MATH 2XX3 - Advanced Calculus II

Sang Woo Park

January 23, 2017

Contents

1 Introduction

In this course, we wish to study calculus using the concepts from linear algebra.

1.1 Vector norm

Definition 1.1. Euclidean norm *of* $\vec{x} = (x_1, x_2, \dots, x_n)$ *is given as*

$$\|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{\sum_{j=1}^{n} x_j^2}$$

Theorem 1.1 (Properties of a norm)**.**

1. $\|\vec{x}\| \geq 0$ *and* $\|\vec{x}\| = 0$ *iff* $\vec{x} = \vec{0} = (0, 0, \dots, 0)$.
2. *For all scalars* $a \in \mathbb{R}$, $\|a\vec{x}\| = |a| \cdot \|\vec{x}\|$.
3. *(Triangle inequality)* $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$.

We say that this is a property of a norm *because there are other norms, which measure* distance *in* $\mathbb{R}^n$ *in different ways!*

Example 1.1.1 (A non-pythagorian norm - *The Taxi Cab Norm*)**.** Consider the following vector $\vec{p} = (p_1, p_2) \in \mathbb{R}^2$. The euclidean norm gives the length of the diagonal line. On the other hand,

$$\|\vec{p}\|_1 = |p_1| + |p_2|$$

gives us the total distance in a rectangular grid system.

For $\vec{p} = (p_1, p_2, \dots, p_n) \in \mathbb{R}^n$, $\|\vec{p}\|_1 = \sum_{j=1}^{n} |p_j|$. Note that the Taxi Cab norm is a valid norm because it satisfies all properties of a norm above. So it also gives us a valid alternative way to measure *distance* in $\mathbb{R}^n$, $\mathrm{dist}(\vec{p}, \vec{q}) = \|\vec{p} - \vec{q}\|$. This way of measuring distance gives $\mathbb{R}^n$ a *different geometry.*

Definition 1.2. Neighborhood *of a point* $\vec{p}$*, or disks centered at* $\vec{p}$ *is defined as*

$$D_r(\vec{p}) = \left\{ \vec{x} \in \mathbb{R}^n \middle| \|\vec{x} - \vec{p}\| < r \right\}$$

Remark. The neighborhood around $\vec{a}$ of radius r may be written using any of the following notations:

$$D_r(\vec{a}) = B_r(\vec{a}) = B(\vec{a}, r)$$

Definition 1.3. *Sphere is defined as*

$$S_r(\vec{p}) = \left\{ \vec{x} \in \mathbb{R}^n \middle| \|\vec{x} - \vec{p}\| = r \right\}$$

What neighboorhood and sphere look like depends on which norm you choose. First, let's start with the familiar euclidean norm. Then, the sphere is given by

$$\begin{aligned}
&\|\vec{x}-\vec{p}\| = r \\
\iff &\sqrt{\sum_{j=1}^{n}(x_j - p_j)^2} = r
\end{aligned}$$

Then, we have

$$\sum_{j=1}^{n}(x_j - p_j)^2 = r^2$$

If $n = 3$, we have $(x_1 - p_1)^2 + (x_2 - p_2)^2 + (x_3 - p_3)^2 = r^2$, usual sphere in $\mathbb{R}^3$ with center $\vec{p} = (p_1, p_2, p_3)$

If $n = 2$, we have $(x_1 - p_1)^2 + (x_2 - p_2)^2 = r^2$, usual circle in $\mathbb{R}^n$ with center $\vec{p} = (p_1, p_2)$.

If we replace Euclidean norm by the Taxi Cab norm (for simplicity, take $\vec{p} = \vec{0}$), we have

$$S_r^{\text{taxi}}(\vec{0}) = \left\{\vec{x} \in \mathbb{R}^n \middle| \|\vec{x} - \vec{0}\|_1 = r\right\}$$

In other words, we have

$$\vec{x} \in S_r^{\text{taxi}}(\vec{0}) \iff \sum_{j=1}^{n} |x_j| = r$$

Looking at it in $\mathbb{R}^2$, we have $\vec{x} = (x_1, x_2)$. Then, $r = |x_1| + |x_2|$. This, in fact, is a diamond.

Remark. Note that $|x_1| + |x_2| = r$ is a *circle* in $\mathbb{R}^2$ under the Taxi Cab norm. Then, we have

$$\pi = \frac{\text{circumference}}{\text{diameter}} = \frac{8r}{2r} = 4$$

1.2 Subset

Let's introduce some properties of *subsets* in $\mathbb{R}^n$. $A \subset \mathbb{R}^n$ means A is a *collection* of points $\vec{x}$, drawn from $\mathbb{R}^n$.

