9. Multivariable calculus

In this chapter we briefly consider functions of more than one variable. We will focus on the case of two variables, but the case of more than two variables is very similar.

9.1 Topology of $\mathbb{R}^2$

$\mathbb{R}^2$ is what is called a **vector space**: we can add two elements (componentwise) and0 we can scale any element by multiplying its components with a fixed real number. We leave the details of this definition to linear algebra.

There is a **zero vector**, denoted 0 or $\mathbf{0}$ such that $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$ and of course $\mathbf{0} = (0,0)$.

Let $\boldsymbol{x} \in \mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$. A **norm** on $\mathbb{R}^2$ is a function

$$\|\cdot\|: \mathbb{R}^2 \to \mathbb{R}$$

such that

1. $\|\boldsymbol{x}\| \geq 0$ for all $\boldsymbol{x}$
2. $\|\boldsymbol{x}\| = 0$ iff $\boldsymbol{x} = (0,0)$
3. $\|c\boldsymbol{x}\| = |c|\|\boldsymbol{x}\|$
4. $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

The usual consequences similar to the absolute value for real numbers follow: For example, for all $\boldsymbol{x}, \boldsymbol{y}$ we have

$$\|\boldsymbol{x} + \boldsymbol{y}\| \geq |\|\boldsymbol{x}\| - \|\boldsymbol{y}\||$$

Indeed, 3. and 4. imply that $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$. Then

$$\|\boldsymbol{x}\| \geq \|\boldsymbol{x} - \boldsymbol{y}\| - \|\boldsymbol{y}\|$$

This holds for all $\boldsymbol{x}, \boldsymbol{y}$, so we may replace $\boldsymbol{x}$ with $\boldsymbol{x} + \boldsymbol{y}$ get

$$\|\boldsymbol{x} + \boldsymbol{y}\| \geq \|\boldsymbol{x}\| - \|\boldsymbol{y}\|$$

Exchanging $\boldsymbol{x}, \boldsymbol{y}$, we also get $\|\boldsymbol{x} + \boldsymbol{y}\| \geq \|\boldsymbol{y}\| - \|\boldsymbol{x}\|$, and therefore the claim.

Example

The following three examples are the ones used most often:

1. The maximum norm $\|\boldsymbol{x}\|_\infty \coloneqq \max\{|x|, |y|\}$
2. $\|\boldsymbol{x}\|_1 \coloneqq |x| + |y|$
3. The Euclidean norm $|\boldsymbol{x}| \coloneqq \|\boldsymbol{x}\|_2 := \sqrt{x^2 + y^2}$

Proposition

All norms on $\mathbb{R}^2$ are **equivalent**. That is, if $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are two norms, there exist positive constants A, B such that for all $\boldsymbol{x}$ we have

$$A\|\boldsymbol{x}\|_\alpha \leq \|\boldsymbol{x}\|_\beta \leq B\|\boldsymbol{x}\|_\alpha$$

EOP.

Proof. It is enough to show hat A, B exist in case $\|\cdot\|_\alpha = \|\cdot\|_\infty$. Let $\|\cdot\|$ be any norm. Then $\|\boldsymbol{x}\| = \|xe_1 + ye_2\|$ where $e_1 = (1,0)$ and $e_2 = (0,1)$.

$$\|xe_1 + ye_2\| \le |x|\|e_1\| + |y|\|e_2\| \le B\max\{|x|,|y|\} = B\|\boldsymbol{x}\|_\infty$$

where $B = 2\max\{\|e_1\|,\|e_2\|\}$.

On the other hand let $A = \inf\{\,\|\boldsymbol{x}\| \mid \|\boldsymbol{x}\|_\infty = 1\}$. Note A exists since $\|\boldsymbol{x}\| > 0$ for all $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_\infty = 1$.

If $A > 0$, then for any $\boldsymbol{x} \neq \boldsymbol{0}$ we have $\|c\boldsymbol{x}\|_\infty = 1$ where $c = \|\boldsymbol{x}\|_\infty^{-1}$, and therefore $\|c\boldsymbol{x}\| \ge A$ and thus $\|\boldsymbol{x}\| \ge Ac^{-1} = A\|\boldsymbol{x}\|_\infty$.

Let $S = \{\,\boldsymbol{x} \mid \|\boldsymbol{x}\|_\infty = 1\,\}$, and let $\boldsymbol{x}_n \in S$ be a sequence such that $\|\boldsymbol{x}_n\| \to A$.

The entries of $\boldsymbol{x}_n = (x_n, y_n)$ are all bounded because $\|\boldsymbol{x}_n\|_\infty = 1$. Thus, we may choose a subsequence such that both x_n, y_n converge (see the last part of the proof of the Fundamental Theorem of Algebra), to x_0, y_0 respectively. Let $\boldsymbol{x}_0 = (x_0, y_0)$. Then $\max\{|x_n - x_0|, |y_n - y_0|\}$ is a zero sequence, that is, $\|\boldsymbol{x}_n - \boldsymbol{x}_0\|_\infty \to 0$.

Then $\|\boldsymbol{x}_n - \boldsymbol{x}_0\| \le B\|\boldsymbol{x}_n - \boldsymbol{x}_0\|_\infty \to 0$. But then by the discussion above

$$|\|\boldsymbol{x}_n\| - \|\boldsymbol{x}_0\|| \le \|\boldsymbol{x}_n - \boldsymbol{x}_0\| \to 0$$

On the other hand $\|\boldsymbol{x}_n\| \to A$. Therefore, $\|\boldsymbol{x}_0\| = A$.

But $\|\boldsymbol{x}_0\|_\infty = 1$ (because $\|\boldsymbol{x}_n\| = 1$ for all n) and therefore $\boldsymbol{x}_0 \neq \boldsymbol{0}$. Thus $A > 0$. QED.

Definition
Let $\boldsymbol{x}_n$ be a sequence of elements of $\mathbb{R}^2$. We say $\boldsymbol{x}_n$ converges to $\boldsymbol{x}_0$ if
$$\lim_{n\to\infty} |\boldsymbol{x}_n - \boldsymbol{x}_0| = 0$$
EOD.

We have chosen the Euclidean norm for this definition. But as all norms are equivalent, convergence (and limits) do not depend on the norm chosen.

Lemma
$\boldsymbol{x}_n = (x_n, y_n) \to \boldsymbol{x}_0 = (a, b)$ iff $x_n \to a$ and $x_n \to b$. EOL.

