by the chain-rule,

$$g'(t) = f'(\varphi(t))\varphi'(t)$$

Note $g : [0,1] \xrightarrow{\varphi} U \xrightarrow{f} \mathbb{R}$ is differentiable on $[0,1]$ thus the MVT gives $c_o \in [0,1]$ such that $g(1) - g(0) = g'(c_o)$.

Thus,

$$f(b) - f(a) = g(1) - g(0) = g'(c_o) = f'(\varphi(c_o))\varphi'(c_o) = f'(c)\cdot(b-a).$$

$$c = \varphi(c_o).$$

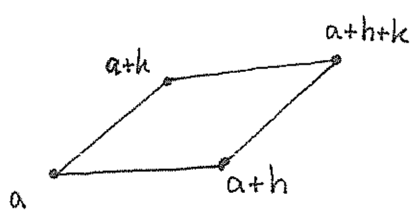
Definition 6.4.12. *(higher derivatives)*

We define nested directional derivatives in the natural way:

$$D_k D_h f(x) = D_k(D_h f(x)) = \lim_{t \to 0} \frac{D_h f(x + tk) - D_h f(x)}{t}$$

Furthermore, the **second difference** is defined by

$$\Delta^2 f_a(h, k) = f(a + h + k) - f(a + h) - f(a + k) + f(a)$$



note by MVT: $f(a+h) - f(a) = f'(a + \alpha h)h$

$f(a+h+k) - f(a+k) = f'(a+k+\beta h)h$

$$\Delta^2 f_a(h,k) = \Delta f_a(h) - \Delta f_{a+k}(h)$$

This is Lemma 3.5 on page 86 of Edwards.

Proposition 6.4.13.

Suppose U us an open set and $f : U \to \mathbb{R}$ which is differentiable on U with likewise differentiable directional derivative function on U. Suppose that $a, a+h, a+k, a+h+k$ are all in U then there exist $\alpha, \beta \in (0,1)$ such that

$$\Delta^2 f_a(h,k) = D_k D_h f(a + \alpha h + \beta k).$$

The proof is rather neat. The α and β stem from two applications of the MVT, once for the function then once for its directional derivative.

Let $g(x) = f(x+k) - f(x)$ then $dg_x = df_{x+k} - df_x$.
Furthermore, using $\Delta^2 f_a(h,k) = f(a+h+k) - f(a+h) - f(a+k) + f(a)$ note,

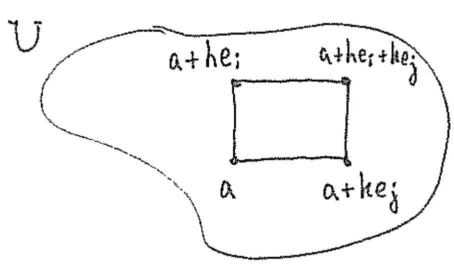
following Edwards pg. 87.

$$\begin{aligned}
\Delta^2 f_a(h,k) &= g(a+h) - g(a) \\
&= g'(a+\alpha h)h \quad : \text{by MVT } \exists\, \alpha \in (0,1). \\
&= (D_h g)(a+\alpha h) \quad : \text{diff} \Rightarrow \text{directional derivative exist.} \\
&= dg_{a+\alpha h}(h) \\
&= df_{a+\alpha h+k}(h) - df_{a+\alpha h}(h) \\
&= D_h f(a+\alpha h+k) - D_h f(a+\alpha h) \\
&= (D_h f)'(a+\alpha h+\beta k)(k) \quad \text{for some } \beta \in (0,1) \text{ by MVT again} \\
&= (D_k D_h f)(a+\alpha h+\beta k) \quad : \text{unraveling notation.}
\end{aligned}$$

Proposition 6.4.14.