Definition 1.4. *Let $A \subset \mathbb{R}^n$, and $\vec{p} \in A$. We say $\vec{p}$ is an interior point of A if there exists a neighbourhood of $\vec{p}$, i.e. an open disk disk, which is entirely contained in A:*

$$D_r(\vec{p}) \subset A.$$

Example 1.2.1.

$$A = \left\{ \vec{x} \in \mathbb{R}^n | \vec{x} \neq \vec{0} \right\}$$

Take any $\vec{p} \in A$, so $\vec{p} \neq \vec{0}$. Then, let $r = \|\vec{p} - \vec{0}\| > 0$, and $D_r(\vec{p}) \subset A$, since $\vec{0} \notin D_r(\vec{p})$. (Notice: any smaller disk, $D_s(\vec{p}) \subset D_r(\vec{p}) \subset A$, where $0 < s < r$ works to show that $\vec{p}$ is an interior point).

So every $\vec{p} \in A$ is an interior point to A.

Definition 1.5. *If every $\vec{p} \in A$ is an interior point, we cal A an* open set*.*

Example 1.2.2. $A = \left\{ \vec{x} \in \mathbb{R}^n | \vec{x} \neq \vec{0} \right\}$ is an open set.

Example 1.2.3. $A = D_R(\vec{0})$ is an open set.

Proof. If $\vec{p} = \vec{0}$, $D_r(\vec{0}) \subseteq A = D_R(\vec{0})$ provided $r \leq R$. So $\vec{p} = \vec{0}$ is interior to A. Consider any other $\vec{p} \in A$. It's evident that $D_r(\vec{p}) \subset A = D_R(\vec{0})$ provided that $0 \leq r \leq R - \|\vec{p}\|$. Therefore, $A = D_R(\vec{0})$ is an open set. □

Example 1.2.4. Suppose we use Taxi Cab disks instead of Euclidean disk. It does not change which points are interior to A since the diamond is inscribed in a circle. In other words,

$$D_r^{\text{taxi}}(\vec{p}) \subset D_r^{\text{Euclid}}(\vec{p})$$

Definition 1.6. *The complement of set A is*

$$A^c = \{\vec{x} | \vec{x} \notin A\}$$

Definition 1.7. *$\vec{b}$ is a* boundary point *of A if for every $r > 0$, $D_r(\vec{b})$ contains both points in A and points not in A:*

$$D_r(\vec{b}) \cap A \neq \emptyset \textit{ and } D_r(\vec{b}) \cap A^c \neq \emptyset$$

In the example 1.2.3, the set of all boundary points of $A = D_R(\vec{0})$

$$\left\{ \vec{b} \,\middle|\, \|\vec{b}\| = R \right\}$$

is a sphere of radius R.

Definition 1.8. *A set A is closed if A^c is open.*

Theorem 1.2. *A is cloased if and only if A contains all its boundary points.*

Example 1.2.5. Consider the following set:

$$A = \{(x_1, x_2) \in \mathbb{R}^2 \big| x_1 \geq 0, x_2 > 0\}$$

If $\vec{p_1} = (p_1, p_2)$, where $p_1 > 0, p_2 > 0$, then $\vec{p_1}$ is an interior point. Take $r = \min\{p_1, p_2\}$. Then, $D_r(\vec{p}) \subset A$. On the other hand, any $\vec{p}$ that lies on either axes (including $\vec{0}$) is a boundary point. Since there are boundary points in A, A can't be open.

2 Functions

2.1 Limits and continuity

In this section, we will be considering vector values functions such that

$$F : A \subseteq \mathbb{R}^n \to \mathbb{R}^k.$$

We can rewrite this using a matrix notation:

$$F(x_1, x_2, \ldots, x_n) = \begin{bmatrix} F_1(x_1, x_2, \ldots, x_n) \\ F_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ F_k(x_1, x_2, \ldots, x_n) \end{bmatrix}$$

Example 2.1.1. For a $(k \times n)$ matrix M,

$$F(\vec{x}) = M\vec{x}$$

First, we wish to study limits. What does $\lim_{\vec{x}\to\vec{a}} F(\vec{x}) = \vec{L}$ mean? Note that it's not enough to treat the variables $x_1, x_2, \ldots x_n$ separately.

