MATH 2XX3 - Advanced Calculus II

Sang Woo Park

March 2, 2017

Contents

1 Introduction

In this course, we are going 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 $\mathbb{R}^n$, $\text{dist}(\vec{p}, \vec{q}) = \|\vec{p} - \vec{q}\|$. 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. $\square$

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:*

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

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 $\delta \leq \varepsilon/2$ $\square$

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

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

$\square$

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 variables, $x_1, x_2, \ldots, x_n$, 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 variable, 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:

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

Then,

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

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:

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

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, $H = G \circ F$ 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

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

3 Paths and Curves

3.1 Directional derivatve

Definition 3.1. *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 3.1.1. 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 3.1.2 (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 3.2. *$f : \mathbb{R}^n \to \mathbb{R}$ 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 value for $h'(0)$ for *any* path $\vec{c}(t)$ going through $\vec{a}$ with velocity $\vec{c}\,'(t) = \vec{v}$. In other words, $h'(0)$ says something about f at $\vec{a}$, and not the path $\vec{c}(t)$.

Definition 3.3 (Directional derivative)**.** *The directional derivative of f at $\vec{a}$ in direction $\vec{v}$ is given by*

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

Now, we can make some observations. Using the Chain Rule, directional derivatives can be rewritten 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 $\vec{v} = \vec{e}_j$.

Second, notice that $D_{2\vec{v}}f(\vec{a}) = \nabla f(\vec{a}) \cdot (2\vec{v}) = 2D_{\vec{v}}f(\vec{a})$. To get the information on how fast f is changing at $\vec{a}$, we need to restrict to unit vectors $\|\vec{v}\| = 1$.

Directional derivatives also gives a geometrical interpretation of the gradient vector, $\nabla f(\vec{a})$. We use the Cauchy-Schwartz Inequality[2] to do so. By applying the Cauchy-Schwartz inequality, we get:

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

Therefore, we can conclude that the length of $\|\nabla f(\vec{a})\|$ is the largest of $D_{\vec{v}}f(\vec{a})$ among all choices of unit directions $\vec{v}$. In other words, the direction $\vec{v}$ in which $f(\vec{x})$ increases most rapidly is the direction of $\nabla f(\vec{a})$, i.e.

$$\vec{v} = \frac{\nabla f(\vec{a})}{\|\nabla f(\vec{a})\|},$$

provided that $\nabla f(\vec{a}) \neq \vec{0}$.

Similarly, $-\nabla f(\vec{a})$ points in the direction of largest of $f(\vec{x})$, i.e.

$$\vec{v} = -\frac{\nabla f(\vec{a})}{\|\nabla f(\vec{a})\|},$$

gives the most negative directional derivative.

3.2 Parameterized curve

A path, $\vec{c}(t)$, is {continuous, differentiable, and C^1} provided that each component $c_j(t)$, $j = 1, 2, \ldots, n$ are. Note that $\{\vec{c}(t) : t \in [a, b]\}$ traces out a curve in $\mathbb{R}^n$, with initial endpoint, $\vec{a}$, and final endpoint, $\vec{b}$. Therefore, the path $\vec{c}(t)$ *parameterizes* the curve drawn out.

Recall that for any function $F : \mathbb{R}^k \to \mathbb{R}^n$, differentiability means that tangent (i.e. linearization) makes a good aproximation. For a differentiable path, $\vec{c}\,'(t)$ is a tangent vector to the curve drawn out when $\vec{c}\,'(t) \neq 0$. We call $\vec{\mathrm{v}}(t) = \vec{c}\,'(t)$ the velocity vector ($v = \|\vec{\mathrm{v}}\| = \|\vec{c}\,'(t)\|$ is the speed).

Finally, we can define the unit tangent vector:

Definition 3.4. *Unit tangent vector is defined as*

$$\vec{T}(t) = \frac{\vec{\mathrm{v}}}{\|\vec{\mathrm{v}}(t)\|} = \frac{\vec{c}\,'(t)}{\|\vec{c}\,'(t)\|}$$

[2]For any vectors $\vec{u} \cdot \vec{v} \le \|\vec{y}\|\|\vec{v}\|$, and equality holds if and only if $\vec{u} = t\vec{v}$ for a scalar t.

Example 3.2.1. Consider a path, $\vec{c} : \mathbb{R} \to \mathbb{R}^2$:

$$\vec{c}(t) = \left(\cos^3 t, \sin^3 t\right), \ t \in [-\pi, \pi].$$

This is a C^1 path[3] whose velocity vector is given by

$$\vec{c}\,'(t) = \left(-3\cos^2 t \sin t, 3\sin^2 t \cos t\right).$$

To find the unit tangent, we have to find its speed first:

$$\begin{aligned} v &= \left\|\left(-3\cos^2 t \sin t, 3\sin^2 t \cos t\right)\right\| \\ &= 3|\sin t \cos t| \, \|(-\cos t, \sin t)\| \\ &= 3|\sin t \cos t| \end{aligned}$$

Then, its unit tangent is given by

$$\vec{T}(t) = \frac{\vec{\mathrm{v}}(t)}{\|\vec{\mathrm{v}}(t)\|} = \left(-|\cos t|\frac{\sin t}{|\sin t|}, |\sin t|\frac{\cos t}{|\cos t|}\right)$$

Note that its tangent is undefined when $\sin t = 0$ or $\cos t = 0$, i.e. at multiples of $\frac{\pi}{2}$. Worse, $\frac{\sin t}{|\sin t|}, \frac{\cos t}{|\cos t|}$ flip discontinuously as t crosses a multiple of $\pi/2$ from -1 to $+1$, or vice versa. Although the path is C^1, the curve is not smooth! When $\vec{\mathrm{v}}(t) = \vec{c}\,'(t) = 0$, it allows the curve to have cusps.

Note that it is possible to have a nice tangent direction even when $\vec{c}\,'(t) = 0$:

Example 3.2.2. Consider a parameterized straight line:

$$\vec{c} = \vec{a} + \vec{w}t^3$$

Its velocity vector, $\vec{c}\,'(t) = 3\vec{w}t^2$, is equal to $\vec{0}$ when $t = 0$. However, it still has a tangent direction which is parallel to $\vec{w}$.

Definition 3.5. *We say a parameterized curve is* smooth[4] *(or* regular*) if its path is C^1, i.e. if it can be parameterized by a path $\vec{c}(t)$, and $\|\vec{c}\,'(t) \neq 0\|$ for any t.*

If $\vec{c}(t)$ is twice-differentiable, $\vec{c}\,''(t) = \vec{a}(t)$ gives us the acceleration vector.

Theorem 3.1.

1. *If $f : \mathbb{R} \to \mathbb{R}$, $\vec{c} : \mathbb{R} \to \mathbb{R}^n$, both differentiable,*

$$\begin{aligned} \frac{d}{dt}\Big(f(t)\vec{c}(t)\Big) &= f(t)\vec{c}\,'(t) + f'(t)\vec{c}(t) \\ &= \sum_{j=1}^{n} \frac{d}{dt}\Big(f(t)\vec{c}_j(t)\Big)\vec{e}_j \\ &= \sum_{j=1}^{n} \frac{d}{dt}\Big(f'(t)\vec{c}_j(t) + f(t)\vec{c}_j\,'(t)\Big)\vec{e}_j \end{aligned}$$

[3] In fact, it is C^∞, differentiable to all orders!

[4] For a smooth curve, the unit tangent $\vec{T}(t)$ is continuous.

2. *If $\vec{c}, \vec{d} : \mathbb{R} \to \mathbb{R}^n$ are differentiable,*

$$\frac{d}{dt}\Big(\vec{c}(t) \cdot \vec{d}(t)\Big) = \vec{c}\,'(t) \cdot \vec{d}(t) + \vec{c}(t) \cdot \vec{d}\,'(t)$$

3. *If $\vec{c}, \vec{d} : \mathbb{R} \to \mathbb{R}^3$ are differentiable,*

$$\frac{d}{dt}\Big(\vec{c}(t) \times \vec{d}(t)\Big) = \vec{c}(t) \times \vec{d}\,'(t) + \vec{c}\,'(t)\vec{d}(t),$$

where $\vec{c} \times \vec{d} = \sum_{i,j,k=1}^{3} = c_i d_j \vec{e}_k \varepsilon_{ijk}$. ε_{ijk} is defined in the footnotes[5]*.*

Example 3.2.3. Suppose $\vec{c}$ is a twice differentiable path and $\vec{a}(t) = k\vec{c}(t)$ for some constant $k \neq 0$. Show that $\vec{c}(t)$ describes a motion in a fixed plane.

