Analysis II - few selective results

Michael Ruzhansky

December 15, 2008

1 Analysis on the real line

1.1 Chapter: Functions continuous on a closed interval

1.1.1 Intermediate Value Theorem (IVT)

Theorem (Intermediate Value Theorem). *Let f be continuous on a closed interval $[a, b]$. Let ℓ be between $f(a)$ and $f(b)$. Then there is some point $c \in [a, b]$ such that $f(c) = \ell$.*

The following is an example of an application of the IVT.

Theorem. *Let $f : [0, 1] \to [0, 1]$ be continuous. Then f has a fixed point, i.e. there is some point $c \in [0, 1]$ such that $f(c) = c$.*

Proof. First we observe that clearly $f(c) = c$ means $f(c) - c = 0$. This motivates one to introduce function $g(x) = f(x) - x$. We immediately see that g is continuous (on $[0, 1]$) as the difference of two continuous functions. We also have $g(0) = f(0) \geq 0$ and $g(1) = f(1) - 1 \leq 0$ because of the assumption that $0 \leq f(x) \leq 1$ for all $x \in [0, 1]$. But now the IVT assures us that there is some $c \in [0, 1]$ such that $g(c) = 0$, i.e. we have $f(c) = c$ as required.

1.1.2 Continuous functions are bounded

Let I be a set in $\mathbb{R}$. We say that

$$f \quad \text{is bounded on } I \text{ (bounded above, bounded below)}$$

if the set $\{f(x) : x \in I\}$ of values of f on I is bounded (bounded above, bounded below, resp.) in $\mathbb{R}$. In other words, f is bounded above (or below) if there is some finite number $M < \infty$ (or $m > -\infty$) such that $f(x) \leq M$ (or $f(x) \geq m$, resp.) for all $x \in I$. So, f is bounded on I if it is bounded both above and below on I. We will be most interested in I being an open or a closed interval.

Examples.
(1) Function $f(x) = x^2$ is bounded on $I = (-1, 1)$. Indeed, we have $0 \leq f(x) < 1$ for all $x \in I$.
(2) Function $f(x) = 1/x$ on $I = (0, 1)$ is bounded below but not above. Indeed, the

image of I under f is $f(I) = (1, \infty)$, so f on I is bounded below (e.g. by 1) but not above.
(3) Function $f(x) = \tan x$ on $I = (-\pi/2, \pi/2)$ is bounded neither above nor below on I. Indeed, we can easily see that $f(I) = (-\infty, \infty)$.

Thus, we may have all possibilities of functions being bounded or unbounded on open intervals. However, this changes completely if the interval I is closed.

Theorem 2.2. *Let f be a continuous function on a closed interval $[a, b]$. Then f is bounded on $[a, b]$.*

Proof. We need to show that f is bounded both above and below on $[a, b]$. But in fact it is sufficient to only show that f is bounded above because the boundedness below then follows by applying the boundedness above to $-f$ instead of f.

We will argue by contradiction. Suppose f is not bounded above on $[a, b]$. This means, in particular, that for every integer $n \in \mathbb{N}$ there is some point $x_n \in [a, b]$ such that $f(x_n) > n$. The sequence x_n does not have to converge, but it is bounded ($x_n \in [a, b]$ for all n), so we can apply the Bolzano–Weierstrass theorem from M1P1 which said that every bounded sequence of real numbers has a convergent subsequence. Thus, x_n has some convergent subsequence, let us denote it by y_m. So, we have that $\{y_m\}_{m=1}^{\infty} \subset \{x_n\}_{n=1}^{\infty}$, i.e. $y_m = x_n$ for some $n \geq m$. In particular, this implies that

$$f(y_m) = f(x_n) > n \geq m, \tag{1.1}$$

i.e. the sequence $f(y_m)$ is unbounded. Let y be the limit of y_m's. Since $a \leq y_m \leq b$ for all $m \in \mathbb{N}$, we also have $a \leq y \leq b$, i.e. $y \in [a, b]$ and hence f is continuous at y (right or left continuous if $y = a$ or $y = b$, respectively). Now, since $y_m \to y$ as $m \to \infty$ and f is continuous at y, by Theorem 6.5 of M1P1 (see also Theorem 2.0 for one-sided cases), we must also have $f(y_m) \to f(y)$ as $m \to \infty$. But a convergent sequence must be bounded (see problem sheets of M1P1), which contradicts (1.1) that meant that $f(y_m)$ is an unbounded sequence. This completes the proof.

1.1.3 Supremum, infimum, maximum and minimum

In fact, continuous functions on closed intervals not only have to be bounded, but much more is true: they also have to attain their bounds! This will be discussed in the next section.