Example 2.1.2. Consider the following function:

$$F(x, y) = \frac{xy}{x^2 + 4y^2},$$

where $(x, y) \neq (0, 0)$. First, we can attempt to find its limit by considering them separately.

$$\lim_{x\to 0}\left(\lim_{y\to 0} F(x, y)\right) = \lim_{x\to 0}\left(\frac{0}{x^2}\right) = \lim_{x\to 0} = 0$$

Similarly, we have

$$\lim_{y\to 0}\left(\lim_{x\to 0} F(x, y)\right) = 0$$

However, if $(x, y) \to (0, 0)$ along a straight line path with $y = mx$, where m is constant, we have

$$F(x, mx) = \frac{mx^2}{x^2 + 4m^2x^2} = \frac{m}{1 + 4m^2}$$

In this case, we have

$$\lim_{\substack{(x,y)\to(0,0) \\ \text{along } y=mx}} F(x, y) = \frac{m}{1 + 4m^2}$$

Therefore, the values of $F(x, y)$ don't approach any particular value as $(x, y) \to (0, 0)$.

Example 2.1.3 (Worse)**.** Consider the following function:

$$F(x, y) = \frac{y^2}{x^4 + y^2}.$$

If we approach $(0, 0)$ along $y = mx$, limit equals 1. However, if we approach along a parabola, $y = mx^2$, limit equals $m^2/(1 + m^2)$. We get different limits alond different parabolas.

We showed that computing

$$\lim_{\vec{x} \to \vec{a}} = \vec{b}$$

is tricky because $\vec{x} \to \vec{a}$ has to be more precise. It can't depend on the path or direction on which $\vec{x}$ approaches $\vec{a}$, but only on *proximity*. In other words, we want $\|F(\vec{x}) - \vec{b}\|$ to go to zero as $\|\vec{x} - \vec{a}\|$ goes to zero.

Definition 2.1. *We say* $\lim_{\vec{x} \to \vec{a}} F(\vec{x}) = \vec{b}$ *if for any given* $\varepsilon > 0$*, there exists* $\delta > 0$ *such that whenever* $0 < \|\vec{x} - \vec{a}\| < \delta$*, we have* $\|F(x) - \vec{b}\| < \varepsilon$*. Therefore,*

$$\lim_{\vec{x} \to \vec{a}} F(x) = \vec{b} \iff \lim_{\vec{x} \to \vec{a}} \|F(\vec{x}) - \vec{b}\| = 0$$

Remark. Geometrically, for any given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$F(\vec{x}) \in D_\varepsilon(\vec{b}),$$

where $\vec{x} \in D_\delta(\vec{a})$.

Before doing examples, here's a useful observations. Take $\vec{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$. Then, we have

$$\|\vec{v}\| = \sqrt{\sum_{j=1}^{n} v_j^2} \geq \sqrt{v_i^2} = |v_i|$$

for each coordinate $i = 1, 2, \ldots, n$.

Example 2.1.4. Show

$$\lim_{(x,y) \to (0,0)} \frac{2x^2 y}{x^2 + y^2} = 0$$

Proof. Note that $F : \mathbb{R} \setminus \{\vec{0}\} \to \mathbb{R}$, $b = 0$, $\vec{a} = (0, 0)$. Call

$$R = \|\vec{x} - \vec{a}\| = \|\vec{x}\| = \sqrt{x^2 + y^2}$$

Since $F(\vec{x}) \in \mathbb{R}$, we have

$$
\begin{aligned}
\|F(\vec{x}) - \vec{b}\| &= |F(\vec{x}) - b| \\
&= \left|\frac{2x^2y}{x^2+y^2} - 0\right| \\
&= \frac{2|x|^2|y|}{x^2+y^2} \\
&\leq \frac{2 \cdot R^2 \cdot R}{R^2} \\
&= 2R \\
&= 2\|\vec{x} - \vec{a}\|
\end{aligned}
$$

By letting $\|\vec{x} - \vec{a}\| = \|\vec{x}\| < \varepsilon/2$, we get $\|F(\vec{x}) = \vec{b}\| < \varepsilon$. Therefore, definition is satisfied with $\delta \leq \varepsilon/2$ □

Example 2.1.5. Consider the following function, $F : \mathbb{R}^3 \setminus \{\vec{0}\} \to \mathbb{R}$:

$$\frac{3z^2 + 2x^2 + 4y^2 + 6z^2}{x^2 + 2y^2 + 3z^2}.$$

Determine whether

$$\lim_{(x,y,z)\to(0,0,0)} F(x,y,z) = 2.$$

Proof. We have

$$
\begin{aligned}
\|F(x,y,z) - \vec{b}\| &= |F(x,y,z) - 2| \\
&= \left|\frac{3z^3 + 2x^2 + 4y^2 + 6z^2}{x^2 + 2y^2 + 3z^2} - 2\right| \\
&= \frac{3|z|^3}{x^2 + 2y^2 + 3z^2} \\
&\leq \frac{3R^3}{x^2 + y^2 + z^3} \\
&= \frac{3R^3}{R^2} \\
&= 3R
\end{aligned}
$$