Define a vector

$$\vec{n} = \vec{c}(t) \times \vec{\mathrm{v}}(t) = \vec{c}(t) \times \vec{c}\,'(t)$$

Notice $\vec{n} \perp \vec{c}(t)$ and $\vec{\mathrm{v}}(t)$, i.e. $\vec{n}$ is normal to the plane.

$$\begin{aligned}\frac{d\vec{n}}{dt} = \frac{d}{dt}\left(\vec{c}(t) \times \vec{c}\,'(t)\right) &= \vec{c}(t) \times \underbrace{\vec{c}\,''(t)}_{\vec{a}(t)} + \underbrace{\vec{c}\,'(t) \times \vec{c}\,'(t)}_{\vec{0}} \\ &= \vec{c}(t) \times k\vec{c}(t) \\ &= \vec{0}\end{aligned}$$

Therefore, $\vec{n}$ is constant in time!

So $\vec{c}(t)$ and $\vec{\mathrm{v}}(t)$ are, for all t, perpendicular to the constant vector $\vec{n}$. Then,

$$P = \{\vec{w} \mid \vec{w} \cdot \vec{n} = 0\}$$

is the plane through $\vec{0}$. So $\vec{c}(t) \in P$ for all t.

Definition 3.6 (Arclength)**.** *The arclength (or distance travelled along the parameterized curve) for $a \leq t \leq b$ is*

$$\int_a^b \underbrace{\|\vec{c}\,'(t)\|}_{speed}\, dt$$

For a variable time interval, the arclength function

$$s(t) = \int_a^t \|\vec{c}\,'(u)\| du$$

is a distance travelled from time a to time t.

---

[5] $\varepsilon_{ijk} = \begin{cases} 0 & \text{if } i = j \text{ or } j = k \text{ or k = l} \\ 1 & \text{if } (i,j,k) \text{ is positively ordered} \\ -1 & \text{if } (i,j,k) \text{ is negatively ordered} \end{cases}$

Example 3.2.4. Consider the following path:

$$\vec{c}(t) = (3\cos t, 3\sin t, 4t),\ t \in [0, 4\pi].$$

Its velocity vector is given by

$$\vec{\mathrm{v}}(t) = (-3\sin t, 3\cos t, 4).$$

It follows that its speed is exactly equal to 5. Then, we can compute the arclength:

$$s(t) = \int_0^t v(t)dt = \int_0^t 5du = 5t$$

Therefore, total arclength is $s(4\pi) = 20\pi$.

Definition 3.7. *When the path $\vec{c}(t)$ traces out the curve with speed $\|\vec{\mathrm{v}}(t)\| = 1$ for all t, we say that the curve is* arclength parameterized.

If a curve is arclength parameterized, arclength function becomes

$$s(t) = t$$

Then, we can use s instead of t as a parameterization in the path.

Example 3.2.5. In example 3.2.4, helix is not arclength parameterized but we can re-parameterize it so that it is. To do so, we need to solve for $t = \varphi(s)$ to invert the function, $s(t)$.

Going back the example, we had $s(t) = 5t$. It follows that $t = \frac{1}{5}s$. Then,

$$\vec{c}(s) = \vec{c}(\varphi(s)) = \vec{c}\left(\frac{s}{5}\right) = \left(3\cos\left(\frac{s}{5}\right), 3\sin\left(\frac{s}{5}\right), \frac{4s}{5}\right)$$

is an arclength parameterization of the original helix, i.e. $\|\vec{c}\,'(s)\| = 1$, $\forall s$.

3.3 Geometry of curves in $\mathbb{R}^3$

Path,

$$\vec{c}(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = (x, y, z)(t),$$

traces out a curve, for $t \in [a, b]$, in space, and its velocity vector and speed are given by $\vec{c}\,'(t)$ and $\|\vec{c}\,'(t)\|$, respectively. This is a smooth parameterization if $\vec{c} \in C^1$ and $\|\vec{c}\,'(t)\| \neq 0$ for any $t \in [a, b]$.

We introduced the arclength function,

$$s(t) = \int_a^t \|\vec{c}\,'(u)\| du,$$

the total distance along the curve up to time t.

We also introduced the idea of arclength parameterization, where $s(t) = t$. Then, since

$$\frac{ds}{dt} = \|\vec{c}\,'(t)\|,$$

arclength parameterization is a path that travels along the curve with unit speed, $ds/dt = 1$, throughout. Therefore, any path with $\|\vec{c}\,'(t)\| \neq 0$ can be parameterized by arclength by inverting $s = s(t)$ such that $t = \varphi(s)$. Note that we can always do this for a smooth path ($ds/dt > 0$ so $s(t)$ is monotonically increasing). In practice, however, you may not be able to find an explicit formula for the arclength parameterization!

Example 3.3.1. Consider the following path:

$$\vec{c}(t) = (t, \frac{1}{2}t^2):$$

Since $y = x^2/2$, it's a parabola. Then, we observe that

$$\vec{c}\,'(t) = (1, t), \|\vec{c}\,'(t)\| = \sqrt{1+t^2} \geq 1 > 0.$$

So the path is smooth. Then, we have

$$s(t) = \int_0^t \|\vec{c}\,'(u)\| du = \int_0^t \sqrt{1+u^2} du = \frac{1}{2}\left(\ln\left|\sqrt{1+t^2}+t\right| + t\sqrt{1+t^2}\right).$$

Clearly, there's no way we can solve for t as a function of s. The way out of this trouble is to treat all $\vec{c}$ as if they were parameterized by arclength and use Chain rule with $ds/dt = \|\vec{c}\,'(t)\|$ to compensate.

Recall that unit tangent vector to $\vec{c}(t)$ is

$$\vec{T}(t) = \frac{\vec{c}\,'(t)}{\|\vec{c}\,'(t)\|}.$$

We wish to understand how direction of the curve changes over time:

Definition 3.8. *The* curvature *of a curve is defined as rate of change of unit tangent:*

$$\kappa = \left\|\frac{d\vec{T}}{ds}\right\|.$$

By chain rule,

$$\frac{d\vec{T}}{dt} = \frac{d\vec{T}}{ds} \cdot \frac{ds}{dt}$$

So, in the original time parameter, t,

$$\kappa(t) = \left\|\frac{1}{\frac{ds}{dt}}\frac{d\vec{T}}{dt}\right\| = \frac{\|\vec{T}\,'(t)\|}{\|\vec{c}\,'(t)\|}$$

Example 3.3.2. Consider a circle of radius $R > 0$ in xy-plane:

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

Now, we can easily find its velocity vector and speed:

$$\begin{aligned} \vec{c}\,'(t) &= (R\cos t, -R\sin t) \\ \|\vec{c}\,'(t)\| &= R \end{aligned}$$

Notice that this travels with constant speed but is not arclength parameterized.

We can also find its unit tangent:

$$\vec{T}(t) = \frac{\vec{c}\,'(t)}{\|\vec{c}\,'(t)\|} = \frac{\vec{c}\,'(t)}{R} = (\cos t, -\sin t)$$

Then,

$$\vec{N}(t) = \vec{T}\,'(t) = (-\sin t, -\cos t)$$

Again, notice that $\vec{N}(t)$ is perpendicular to $\vec{T}(t)$.

Finally, we have

$$\kappa(t) = \frac{\|\vec{T}\,'(t)\|}{\|\vec{c}\,'(t)\|} = \frac{1}{R}.$$

Therefore, circle with large radius has less curvature.

Example 3.3.3. Consider the following helix:

$$\vec{c}(t) = (3\cos t, 3\sin t, 4t).$$

Following the same approach as shown in the previous example, we get

$$\begin{aligned} &\vec{c}\,'(t) = (-3\sin t, 3\cos t, 4) \\ &\|\vec{c}\,'(t)\| = 5 \\ &\vec{T}(t) = \left(-\frac{3}{5}\sin t, \frac{3}{5}\cos t, \frac{4}{5}\right) \\ &\vec{T}\,'(t) = \left(-\frac{3}{5}\cos t, -\frac{3}{5}\sin t, 0\right) \end{aligned}$$

Then,

$$\kappa(t) = \frac{\|\vec{T}\,'(t)\|}{\|\vec{c}\,'(t)\|} = \frac{3/5}{5} = \frac{3}{25}$$

This curve also has a constant curvature.