Let U be an open subset of $\mathbb{R}^n$. If $f : U \to \mathbb{R}$ is a function with continuous first and second partial derivatives on U then for all $i, j = 1, 2, \ldots, n$ we have $D_i D_j f = D_j D_i f$ on U;

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$



$$\Delta^2 f_a(he_i, ke_j) = D_{ke_j} D_{he_i} f(a + \alpha_1 he_i + \beta_1 ke_j)$$

$$\Delta^2 f_a(ke_j, he_i) = D_{he_i} D_{ke_j} f(a + \alpha_2 ke_j + \beta_2 he_i)$$

Then $\Delta^2 f_a(he_i, ke_j) = \Delta^2 f_a(ke_j, he_i)$

$$\Rightarrow D_j D_i f(a) = D_i D_j f(a)$$

(pull out h, k using homogeneity of $D_{cv} f = c D_v f$)
take limit $h, k \to 0$ to drop the h, k inside the $Df(\_)$ terms. Can do this by continuity of partial derivatives near a.

6.5 differential forms and differentials

Definition 6.5.1.

A **form field** on $\mathbb{R}$ is a function from $\mathbb{R}$ to the space of all linear maps from $\mathbb{R}$ to $\mathbb{R}$. In other words, a form field assigns a dual vector at each point in $\mathbb{R}$. Remember that $\mathbb{R}^* = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is a linear function}\}$. We call α a **differential one-form** or **differential form** if α can be written as $\alpha = \alpha_1 dx$ for some smooth function α_1.

The definition above is probably unecessary for this section. I give it primarily for the sake of making a larger trend easier to grasp later on. Feel free to ignore it for now.

6.5.1 differential form notation

Let $g(x) = x$ for all $x \in \mathbb{R}$. Note that $g'(x) = 1$ and it follows that $dg_a(x) = 1 \cdot x = x$ for all $x \in \mathbb{R}$. Therefore, $dg = g$. If we denote $g = x$ so that $dx = x$ in this notation. Note then we can write the differential in terms of the derivative function:

$$df(a)(h) = df_a(h) = f'(a)h = f'(a)dx_a(h) \qquad \text{for all } h \in \mathbb{R}$$

Hence $df(a) = f'(a)dx_a$ for all $a \in \mathbb{R}$ hence $df = f'dx$ or we could denote this by the deceptively simple formula $df = \frac{df}{dx}dx$. Thus the differential notation introduced in this section is in fact consistent with our usual notation for the derivative from calculus I. However, df and dx are actually differential forms in this viewpoint so I'm not so sure that df/dx really makes sense anymore. In retrospect, the main place we shift differentials around as if they are tiny real numbers is in the calculations of u-substitution or separation of variables. In both of those cases the differential notation serves as a shorthand for the application of a particular theorem. Just as in calculus III the differentials dx, dy, dz in the line integral $\int_C pdx + qdy + rdz$ provide a notational shortand for the rigorous definition in terms of a path covering the curve C.

Differentials are notational devices in calculus, one should be careful not to make more of them then is appropriate for a given context. That said, if you adopt the view point that dx, dy, dz are differential forms and their product is properly defined via a wedge product then the wedge product together with the total differential (to be discussed in the next section) will generate the formulas for coordinate change. Let me give you a taste:

$$\begin{aligned} dx \wedge dy &= d(r\cos(\theta)) \wedge d(r\sin(\theta)) \\ &= [\cos(\theta)dr - r\sin(\theta)d\theta] \wedge [\sin(\theta)dr + r\cos(\theta)d\theta] \\ &= r\cos^2(\theta)dr \wedge d\theta - r\sin^2(\theta)d\theta \wedge dr \\ &= rdr \wedge d\theta \end{aligned}$$

where I used that $dr \wedge d\theta = -d\theta \wedge dr$, $dr \wedge dr = 0$ and $d\theta \wedge d\theta = 0$ because of the antisymmety of the **wedge product** $\wedge$. In calculus III we say for polar coordinates the Jacobian is $\frac{\partial(x,y)}{\partial(r,\theta)} = r$. The determinant in the Jacobian is implicitly contained in the algebra of the wedge product. If you want to change coordinates in differential form notation you just substitute in the coordinate change formulas and take a few total differentials then the wedge product does the rest. In other words, the Jacobian change of coordinates formula is naturally encoded in the langauge of differential forms.

6.5.2 linearity properties of the derivative

Proposition 6.5.2.