Let us recall some definitions first. Let f be a function on some set $I \subset \mathbb{R}$, $f : I \to \mathbb{R}$. The *supremum* of f on I, denoted by $\sup_{x \in I} f(x)$, is defined as

$$\sup_{x \in I} f(x) = \text{ the smallest } M \text{ such that } f(x) \leq M \text{ for all } x \in I.$$

If the function is bounded above, its supremum is finite. If not, there is not M as above, and we set the supremum to be $+\infty$. Similarly, the *infimum* of f on I, denoted by $\inf_{x \in I} f(x)$, is defined as

$$\inf_{x \in I} f(x) = \text{ the largest } m \text{ such that } f(x) \geq m \text{ for all } x \in I.$$

Again, a bounded below function has a finite supremum. If f is not bounded below on I, we set $\inf_{x \in I} f(x) = -\infty$.

We will say that f *attains its upper bound* if there is some $c \in I$ such that $f(c) = \sup_{x\in I} f(x)$. In this case we will also write $\max_{x\in I} f(x)$ for $\sup_{x\in I} f(x)$. Similarly, we will say that f *attains its lower bound* if there is some $d \in I$ such that $f(d) = \inf_{x\in I} f(x)$. In this case we will also write $\min_{x\in I} f(x)$ for $\inf_{x\in I} f(x)$. We will say that $f(c)$ and $f(d)$ are the *maximum* and *minimum*, respectively, of function f on the set I.

However, in many situations it is important to talk about local maximum or a local minimum since often local properties of functions are the ones we are interested in. Thus, we can say that f has a *local maximum* at c if there is some interval containing c such that f attains its maximum in this interval in the point c. Note that the interval may be very small so no information about the behaviour of f outside this interval is included in the definition. To be rigorous, we define $f(c)$ to be the *local maximum* of f if

$$\text{there is } \delta > 0 \text{ such that } |x - c| < \delta \text{ implies } f(x) \le f(c).$$

Similarly, we will say that f has a *local minimum* at d if

$$\text{there is } \delta > 0 \text{ such that } |x - d| < \delta \text{ implies } f(x) \ge f(d).$$

1.1.4 Extreme Value Theorem (EVT)

The following (extreme value) theorem says that continuous functions on closed intervals are not only bounded, but they also attain their upper and lower bounds.

Extreme Value Theorem. *Let f be a continuous function on a closed interval $[a, b]$. Then f attains its upper and lower bounds on $[a, b]$, i.e. there exist some points $c, d \in [a, b]$ such that $f(d) \le f(x) \le f(c)$ for all $x \in [a, b]$.*

Clearly, this theorem implies Theorem 2.2. However, we will actually use Theorem 2.2 in the proof of the extreme value theorem. In the notation of 1.1.3, we may also write

$$f(c) = \max_{x\in I} f(x) \text{ and } f(d) = \min_{x\in I} f(x).$$

Let us now prove the extreme value theorem.

Proof. It is enough to prove only half of the statement, for example the existence of $c \in [a, b]$ such that $f(x) \le f(c)$ for all $x \in [a, b]$. Indeed, if we apply this to $-f$ instead of f, we get that there exists some $d \in [a, b]$ such that $-f(x) \le -f(d)$ for all $x \in [a, b]$, which would mean that $f(d) \le f(x)$ for all $x \in [a, b]$.

By Theorem 2.2 function f is bounded above on $[a, b]$, i.e. the supremum of f on I is finite. Let us denote $M = \sup_{x\in[a,b]} f(x)$. If $f(c) = M$ for some $c \in [a, b]$, the proof would be finished, so let us assume that $f(x) < M$ for all $x \in [a, b]$. Let us then consider the function

$$g(x) = \frac{1}{M - f(x)}$$

Clearly, g is well defined on $[a, b]$ since $f(x) < M$ for all $x \in [a, b]$. Moreover, g is continuous on $[a, b]$ by Theorem 6.7 of M1P1 as a quotient of two continuous functions with non-vanishing denominator. Now, by the definition of the supremum, M is the smallest upper bound for f. This means that for any $\epsilon > 0$ the number $M - \epsilon$ is not the upper bound for f on $[a, b]$, i.e. for every $\epsilon > 0$ there exists some $x \in [a, b]$ such that $f(x) > M - \epsilon$. But then it follows that $g(x) > 1/\epsilon$, and hence g is not bounded above on $[a, b]$. Indeed, for each $n \in \mathbb{N}$, taking $\epsilon = 1/n$ we would get that there is some $x \in [a, b]$ with $g(x) > n$, implying that g is not bounded above. On the other hand, as we have seen, g is a continuous function on a closed interval $[a, b]$, and hence it should be bounded by Theorem 2.2. This contradiction shows that value M has to be attained by f completing the proof of the extreme value theorem.