Definition 3.9 (Principal normal vector)**.**

$$\vec{N} = \frac{\vec{T}\,'(s)}{\|\vec{T}\,'(s)\|} = \frac{\vec{T}\,'(t)}{\|\vec{T}\,'(t)\|}$$

Since $\|\vec{T}(s)\| = 1$ for all s, $\vec{T}(s) \cdot \vec{T}(s) = \|\vec{T}(s)\|^2 = 1$. By implicit differentiation, we have

$$\begin{aligned} \frac{d}{ds}(1) &= \frac{d}{ds}(\vec{T}(s) \cdot \vec{T}(s)) \\ 0 &= \vec{T}\,'(s) \cdot \vec{T}(s) + \vec{T}(s) \cdot \vec{T}\,'(s) \\ &= 2\vec{T}\,'(s) \cdot \vec{T}(s) \end{aligned}$$

Therefore, $\vec{T}\,'(s) \perp \vec{T}(s)$ for all s. So as long as $\vec{T}\,'(s) \neq 0$, i.e. $\kappa \neq 0$, we have $\vec{N}(s) \perp \vec{T}(s)$. In fact, $\vec{T}\,'(s) = \|\vec{T}\,'(s)\|\vec{N} = \kappa\vec{N}$, so the tangent turns in the direction of $\vec{N}$. For motion in a line, where $\kappa(s) = 0$ for all s, $\vec{N}$ cannot be defined!

$\vec{T}, \vec{N}$ determines a plane in $\mathbb{R}^3$, the osculating plane. The normal vector to the osculating plane is given by

$$\vec{B} = \vec{T} \times \vec{N}.$$

Definition 3.10 (Binormal vector)**.** $\vec{B} = \vec{T} \times \vec{N}$

We observe that $\vec{B} \perp \vec{T}$, $\vec{B} \perp \vec{N}$, and

$$\|\vec{B}\| = \|\vec{T}\|\|\vec{N}\||\sin\theta| = 1 \cdot 1 \cdot \sin(\pi/2) = 1$$

Therefore, $\{\vec{T}(s), \vec{N}(s), \vec{B}(s)\}$ is a moving *orthonormal basis* for $\mathbb{R}^3$ at each point along the curve. This plane is also referred to as *moving frame* or *frenet frame.* Now, we introduce the following consequences:

(1). If curvature $\kappa(s) = 0$ for all s, then the curve is a straight line.

To see this, $\vec{T}\,'(s) = \kappa\vec{N}(s) = 0$ for all s. Therefore, $\vec{T}(s) = \vec{u}$ is a constant vector and

$$\vec{r}(s) = \vec{u}s + \vec{p},$$

a line thorugh $\vec{p} = \vec{r}(0)$ with direction vector $\vec{u}$.

(2). When $\kappa = 0$, $\vec{N}$ and $\vec{B}$ cannot be defined.

(3). If $\vec{B}(s)$ is a constant vector, then $\vec{c}(t)$ ($\vec{r}(t)$) move in a fixed plane, with normal vector $\vec{B}$.

Now, suppose $\vec{B}(s)$ isn't constant. First, $\|\vec{B}(s)\| = 1$ for all s. Then,

$$1 = \|\vec{B}(s)\|^2 = \vec{B}(s) \cdot \vec{B}(s)$$

holds for all s. So we can apply implicit differentiation:

$$0 = \frac{d}{dS}(1) = \frac{d}{dS}\left(\vec{B} \cdot \vec{B}\right) = 2\vec{B}\,' \cdot \vec{B}.$$

Then, it follows that $\vec{B}\,' \perp \vec{B}$, for every s.

Next, since $\vec{B}(s) \perp \vec{T}(s)$ for all s, we have $\vec{B} \cdot \vec{T} = 0$ for all s. Then,

$$\frac{d}{ds}\left(\vec{B} \cdot \vec{T}\right) = \vec{B}\,'(s) \cdot \vec{T}(s) + \vec{B}(s) \cdot \vec{T}\,'(s) = 0.$$

Since $\vec{T}\,' = \kappa \vec{N}$ and $\vec{B} \cdot \vec{N} = 0$, it follows that

$$\vec{B}\,'(s) \cdot \vec{T}(s) = 0 \iff \vec{B}\,'(s) \perp \vec{T}(s)$$

Since $\left\{\vec{T}, \vec{N}, \vec{B}\right\}$ form a orthonormal basis for $\mathbb{R}^3$, we must have $\vec{B}\,'(s)$ parallel to $\vec{N}$. Therefore,

$$\vec{B}\,'(s) = -\tau(s)\vec{N}(s)$$

for a function $\tau(s)$ called the *torsion*. Since $\tau = \|d\vec{B}/ds\|$, it measures how fast the normal $\vec{B}$ to the osculating plane is twisting.

Definition 3.11 (Torsion)**.**

$$\tau = \left\|\frac{d\vec{B}}{ds}\right\| = \frac{\|\vec{B}\,'(t)\|}{\|\vec{c}\,'(t)\|}$$

Putting all the information together we get *Frenet formulas*:

Theorem 3.2 (Frenet formula)**.**

$$\begin{cases} \frac{d\vec{T}}{ds} = \kappa \vec{N} \\ \frac{d\vec{B}}{ds} = -\tau \vec{N} \\ \frac{d\vec{N}}{ds} = -\kappa \vec{T} + \tau \vec{B} \end{cases}$$

Example 3.3.4. Consider the following helix:

$$\vec{c}(t) = (3\cos t, 3\sin t, 4t)$$

Then, we have

$$\begin{aligned}
&\|\vec{c}\,'(t)\| = 5, \\
&\vec{T}(t) = \left(-\frac{3}{5}\sin t, \frac{3}{5}\cos t, \frac{4}{5}\right), \\
&\vec{T}\,'(t) = \left(-\frac{3}{5}\cos t, -\frac{3}{5}\sin t, 0\right), \\
&\kappa = \frac{3}{25}, \\
&\vec{N}(t) = (-\cos t, -\sin t, 0), \\
&\vec{B}(t) = \vec{T} \times \vec{N} = \left(\frac{4}{5}\sin t, -\frac{4}{5}\cos t, \frac{3}{5}\right), \\
&\vec{B}\,' = \left(\frac{4}{5}\cos t, \frac{4}{5}\sin t, 0\right), \\
&\tau = \frac{4}{25}.
\end{aligned}$$

3.4 Dynamics

How do these quantities relate to dynamical quantities? Given, $\vec{c}(t)$, a position vector along the curve, $\vec{c}\,'(t) = \vec{\mathrm{v}}(t) = \vec{T}(t) \cdot ds/dt$ is its velocity vector and $\|\vec{c}\,'(t)\| = ds/dt$ is its speed.

Definition 3.12 (Acceleration)**.** $\vec{a}(t) = \vec{\mathrm{v}}\,'(t) = \vec{c}\,''(t)$

First, observe that

$$\vec{\mathrm{v}}(t) = \vec{c}\,'(t) = \frac{ds}{dt} \cdot \vec{T}(t)$$

Then,

$$\begin{aligned}\vec{a}(t) &= \frac{d}{dt}\left(\frac{ds}{dt} \cdot \vec{T}(t)\right) = \frac{d^2s}{dt^2} \cdot \vec{T}(t) + \frac{ds}{dt} \cdot \vec{T}\,'(t)\\ &= \frac{d^2s}{dt^2} \cdot \vec{T} + \frac{ds}{dt} \cdot \left(\frac{d\vec{T}}{ds} \cdot \frac{ds}{dt}\right)\end{aligned}$$

So we have

$$\vec{a}(t) = \underbrace{\frac{d^2s}{dt^2} \cdot \vec{T}}_{\text{Linear acceleration}} + \underbrace{\kappa \left(\frac{ds}{dt}\right)^2 \vec{N}}_{\text{Steering-term}}$$

By looking at the steering term, we see that acceleration to turn on a curve is proportional to the curvature and (speed)^2.

Example 3.4.1. Consider the following path

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

that drwas a spiral in xy direction and monotonically increases along z coordinate.