Proof. Suppose $\boldsymbol{x}_n \to \boldsymbol{x}_0$. Then $\|\boldsymbol{x}_n - \boldsymbol{x}_0\|_\infty \to 0$. Therefore $\max\{|x_n - a|, |y_n - b|\} \to 0$ and therefore both $x_n \to a$ and $y_n \to b$.

Conversely suppose $x_n \to a$ and $y_n \to b$, then $(x_n - a)^2 + (b_n - a)^2 \to 0$ (sum and products of limits) and as $\sqrt{x}$ is continuous we get $\sqrt{(x_n - a)^2 + (y_n - b)^2} \to 0$. QED.

Definition
A subset A of $\mathbb{R}^2$ is called **closed** if A contains the limit of any convergent sequence in A.
EOD.

Definition
As subset U of $\mathbb{R}^2$ is called **open** if for every $\boldsymbol{x} \in U$ there is $\varepsilon > 0$ such that
$$B_\varepsilon(\boldsymbol{x}) = \{\,\boldsymbol{y} \in \mathbb{R}^2 \mid |\boldsymbol{y} - \boldsymbol{x}| < \varepsilon\,\} \subseteq U$$

EOD.

Exercise
Show that U is open if and only if $\mathbb{R}^2 \setminus U$ is closed. EOE.

Since convergence (and hence being closed) does not depend on the norm, "open" does also not depend on the norm chosen.

If $a < b$ and $c < d$, then $(a, b) \times (c, d)$ is open.

9.1.1 Limits and continuous functions

Let $f: D \to \mathbb{R}$ be a function, where D is open.

Definition
Let $\boldsymbol{x}_0$ be an accumulation point of D. That is, there exists a sequence $\boldsymbol{x}_n \in D$ with $\boldsymbol{x}_n \neq \boldsymbol{x}_0$ for all n such that $\boldsymbol{x}_n \to \boldsymbol{x}_0$.
If $\alpha \in \mathbb{R}$, we write $\lim\limits_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = \alpha$ if $\lim\limits_{n\to\infty} f(\boldsymbol{x}_n) = \alpha$ for every sequence $\boldsymbol{x}_n \in D$ such that $\boldsymbol{x}_n \neq \boldsymbol{x}_0$ and $\lim \boldsymbol{x}_n = \boldsymbol{x}_0$.
Equivalently, for every $\varepsilon > 0$ there is $\delta > 0$ such that for all $\boldsymbol{x} \in B_\delta(\boldsymbol{x}_0) \cap D$ and $\boldsymbol{x} \neq \boldsymbol{x}_0$ we have $|f(\boldsymbol{x}) - \alpha| < \varepsilon$. EOD.

It should be clear what it means for a limit to be $\pm\infty$.

Definition
Let $\boldsymbol{x}_0 \in D$. f is called **continuous** at $\boldsymbol{x}_0$ if for every sequence $\boldsymbol{x}_n \in D$ with $\boldsymbol{x}_n \to \boldsymbol{x}_0$ we have $f(\boldsymbol{x}_n) \to f(\boldsymbol{x}_0)$. Equivalently, for every $\varepsilon > 0$, there is $\delta > 0$ such that for all $\boldsymbol{x} \in B_\delta(\boldsymbol{x}_0) \cap D$, $|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| < \varepsilon$. EOD.

For example, every norm is continuous.

Example

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

is *not* continuous at 0.
While for every zero sequences x_n, y_n we have $f(x_n, 0) \to 0$ and $f(0, y_n) \to 0$ in general we do not have $f(x_n, y_n) \to 0$. Consider $x_n = y_n = \frac{1}{n}$, then

$$f\left(\frac{1}{n}, \frac{1}{n}\right) = \frac{\frac{1}{n^2}}{\frac{2}{n^2}} = \frac{1}{2} \to \frac{1}{2} \neq 0$$

This is an example where the limit does not exist (even though $\lim\limits_{x\to 0} \lim\limits_{y\to 0} f(x, y) = \lim\limits_{y\to 0} \lim\limits_{x\to 0} f(x, y) = 0$).
EOE.

9.2 Differentiation

9.2.1 Directional derivatives

In the following let f be a function defined on an open subset $D \subseteq \mathbb{R}^2$.

Let $\boldsymbol{x}_0 \in D$ and $\boldsymbol{v} \in \mathbb{R}^2$. There is $\varepsilon > 0$ such that $\boldsymbol{x}_0 + t\boldsymbol{v} \in D$ for all $t \in (-\varepsilon, \varepsilon)$. (This uses that D is open.) Consider the function $g(t) = f(\boldsymbol{x}_0 + t\boldsymbol{v})$ defined on $(-\varepsilon, \varepsilon)$.

Definition
The **directional derivative** of f at $\boldsymbol{x}_0$ with respect to $\boldsymbol{v}$ is defined as $g'(0)$ if g is differentiable at 0, and denoted by $\partial_{\boldsymbol{v}} f(\boldsymbol{x}_0)$. EOD.

You will also find notation such as $\frac{\partial}{\partial v}\Big|_{\boldsymbol{x}=\boldsymbol{x_0}} f(\boldsymbol{x})$ or similar. Often it is also assumed that $\|\boldsymbol{v}\|_2 = 1$.

$$\partial_{\boldsymbol{v}} f(\boldsymbol{x}_0) = \lim_{t \to 0} \frac{f(\boldsymbol{x_0} + t\boldsymbol{v}) - f(\boldsymbol{x}_0)}{t}$$

Example
Let $\boldsymbol{v} = (1,2)$ and $f(x, y) = e^{x-y^2}$.
Then $g(t) = e^{x_0+t-(y_0+2t)^2} = e^{x_0-y_0^2} e^{t+4y_0t-4t^2}$ and $g'(t) = e^{x_0-y_0^2} e^{(1+4y_0)t+4t^2}(1 + 4y_0 + 8t)$
Thus, $\partial_{(1,2)} f(x_0, y_0) = e^{x_0-y_0^2}(1 + 4y_0)$. EOE.

9.2.2 Partial derivatives

Consider $e_1 = (1,0)$ and $e_2 = (0,1)$. Then the **partial derivatives** of f are defined as the directional derivatives $\partial_{e_1} f$ and $\partial_{e_2} f$.