Then,

$$\|F(x,y,z) - \vec{b}\| < 3R < \varepsilon$$

provided that

$$R = \|\vec{x} - \vec{0}\| < \delta = \frac{\varepsilon}{3}$$

□

Definition 2.2. *Consider a function $F : \mathbb{R}^n \to \mathbb{R}^k$ with domain $A \subseteq \mathbb{R}^n$. For $\vec{a} \in A$, we say F is* continuous *at $\vec{a}$ in the domain of F iff*

$$F(a) = \lim_{\vec{x} \to \vec{a}} F(\vec{a})$$

Example 2.1.6. Going back the example 2.1.5, if we redefine F as follows,

$$F = \begin{cases} \frac{3z^2+2x^2+4y^2+6z^2}{x^2+2y^2+3z^2} & (x,y,z) \neq (0,0,0) \\ 2 & (x,y,z) = (0,0,0) \end{cases}$$

then F is continuous at $(0,0,0)$ (and in fact at all $\vec{x} \in \mathbb{R}$).

If F is continuous at every $\vec{a} \in A$, $(\forall \vec{x} \in A)$, we say F is continuous on the set A. Continuity is always preserved by the usual algebraic operations: sum. product, quotient, and composition of continuous functions is continuous [1] .

2.2 Differentiability

Definition 2.3. *For a function $f : \mathbb{R} \to \mathbb{R}$, its derivative is defined as*

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

If it exists, we say f is differentiable *at x.*

Theorem 2.1. *If f is differentiable at x, $f(x)$ is also continuous at x.*

Note that differentiable functions, $f(x)$, are well approximated by theire tangent lines (also known as linearization). We wish to extend this idea to $F : \mathbb{R}^n \to \mathbb{R}^m$.

First, we try dealing with the independent varaibles, $x_1, x_2, \ldots, x_n$, one at a time by using partial derivatives. We start by introducing the standard basis in $\mathbb{R}^n$:

$$\begin{aligned} \vec{e}_1 &= (1,0,0,\ldots,0) \\ \vec{e}_2 &= (0,1,0,\ldots,0) \\ &\vdots \\ \vec{e}_n &= (0,0,0,\ldots,1) \end{aligned}$$

(In $\mathbb{R}^3$, $\vec{e}_1 = \vec{i}, \vec{e}_2 = \vec{j}, \vec{e}_3 = \vec{k}$).

For any $\vec{x} \in \mathbb{R}^n$, and $h \in \mathbb{R}$, $(\vec{x} + h\vec{e}_j)$ moves from $\vec{x}$ parallel to the x_j axis by distance h. In other words,

$$\vec{x} + h\vec{e}_j = (x_1, x_2, \ldots, x_j + h, x_{j+1}, \ldots, x_n).$$

[1] Provided we remain in the domain of continuity of both functions and denominators aren't zero

Definition 2.4. *Partial derivatives of $f(x)$ is defined as*

$$\frac{\partial f}{\partial x_j}(\vec{x}) = \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e}_j) - f(\vec{x})}{h},$$

for all $j = 1, 2, \ldots, n$.

Partial derivatives calculate the derivatives of f, treating of $\vec{x}_j$ as the only varaible, and all others treated as constants. For a vector valued function $F : \mathbb{R}^n \to \mathbb{R}^m$:

$$F(\vec{x}) = \begin{bmatrix} F_1(\vec{x}) \\ F_2(\vec{x}) \\ \vdots \\ F_m(\vec{x}) \end{bmatrix},$$

we treat each component $F_i(\vec{x}) : \mathbb{R}^n \to \mathbb{R}$ separately as a real valued function. Each has n partial derivatives, and so $F : \mathbb{R}^n \to \mathbb{R}^m$ has $(m \times n)$ partial derivatives, which form an $(m \times n)$ matrix:

$$\left(\frac{\partial F_i}{\partial x_j} \right)_{\substack{i=1,2,\ldots,m \\ j=1,2,\ldots,n}}.$$

We call this the derivative matrix or *Jacobian matrix*, $DF(\vec{x})$.