First, observe that

$$\begin{aligned}\vec{\mathrm{v}}(t) &= \vec{c}\,'(t) = (-e^t \sin + e^t \cos t, e^t \cos t + e^t \sin t, e^t)\\ \frac{ds}{dt} &= \|\vec{c}\,'(t)\| = \sqrt{3}e^t\end{aligned}$$

Then, we have

$$\begin{aligned}\vec{T}(t) &= \frac{\vec{c}\,'(t)}{\|\vec{c}\,'(t)\|} = \frac{1}{\sqrt{3}}(-\sin t + \cos t, \cos t + \sin t, 1),\\ \vec{T}\,'(t) &= \frac{\vec{c}\,'(t)}{\|\vec{c}\,'(t)\|} = \frac{1}{\sqrt{3}}(-\cos t - \sin t, -\sin t + \cos t, 0).\end{aligned}$$

Since $\|\vec{T}\,'(t)\| = \sqrt{2/3}$, we can easily find the principal normal vector:

$$\vec{N}(t) = \frac{\vec{T}\,'(t)}{\|\vec{T}\,'(t)\|} = \frac{1}{\sqrt{2}}(-\cos t - \sin t, -\sin t + \cos t, 0)$$

Then,

$$\kappa = \frac{\|\vec{T}\,'(t)\|}{\|\vec{c}\,'(t)\|} = \frac{\sqrt{2}}{3}e^{-t}.$$

Furthermore,

$$\vec{B}(t) = \vec{T}(t) \times \vec{N}(t) = \cdots = \frac{1}{\sqrt{6}}(\cos t - \sin t, -\sin t - \cos t, 2)$$
$$\vec{B}\,'(t) = \frac{1}{\sqrt{6}}(-\sin t - \cos t, -\cos t + \sin t, 0)$$

Therefore, torsion of the curve is given by

$$\tau(t) = \frac{\|\vec{B}\,'(t)\|}{\|\vec{c}\,'(t)\|} = \frac{1}{3}e^{-t}$$

We can then veriy formula for $\vec{a}$ in terms of $\vec{T}, \vec{N}, \kappa$, (and verify that it agrees with $\vec{a} = \vec{v}\,'(t)$ calculated directly).

Now, we present an alternative equation for curvature using dynamical quantities:

Theorem 3.3.

$$\kappa(t) = \frac{\|\vec{c}\,'(t) \times \vec{c}\,''(t)\|}{\|\vec{c}\,'(t)\|^3} = \frac{\|\vec{v}(t) \times \vec{a}(t)\|}{\|\vec{v}(t)\|^3}$$

Proof. To verify it, we use the decomposition of $\vec{a}$:

$$\begin{aligned}
\vec{v} \times \vec{a} &= \vec{v} \times \left(\frac{d^2s}{dt^2} \cdot \vec{T} + \kappa\left(\frac{ds}{dt}\right)^2 \vec{N}\right) \\
&= \frac{d^2s}{dt^2}\left(\vec{v} \times \vec{T}\right) + \kappa\left(\frac{ds}{dt}\right)^2\left(\vec{v} \times \vec{N}\right) \\
&= \kappa\left(\frac{ds}{dt}\right)^3\left(\vec{T} \times \vec{N}\right) \\
&= \kappa\left(\frac{ds}{dt}\right)^3 \vec{B}
\end{aligned}$$

Then, $\kappa(ds/dt)^3\|\vec{B}\| = \|\vec{v} \times \vec{a}\|$. Since $\vec{B}$ is a unit vector, the desired result has been achieved. $\square$

4 Implicit functions

4.1 The Implicit Function Theorem I

Often, we have an *implicit* relationship between variables,

$$F(x_1, x_2, \ldots, x_n) = 0,$$

rather than an *explicit* function relation, such as

$$x_n = f(x_1, x_2, \ldots, x_{n-1}).$$

Example 4.1.1. Look at a familiar example in $\mathbb{R}^2$,

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

This fails vertical line test ($y \neq f(x)$) as well as horizontal line test ($x \neq g(y)$); globally, this relation does not define a function. Locally, we can write this as a function, i.e. by restricting attention to small pieces of the curve.

First, define

$$F(x, y) = x^2 + y^2 - 1$$

If $y_0 > 0$, $x_0^2 + y_0^2 = 1$, i.e. $F(x_0, y_0) = 0$, and we look at a window (or *neighborhood*) around (x_0, y_0), which lies entirely in the upper half plane, we can solve for $y = f(x)$,

$$y = \underbrace{\sqrt{1 - x^2}}_{f(x)}$$

We could calculate $y' = f'(x)$ from the explicit formula but we can also get it via *implicit differentiation*:

$$\begin{aligned} \frac{d}{dx}\left(F(x, f(x))\right) &= \frac{\partial F}{\partial x} \cdot \frac{dx}{dx} + \frac{\partial F}{\partial y} \cdot f'(x) \\ &= 2x + 2yf'(x), \end{aligned}$$

so $f'(x) = -x/y$.

For a general $F(x, y) = 0$, we can solve for $f'(x)$ where its coefficient

$$\frac{\partial F}{\partial y}(x_0, y_0) \neq 0,$$

where y is the variable we want to solve for. This gives the limitation on which we can solve for $y = f(x)$ locally! For the circle example,

$$\frac{\partial F}{\partial y} = 2y.$$

When $y = 0$, the vertical line test fails in every neighborhood of $(x_0, y_0) = (\pm 1, 0)$.

In general, suppose we have a C^1 function,

$$F : \mathbb{R}^{n+1} \to \mathbb{R},$$

and consider all functions of $F(\vec{x}, y) = 0$. In order that $y = g(\vec{x})$, i.e. we can solve for y as a differentiable function of $\vec{x}$, we do the same implicit differentiation, with the chain rule,

$$\frac{\partial}{\partial x_i} \left(F \left(x_1, x_2 \ldots . . x_n, f(\vec{x}) \right) \right) = \frac{\partial F}{\partial x_i} + \frac{\partial F}{\partial y} \frac{\partial f}{\partial x_i}$$

for each $i = 1, 2, \ldots, n$. We can then solve for each

$$\frac{\partial f}{\partial x_i} = \frac{-\dfrac{\partial F}{\partial x_i}}{\dfrac{\partial F}{\partial y}},$$

provided $\partial F / \partial y \neq 0$. This is a sufficient condition to solve for $y = f(\vec{x})$.

Theorem 4.1 (Implicit Function Theorem I)**.** *Assume $F : \mathbb{R}^{n+1} \to \mathbb{R}$ is C^1 in a neighborhood of $(\vec{x}_0, y_0)$ with $F(\vec{x}_0, y_0) = 0$. If $\frac{\partial F}{\partial y}(\vec{x}_0, y_0) \neq 0$, then there exists neighborhood U of $\vec{x}_0$ and V of y_0 and a C^1 function*

$$f : U \subset \mathbb{R}^n \to V \subset \mathbb{R},$$

for which $F(\vec{x}, f(\vec{x})) = 0$ for all $\vec{x} \in U$. In addition,

$$Df(\vec{x}) = \frac{-1}{\dfrac{\partial F}{\partial y}(\vec{x}, y)} D_{\vec{x}} F(\vec{x}, y),$$

where

$$D_{\vec{x}} F(\vec{x}, y) = \begin{bmatrix} \frac{\partial F}{\partial x_1} & \frac{\partial F}{\partial x_2} & \cdots & \frac{\partial F}{\partial x_n} \end{bmatrix}.$$

Example 4.1.2. Consider the following function:

$$xy + y^2 z + z^3 = 1.$$

For which parts on this surface ca we write $z = f(x, y)$, i.e.

$$F : \mathbb{R}^3 \to \mathbb{R}, F(x, y, z) = xy + y^2 z + z^3 - 1$$

is C^1?

We want to solve for z, so we look at

$$\frac{\partial F}{\partial z} = y^2 + 3z^2$$

We observe that $\partial F / \partial z = 0$ iff $y = 0$ and $z = 0$. However, $y = 0$ and $z = 0$ is not defined on this surface. At all points on this surface, $\partial F / \partial z \neq 0$. So at every (x_0, y_0, z_0) with $F(x_0, y_0, z_0) = 0$, we can solve for $z = f(x, y)$ locally near (x_0, y_0)!

We can then use the implicit differentiation formula in the theorem to calculate $Df(x, y)$:

$$D_{(x,y)}F = \begin{bmatrix} y & (x + 2yz) \end{bmatrix},$$

so we get

$$Df(x, y) = \frac{-D_{(x,y)}F}{\partial F/\partial z} = \begin{bmatrix} -\frac{y}{y^2 + 3z^2} & -\frac{x + 2yz}{y^2 3z^2} \end{bmatrix}$$

or

$$\nabla f(x, y) = \left(-\frac{y}{y^2 + 3z^2}, -\frac{x + 2yz}{y^2 3z^2} \right).$$

Example 4.1.3. Consider the following equation:

$$x^4 + xz^2 + z^4 = 1.$$

Show that we can solve for $z = g(x)$ near $(x_1, z_1) = (-1, 1)$ but not near $(x_2, z_2) = (1, 0)$.

Proof. First, let

$$F(x, z) = x^4 + xz^2 + z^4 - 1.$$

Clearly, $F : \mathbb{R}^2 \to \mathbb{R}$ is C^1 for all $(x, z) \in \mathbb{R}^2$. Observe that

$$\frac{\partial F}{\partial z} = 2xz + 4z^3,$$

and so $\partial F(-1, 1)/\partial z \neq 0$.