Definition
We define $\frac{\partial f}{\partial x}(\boldsymbol{x}_0) \coloneqq \partial_{e_1} f(\boldsymbol{x}_0)$ and $\frac{\partial f}{\partial y}(\boldsymbol{x}_0) \coloneqq \partial_{e_2} f(\boldsymbol{x}_0)$. EOD.

For example,

$$\frac{\partial f}{\partial x}(x_0, y_0) = \frac{d}{dt}\Big|_{t=0} f(x_0 + t, y_0) = \lim_{t \to 0} \frac{f(x_0 + t, y_0) - f(x_0, y_0)}{t}$$

If the partial derivatives exist everywhere on D, then we obtain functions $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$.

9.2.3 C^1-functions

Definition
$f: D \to \mathbb{R}$ is called a C^1-function at $\boldsymbol{x}_0 \in D$, if both partial derivatives of f exist in an open neighborhood U of $\boldsymbol{x}_0$ and are continuous at $\boldsymbol{x}_0$.
f is a C^1-function if it is C^1 at all points of D. EOD.

We can iteratively define C^m functions for $m > 1$: we say f is C^m at $\boldsymbol{x}_0$ if f is C^1 at $\boldsymbol{x}_0$ and the partial derivatives of f are $\boldsymbol{C}^{m-1}$ at $\boldsymbol{x}_0$. And $C^m(D)$ is then the set of functions defined on D that are C^m.

We say a function is C^0 if it is continuous at $\boldsymbol{x}_0$, and consequently $C^0(D)$ denotes the set of continuous functions on D.

The existence of partial derivatives may be useful in some cases, but in general, there is nothing that connects $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$, and we cannot draw a lot of meaningful conclusions about the function f. For instance, suppose f is a function where the partial derivatives exist at $\boldsymbol{0}$. This may tell us something about f along the x- and y-axes, but tells us certainly nothing about anything else.

This changes completely for $\mathcal{C}^1$-functions. The point of derivatives is to give meaningful insights into the change of value of a function.

Suppose f is $\mathcal{C}^1$ at $\boldsymbol{x}_0$. In the following all vectors $\boldsymbol{h} = (h, k)$ are "small" enough such that $\boldsymbol{x}_0 + \boldsymbol{h} \in D$, but also $\boldsymbol{x}_0 + (r, s) \in D$ for all (r, s) with $|r| \leq |h|$ and $|s| \leq |k|$.

We assume that both h, k are nonzero, otherwise the discussion reduces to a one-variable scenario.

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) = \big(f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0 + ke_2)\big) + \big(f(\boldsymbol{x}_0 + ke_2) - f(\boldsymbol{x}_0)\big)$$

We treat each of the brackets on the right separatedly.

Let $\boldsymbol{x}_0 = (x_0, y_0)$. Then

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0 + ke_2) = f(x_0 + h, y_0 + k) - f(x_0, y_0 + k)$$

Viewing this as a function of the first variable (and here then h), the MVT tells us that there is h' with $|h'| < |h|$ such that

$$f(x_0 + h, y_0 + k) - f(x_0, y_0 + k) = \frac{\partial f}{\partial x}(x_0 + h', y_0 + k)h$$

Therefore

$$f(x_0 + h, y_0 + k) - f(x_0, y_0 + k) = \frac{\partial f}{\partial x}(x_0, y_0)h + \left(\frac{\partial f}{\partial x}(x_0 + h', y_0 + k)h - \frac{\partial f}{\partial x}(x_0, y_0)h\right)$$
$$= \frac{\partial f}{\partial x}(x_0, y_0)h + \varrho_1(h, k)h$$

where $\varrho_1(h, k) = \frac{\partial f}{\partial x}(x_0 + h', y_0 + k) - \frac{\partial f}{\partial x}(x_0, y_0)$ (keeping in mind that $h' = h'(h)$ depends on h).

Note this equation still holds, if $h = 0$ if we put $\varrho_1(h, k) = 0$.

Similarly,

$$f(x_0, y_0 + k) - f(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0)k + \left(\frac{\partial f}{\partial y}(x_0, y_0 + k')k - \frac{\partial f}{\partial y}(x_0, y_0)k\right)$$
$$= \frac{\partial f}{\partial y}(x_0, y_0)k + \varrho_2(k)k$$

with $\varrho_2(k) = \frac{\partial f}{\partial y}(x_0, y_0 + k') - \frac{\partial f}{\partial y}(x_0, y_0)$ and again keeping in mind that $k' = k'(k)$ depends on k. Again, this holds if $k = 0$ if we put $\varrho_2(k) = 0$.

Notice that because f is $\mathcal{C}^1$ (and this is where we use this hypothesis), we can conclude that

$$\lim_{(h,k)\to 0} \varrho_1(h, k) = 0$$

Similarly, $\lim_{(h,k)\to 0} \varrho_2(k) = 0$. If we define $R(h, k) \coloneqq \varrho_1(h, k)h + \varrho_2(k)k$ we find

$$\frac{R(h, k)}{|h| + |k|} = \varrho_1(h, k)\left(\frac{h}{|h| + |k|}\right) + \varrho_2(k)\left(\frac{k}{|h| + |k|}\right) \to 0$$

for $(h,k) \to 0$. As all norms are equivalent, we have shown that

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0) = \frac{\partial f}{\partial x}(\boldsymbol{x}_0)h + \frac{\partial f}{\partial y}(\boldsymbol{x}_0)k + R(\boldsymbol{h})$$

with $\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{R(\boldsymbol{h})}{\|\boldsymbol{h}\|} = 0$.

This should remind you of the definition of differentiability in one variable.

9.2.4 Differentiation

Definition

f is **differentiable** at $\boldsymbol{x}_0$ if there is a linear transformation[1] $T: \mathbb{R}^2 \to \mathbb{R}$ such that

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) = f(\boldsymbol{x}_0) + T(\boldsymbol{h}) + R(\boldsymbol{h})$$

where $\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{R(\boldsymbol{h})}{\|\boldsymbol{h}\|} = 0$. f is differentiable if it is differentiable at every point in its domain.
If f is differentiable at $\boldsymbol{x}_0$ we call $f'(\boldsymbol{x}_0) \coloneqq T$ its derivative. EOD.

Below we will show that T is uniquely determined if it exists so this definition makes sense.

As an immediate consequence of the definition we obtain:

Lemma

If f is differentiable at $\boldsymbol{x}_0$, it is continuous at $\boldsymbol{x}_0$. EOL.