An **alternative proof to the extreme value theorem** is possible, similar to Theorem 2.2. Let us denote again $M = \sup_{x\in[a,b]} f(x)$. Then there exists a sequence of points $x_n \in [a, b]$ such that $f(x_n) \to M$ as $n \to \infty$. Indeed, for each $n \in \mathbb{N}$ the number $M - 1/n$ is not an upper bound for f on $[a, b]$, hence there exists some $x_n \in [a, b]$ such that $M - 1/n < f(x_n) \leq M$. Passing to the limit we must have $f(x_n) \to M$ as $n \to \infty$. This sequence $\{x_n\} \subset [a, b]$ is a bounded sequence and by the Bolzano–Weierstrass theorem there exists a convergent subsequence which we denote by y_m. Thus, $y_m \to c$ as $m \to \infty$, for some $c \in [a, b]$. Since f is continuous at c, we must also have $f(y_m) \to f(c)$ as $m \to \infty$. On the other hand, $f(y_m) \to M$ as $m \to \infty$ since $\{y_m\}$ is a subsequence of $\{x_n\}$. It follows from the uniqueness of limits that $f(c) = M$ and hence the supremum of f on $[a, b]$ is attained at the point c.

1.1.5 Inverse Function Theorem (IFT)

Let $f : [a, b] \to \mathbb{R}$ be a strictly increasing continuous function. By definition, this means that if $x_1, x_2 \in [a, b]$ and $x_1 < x_2$, then $f(x_1) < f(x_2)$. Let $c = f(a)$ and $d = f(b)$, so that $c < d$. We claim first that f is a bijection from $[a, b]$ to $[c, d]$ (bijective means injective and surjective). Clearly, $f(x_1) = f(x_2)$ implies $x_1 = x_2$ (this means that f is injective). Indeed, if either $x_1 < x_2$ or $x_2 < x_1$, then we would have $f(x_1) < f(x_2)$ or $f(x_2) < f(x_1)$, resp., because f is strictly increasing. On the other hand, by the intermediate value theorem f should take all values between c and d (this means that f is surjective). To summarise, we showed that a strictly increasing function $f : [a, b] \to [c, d]$ with $c = f(a)$ and $d = f(b)$ takes each value in $[c, d]$ exactly once, so we may define its inverse $g : [c, d] \to [a, b]$ by setting

$$x = g(y) \quad \text{if and only if} \quad y = f(x).$$

Inverse Function Theorem. *Let f be a strictly increasing continuous function on a closed interval $[a, b]$, let $c = f(a), d = f(b)$, and let $g : [c, d] \to [a, b]$ be its inverse. Then g is a strictly increasing continuous function on $[c, d]$.*

Proof. Let us first show that g is strictly increasing. Let $y_1, y_2 \in [c, d]$ be such that $y_1 < y_2$. Let $x_1 = g(y_1)$ and $x_2 = g(y_2)$, so that $y_1 = f(x_1)$ and $y_2 = f(x_2)$. We then have either $x_1 < x_2$, or $x_1 = x_2$, or $x_1 > x_2$. Let us show that only the first case is possible. Indeed, if $x_1 = x_2$ or $x_1 > x_2$, then $f(x_1) = f(x_2)$ or $f(x_1) > f(x_2)$, since f is strictly increasing. But this contradicts the assumption $y_1 < y_2$. Hence the only possibility left is $x_1 < x_2$, which shows that g is strictly increasing.

Let us now show that g is continuous on $[c, d]$. Let $\beta \in (c, d)$ and let $\alpha = g(\beta)$, so that $\beta = f(\alpha)$. Let us prove that g is continuous at β. For every $\epsilon > 0$ let us take $\delta = \min\{f(\alpha + \epsilon) - \beta, \beta - f(\alpha - \epsilon)\}$. Then one can readily see that $|y - \beta| < \delta$ implies that $f(\alpha-\epsilon) < y < f(\alpha+\epsilon)$, and hence also $\alpha-\epsilon < g(y) < \alpha+\epsilon$ because g is strictly increasing. If we recall that $\alpha = g(\beta)$, we see that we proved that for each $\epsilon > 0$ there is some $\delta > 0$ such that $|y-\beta| < \delta$ implies $|g(y) - g(\beta)| < \epsilon$. This means precisely that g is continuous at β. In the case of β being an endpoint of $[c, d]$, a simple modification of this argument for one-sided continuity implies that g is also continuous at the endpoints, completing the proof of the inverse function theorem.

1.1.6 Derivative

We will first motivate the definition by some informal discussion of what we want to have.