By the Implicit Function Theorem, we can solve for $z = g(x)$ locally near $(x_1, z_1) = (-1, 1)$. In addition, we can get an explicit formula for its derivative:

$$Dg(x) = g'(x) = -\frac{\partial F(x, z)/\partial x}{\partial F/\partial z} = -\frac{4x^3 + z^2}{2xz + 4z^3}$$

Finally, since $\partial F(1, 0)/\partial z = 0$, the Implicit Function Theorem does not apply near $(1, 0)$. ☐

Example 4.1.4. Consider the following equation:

$$x - z^3 = 0$$

Clearly, $F(x, z) = x - z^3$ is C^1 for all $(x, z) \in \mathbb{R}^2$. Note that

$$\frac{\partial F}{\partial z} = -3z^2.$$

Clearly, $\partial F/\partial z = 0$ at $(x, z) = (0, 0)$. However, we can write $z = x^{1/3}$ globally. So $z = g(x) = x^{1/3}$ exists but isn't differentiable at $(x_0, z_0) = (0, 0)$.

Example 4.1.5. Suppose we have a system of equations with more unknowns:

$$\begin{cases} u^2 - v^2 - x^3 = 0 \\ 2uv - y^5 = 0 \end{cases}$$

Can we solve for (u, v) as functions of (x, y)?

First, consider a C^1 function, $F : \mathbb{R}^4 \to \mathbb{R}^2$, that is defined as follows:

$$\begin{cases} F_1(x, y, u, v) = u^2 - v^2 - x^2 = 0 \\ F_2(x, y, u, v) = 2uv - y^5 = 0 \end{cases}$$

Following what we did before, we can assume $(u, v) = g(x, y)$ and see when we can calculate Dg. Note that

$$\begin{aligned} 0 &= \frac{\partial}{\partial x} F_1(x, y, u(x, y), v(x, y)) = \frac{\partial F_1}{\partial x} + \frac{\partial F_1}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F_1}{\partial v}\frac{\partial v}{\partial x} \\ 0 &= \frac{\partial}{\partial x} F_2(x, y, u(x, y), v(x, y)) = \frac{\partial F_2}{\partial x} + \frac{\partial F_2}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial F_2}{\partial v}\frac{\partial v}{\partial x} \end{aligned}$$

Then, we can solve for $\partial u/\partial x$ and $\partial v/\partial x$. Rearranging,

$$\begin{bmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial x} \end{bmatrix} = \begin{bmatrix} -\frac{\partial F_1}{\partial x} \\ -\frac{\partial F_2}{\partial x} \end{bmatrix}$$

This can be solved if $D_{(u,v)}F$ is invertible, i.e. $\det\left[D_{(u,v)}F\right] \neq 0$.

Simiarly, we can also solve for $\partial u/\partial y$ and $\partial v/\partial y$. As a result, we get a different linear system to solve but with the same matrix $\left[D_{(u,v)}F\right]$. The second version of the Implicit Function Theorem says that this is the correct condition to solve for $g(x)$ in this setting.

4.2 The Implicit Function Theorem II

Implicit differentiation allows us to look at an underdetermined system of (nonlinear) equations. Given a following function,

$$\begin{gathered} F_1(x_1, \ldots, x_n, u_1, \ldots u_m) = 0 \\ F_2(x_1, \ldots, x_n, u_1, \ldots u_m) = 0 \\ \vdots \\ F_m(x_1, \ldots, x_n, u_1, \ldots u_m) = 0 \end{gathered}$$

we want to solve for $\vec{u} = (u_1, \ldots, u_m) \in \mathbb{R}^m$ as a function, $\vec{u} = g(\vec{x})$, of $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Via implicit differentiation, for the case of $n = m = 2$, we arrived at an appropriate condition where this is possible.

Theorem 4.2 (Implicit Function Theorem II - General Form)**.** *Let*

$$F : \mathbb{R}^{n+m} \to \mathbb{R}^m$$

be a C^1 function in a neighborhood of $(\vec{x}_0, \vec{u}_0) \in \mathbb{R}^{n+m}$, with $F(\vec{x}_0, \vec{u}_0) = \vec{0}$. If, in addition, $D_{\vec{u}}F(\vec{x}_0, \vec{u}_0)$ is invertible, then there exists neighborhoods $\mathcal{V}$ of $\vec{x}_0$ and $\mathcal{U}$ of $\vec{u}_0$, for which solutions of $F(\vec{x}, \vec{u}) = \vec{0}$ lie on a C^1 graph, $\vec{u} = g(\vec{x})$,

$$g : \mathcal{V} \subset \mathbb{R}^n \to \mathcal{U} \in \mathbb{R}^m$$

Example 4.2.1. Consider the following set of equations:

$$\begin{cases} 2xu^2 + yv^4 = 2 \\ xy(u^2 - v^2) = 0 \end{cases}$$

Can we solve for $(u, v) = g(x, y)$ near $(x_0, y_0, y_0, v_0) = (1, 1, -1, -1)$?

Let

$$F = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}, \vec{x} = (x, y), \vec{u} = (u, v),$$

where F is defined as

$$\begin{aligned} F_1(\vec{x}, \vec{u}) &= 2xu^2 + yv^4 - 2 = 0 \\ F_2(\vec{x}, \vec{u}) &= xy(u^2 - v^2 = 0) \end{aligned}$$

Then, we get the following Jacobian

$$\begin{aligned} D_{\vec{u}}F &= \frac{\partial(F_1, F_2)}{\partial(u, v)} \\ &= \begin{bmatrix} \dfrac{\partial F_1}{\partial u} & \dfrac{\partial F_1}{\partial v} \\ \dfrac{\partial F_2}{\partial u} & \dfrac{\partial F_2}{\partial v} \end{bmatrix} \\ &= \begin{bmatrix} 4xu & 4yv^3 \\ 2uxy & -2vxy \end{bmatrix} \end{aligned}$$

Substituting the given values, we have

$$D_{\vec{u}}F(1, 1, -1, -1) = \begin{bmatrix} -4 & -4 \\ -2 & 2 \end{bmatrix}$$

Since $\det D_{\vec{u}}F = -16 \neq 0$, the Implicit Function Theorem does apply, and we can solve for $\vec{u} = (u, v) = g(\vec{x}) = g(x, y)$ near $(x_0, y_0, u_0, v_0) = (1, 1, -1, -1)$.

Remark. In general, we can't get an explicit formula for g, but we can get a formula for $Dg(x, y)$, /ie its partial derivatives, using implicit differentiation.

Example 4.2.2. Consider the following set of equations:

$$\begin{cases} x = u^2 - v^2 \\ y = 2uv \end{cases}$$

Note that this example fits the form of the Implicit Function Theorem, but it's a special case. We want to invert this relationtion, i.e. given, $\vec{x} = f(\vec{u})$, we want to solve for $\vec{u} = g(\vec{x})$.

To get a nice theorem for this special case, we can use the framework of the Implicit Function Theorem:

$$F_1(\vec{x}, \vec{u}) = f_1(\vec{u}) - x = 0 F_2(\vec{x}, \vec{u}) = f_2(\vec{u}) - x = 0$$

Since

$$D_{\vec{u}} F(\vec{x}, \vec{u}) = Df(\vec{u}),$$

we can do this locally near a point $(\vec{x}_0, \vec{u}_0)$ provided that

$$\det (Df(\vec{u})) \neq 0$$

Note that if we had a linear system, $\vec{x} = M\vec{u}$, we can solve $\vec{u} = M^{-1}\vec{x}$ provided $\vec{M} \neq 0$. This is why we call this derivative matrix, $Df(\vec{x})$ the *linearization* of $f(\vec{u})$.

4.3 Inverse Function Theorem

In general, suppose we have a C^1 function, $f : \mathbb{R}^n \to \mathbb{R}^n$, where $\vec{x} = f(\vec{u})$. How do we solve for $\vec{u} = g(\vec{x})$?

First, let's In single-variable calculus, a function $f : \mathbb{R} \to \mathbb{R}$ is *one-to-one* on an interval $[a, b]$ if and only if f is strictly monotone on $[a, b]$. For these functions, f has an inverse $g = f^{-1}$,

$$g(f(x)) = x, \quad \forall x \in [a, b]$$

If f is differentiable on $[a, b]$, and $f'(x) > 0$ on $[a, b]$ (or $f'(x) < 0$ on $[a, b]$), then the inverse $g(x)$ is also differentiable, and

$$g'(f(x)) = \frac{1}{f'(x)}, \quad \forall x \in [a, b]$$

If, for example, $f'(x) > 0$ for all $x \in \mathbb{R}$, then it's globally invertible, i.e. $g(f(x)) = x$ for all $x \in \mathbb{R}$. How do we apply this for $f : \mathbb{R}^n \to \mathbb{R}^n$ with $n \geq 2$?