Example 2.2.1. Consider a function $F : \mathbb{R}^2 \to \mathbb{R}^3$:

$$F(\vec{x}) = \begin{bmatrix} x_1^2 \\ x_1 x_2 \\ x_2^4 \end{bmatrix}.$$

Jacobian of the function is given by

$$\begin{aligned} DF(\vec{x}) &= \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} \\ \frac{\partial F_3}{\partial x_1} & \frac{\partial F_3}{\partial x_2} \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 & 0 \\ x_2 & x_1 \\ 0 & 4x_2^3 \end{bmatrix} \end{aligned}$$

Do we get the same properties for $DF(\vec{x})$ as we did for single-value calculus?

Example 2.2.2. Consider the following function:

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

Do the partial derivatives exist at $(0, 0)$?

By definition,

$$\begin{aligned}\frac{\partial f}{\partial x}(0,0) &= \lim_{h\to 0}\frac{f(0+h,0)-f(0,0)}{h}\\ &= \lim_{h\to 0}\frac{\frac{h\cdot 0}{(h^2+0^2)^2}-0}{h}\\ &= \lim_{h\to 0}\frac{0}{h}=0\end{aligned}$$

Similarly, $\frac{\partial f}{\partial y}(0,0)=0$ (symmetry of x,y). Therefore,

$$Df(0,0)=\begin{bmatrix}0 & 0\end{bmatrix}$$

Although partial derivatives exist, f is not cotinuous at $(0,0)$! (For example, $f(x,mx)\to\pm\infty$ as $x\to 0^{\pm}$ for $m\neq 0$).

To get reasonable information from $Df(\vec{x})$, we need to say more. First, let's go back to $f:\mathbb{R}\to\mathbb{R}$. Note

$$\begin{aligned}f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}\\ \iff &\lim_{h\to 0}\left(\frac{f(x+h)-f(x)}{h}-f'(x)\right)=0\\ \iff &\lim_{h\to 0}\left(\frac{f(x+h)-[\overbrace{f(x)+hf'(x)}^{L}]}{h}\right)=0\end{aligned}$$

Numerator is the difference between $f(x+h)$ and its linear approximation, L (i.e. the tangent line). So f is differentiable at x if its linear approximation gives an estimate of the values $f(x+h)$ within an error which is small compared to $\Delta x=h$. More precisely, the linearization of $f(x)$ at $x=a$ (or the tangent line) is given by

$$L_a(x)=f(a)+f'(a)(x-a)$$

We wish to extend this idea to higher dimensions. For $F:\mathbb{R}^n\to\mathbb{R}^m$, $F(\vec{x})$ has $(m\times n)$ partial derivates (see definition 2.4). Then, the linearization of F at $\vec{a}$ is

$$L_{\vec{a}}(\vec{x})=\underbrace{F(\vec{a})}_{m\times 1}+\underbrace{DF(\vec{a})}_{m\times n}(\underbrace{\vec{x}-\vec{a}}_{n\times 1}).$$

So, $L:\mathbb{R}^n\to\mathbb{R}^m$, just like F. The derivative matrix $DF(\vec{a})$ is a *linear transformation* of $\mathbb{R}^n\to\mathbb{R}^m$.

Notice that when $n=2$ and $m=1$, For $F:\mathbb{R}^2\to\mathbb{R}$, we have

$$DF(\vec{a})=\begin{bmatrix}\frac{\partial F}{\partial x_1}(\vec{a}) & \frac{\partial F}{\partial x_2}(\vec{a})\end{bmatrix},$$

a (1×2) row vector and

$$\vec{x}-\vec{a}=\begin{bmatrix}x_1-a_1\\x_2-a_2\end{bmatrix},$$

so we have

$$L_{\vec{a}}(\vec{x})=F(\vec{a})+\frac{\partial F}{\partial x_1}(x_1-a_1)+\frac{\partial F}{\partial x_2}(x_2-a_2),$$

a familiar equation of the tangent plane to $z=F(x_1,x_2)$.

Finally, we can introduce the idea of differentiable:

Definition 2.5 (Differentiability)**.** *We say $F:\mathbb{R}^n\to\mathbb{R}^m$ is* differentiable *if*

$$\lim_{\vec{x}\to\vec{a}}\frac{\|F(\vec{x})-F(\vec{a})-DF(\vec{a})(\vec{x}-\vec{a})\|}{\|\vec{x}-\vec{a}\|}=0.$$

Equivalently,

$$\lim_{\vec{h}\to\vec{0}}\frac{\|F(\vec{x}+\vec{h})-F(\vec{x})=DF(\vec{x})\vec{h}\|}{\|\vec{h}\|}=0$$