Suppose that f, g are functions such that their derivative functions f' and g' share the same domain U then $(f+g)' = f' + g'$ and $(cf)' = cf'$. Moreover, the differentials of those functions have

$$d(f+g) = df + dg \qquad \text{and} \qquad d(cf) = cdf$$

Proof: The proof that $(f+g)' = f' + g'$ and $(cf)' = cf'$ follows from earlier general arguments in this chapter. Consider that,

$$\begin{aligned} d(f+g)_a(h) &= h(f+g)'(a) && \text{def. of differential for } f+g \\ &= h(f'(a)+g'(a)) && \text{using linearity of derivative.} \\ &= df_a(h) + dg_a(h) && \text{algebra and def. of differential for } f \text{ and } g. \\ &= (df+dg)_a(h) && \text{def. of sum of functions.} \end{aligned}$$

thus $d(f+g) = df + dg$ and the proof that $d(cf) = cdf$ is similar. □.
We see that properties of the derivative transfer over to corresponding properties for the differential. Problem 1.7 on pg 62-63 of Edwards asks you to work out the product and chain rule for differentials.

6.5.3 the chain rule revisited

Proposition 6.5.3.

Suppose that $f : dom(f) \to range(f)$ and $g : dom(g) \to range(g)$ are functions such that g is differentiable on U and f differentiable on $g(U)$ then

$$(f \circ g)'(a) = g'(a)f'(g(a))$$

for each $a \in U$ and it follows $d(f \circ g)_a = df_{g(a)} \circ dg_a$.

An intuitive proof is this: the derivative of a composite is the slope of the tangent line to the composite. However, if f_1 and f_2 are linear functions with slopes m_1 and m_2 then $f_1 \circ f_2$ is a linear function with slope $m_1 m_2$. Therefore, the derivative of a composite is the product of the derivatives of the inside and outside function and we are forced to evaluate the outside function at $g(a)$ since that's the only thing that makes sense[5]. Finally,

$$d(f \circ g)(a)(h) = h(f \circ g)'(a) = hg'(a)f'(g(a)) = df_{g(a)}(hg'(a)) = df_{g(a)}(dg_a(h)) = (df_{g(a)} \circ dg_a)(h)$$

Therefore we find $d(f \circ g)_a = df_{g(a)} \circ dg_a$.

Proof: Let $a \in U$ then $g'(a) = \lim_{h \to 0} \frac{g(a+h)-g(a)}{h}$ thus $\lim_{h\to 0} g(a+h) = \lim_{h \to 0} g(a) + hg'(a)$. In other words, the function g and it's tangent line are equal in the limit you approach the point

[5]this is argument by inevitability, see Agent Smith for how this turns out as a pattern of deduction.

of tangency. Likewise, $f'(g(a)) = \lim_{\delta \to 0} \frac{f(g(a)+\delta)-f(g(a))}{\delta}$ hence $\lim_{\delta \to 0} f(g(a)+\delta) = f(g(a)) + \delta f'(g(a))$. Calculate then,

$$\begin{aligned}
(f \circ g)'(a) &= \lim_{h \to 0} \tfrac{(f \circ g)(a+h)-(f \circ g)(a)}{h} && \text{defn. of derivative} \\
&= \lim_{h \to 0} \tfrac{f(g(a+h))-f(g(a))}{h} && \text{defn. of } f \circ g \\
&= \lim_{h \to 0} \tfrac{f(g(a)+hg'(a))-f(g(a))}{h} && \text{since } g(a+h) \approx g(a) + hg'(a) \\
&= g'(a) \lim_{\delta \to 0} \tfrac{f(g(a)+\delta)-f(g(a))}{\delta} && \text{made subst. } \delta = g'(a)h \\
&= g'(a) \lim_{\delta \to 0} \tfrac{f(g(a))+\delta f'(g(a))-f(g(a))}{\delta} && \text{as } f(g(a)+\delta) \approx f(g(a)) + \delta f'(g(a)) \\
&= g'(a) f'(g(a)) && \text{limit of constant is just the constant.}
\end{aligned}$$

I have used the notation $\approx$ to indicate that those equations were not precisely true. However, the error is small when h or δ are close to zero and that is precisely the case which we were faced with in those calculations. Admittably we could give a more rigoruous proof in terms of ϵ and δ but this proof suffices for our purposes here. The main thing I wanted you to take from this is that the chain rule is a consequence of the tangent line approximation. □

Notice that most of the work I am doing here is to prove the result for the derivative. The same was true in the last subsection. In your homework I say you can assume the product and quotient rules for functions so that problem shouldn't be too hard. You just have to pay attention to how I defined the differential and how it is related to the derivative.