Proof. $\lim_{\boldsymbol{h}\to\boldsymbol{0}}(f(\boldsymbol{x}_0 + \boldsymbol{h}) - f(\boldsymbol{x}_0)) = \lim_{\boldsymbol{h}\to\boldsymbol{0}}(f'(\boldsymbol{x}_0)(\boldsymbol{h}) + R(\boldsymbol{h})) = 0$. QED.

Remark

A linear transformation on $\mathbb{R}^2$ can be described by its standard matrix. Here, this means T is determined by two numbers a, b and we then write[2] $T = [a, b]$ such that

$$T\begin{pmatrix} x \\ y \end{pmatrix} = [a, b]\begin{bmatrix} x \\ y \end{bmatrix} = ax + by$$

It is common to write the elements of $\mathbb{R}^2$ here as **column vectors**. EOR.

This begs the question how to determine a and b.

Proposition

If f is differentiable at $\boldsymbol{x}_0$ then all partial derivatives at $\boldsymbol{x}_0$ exist, and

$$f'(\boldsymbol{x}_0) = \left[\frac{\partial f}{\partial x}(\boldsymbol{x}_0), \frac{\partial f}{\partial y}(\boldsymbol{x}_0)\right]$$

More generally, all directional derivatives at $\boldsymbol{x}_0$ exist and $\partial_{\boldsymbol{v}} f(\boldsymbol{x}_0) = f'(\boldsymbol{x}_0)(\boldsymbol{v})$.
EOP.

Proof. Let $\boldsymbol{v} \in \mathbb{R}^2$. Then

$$g(t) \coloneqq f(\boldsymbol{x}_0 + t\boldsymbol{v}) = f(\boldsymbol{x}_0) + f'(\boldsymbol{x}_0)(t\boldsymbol{v}) + R(t\boldsymbol{v}) = f(\boldsymbol{x}_0) + tf'(\boldsymbol{x}_0)(\boldsymbol{v}) + R(t\boldsymbol{v})$$

And

---

[1] A linear transformation is a function such that $T(\boldsymbol{x} + \boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$ and $T(c\boldsymbol{x}) = cT(\boldsymbol{x})$.

[2] This is very sloppy: it depends on the choice of a basis for $\mathbb{R}^2$. We usually choose (e_1, e_2).

$$\lim_{t\to 0}\frac{R(t\boldsymbol{v})}{t}=0$$

Because $\lim_{t\to 0}\frac{R(t\boldsymbol{v})}{\|t\boldsymbol{v}\|}=0$ and $\frac{R(t\boldsymbol{v})}{\|t\boldsymbol{v}\|}=\frac{1}{\|\boldsymbol{v}\|}\frac{R(t\boldsymbol{v})}{|t|}$.

But this means $g'(0)$ exists and is equal to $f'(\boldsymbol{x}_0)(\boldsymbol{v})$. (This was one of the three equivalent definitions for a function to be differentiable.)

But this means $\partial_{e_1}f(\boldsymbol{x}_0)=f'(\boldsymbol{x}_0)(e_1)$ and $\partial_{e_2}f(\boldsymbol{x}_0)=f'(\boldsymbol{x}_0)(e_2)$. Together

$$f'(\boldsymbol{x}_0)(\boldsymbol{h})=\frac{\partial f}{\partial x}(\boldsymbol{x}_0)h_1+\frac{\partial f}{\partial y}(\boldsymbol{x}_0)h_2=\left[\frac{\partial f}{\partial x}(\boldsymbol{x}_0),\frac{\partial f}{\partial y}(\boldsymbol{x}_0)\right]\begin{bmatrix}h_1\\h_2\end{bmatrix}$$

QED.

Note that this also shows that $f'(\boldsymbol{x}_0)$ is uniquely determined.

By the previous section, we conclude that if f is $\mathcal{C}^1$ at $\boldsymbol{x}_0$, then f is differentiable at $\boldsymbol{x}_0$.

The converse is not true.

Exercise
Consider

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

Show that f is differentiable at $(0,0)$ but not $\mathcal{C}^1$. EOE.

On the other hand, the existence of all directional derivatives is not enough for a function to be differentiable:

Exercise
Consider

$$f(x,y)=\vdash\begin{cases}\frac{xy^2}{x^2+y^4} & (x,y)\neq(0,0)\\ 0 & (x,y)=(0,0)\end{cases}$$

Show that all directional derivatives exist, but that f is not differentiable at $(0,0)$. (Check continuity.) EOE.

9.2.5 The gradient of a function

If f is defined on $D\subseteq\mathbb{R}^2$, and the partial derivatives exist at $\boldsymbol{x}_0\in D$, then

$$\operatorname{grad}(f)(\boldsymbol{x}_0):=\begin{bmatrix}\frac{\partial f}{\partial x}(\boldsymbol{x}_0)\\ \frac{\partial f}{\partial y}(\boldsymbol{x}_0)\end{bmatrix}$$

is the column vector given by the partial derivatives. If f is differentiable at $\boldsymbol{x}_0$, then it is the transpose of $f'(\boldsymbol{x}_0)$.

If for $\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^2$ we define $\boldsymbol{v}\cdot\boldsymbol{w}:=v_1w_1+v_2w_2$ (here $\boldsymbol{v}=\begin{bmatrix}v_1\\v_2\end{bmatrix}$ and $\boldsymbol{w}=\begin{bmatrix}w_1\\w_2\end{bmatrix}$), then

$$\|\boldsymbol{v}\|_2 = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$$

This operation is often referred to as the dot-product. It should then not be surprising that there is a relationship between the dot-product of two vectors, and their Euclidean norms.

In the following we assume that $\|\cdot\| = \|\cdot\|_2$ is the Euclidean norm.

This is known as the Cauchy-Schwarz inequality:

$$|\boldsymbol{v} \cdot \boldsymbol{w}| \leq \|\boldsymbol{v}\|\|\boldsymbol{w}\|$$

Because of this, assuming both $\boldsymbol{v}, \boldsymbol{w}$ are nonzero there is a unique $\alpha \in [0, \pi]$ such that

$$\boldsymbol{v} \cdot \boldsymbol{w} = \|\boldsymbol{v}\|\|\boldsymbol{w}\| \cos \alpha$$

This α is called the **angle between $\boldsymbol{v}, \boldsymbol{w}$**. It coincides with the intuitive notion of angles between two vectors (if one finds a way to properly define that).