In summary, F is differentiable if $\|F(\vec{x})-L_{\vec{a}}(\vec{x})\|$ is small compared to $\|\vec{x}-\vec{a}\|$. Or, $F(\vec{x})$ is approximated by $L_{\vec{a}}(\vec{x})$ with and error which is much smaller than $\|\vec{x}-\vec{a}\|$. Note that we write $o\left(\|\vec{x}-\vec{a}\|\right)$ "little-oh" for quantity which is small compred to $\|\vec{x}-\vec{a}\|$. Using this notation, differentiability can be written as

$$\|F(\vec{x})-F\left(\vec{a}-Df(\vec{a})(\vec{x}-\vec{a})\right)\|=o(\|\vec{x}-\vec{a}\|).$$

Example 2.2.3. Is the following function differentiable at $\vec{a}=\vec{0}$?

$$F(x_1,x_2)=\begin{cases}\frac{x_2^2\sin x_1}{\sqrt{x_1^2+x_2^2}}, & \vec{x}\neq\vec{0}\\ 0, & \vec{x}=\vec{0}\end{cases}$$

First, we have

$$\begin{aligned}\frac{\partial F}{\partial x_1}(\vec{0})&=\lim_{h\to 0}\frac{F(\vec{0}+h\vec{e}_1)-F(\vec{0})}{h}\\&=\lim_{h\to 0}\frac{0-0}{h}=0\end{aligned}$$

Similarly, we have

$$\frac{\partial F}{\partial x_2}(\vec{0})=0$$

So we have

$$DF(\vec{0})=\begin{bmatrix}\frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_2}\end{bmatrix}=\begin{bmatrix}0 & 0\end{bmatrix}$$

For differentiability, we have to look at:

$$\left| \frac{x_2^2 \sin x_1}{\sqrt{x_1^2 + x_2^2}} - 0 - \begin{bmatrix} 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right|$$
$$= \frac{x_2^2 |\sin x_1|}{\sqrt{x_1^2 + x_2^2}}$$

Then,

$$\begin{aligned} \frac{|F(\vec{x}) - L_{\vec{0}}(\vec{x})|}{\vec{x} - \vec{0}} &= \frac{x_2^2 |\sin x_1|}{\left(\sqrt{x_1^2 + x_2^2}\right)^2} = \frac{x_2^2 |\sin x_1|}{x_1^2 + x_2^2} \\ &\leq \frac{R^2 \cdot R}{R^2} = R = \|\vec{x} - \vec{0}\| \end{aligned}$$

By squeeze theorem, we have

$$\lim_{\vec{x} \to \vec{0}} \frac{|F(\vec{x}) - L_{\vec{0}}(\vec{x})|}{\|\vec{x} - \vec{0}\|} = 0$$

Therefore, F is differentiable at $\vec{x} = \vec{0}$

Example 2.2.4. Verify that F is differentiable at $\vec{a} = \vec{0}$.

$$F(\vec{x}) = \begin{bmatrix} 1 + x_1 + x_2^2 \\ 2x_2 - x_1^2 \end{bmatrix}$$

First, note that

$$F(\vec{a}) = F(\vec{0}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We also need to compute the Jacobian at $\vec{0}$:

$$DF(\vec{0}) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Then, we get the following linearization of the function:

$$\begin{aligned} L_{\vec{0}}(\vec{x}) &= F(\vec{0}) + DF(\vec{x})(\vec{x} - \vec{0}) \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 + x_1 \\ 2x_2 \end{bmatrix} \end{aligned}$$

Then, look at

$$\frac{\|F(\vec{x}) - L_{\vec{0}}(\vec{x})\|}{\|\vec{x} - \vec{0}\|} = \frac{\left\| \begin{bmatrix} x_2^2 \\ -x_1^2 \end{bmatrix} \right\|}{\|\vec{x}\|} = \frac{\sqrt{x_2^4 + x_1^4}}{\sqrt{x_1^2 + x_2^2}} \leq \frac{R^4 + ^4}{R} = \sqrt{2} R = \sqrt{2} \|\vec{x} - \vec{0}\|$$

As $\vec{x} \to \vec{0}$, $\|\vec{x} - \vec{0}\| = R \to 0$, so by the squeeze theorem, the desired limit goes to 0. Therefore, F is differentiable at $\vec{0}$.