Theorem 4.3 (Inverse Function Theorem)**.** *Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ which is C^1 in a neighborhood of $\vec{u}_0$, with $f(\vec{u}_0) = \vec{x}_0$. If $\det (Df(\vec{u}_0)) \neq 0$, then there exist neighborhoods $\mathcal{U}$ of $\vec{u}_0$ and $\mathcal{V}$ of $\vec{x}_0$ and a C^1 function $g : \mathcal{V} \to \mathcal{U}$, with*

$$\underset{\text{with } \vec{u} \in \mathcal{U}}{\vec{x} = f(\vec{u})} \iff \underset{\text{with } \vec{x} \in \mathcal{V}}{\vec{u} = g(\vec{x})},$$

i.e. near $\vec{x}_0$ and $\vec{u}_0$, g is the inverse of f.

Example 4.3.1. Apply the Inverse Function Theorem to the function that was defined in the previous example:

$$\begin{cases} x = u^2 - v^2 \\ y = 2uv \end{cases}$$

Observe that

$$\det\left(Df(u,v)\right) = \det \begin{bmatrix} 2u & -2v \\ 2v & 2u \end{bmatrix} = 4u^2 + 4v^2 \neq 0$$

as long as $(u_0, v_0) \neq (0,0)$. So we can invert the variables and solve for $(u,v) = g(x,y)$, locally near any $(u_0, v_0) \neq (0,0)$.

Notice that

$$f_1(-u,-v) = x = f_1(u,v)$$
$$f_2(-u,-v) = y = f_2(u,v)$$

So in any neighborhood of $(0,0)$ there are 2 values of (u,v) corresponding to each (x,y). So f is not invertible near $(u,v) = (0,0)$.

Example 4.3.2. Consider the following equations:

$$\begin{cases} x = e^y \cos v \\ y = e^u \sin v \end{cases}$$

For which (u,v,x,y) can we solve for u,v as functions of x,y?

Call

$$f(u,v) = \begin{bmatrix} e^u \cos v \\ e^u \sin v \end{bmatrix}.$$

Then, we have

$$Df(u,v) \begin{bmatrix} e^u \cos v & -e^u \sin v \\ e^u \sin v & e^u \cos v \end{bmatrix}$$

Then, we can compute $\det(Df(u,v))$, (or $\det\left(\frac{\partial(x,y)}{\partial(u,v)}\right)$):

$$\det(Df(u,v)) = e^{2u}.$$

Clearly, $\det(Df(u,v)) > 0$ for all u,v. By the Inverse Function Theorem, we can invert and solve for $(u,v) = g(x,y)$, near any (u_0,v_0).

We can invert locally near any point; can we find a global inverse, i.e. a g for which $(u,v) = g(x,y)$ for every $(u,v) \in \mathbb{R}^2$? If so, then f would have to be a one-to-one function. However,

$$f(u, v + 2\pi k) = f(u,v)$$

for all $k \in \mathbb{Z}$. Therefore, f can't be globally inverted.

Example 4.3.3. Consider the following equations:

$$\begin{cases} x = f_1(u,v) = u^3 - 3uv^2 \\ y = f_2(u,v) = -v^3 + 3u^2v \end{cases}$$

Since they're polynomials, $f : \mathbb{R}^2 \to \mathbb{R}^2$ is C^1. Then, we have

$$\begin{aligned} \frac{\partial(x,y)}{\partial(u,v)} &= \begin{bmatrix} 3u^2 - 3v^2 & -6uv \\ 6uv & -3v^2 + 3u^2 \end{bmatrix} \\ \det\left(\frac{\partial(x,y)}{\partial(u,v)}\right) &= (3u^2 - 3v^2) + (6uv)^2 \end{aligned}$$

Clearly, $\det(\partial(x,y)/\partial(u,v)) = 0$ iff $u = v = 0$. So, Inverse Function Theorem holds for all $(u_0, v_0) \neq (0,0)$, and we can solve for $(x,y) = g(u,v)$ around any $(u_0, v_0) \neq (0,0)$.

5 Taylor's Theorem

5.1 Taylor's Theorem in one dimension

Consider a one-dimensional function:

$$g : \mathbb{R} \to \mathbb{R},$$

which is C^{k+1}, i.e. it is $(k+1)$ times continuously differentiable; i.e., each derivative,

$$\frac{d^j g}{dx^j}(x),\ j = 1, 2, \ldots, k+1,\ \text{(of order up to and including the } (k+1)^{\text{st}}\text{)},$$

exists and is a continuous function (in some interval). Then, we can approximate $g(x)$ locally near $x = a$ by a polynomial of degree k, *Tayloer's polynomial*, $P_k(x)$:

$$P_k(x) = g(a) + g'(a)(x-a) + \frac{1}{2!}g''(a)(x-a)^2 + \cdots + \frac{1}{k!}\frac{d^k g}{dx^k}(a)(x-a)^k$$

There are chosen to match $g(x)$ up to the k^{th} derivative at $x = 0$,

$$\frac{d^j P_k}{dx^j}(a) = \frac{d^j g}{dx^j}(a),\ j = 0, 1, 2, \ldots, k.$$

For example, $P_1(x) = g(a) + g'(a)(x-a)$ is the tangent line. Since we know that g is differentiable,

$$\lim_{x \to a} \frac{|g(x) - P_1(x)|}{|x-a|} = 0 \text{ or } g(x) = P_1(x) + o(|x-a|)$$

Theorem 5.1 (Taylor's Theorem)**.** *Assume $g : \mathbb{R} \to \mathbb{R}$ is C^{k+1} in a neighborhood $\mathcal{N}$ around $x = a$. Then, for each $x \in N$, there is a c between a and x for which*

$$g(x) = P_k(x) + \underbrace{\frac{1}{(k+1)!}\frac{d^{k+1}g}{dx^{k+1}}(c)(x-a)^{k+1}}_{\textit{Remainder term } R_k(a,x)}$$

Since we assume g is continuous, we have

$$\lim_{x \to a} \frac{R_k(a,x)}{|x-a|^k} = 0,$$

i.e. $R_k(a,x) = o\left(|x-a|^k\right)$. So $R_k(a,x)$ is small compared with each of the terms in $P_k(x)$.

Remark. Locally, $g(x)$ is well approximated by its Tayloer polynomials, but only near $x = a$.

Example 5.1.1. Notice that

$$g(x) = \cos x = \underbrace{1 - \frac{1}{2}x^2 + 0x^3}_{P_3(x)} + o(x^4)$$

This tells us that $\cos x$ is quadratic near $a = 0$. However, this clearly doesn't work for x that is not near $a = 0$.

5.2 Taylor's Theorem in higher dimensions

Can we apply Taylor's Theorem for $f : \mathbb{R}^n \to \mathbb{R}$, i.e., approximate a smooth function locally near $\vec{x} = \vec{a}$ via polynomial? We can do so by restricting our attention to each line, $\vec{x} = \vec{a} + t\vec{u}$, through $\vec{a}$ in direction $\vec{u}$.

Assume $f \in C^3$ near $\vec{x}_0 \in \mathbb{R}^n$. Let's sample $f(\vec{x})$ along a line running through x_0. Take a unit vector $\vec{u}$, $\|\vec{u}\| = 1$, and the line,

$$\vec{l}(t) = \vec{x}_0 + t\vec{u},$$

that goes through $\vec{x}_0$ at $t = 0$ in the direction of $\vec{u}$. Then, we get

$$g(t) = f(\vec{l}(t)) = f(\vec{x}_0 + t\vec{u}),$$

so $g : \mathbb{R} \to \mathbb{R}$.

By chain rule, if f is C^3 near $\vec{x}_0$, then g is C^3 near $t = 0$.So we use Taylors Theorem in g:

$$\begin{aligned} g(0) &= f(\vec{x}_0), \\ g'(t) &= Df(\vec{x}_0 + t\vec{u}) \cdot \vec{l}\,'(t) = Df(\vec{x}_0 + t\vec{u})\vec{u}. \end{aligned}$$

So $g'(0) = Df(\vec{x}_0)\vec{u} = \nabla f(\vec{x}_0) \cdot \vec{u}$. Using coordinates,

$$g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\vec{x}_0 + t\vec{u})u_i$$

so

$$g''(t) = \sum_{i=1}^{n} \underbrace{\frac{d}{dt}\left(\frac{\partial f}{\partial x_i}(\vec{x}_0 + t\vec{u})\right)}_{D\frac{\partial f}{\partial x_i} \cdot \vec{u}} u_i = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}_0 + t\vec{u})u_j u_i$$

Therefore,

$$g''(0) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}_0)u_j u_i$$

Now, call

$$H(\vec{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\vec{x}_0)\right]_{i,j=1,\ldots,n},$$

the *Hessian matrix* of f at $\vec{x}_0$. For f, a C^2 function, $f_{x_i x_j} = f_{x_j x_i}$, so $H(\vec{x}_0)$ is a symmetric matrix.