6.6 special product rules

In this section I gather together a few results which are commonly needed in applications of calculus.

6.6.1 calculus of paths in $\mathbb{R}^3$

A **path** is a mapping from $\mathbb{R}$ to $\mathbb{R}^m$. We use such mappings to model position, velocity and acceleration of particles in the case $m = 3$. Some of these things were proved in previous sections of this chapter but I intend for this section to be self-contained so that you can read it without digging through the rest of this chapter.

Proposition 6.6.1.

If $F, G : U \subseteq \mathbb{R} \to \mathbb{R}^m$ are differentiable vector-valued functions and $\phi : U \subseteq \mathbb{R} \to \mathbb{R}$ is a differentiable real-valued function then for each $t \in U$,

1. $(F + G)'(t) = F'(t) + G'(t)$.
2. $(cF)'(t) = cF'(t)$.
3. $(\phi F)'(t) = \phi'(t)F(t) + \phi(t)F'(t)$.
4. $(F \cdot G)'(t) = F'(t) \cdot G(t) + F(t) \cdot G'(t)$.
5. provided $m = 3$, $(F \times G)'(t) = F'(t) \times G(t) + F(t) \times G'(t)$.
6. provided $\phi(U) \subset dom(F')$, $(F \circ \phi)'(t) = \phi'(t)F(\phi(t))$.

We have to insist that $m = 3$ for the statement with cross-products since we only have a standard cross-product in $\mathbb{R}^3$. We prepare for the proof of the proposition with a useful lemma. Notice this lemma tells us how to actually calculate the derivative of paths in examples. The derivative of component functions is nothing more than calculus I and one of our goals is to reduce things to those sort of calculations whenever possible.

Lemma 6.6.2.

If $F : U \subseteq \mathbb{R} \to \mathbb{R}^m$ is differentiable vector-valued function then for all $t \in U$,

$$F'(t) = (F_1'(t), F_2'(t), \dots, F_m'(t))$$

We are given that the following vector limit exists and is equal to $F'(t)$,

$$F'(t) = \lim_{h \to 0} \frac{F(t+h) - F(t)}{h}$$

then by Proposition 3.2.10 the limit of a vector is related to the limits of its components as follows:

$$F'(t) \cdot e_j = \lim_{h \to 0} \frac{F_j(t+h) - F_j(t)}{h}.$$

Thus $(F'(t))_j = F_j'(t)$ and the lemma follows[6]. ∇

Proof of proposition: We use the notation $F = \sum F_j e_j = (F_1, \dots, F_m)$ and $G = \sum_i G_i e_i = (G_1, \dots, G_m)$ throughout the proofs below. The $\sum$ is understood to range over $1, 2, \dots m$. Begin with (1.),

$$\begin{aligned}
[(F+G)']_j &= \tfrac{d}{dt}[(F+G)_j] && \text{using the lemma}\\
&= \tfrac{d}{dt}[F_j + G_j] && \text{using def. } (F+G)_j = F_j + G_j\\
&= \tfrac{d}{dt}[F_j] + \tfrac{d}{dt}[G_j] && \text{by calculus I, } (f+g)' = f' + g'.\\
&= [F' + G']_j && \text{def. of vector addition for } F' \text{ and } G'
\end{aligned}$$