Theorem 2.2. *Suppose $F : \mathbb{R}^n \to \mathbb{R}^m$, and $\vec{a} \in \mathbb{R}^n$. If there exists a disk $D_r(\vec{a})$ in which all the partial derivatives $\partial(F_i(\vec{x}))/\partial x_j$ exist and are continuous, then F is differentiable at $\vec{x} = \vec{a}$.*

Definition 2.6. *A function which satisfies Theorem 2.2 is called* continuously differentiable*, of C^1.*

So far as our example, we calculate the partial for $\vec{x} \neq \vec{0}$:

$$\begin{aligned}\frac{\partial F}{\partial x_1} &= x_2^2 \left(\cos x_1 \left(x_1^2 + x_2^2\right)^{-\frac{1}{2}} + \left(-\frac{1}{2} \left(x_1^2 + x_2^2\right)^{-\frac{3}{2}} 2x_1 \right) \sin x_1 \right) \\ &= \frac{x_2^2}{(x_1^2 + x_2^2)^{3/2}} \left[\cos x_1 \left(x_1^2 + x_2^2\right) - x_1 \sin x_1 \right]\end{aligned}$$

which is continuous as long as $\vec{x} \neq \vec{0}$. We do the same for $\frac{\partial F}{\partial x_2}$ and conclude that F is C^1 at all $\vec{x} \neq \vec{0}$. We summarize these ideas in the figure below:

2.3 Chain rule

Definition 2.7. *Suppose $A \subseteq \mathbb{R}^n$ is open, and we have a function*

$$F : A \subseteq \mathbb{R}^n \to \mathbb{R}^m.$$

Similarly, supposed $B \subseteq \mathbb{R}^m$ is open, and we have a function

$$G : B \subseteq \mathbb{R}^m \to \mathbb{R}^p.$$

Assume $\vec{a} \in A$ and $F(\vec{a}) = \vec{b} = B$. The composition

$$H(\vec{x}) = G \circ F(\vec{x}) = G\left(F(\vec{x})\right)$$

is a function $\mathbb{R}^n \to \mathbb{R}^p$.

Example 2.3.1. Consider the following linear functions:

$$\begin{cases} F(\vec{x}) = M\vec{x} & M \text{ an } (m \times n) \text{ matrix} \\ G(\vec{y}) = N\vec{y} & N \text{ an } (p \times m) \text{ matrix} \end{cases}$$

Then,

$$H(\vec{x}) = G\left(F(\vec{x})\right) = NM\vec{x}$$

is also a linear and represented by the product NM.

Theorem 2.3. *Assume $F : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\vec{x} = \vec{a}$ and $G : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $\vec{b} = F(\vec{a})$. Then, $H = F \circ G$ is differentiable at $\vec{x} = \vec{a}$ and*

$$DH(\vec{a}) = \underbrace{DG(\vec{b})}_{DG(F(\vec{a}))} \; DF(\vec{a})$$

Note that all of the various forms of Chain Rule done in first year calculus can be derived directly from this general formula.

Example 2.3.2. Consider the following functions, $F : \mathbb{R}^3 \to \mathbb{R}^2$ and $G : \mathbb{R}^2 \to \mathbb{R}^2$:

$$F(\vec{x}) = \begin{bmatrix} x_1^2 + x_2 x_3 \\ x_1^2 + x_3^2 \end{bmatrix}, G(\vec{y}) = \begin{bmatrix} -y_2^3 \\ y_1 + y_2 \end{bmatrix}$$

Let $H = G \circ F(\vec{x})$. Find $DH(\vec{a})$ where $a = (1, -1, 0)$.

First, we have

$$DF(\vec{x}) = \begin{bmatrix} 2x_1 & x_3 & x_2 \\ 2x_1 & 0 & 2x_3 \end{bmatrix}, DF(1, -1, 0) = \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & 0 \end{bmatrix}$$

Similarly, we have

$$DG(\vec{y}) = \begin{bmatrix} 0 & -3y_2 \\ 1 & 1 \end{bmatrix}, DG(1, 1) = \begin{bmatrix} 0 & -3 \\ 1 & 1 \end{bmatrix}$$

By Chain Rule, we get

$$\begin{aligned} DH(1, -1, 0) &= DG(1, 1) DF(1, -1, 0) \\ &= \begin{bmatrix} 0 & -3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -6 & 0 & 0 \\ 4 & 0 & -1 \end{bmatrix} \end{aligned}$$

2.4 Directional derivatve

Definition 2.8. *A* path *is $\vec{C} : \mathbb{R} \to \mathbb{R}^n$ is a vector-valued function of a scalar independent variable, usually, t:*

$$\vec{c}(t) = \begin{bmatrix} c_1(t) \\ c_2(t) \\ \vdots \\ c_n(t) \end{bmatrix}$$

$\vec{c}(t)$ can be thought of as a moving vector. It takes out a curve in $\mathbb{R}^n$ as t increases. Basically, path is a way of describing a curve using functions. Note that this is not the only way to describe a curve.