So $g''(0) = \vec{u} \cdot H(\vec{x}_0)\vec{u}$. Using Taylor's Theorem, for g, we get:

Theorem 5.2 (Second order Taylor's approximation)**.** *Assume $f : \mathbb{R}^n \to \mathbb{R}$ and is C^3 in a neighborhood of $\vec{x}_0$. Then,*

$$\underbrace{f(\vec{x})}_{g(\vec{x}_0+t\vec{u})} = \underbrace{f(\vec{x}_0)}_{g(0)} + \underbrace{Df(\vec{x}_0)(\vec{x}-\vec{x}_0)}_{g'(0)(t-0)} + \frac{1}{2!}\underbrace{(\vec{x}-\vec{x}_0)\cdot H(\vec{x}_0)(\vec{x}-\vec{x}_0)}_{g''(0)(t-0)^2} + R_2(\vec{x}_0,\vec{x}),$$

where

$$H(\vec{x}_0) = D^2 f(\vec{x}_0) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{i,j=1,2,\ldots,n}$$

is the hessian and

$$\lim_{\vec{x}\to\vec{x}_0} \frac{R_2(\vec{x}_0,\vec{x})}{\|\vec{x}-\vec{x}_0\|^2} = 0,$$

i.e. $R_2(\vec{x}_0\vec{x}) = o(\|\vec{x}-\vec{x}_0\|^2)$. *Alternatively, the second order Taylor's approximation can be written as*

$$f(\vec{a}+\vec{h}) = f(\vec{a}) + Df(\vec{a})\vec{h} + \frac{1}{2}\vec{h}\cdot D^2 f(\vec{a})\vec{h} + R_2(\vec{a},\vec{h}),$$

with

$$\lim_{\vec{h}\to\vec{0}} = \frac{R_2(\vec{a},\vec{h})}{\|\vec{h}\|^2} = 0.$$

Example 5.2.1. Find the second order Taylor polynomial of the following functions:

$$f(x,y) = \cos\left(xy^2\right)$$

near $\vec{a} = (\pi, 1)$.

First, we compute the derivatives:

$$\begin{aligned}
f(\vec{a}) &= f(\pi,1) = \cos(\pi) = 1,\\
\frac{\partial f}{\partial x} &= -y^2 \sin\left(xy^2\right),\\
\frac{\partial f}{\partial y} &= -2xy \sin\left(xy^2\right),\\
\frac{\partial^2 f}{\partial x^2} &= -y^2 \cos\left(xy^2\right) y^2,\\
\frac{\partial^2 f}{\partial x \partial y} &= -2y \sin\left(xy^2\right) - 2xy^3 \cos\left(xy^2\right),\\
\frac{\partial^2 f}{\partial y^2} &= -2x \sin\left(xy^2\right) - 2xy \cos\left(xy^2\right).
\end{aligned}$$

Then, at $\vec{a} = (\pi, 1)$, we find that

$$\begin{aligned}
Df(\vec{a}) &= \begin{bmatrix} 0 & 0 \end{bmatrix}\\
D^2 f(\vec{a}) &= \begin{bmatrix} 1 & 2\pi \\ 2\pi & 4\pi^2 \end{bmatrix}
\end{aligned}$$

So, we have

$$\begin{aligned}
f(\vec{a}+\vec{h}) &= -1 + \frac{1}{2}\vec{h}\cdot\begin{bmatrix}1 & 2\pi\\ 2\pi & 4\pi^2\end{bmatrix}\vec{h} + R_2\\
f(\pi+h_1, 1+h_2) &= -1 + \frac{1}{2}\begin{bmatrix}h_1\\ h_2\end{bmatrix}\cdot\begin{bmatrix}h_1 + 2\pi h_2\\ 2\pi h_1 + 4\pi^2 h_2\end{bmatrix}\\
&= -1 + \frac{1}{2}\left(h_1^2 + 4\pi h_1 h_2 + 4\pi^2 h_2^2\right) + o(\|\vec{h}^2\|)
\end{aligned}$$

In terms of a point $\vec{x}$ (near $\vec{a}$), we can write $\vec{x} = \vec{a}+\vec{h}$, so $\vec{h} = \vec{x}-\vec{a}$, and then

$$\cos\left(xy^2\right) = -1 + \frac{1}{2}\left((x-a_1)^2 + 4\pi(x-a_1)(y-a_2) + 4\pi^2(y-a_2)^2\right) + R_2.$$

Advantage to the $f(\vec{a}+\vec{h})$ form is that it makes it easier to guess the behaviour of $f(\vec{x})$ near $\vec{x}=\vec{a}$.

5.3 Local minima/maxima

Definition 5.1. *We say $\vec{a}$ is a* local minimum *for f if there exists an open disk $D_r(\vec{a})$ for which*

$$f(\vec{a}) \leq f(\vec{x})$$

for all $\vec{x} \in D_r(\vec{a})$. $\vec{a}$ is a strict local minimum *if*

$$f(\vec{a}) < f(\vec{x})$$

for all $\vec{x} \neq \vec{a}, \vec{x} \in D_r(\vec{a})$.

Definition 5.2. *We say $\vec{a}$ is a local maximum for f if $\exists r > 0$ with $f(\vec{a}) \geq f(\vec{x})$, $\forall \vec{x} \in D_r(\vec{a})$. $\vec{a}$ is a strict local max if $f(\vec{a}) > f(\vec{x})$, $\forall \vec{x} \in D_r(\vec{a}) \setminus \{\vec{a}\}$.*

Note that if f is differentiable, we have a necessary condition for local maxima and minima.

Theorem 5.3. *If f has a local maxima or minima at $\vec{a}$ and is differentiable at $\vec{a}$, then $Df(\vec{a}) = \vec{0}$.*

Proof. Again, we start by restricting to line through $\vec{a}$:

$$g(t) = f(\vec{a}+t\vec{u}),$$

where $\vec{u}$ is a unit vector. If f has a local minima at $\vec{a}$, then

$$g(0) = f(\vec{a}) \leq f(\vec{a}+t\vec{u}) = g(t),$$

for all t with $|t| < r$. So $g(t)$ has a local minima at $t=0$. By a calculus theorem, $g'(0) = 0$. But,

$$0 = g'(0) = Df(\vec{a})\vec{u},$$

for all $\vec{u}$. Then, by taking $\vec{u} = \vec{e}_1, \vec{e}_2 \ldots, \vec{e}_n$, we get

$$\frac{\partial f}{\partial x_j}(\vec{a}) = 0,$$

for each $j = 1, 2, 3, \ldots, n$. Therfore, $Df(\vec{a}) = 0$. □

Definition 5.3. *An $\vec{a}$ for which $Df(\vec{a}) = 0$ is called a* critical point.

Example 5.3.1. In the previous example, $\vec{a} = (\pi, 1)$ was a critical point.

Now, we want to combine Taylor's Theorem and linear algebra to classify critical points as local minima, maxima, or others[6]. Taylor's theorem states that for $\vec{x} = \vec{a} + \vec{h}$, if $\|\vec{h}\|$ is small,

$$f(\vec{x}) = f(\vec{a} + \vec{h}) = f(\vec{a}) + \underbrace{Df(\vec{h})}_{0} + \underbrace{\frac{1}{2}\vec{h} \cdot D^2 f(\vec{a})\vec{h}}_{\text{gradient form}} + \underbrace{R_2(\vec{a}, \vec{h})}_{o(\|\vec{h}\|^2)}$$

So we expect the behaviour of $f(\vec{x})$ near $\vec{a}$ to be determined by the gradient term.

Notice that the Hessian matrix, $M = D^2 f(\vec{a})$, is a symmetric matrix. This allows us to apply the following theorem:

Theorem 5.4 (Spectral Theorem)**.** *Assume M is a symmetric $(n \times n)$ matrix. Then,*

- *All eigenvalues of M are real, $\lambda_i \in \mathbb{R}\, \forall i = 1, 2, \ldots, n$.*
- *There is an orthonormal basis composed ot eigenvalues of M,*

$$\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}, M\vec{u}_i = \lambda_i \vec{u}_i, \|\vec{u}_i\| = 1, \vec{u}_i \cdot \vec{u}_j = 0 \textit{ for } i \neq j$$

- *In the basis of eigenvalues, M is a diagonal matrix. In other words, if we let U be the matrix whose columns are the $\vec{u}_i$; then*

$$MU = U\Lambda,$$

  where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

[6] **Cramer's Rule.** Given a system of linear equations that is represented by 2×2 matrices,

$$\begin{cases} ax + by = s \\ cx + dy = t \end{cases},$$

solution of the system is given by

$$x = \frac{\det \begin{pmatrix} s & b \\ t & d \end{pmatrix}}{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}},\; y = \frac{\det \begin{pmatrix} a & s \\ c & t \end{pmatrix}}{\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}}$$

Remark. Note that since the eigenvalues are real, they can be ordered, smallest to largest:

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

However, they may not be necessarily distinct.