Hence $(F \times G)' = F' \times G + F \times G'$.The proofs of 2,3,5 and 6 are similar. I'll prove (5.),

$$\begin{aligned}
[(F\times G)']_k &= \tfrac{d}{dt}[(F\times G)_k] && \text{using the lemma}\\
&= \tfrac{d}{dt}[\textstyle\sum \epsilon_{ijk} F_i G_j] && \text{using def. } F \times G\\
&= \textstyle\sum \epsilon_{ijk} \tfrac{d}{dt}[F_i G_j] && \text{repeatedly using, } (f+g)' = f' + g'\\
&= \textstyle\sum \epsilon_{ijk} [\tfrac{dF_i}{dt} G_j + F_i \tfrac{dG_j}{dt}] && \text{repeatedly using, } (fg)' = f'g + fg'\\
&= \textstyle\sum \epsilon_{ijk} \tfrac{dF_i}{dt} G_j \sum \epsilon_{ijk} F_i \tfrac{dG_j}{dt}] && \text{property of finite sum } \textstyle\sum\\
&= (\tfrac{dF}{dt} \times G)_k + (F \times \tfrac{dG}{dt})_k) && \text{def. of cross product}\\
&= \left(\tfrac{dF}{dt} \times G + F \times \tfrac{dG}{dt} \right)_k && \text{def. of vector addition}
\end{aligned}$$

Notice that the calculus step really just involves calculus I applied to the components. The ordinary product rule was the crucial factor to prove the product rule for cross-products. We'll see the same for the dot product of mappings. Prove (4.)

$$\begin{aligned}
(F\cdot G)'(t) &= \tfrac{d}{dt}[\textstyle\sum F_k G_k] && \text{using def. } F \cdot G\\
&= \textstyle\sum \tfrac{d}{dt}[F_k G_k] && \text{repeatedly using, } (f+g)' = f' + g'\\
&= \textstyle\sum [\tfrac{dF_k}{dt} G_k + F_k \tfrac{dG_k}{dt}] && \text{repeatedly using, } (fg)' = f'g + fg'\\
&= \tfrac{dF}{dt} \cdot G + F \cdot \tfrac{dG}{dt}. && \text{def. of dot product}
\end{aligned}$$

The proof of (3.) follows from applying the product rule to each component of $\phi(t)F(t)$. The proof of (2.) follow from (3.) in the case that $phi(t) = c$ so $\phi'(t) = 0$. Finally the proof of (6.) follows from applying the chain-rule to each component. □

[6] this notation I first saw in a text by Marsden, it means the proof is partially completed but you should read on to finish the proof

6.6.2 calculus of matrix-valued functions of a real variable

Definition 6.6.3.

A matrix-valued function of a real variable is a function from $I \subseteq \mathbb{R}$ to $\mathbb{R}^{m \times n}$. Suppose $A : I \subseteq \mathbb{R} \to \mathbb{R}^{m\times n}$ is such that $A_{ij} : I \subseteq \mathbb{R} \to \mathbb{R}$ is differentiable for each i, j then we define

$$\frac{dA}{dt} = \left[\frac{dA_{ij}}{dt}\right]$$

which can also be denoted $(A')_{ij} = A'_{ij}$. We likewise define $\int A dt = [\int A_{ij} dt]$ for A with integrable components. Definite integrals and higher derivatives are also defined component-wise.

Example 6.6.4. *Suppose* $A(t) = \begin{bmatrix} 2t & 3t^2 \\ 4t^3 & 5t^4 \end{bmatrix}$. *I'll calculate a few items just to illustrate the definition above. calculate; to differentiate a matrix we differentiate each component one at a time:*

$$A'(t) = \begin{bmatrix} 2 & 6t \\ 12t^2 & 20t^3 \end{bmatrix} \qquad A''(t) = \begin{bmatrix} 0 & 6 \\ 24t & 60t^2 \end{bmatrix} \qquad A'(0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

Integrate by integrating each component:

$$\int A(t)dt = \begin{bmatrix} t^2 + c_1 & t^3 + c_2 \\ t^4 + c_3 & t^5 + c_4 \end{bmatrix} \qquad \int_0^2 A(t)dt = \begin{bmatrix} t^2\big|_0^2 & t^3\big|_0^2 \\ t^4\big|_0^2 & t^5\big|_0^2 \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 16 & 32 \end{bmatrix}$$

Proposition 6.6.5.