If $\boldsymbol{v}$ is a *unit vector*, that is a vector for which $\|\boldsymbol{v}\| = 1$, then $\boldsymbol{v} \cdot \boldsymbol{w}$ is maximal if and only if $\alpha = 0$. One can show that this happens if and only if $\boldsymbol{w} = c\boldsymbol{v}$ where $c > 0$. And then $|\boldsymbol{v} \cdot \boldsymbol{w}| = |c| = \|\boldsymbol{w}\|$.

Summarizing the discussion above, if f is differentiable at $\boldsymbol{x}_0$, then for any $\boldsymbol{v}$ we have

$$\partial_{\boldsymbol{v}} f(\boldsymbol{x}_0) = \mathrm{grad}(f)(\boldsymbol{x}_0) \cdot \boldsymbol{v}$$

Assuming $\|\boldsymbol{v}\| = 1$, there is a unique "direction" where this derivative is maximal, namely the direction of the gradient. Its value is then $\|\mathrm{grad}(f)(\boldsymbol{x}_0)\|$. Thus, the gradient points into the direction of "steepest growth" of f. If on the other hand $\boldsymbol{v}$ is tangent to the level set $S \coloneqq f^{-1}(c)$ where $c = f(\boldsymbol{x}_0)$, then $\partial_{\boldsymbol{v}}(f)(\boldsymbol{x}_0) = 0$. Here we call $\boldsymbol{v}$ tangent to S, if there is a function $\gamma: I \to S$ such that $I = (a, b)$ contains 0 and $\gamma(t) = \begin{bmatrix} \gamma_1(t) \\ \gamma_2(t) \end{bmatrix}$, $\gamma(0) = \boldsymbol{x}_0$ and $\gamma'(0) = \begin{bmatrix} \gamma_1'(0) \\ \gamma_2'(0) \end{bmatrix} = \boldsymbol{v}$.

This follows from the fact hat $f(\gamma(t))$ is constant, and one easily checks that $\frac{d}{dt} f(\gamma(t)) = f'(\boldsymbol{x}_0)\gamma'(0) = \partial_{\boldsymbol{v}} f(\boldsymbol{x}_0)$ (see the chain rule below).

9.3 Differentiation in $\mathbb{R}^n$

The situation does not change significantly if we pass to functions on open subsets (defined in the same way) of $\mathbb{R}^n$.

But what if we also change the co-domain from $\mathbb{R}$ to $\mathbb{R}^m$ (say)?

Let $f: D \to \mathbb{R}^m$ be a function where $D \subseteq \mathbb{R}^n$. We write elements of $\mathbb{R}^m$ as column vectors so that

$$f(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$$

It turns out f is *continuous, differentiable, $\mathcal{C}^m$ at $\boldsymbol{x}_0$* if each f_i is (see below).

In particular, if f is differentiable at $\boldsymbol{x}_0$, $f'(\boldsymbol{x}_0)$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$. It corresponds to an $m \times n$ matrix, namely the matrix

$$\begin{bmatrix} f_1'(\boldsymbol{x}_0) \\ f_2'(\boldsymbol{x}_0) \\ \vdots \\ f_m'(\boldsymbol{x}_0) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

This matrix is often called the **Jacobi** matrix of f.

Then $f(\boldsymbol{x}_0 + \boldsymbol{h}) = f(\boldsymbol{x}_0) + f'(\boldsymbol{x}_0)(\boldsymbol{h}) + R(\boldsymbol{h})$

where $R(\boldsymbol{h})$ is now a vector-valued function and

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{R(\boldsymbol{h})}{\|\boldsymbol{h}\|} = \boldsymbol{0}$$

as before.

Equivalently:

Definition
A function $f: D \subseteq \mathbb{R}^n \to \mathbb{R}^m$, where D is open, is called **differentiable** at $\boldsymbol{x}_0 \in D$ if

$$f(\boldsymbol{x}_0 + \boldsymbol{h}) = f(\boldsymbol{x}_0) + T(\boldsymbol{h}) + R(\boldsymbol{h})$$

where $T: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{\|R(\boldsymbol{h})\|}{\|\boldsymbol{h}\|} = 0$$

The linear transformation T is called the **derivative** of f at $\boldsymbol{x}_0$.
f is called **differentiable** if it is differentiable everywhere in D. EOD.

Exercise
Show that $f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$ is differentiable if and only if it each f_i is. And that $f'(\boldsymbol{x}_0) = \begin{bmatrix} f_1'(\boldsymbol{x}_0) \\ f_2'(\boldsymbol{x}_0) \\ \vdots \\ f_m'(\boldsymbol{x}_0) \end{bmatrix}$. EOE.

Remark
The linear transformation T is uniquely determined by the exercise. Indeed, it shows that that if we write T as a matrix the entry at position (i,j) is $\frac{\partial f_i}{\partial x_j}(\boldsymbol{x}_0)$. EOR.
The usual rules of differentiation apply, with the proofs essentially unchanged from the one-variable case.

The most interesting case is the chain rule:

Lemma
Let $f: D \to \mathbb{R}^p$ and $g: E \to \mathbb{R}^q$ be functions where $D \subset \mathbb{R}^n$ is open and $E \subset \mathbb{R}^p$ is open such that $f(D) \subset E$. Then $g \circ f$ is defined.
If f is differentiable at $\boldsymbol{x}_0$ and g is differentiable at $\boldsymbol{y}_0 = f(\boldsymbol{x}_0)$, then $g \circ f$ is differentiable at $\boldsymbol{x}_0$ and

$$(g \circ f)'(\boldsymbol{x}_0) = g'(\boldsymbol{y}_0) f'(\boldsymbol{x}_0)$$

where the product on the right is the composition of linear transformations (or the product of matrices). EOL.

Proof. Again, this is essentially the same proof as in the one-variable case. The only additional observation needed is that if $S: \mathbb{R}^n \to \mathbb{R}^p$ and $T: \mathbb{R}^p \to \mathbb{R}^q$ are linear transformations then $T \circ S: \mathbb{R}^n \to \mathbb{R}^q$ is a linear transformation with standard matrix $[T][S]$ (see below). QED.

The chain rule is often relevant in the following setting:

Let $f: D \to \mathbb{R}$ be a differentiable function defined on an open subset D of $\mathbb{R}^n$. Consider a differentiable function $g: I \to D$ where I is an interval in $\mathbb{R}$.