Written in the orthonormal basis of eigenvalues, the quadratic form, $\vec{h} \cdot M\vec{h}$, has an easy expression. Firt, we write

$$\vec{h} = \sum_{i=1}^{n} c_i \vec{u}_i,\ c_i \in \mathbb{R},\ \forall i = 1, 2, \ldots, n$$

Notice that

$$\|\vec{h}\| = \sqrt{\sum_{i=1}^{n} c_i^2}$$

Then, we have:

$$\begin{aligned}
\vec{h} \cdot M\vec{h} &= \vec{h} \cdot \sum_{i=1}^{n} c_i M\vec{u}_i \\
&= \vec{h} \cdot \sum_{i=1}^{n} \lambda_i c_i \vec{u}_i \\
&= \sum_{i=1}^{n} \lambda_i c_i (\underbrace{\vec{h} \cdot \vec{u}_i}_{c_i}) \\
&= \sum_{i=1}^{n} \lambda_i c_i^2
\end{aligned}$$

Theorem 5.5. *Suppose M is a symmetric matrix with eigenvalues*

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

Then,

$$\lambda_1 \|\vec{h}\|^2 \leq \vec{h} \cdot M\vec{h} \leq \lambda_n \|\vec{h}\|^2$$

Proof. First, we have

$$\begin{aligned}
\vec{h} \cdot M\vec{h} &= \sum_{i=1}^{n} \lambda_i c_i^2 \\
&\leq \sum_{i=1}^{n} \lambda_n c_i^2 \\
&= \lambda_n \sum_{i=1}^{n} c_i^2 = \lambda_n \|\vec{h}\|^2,
\end{aligned}$$

which proves the right hand inequality. For the let hand one,

$$\vec{h} \cdot M\vec{h} \geq \sum_{i=1}^{n} \lambda_1 c_i^2 = \lambda_1 \|\vec{h}\|^2.$$

This proves both sides of the inequality. $\square$

Now we apply this idea to the Hessian via Taylor's Theorem to get the following theorem:

Theorem 5.6 (Second derivative test)**.** *Suppose f is C^3 in a neighborhood of a critical point $\vec{a}$. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $D^2 f(\vec{a})$. Then,*

1. *If all eigenvalues are positive, then $\vec{a}$ is a strict local minima of f.*

2. *If all eigenvalues are negative, then $\vec{a}$ is a strict local maxima of f.*

3. *If $D^2 f(\vec{a})$ has at least one positive and at least one negative eigenvalue, then $\vec{a}$ is a saddle point. In other words, $\exists r_0 > 0$ for which in $D_r(\vec{a}), 0 < r < r_0$, there are points with $f(\vec{x}) > f(\vec{a})$ and points with $f(\vec{x}) < f(\vec{a})$.*

Proof. Let's verify (1). By Taylor's Theorem, with $\vec{x} = \vec{a} + \vec{h}$, we have

$$f(\vec{x}) = f(\vec{a} + \vec{h}) = f(\vec{a}) + \frac{1}{2}\vec{h} \cdot D^2 f(\vec{a})\vec{h} + R_2(\vec{a}, \vec{h})$$

Notice that

$$\frac{1}{2}\vec{h} \cdot D^2 f(\vec{a})\vec{h} \geq \frac{1}{2}\lambda_1 \|\vec{h}\|^2,$$

where λ_1 is the smallest eigenvalue.

Now, we look at the rectangular term: $R_2(\vec{a}, \vec{h}) = o(\|\vec{h}\|^2)$. By taking $\epsilon = \lambda_1/4 > 0$, there exists $\delta > 0$ for which

$$\frac{|R_1|}{\|\vec{h}\|^2} < \epsilon = \frac{1}{4}\lambda_1,$$

if $0 < \|\vec{h}\| < \delta$, i.e. if $\vec{h} \in D_\delta(\vec{0})$, then $|R_2| < \frac{1}{4}\lambda_1 \|\vec{h}\|^2$. This implies that $R_2 > -\frac{1}{4}\lambda_1 \|\vec{h}\|^2$.

Combining these two results with Taylor expansion, if $\vec{x} \in D_\delta(\vec{a}), \vec{h} \in D_\delta(\vec{0})$, we get

$$\begin{aligned} f(\vec{x}) &\geq f(\vec{a}) + \frac{1}{2}\lambda_1 \|\vec{h}\|^2 - \frac{1}{4}\lambda_1 \|\vec{h}\|^2 \\ &\geq f(\vec{a}) + \frac{1}{4}\lambda_1 \|\vec{h}\|^2 \\ &> f(\vec{a}) \end{aligned}$$

if $\vec{h} \neq 0, \vec{h} \in D_\delta(\vec{0})$, i.e. $\vec{x} \in D_\delta(\vec{a})$. $\square$

Remark. When $D^2 f(\vec{a})$ has zero as an eigenvalue, things can get complicated. For example, if $\lambda_i \geq 0$ for all i, you *could* still have a local minima. In this case, the behaviour would be determined by higher order terms in Taylor Series. We call this *Degenerate critical point.*

Example 5.3.2. Consider

$$f(x,y,z) = x^3 - 3xy + y^3 + \cos z$$

Find all critical points and classify them using the Hessian.

First, observe that

$$\begin{cases} \frac{\partial f}{\partial x} = 3x^2 - 3y \\ \frac{\partial f}{\partial y} = -3x + 3y^2 \\ \frac{\partial f}{\partial z} = -\sin z \end{cases}$$

Critical points are defined as $\nabla f(\vec{a}) = \vec{0}$ so we get the following critical points

$$(0,0,n\pi), (1,1,n\pi),$$

where $n \in \mathbb{Z}$.

Then, we want to compute the Hessian at each point.

$$D^2 f(\vec{a}) = \begin{bmatrix} 6x & -3 & 0 \\ -3 & 6y & 0 \\ 0 & 0 & -\cos z \end{bmatrix}$$

Notice that at $(0,0,n\pi)$, we get

$$D^2 f(0,0,2k\pi) = \begin{bmatrix} 0 & -3 & 0 \\ -3 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$
$$D^2 f(0,0,(2k+1)\pi) = \begin{bmatrix} 0 & -3 & 0 \\ -3 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

When n is even, we find that its eigenvalues are

$$\lambda = -3, -1, 3$$

so we get a saddle at $(0,0,2k\pi), k \in \mathbb{Z}$. Similarly, when n is odd, we find that its eigenvalues are

$$\lambda = -3, 1, 3$$

which is also a saddle. Thus, we get a saddle at $(0,0,n\pi)$ for all $n \in \mathbb{Z}$.

At $(1,1,n\pi)$, we get

$$D^2 f(1,1,n\pi) = \begin{bmatrix} 6 & -3 & 0 \\ -3 & 6 & 0 \\ 0 & 0 & (-1)^{n+1} \end{bmatrix}$$

By observation, we find that $\vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is an eigenvector with $\lambda = (-1)^{n+1}$.

Then, the two eigenvalues are eigenvalues of $\begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}$. Since its trace is 12 and determinant is 27, its characteristic equation is given by

$$\lambda^2 - 12\lambda + 27 = 0.$$

So we find that two other eigenvalues are $\lambda = 3, 9$. Therefore, $(1, 1, (2k+1)\pi)$ is a local minima, and $(1, 1, 2k\pi)$ are saddles.

Example 5.3.3. Consider

$$f(x, y) = x^2 + y^4$$

We find that

$$\nabla f(x, y) = \begin{bmatrix} 2x \\ 4y^3 \end{bmatrix}$$

so we get only one critical point, $(x, y) = (0, 0)$. Ntice that

$$D^2 f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 12y^2 \end{bmatrix} \text{ so } D^2 f(0, 0) \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

So we get $\lambda = 2, 0$. Since the quadractic doesn't dominate the remainder, we call this a *degenerate* case.

Still, $f(0, 0) < f(x, y)$ for all $(x, y) \neq (0, 0)$ so its a minima even if the Hessian test doesn't tell us so.

Example 5.3.4. Consider

$$g(x, y) = x^2 - y^4$$

This has the same second order Taylor expansion as the previous example but has a different ramainder, $R_2 = -y^4$. This is a degenerate saddle.