Suppose A, B are matrix-valued functions of a real variable, f is a function of a real variable, c is a constant, and C is a constant matrix then

1. $(AB)' = A'B + AB'$ (product rule for matrices)
2. $(AC)' = A'C$
3. $(CA)' = CA'$
4. $(fA)' = f'A + fA'$
5. $(cA)' = cA'$
6. $(A + B)' = A' + B'$

where each of the functions is evaluated at the same time t and I assume that the functions and matrices are differentiable at that value of t and of course the matrices A, B, C are such that the multiplications are well-defined.

Proof: Suppose $A(t) \in \mathbb{R}^{m\times n}$ and $B(t) \in \mathbb{R}^{n\times p}$ consider,

$$\begin{aligned}
(AB)'_{ij} &= \tfrac{d}{dt}((AB)_{ij}) && \text{defn. derivative of matrix} \\
&= \tfrac{d}{dt}(\textstyle\sum_k A_{ik}B_{kj}) && \text{defn. of matrix multiplication} \\
&= \textstyle\sum_k \tfrac{d}{dt}(A_{ik}B_{kj}) && \text{linearity of derivative} \\
&= \textstyle\sum_k \left[\tfrac{dA_{ik}}{dt}B_{kj} + A_{ik}\tfrac{dB_{kj}}{dt}\right] && \text{ordinary product rules} \\
&= \textstyle\sum_k \tfrac{dA_{ik}}{dt}B_{kj} + \sum_k A_{ik}\tfrac{dB_{kj}}{dt} && \text{algebra} \\
&= (A'B)_{ij} + (AB')_{ij} && \text{defn. of matrix multiplication} \\
&= (A'B + AB')_{ij} && \text{defn. matrix addition}
\end{aligned}$$

this proves (1.) as i, j were arbitrary in the calculation above. The proof of (2.) and (3.) follow quickly from (1.) since C constant means $C' = 0$. Proof of (4.) is similar to (1.):

$$\begin{aligned}
(fA)'_{ij} &= \tfrac{d}{dt}((fA)_{ij}) && \text{defn. derivative of matrix} \\
&= \tfrac{d}{dt}(fA_{ij}) && \text{defn. of scalar multiplication} \\
&= \tfrac{df}{dt}A_{ij} + f\tfrac{dA_{ij}}{dt} && \text{ordinary product rule} \\
&= (\tfrac{df}{dt}A + f\tfrac{dA}{dt})_{ij} && \text{defn. matrix addition} \\
&= (\tfrac{df}{dt}A + f\tfrac{dA}{dt})_{ij} && \text{defn. scalar multiplication.}
\end{aligned}$$

The proof of (5.) follows from taking $f(t) = c$ which has $f' = 0$. I leave the proof of (6.) as an exercise for the reader. □.

To summarize: the calculus of matrices is the same as the calculus of functions with the small qualifier that we must respect the rules of matrix algebra. The noncommutativity of matrix multiplication is the main distinguishing feature.

6.6.3 calculus of complex-valued functions of a real variable

Differentiation of functions from $\mathbb{R}$ to $\mathbb{C}$ is defined by splitting a given function into its real and imaginary parts then we just differentiate with respect to the real variable one component at a time. For example:

$$\begin{aligned}
\frac{d}{dt}(e^{2t}\cos(t) + ie^{2t}\sin(t)) &= \frac{d}{dt}(e^{2t}\cos(t)) + i\frac{d}{dt}(e^{2t}\sin(t)) \\
&= (2e^{2t}\cos(t) - e^{2t}\sin(t)) + i(2e^{2t}\sin(t) + e^{2t}\cos(t)) \qquad (6.1) \\
&= e^{2t}(2+i)(\cos(t) + i\sin(t)) \\
&= (2+i)e^{(2+i)t}
\end{aligned}$$

where I have made use of the identity[7] $e^{x+iy} = e^x(\cos(y) + i\sin(y))$. We just saw that $\frac{d}{dt}e^{\lambda t} = \lambda e^{\lambda t}$ which seems obvious enough until you appreciate that we just proved it for $\lambda = 2 + i$.

[7] or definition, depending on how you choose to set-up the complex exponential, I take this as the definition in calculus II