Then $h(t) \coloneqq f \circ g(t) = f(g_1(t), g_2(t), \dots, g_n(t))$ is differentiable on I and

$$h'(t) = \frac{\partial f}{\partial x_1}\big(g_1(t)\big)g_1'(t) + \frac{\partial f}{\partial x_2}\big(g_2(t)\big)g_2'(t) + \cdots + \frac{\partial f}{\partial x_n}\big(g_n(t)\big)g_n'(t)$$

As a consequence, one also obtains that if $g: E \to D$ is a function defined on some open subset $E \subset \mathbb{R}^p$, such that a certain partial derivative $\frac{\partial g}{\partial z_i}$ exists then $f \circ g$ is partially differentiable in z_i and

$$\frac{\partial (f \circ g)}{\partial z_i} = \frac{\partial f}{\partial x_1}\frac{\partial g_1}{\partial z_i} + \frac{\partial f}{\partial x_2}\frac{\partial g_2}{\partial z_i} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial g_n}{\partial z_i}$$

9.4 *Excursion: The general picture

9.4.1 Some notions from linear algebra

To understand this section, you will need a basic understanding of linear algebra, in particular the notion of vector spaces and linear transformations.

Suppose $X \subseteq \mathbb{R}^n$ is an open subset and $f: X \to \mathbb{R}^m$ is a function differentiable everywhere. By the above, the derivative of f at $\boldsymbol{x}_0$ is a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ depending on the "base point" $\boldsymbol{x}_0$.

If V, W are **vector spaces**, we write $\mathrm{Hom}(V, W)$ for the set of all linear transformations $T: V \to W$.

Here T is a **linear transformation**, if for all $v, w \in V$ and $c \in \mathbb{R}$ we have $T(v + w) = T(v) + T(w)$ and $T(cv) = cT(v)$.

$\mathrm{Hom}(V, W)$ is itself a vector space by pointwise addition and scalar multiplication. Suppose $\dim V = n$ and $\dim W = m$, and let $\beta = (v_1, v_2, \dots, v_n)$ be a basis for V and $\gamma = (w_1, w_2, \dots, w_m)$ be a basis for W. In particular, this means that for example the elements of V are in one-to-one correspondence with the column vectors in $\mathbb{R}^n$:

Equation 9-1

$$v = c_1 v_1 + v_2 v_2 + \cdots + c_n v_n \in V \leftrightarrow [v]_\beta := \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$$

If $\boldsymbol{x}$ is a column vector in $\mathbb{R}^n$, it is sometimes useful to write $\beta\boldsymbol{x}$ for the vector $x_1v_1 + x_2v_2 + \cdots + x_nv_n \in V$. Then the above means $\beta[v]_\beta = v$.

To any $T \in \mathrm{Hom}(V, W)$ one may then associate a $m \times n$ **matrix** $[T]_\beta^\gamma$. Indeed,

$$[T]_\beta^\gamma = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & \ddots & \vdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix}$$

is uniquely defined by the property that $T(v_j) = c_{1j}w_1 + c_{2j}w_2 + \cdots + c_{mj}w_m$ (that is $T(v_j)$ determines (and is determined by) the jth column of the matrix).

Under the above identification of V with $\mathbb{R}^n$ by means of β, and the corresponding identification of W with $\mathbb{R}^m$ by means of γ, we have

$$[T(v)]_\gamma = [T]_\beta^\gamma [v]_\beta$$

where for any vector v in a vector space with basis δ, we write $[v]_\delta$ for the associated column vector defined by the basis (and v). (see Equation 9-1).

The product here is so-called **matrix multiplication**. Where if $A = [a_{ij}]$ is an $m \times n$ matrix and $\boldsymbol{x} \in \mathbb{R}^n$ is a column vector then $A\boldsymbol{x}$ is the column vector in $\mathbb{R}^m$ with entries $y_i = \sum_{k=1}^n a_{ik}x_k$ ("product of the ith row of A with $\boldsymbol{x}$").

Note by means of a basis we can equip any (finite dimensional) vector space with a notion of convergence:

We say $v_n \to w$ if and only if $[v_n]_\beta \to [w]_\beta$. This allows us to define closed sets and as a consequence open subsets (as their complements[3]). It then turns out that this does not depend on the choice of the basis:

Indeed, if β, γ are two bases for V, there is a unique invertible matrix P such that for all $v \in V$ we have $[v]_\beta = P[v]_\gamma$. If $[v_n]_\gamma \to [v]_\gamma$, then $[v_n]_\beta = P[v_n]_\gamma \to P[v]_\gamma = [v]_\beta$. Conversely, if $[v_n]_\beta \to [v]_\beta$, then $[v_n]_\gamma = P^{-1}[v]_\beta \to P^{-1}[v]_\beta = [v]_\gamma$. In both cases we use use that the function $\boldsymbol{x} \to A\boldsymbol{x}$ is a continuous function $\mathbb{R}^n \to \mathbb{R}^n$ if A is an $n \times n$ matrix.

Alternatively, we could just choose any norm on V and again show that all norms are equivalent, and then define convergence using any norm. To pick a norm, define $\|v\| \coloneqq \|[v]_\beta\|$ where the right hand side is formed using a fixed basis β and any norm on $\mathbb{R}^n$ (if $\dim V = n$).

9.4.2 General differentiation

Let V be a fixed n-dimensional vector space and $X \subseteq V$ an open subset. Let W be a vector space of dimension m and $f: X \to W$ be a function.

Then f is **differentiable** if the corresponding function $[f]_\beta^\gamma: [X]_\beta \to \mathbb{R}^m$ is differentiable. Here we write $[X]_\beta$ for the set of all $[v]_\beta$ with $v \in X$. This is an open subset of $\mathbb{R}^n$, and define $[f]_\beta^\gamma$ by the condition

[3] Equivalently, one could define open subsets by means of a norm transferred from $\mathbb{R}^n$ by means of a basis.

$$f(v) = \gamma\left([f]_\beta^\gamma([v]_\beta)\right)$$

Thus, for every $\boldsymbol{x} \in [X]_\beta$, $[f]_\beta^\gamma(\boldsymbol{x})$ is the coordinate vector of $f(\beta\boldsymbol{x})$.