Example 2.4.1. A unit circle in $\mathbb{R}^2$ described as a path is

$$\vec{c}(t) = (\cos t, \sin t),$$

where $t \in [0, 2\pi)$. But we could also describe the unit circle *non-parametrically* as

$$x^2 + y^2 = 1$$

Note that the same curve can be described by diffrent paths. Going back to unit circle, we can also write

$$\vec{b}(t) = \left(\sin(t^2), \cos(t^2)\right).$$

Using different paths can change (1) time dyanmics and (2) direction of the curve. This curve has a non-constant speed and reversed orientation.

If $\vec{c}$ is differentiable, $D\vec{c}(t)$ is an $(n \times 1)$ matrix. Since each compoenent $\vec{c}_j(t)$ is a real-valued function of only one variable, the *partial-derivative* is the usual derivative:

$$\frac{\partial c_j}{\partial t} = \frac{dc_j}{dt} = c_j'(t) = \lim_{h \to 0} \frac{c_j(t+h) - c_j(t)}{h}$$

So $D\vec{c}(t) = \vec{c}'(t)$ is written as a column vector:

$$\begin{aligned} D\vec{c}(t) &= \begin{bmatrix} c_1'(t) \\ c_2'(t) \\ \vdots \\ c_3'(t) \end{bmatrix} \\ &= \lim_{h \to 0} \frac{\vec{c}(t+h) - \vec{c}(t)}{h} \end{aligned}$$

which is a vector which is tangent to the curve traced out at $\vec{x} = \vec{x}(t)$. Physically, $\vec{c}'(t)$ is the velocity vector for motion along the path.

Example 2.4.2 (Lines)**.** Given two points, $\vec{p}_1, \vec{p}_2 \in \mathbb{R}^n$, there is a unique line connecting them. One path which represents this line is

$$\vec{c}(t) = \vec{p}_1 + t\vec{v},$$

where $\vec{v} = \vec{p}_2 - \vec{p}_1$. Velocity is then given by $\vec{c}'(t) = \vec{v}$, a constant.

Definition 2.9. $f : \mathbb{R}^n \to \mathbb{R}$ *is a* scalar-valued function.

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, $Df(\vec{x})$ is a $(1 \times n)$ matrix:

$$Df(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

We use *paths* $\vec{c}(t)$ to explore $f(x)$ by looking at

$$h(t) = f \circ \vec{c}(t) = f\left(\vec{c}(t)\right).$$

where $h : \mathbb{R} \to \mathbb{R}$, By chain rule,

$$\begin{aligned} Dh(t) = h'(t) &= \underbrace{Df(\vec{c}(t))}_{1\times n}\underbrace{D\vec{c}(t)}_{n\times 1} \\ &= Df\left(\vec{c}(t)\right)\vec{c}\,'(t) \\ &= \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \begin{bmatrix} c_1' \\ c_2' \\ \vdots \\ c_n' \end{bmatrix} \end{aligned}$$

We can think of this as a dot product of $\vec{c}\,'(t)$ with a vector $Df^T = \nabla f$, the gradient vector:

$$h'(t) = \nabla f(\vec{c}(t)) \cdot \vec{c}\,'(t)$$

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\vec{a} \in \mathbb{R}^n$, and we have a path $\vec{c} : \mathbb{R}^n \to \mathbb{R}$ with $\vec{c}(0) = \vec{a}$. Let $\vec{v} = \vec{c}\,'(0)$. Then, $h'(0)$ measures rate of change of f along the path as we cross through $\vec{a}$:

$$\begin{aligned} h'(0) &= \nabla f(\vec{c}(0)) \cdot \vec{c}\,'(0) \\ &= \nabla f(\vec{c}(0)) \cdot \vec{v} \end{aligned}$$

Note that we get the same vlaue for $h'(0)$ for any path $\vec{c}(t)$ going through $\vec{a}$ with velocity $\vec{v}$. In other words, $h'(0)$ says something about f at $\vec{a}$, and not the path $\vec{c}(t)$.

Definition 2.10 (Directional derivative)**.** *The directional derivative of f at $\vec{a}$ in direction $\vec{v}$ is*

$$D_{\vec{v}}f(\vec{a}) = Df(\vec{a})\vec{v} = \nabla f(\vec{a}) \cdot \vec{v}.$$

Using the Chain Rule, this can also be written as

$$D_{\vec{v}}f(\vec{a}) = \lim_{t\to 0} \frac{f(\vec{a} + t\vec{v}) - f(\vec{a})}{t}.$$

Note the similarity to partial derivatives, where $\vec{v} = \vec{e}_j$.