Equivalently, for $v \in X$, and $h \in V$ such that $v + h \in X$ we have

$$f(v+h) = f(v) + T(h) + R(h)$$

such that $T \in \mathrm{Hom}(V, W)$ and $\frac{\|R(h)\|}{\|h\|} \to 0$ for $h \to 0$. Here $\|h\| = \|[h]_\beta\|$ for any norm on $\mathbb{R}^n$, and $\|R(h)\| = \|[R(h)]_\gamma\|$ for any norm on $\mathbb{R}^m$. Again, this does not depend on the choice of bases β, γ (well the actual values may depend on the basis; but the fact that something converges to 0 (or anything else) does not).

Then $T = f'(v)$ is called the derivative of f at v.

Example
Let $T: V \to W$ be a linear transformation. Then T is differentiable everywhere and $T' = T$.
To see this observe that if $v_0 \in V$, then $T(v_0 + h) = T(v_0) + T(h)$. Thus $R(h) = 0$ in this case.

If $f: X \to W$ is a function $x_0 \in X$ and $v \in V$, we can define the **directional derivative** $\partial_v f(x_0)$ analogously to before

$$\partial_v f(x_0) \coloneqq \left.\frac{d}{dt}\right|_0 f(x_0 + tv)$$

which is well defined. As before, if f is differentiable at x_0, then $\partial_v f(x_0) = f'(x_0)(v)$.

Let $\mathcal{D}(X)$ denote the set of all differentiable (real valued) functions on $X \subset V$. Then $\mathcal{D}(X)$ is a vector space, and for $x_0 \in X$ and $v \in V$, the map $\partial_{v,x_0}: f \mapsto \partial_v f(x_0)$ defines a linear transformation.

One can turn this around: for $v \in V$, define $df_{x_0}(v) \coloneqq \partial_v f(x_0) = f'(x_0)(v)$. Then df_{x_0} is a linear transformation $V \to \mathbb{R}$. The assignment Thus, we obtain a function $df: X \to V^*$, $x \mapsto df_x$. This is called a (closed) differential form on X. A general differential form is a function $X \to V^*$. Usually one requires this function to be at least continuous.

9.4.3 Excursion*: "You cannot comb a hedgehog"

Now let $M \subset V$ be a "nice" subset. The prototype of a "nice" subset is a subset of the form $M = \{\, v \in V \mid f_1(v) = f_2(v) = \cdots = f_m(v) = 0 \,\}$ where $f_1, f_2, \dots, f_m: V \to \mathbb{R}$ are differentiable functions such that for all $v \in M$, the derivatives $df_{1v}, df_{2v}, \dots, df_{mv}$ are **linearly independent**. One usually wants that the functions f_i are at least $\mathcal{C}^1$.

We do not want to discuss technical details here. The interested can look up the concept of a "submanifold of $\mathbb{R}^n$". But we will discuss a simple example that encapsulates a lot of the concepts.

Let $S^2 \coloneqq \{v \in \mathbb{R}^3 \mid \|v\|_2 = 1\}$ be the 2-sphere. This is a very interesting geometrical object.

$S^2 = \{\, v \in \mathbb{R}^2 \mid f(v) = 0\}$ where $f(v) = v_x^2 + v_y^2 + v_z^2 - 1$, and v has the three coordinates v_x, v_y, v_z.

Let $x_0 \in S^2$. We call a vector $v \in \mathbb{R}^3$ a **tangent** vector to S^2 at x_0 if $df_{x_0}(v) = 0$. We discussed in a live meeting that this means $\partial_v f(x_0) = 0$, and v is perpendicular to the gradient of f at x_0 (with respect to the dot product), and S^2 is a level set of f (it is after all the set where $f(x) = 0$).

A **vector field** on S^2 is a function $v: S^2 \to \mathbb{R}^3$ such that $v(x)$ is tangent to S^2 for all $x \in S^2$. Typically, one requires at least that v is continuous. So, for a vector field v we have that $v(x) \cdot x = 0$ for all $x \in S^2$. Indeed, the condition that $v(x)$ is tangent to S^2 at x means $df_x\big(v(x)\big) = 0$. Now,

$$df_x\big(v(x)\big) = 2v_1(x)x_1 + 2v_2(x)x_2$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $v(x) = \begin{bmatrix} v_1(x) \\ v_2(x) \end{bmatrix}$.

An important and non-trivial theorem states that if v is a "reasonable" (say, continuous) vector field on S^2, then there is $x_0 \in S^2$ such that $v(x_0) = 0$. See https://en.wikipedia.org/wiki/Hairy_ball_theorem for a discussion.

The "hedgehog" reference stems from the fact that if you accept S^2 as a version of a deformed hedgehog, with its needles being represented as the values of a function $v: S^2 \to \mathbb{R}^3$, you "cannot comb the hedgehog" such that all needles are preserved and are tangent to the hedgehog, so you cannot create a vector field from the needles. It is an entirely different question why anyone would want to comb a hedgehog. As the Wikipedia article states, this is an analogy (?) used in Europe. In North America it is apparently referred to as the "hairy ball theorem."

The theorem has interesting consequences within mathematics: it is known that if a (sufficiently nice) space has a group structure, then there are vector fields on that space that do not vanish anywhere. Thus, as a corollary, we must conclude that S^2 does not allow a "nice" group structure.

This contrasts with the case of the 1-sphere $S^1 = \{v \in \mathbb{R}^2 \mid v_x^2 + v_y^2 = 1\}$, also known as the circle group: it is a group under addition of angles. More precisely, we have seen that $x \in S^1$ if and only if $x = e^{i\alpha}$ for some $\alpha \in \mathbb{R}$, and yes, I am mixing complex and real analysis here. But then $e^{i\alpha}e^{i\beta} = e^{i(\alpha+\beta)}$.

This does turn S^1 into a group.

9.4.4 The second derivative

We start with an open subset $X \subseteq \mathbb{R}^n$ and a differentiable function $f: X \to \mathbb{R}$.

Then $\boldsymbol{x} \mapsto f'(\boldsymbol{x})$ defines a function $X \to W \coloneqq \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$. W isagain a vector space, the so called "dual" space of $\mathbb{R}^n$. If we write the elements of $\mathbb{R}^n$ as column vectors, then the elements of V are naturally row vectors.

Now suppose this map is again differentiable on X. Then its derivative, denoted $f''(\boldsymbol{x})$ is an element of $\mathrm{Hom}(\mathbb{R}^n, W)$. How can we interpret this?

Let $H \in \mathrm{Hom}(\mathbb{R}^n, W) = \mathrm{Hom}(\mathbb{R}^n,\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}))$. Then for any $v \in \mathbb{R}^n$, we get a linear transformation $H(v): \mathbb{R}^n \to \mathbb{R}$. Evaluating this at any $w \in \mathbb{R}^n$, we get a real number $H(v)(w)$, which we write as $H(v, w)$. The fact that both H and $H(v)$ are linear transformations means that

$$(v, w) \mapsto H(v, w)$$

is a **bilinear form**. That means $H(av + bv', w) = aH(v,w) + bH(v',w)$ and $H(v, aw + bw') = aH(v,w) + bH(v,w')$ for all $a, b \in \mathbb{R}$ and $v, v', w, w' \in \mathbb{R}^n$.

Thus, we may interpret $H \coloneqq f''(\boldsymbol{x})$ as a bilinear form.

If the original function f is a C^2 function, then it turns out that H is **symmetric** that is, $H(v,w) = H(w,v)$.

If we choose the **standard basis** $e_1, e_2, \dots, e_n$ for $\mathbb{R}^n$, then we can associate to H a matrix $[H]$ whose entry at position (i,j) is $H(e_i, e_j)$. Then for any $v, w \in \mathbb{R}^n$ we have

$$H(v,w) = v^T[H]w$$

where the multiplication is matrix multiplication and v^T is the *transpose* of v (v written as a row vector).

Can we determine the matrix $[H]$?

Fix $\boldsymbol{x} \in X$. Then $f'(\boldsymbol{x} + \boldsymbol{h}) = f'(\boldsymbol{x}) + f''(\boldsymbol{x})(\boldsymbol{h}) + R(\boldsymbol{h})$. In our notation above $f''(\boldsymbol{x})(\boldsymbol{h}) = H(\boldsymbol{h})$.

Let $\boldsymbol{h} = te_i$. Then

$$f'(\boldsymbol{x} + te_i) = f'(\boldsymbol{x}) + H(te_i) + R(te_i) \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}) = W$$

By means of the standard basis we can identify W with row vectors in $\mathbb{R}^n$, and we get for the jth-entry:

$$\frac{\partial f}{\partial x_j}(\boldsymbol{x} + te_i) = \frac{\partial f}{\partial x_j}(\boldsymbol{x}) + H(te_i)(e_j) + R(te_i)(e_j)$$

Now H is bilinear, so $H(te_i)(e_j) = tH(e_i, e_j)$. But this means $H(e_i, e_j) = \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(\boldsymbol{x})$.

(Because $\lim_{t \to 0} \frac{R(te_i)(e_j)}{t} = 0$. But this would need a proof.)

The upshot is that $[H]$ is the matrix with entry $\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x})$ at position (i,j).

$[H]$ is often called the **Hessian** of f. If f is a C^2-function, then $[H]$ is a **symmetric** matrix.

If we write $f'(\boldsymbol{x}_0)$ as a column vector, so we technically consider $[f'(\boldsymbol{x}_0)]_\beta$, where $\beta = (\lambda_1, \lambda_2, \dots, \lambda_n)$ where $\lambda_i\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = x_i$. Then the derivative of the function $\boldsymbol{x} \mapsto [f'(\boldsymbol{x})]_\beta$ has the matrix $H' \coloneqq \left[h_{i_j}\right]$ where $h_{ij} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\boldsymbol{x}_0)$ at $\boldsymbol{x}_0$. In other words, we obtain the **transpose** of the matrix H above. One needs to be careful how to interpret the matrix. If one chooses the right identifications, there should be no confusion. Also if f is C^2, then $H = H'$.

For such functions, Taylor's Theorem then has the following analogue

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(\boldsymbol{x}) + f'(\boldsymbol{x})(\boldsymbol{h}) + \frac{1}{2}H(\boldsymbol{h}, \boldsymbol{h}) + R(\boldsymbol{h})$$

Where $\lim_{\boldsymbol{h}\to 0} \frac{|R(\boldsymbol{h})|}{\|\boldsymbol{h}\|^2} = 0$.

To avoid the confusion resulting from different identifications, it is useful to avoid the choice of coordinate vectors as follows:

$$\mathrm{Hom}(\mathbb{R}^n, W) \cong (\mathbb{R}^n)^* \otimes W = W \otimes W$$

because $W = (\mathbb{R}^n)^*$. Therefore, our derivative $T \coloneqq f''(\boldsymbol{x}_0) = \sum_{i,j} c_{i,j}\lambda_i \otimes \lambda_j$ for suitable $c_{ij} \in \mathbb{R}$.

To compute c_{ij} consider that $T: \mathbb{R}^n \to W$ and therefore $f'(\boldsymbol{x}_0 + \boldsymbol{h}) = f'(\boldsymbol{x}_0) + T(\boldsymbol{h}) + \cdots$

where $T(\boldsymbol{h}) = \sum_{i,j} c_{ij}\lambda_i(\boldsymbol{h})\lambda_j$. On the other hand $f'(\boldsymbol{x}) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\boldsymbol{x})\lambda_j$, so $f'(\boldsymbol{x}_0 + \boldsymbol{h}) =$

$$\sum \frac{\partial f}{\partial x_j}(\boldsymbol{x}_0)\lambda_j + \sum \left(\frac{\partial f}{\partial x_j}\right)'(\boldsymbol{x}_0)(\boldsymbol{h})\lambda_j + \cdots$$

It now follows that $\sum_i c_{ij}\lambda_i = \left(\frac{\partial f}{\partial x_j}\right)'(\boldsymbol{x}_0)$ (by comparing coefficients of λ_j, and observing that … terms are irrelevant).

But that means that $c_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}_0)$.

One can continue this way and consider third derivatives: the derivative is then naturally an element of $W \otimes W \otimes W$, which has basis $\{\lambda_i \otimes \lambda_j \otimes \lambda_k \mid 1 \le i, j, k \le n\}$, and the coefficient of the derivative in front of $\lambda_i \otimes \lambda_j \otimes \lambda_k$ is

$$\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\boldsymbol{x}_0)$$

Remark
This discussion can be generalized further, so that the nth derivative of a function $f: X \to \mathbb{R}$ corresponds to an n-linear form (ie. a function in n variables, that is linear in each). This requires a bit of knowledge of multi-linear algebra (tensor products also help). EOR.