Informal definition. Imagine that we have drawn the graph of a function f on the usual plane $\mathbb{R}^2$. The derivative of f at a has the meaning of the value of the slope of the line which is tangent to the graph of f at the point $(a, f(a))$. However, we have to do some extra work since $(a, f(a))$ is the only point that we know that lies on this line, and to know the line precisely we need to know at least two points. The way to get around this is to try to approximate the tangent line by lines that pass through two points on the graph - the main point $(a, f(a))$ and the nearby points. Such points are all of the form $(x, f(x))$ because they lie on the graph of f. So, we approximate the tangent line to the graph of f at the point $(a, f(a))$ by lines through points $(a, f(a))$ and $(x, f(x))$ as x approaches a. The slope of the line through two points $(a, f(a))$ and $(x, f(x))$ on the place is given by $\frac{f(x)-f(a)}{x-a}$. So we arrive at the formal definition:

Formal definition. Let function f be defined on some open interval containing a point $a \in \mathbb{R}$. We will say that f is *differentiable at a* if the expression $\frac{f(x)-f(a)}{x-a}$ has a limit as $x \to a$. In this case we will call the limit $f'(a)$, so that we have

$$\boxed{f'(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a}}$$

We can note two things. First, at the point $x = a$ the expression under the sign of the limit is of the form $\frac{0}{0}$, which is undefined. However, we know that we do not look at the value of the function at the limiting point when we take limits, so we only look at $x \neq a$, where the expression under the limit makes perfect sense. Another thing is that it is often convenient to emphasise the fact that we are around point a by writing $\boxed{x = a + h}$ Clearly, x is close to a if and only if h is close to 0, and $x \to a$ if and only if $h \to 0$. We also note that obviously $\frac{f(x)-f(a)}{x-a} = \frac{f(a+h)-f(a)}{h}$. Thus, the definition of the derivative becomes

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

1.1.7 Chain rule

The chain rule is extremely useful when differentiating composed functions, like 2^{3^x}, $\sin(x^7)$, etc. The composition $f \circ g$ of two functions f and g is defined as usual by $(f \circ g)(x) = f(g(x))$.

Theorem (Chain rule). *Let g be differentiable at a and let f be differentiable at $g(a)$. Then $f \circ g$ is also differentiable at a and we have*

$$\boxed{(f \circ g)'(a) = f'(g(a))g'(a)} \qquad \text{(the chain rule)}$$

Proof. We have to show that

$$\frac{(f \circ g)(x) - (f \circ g)(a)}{x - a} \to f'(g(a))g'(a) \quad \text{as} \quad x \to a.$$

For the proof we will introduce two auxiliary functions F and Δ and will rewrite the limit in terms of these functions. Having established basic properties of F and Δ, we will be able to prove the chain rule.

We set $\beta = g(a)$ and define the function

$$F(y) = \begin{cases} \frac{f(y)-f(\beta)}{y-\beta}, & \text{if} \quad y \neq \beta \\ f'(\beta), & \text{if} \quad y = \beta. \end{cases}$$

Then by differentiability of f at β we can conclude that for $y \neq \beta$ we have

$$F(y) = \frac{f(y) - f(\beta)}{y - \beta} \to f'(\beta) = F(\beta) \quad \text{as} \quad y \to \beta,$$

so F is continuous at β. On the other hand we also note the formula

$$(y - \beta)F(y) = f(y) - f(\beta),$$

which holds for all y, including $y = \beta$, when both sides are just zeros. Let us now define another function

$$\Delta(x) = g(x) - g(a).$$

We can easily observe that $\frac{\Delta(x)}{x-a} \to g'(a)$ as $x \to a$, simply by definition of Δ and differentiability of g at a. For $x \neq a$, we can now rewrite the limit we need to investigate as

$$\begin{aligned} \frac{(f \circ g)(x) - (f \circ g)(a)}{x - a} &= \frac{f(g(x)) - f(g(a))}{x - a} \quad \text{(use def of } \Delta(x)) \\ &= \frac{f(g(a) + \Delta(x)) - f(g(a))}{x - a} \quad \text{(use } \beta = g(a)) \\ &= \frac{f(\beta + \Delta(x)) - f(\beta)}{x - a} \quad \text{(use formula for } (y - \beta)F(y)) \\ &= \frac{\Delta(x)F(\beta + \Delta(x))}{x - a} \quad \text{(use } \beta + \Delta(x) = g(x)\text{)} \\ &= F(g(x))\frac{\Delta(x)}{x - a}. \end{aligned}$$

Now, since g is continuous at a (it is differentiable at a, hence also continuous), and F is continuous at $\beta = g(a)$, we get that $F \circ g$ is continuous at a and hence $F(g(x)) \to F(g(a)) = F(\beta) = f'(\beta) = f'(g(a))$ as $x \to a$. Moreover, we already know that $\frac{\Delta(x)}{x-a} \to g'(a)$ as $x \to a$. Therefore, we get

$$\frac{(f \circ g)(x) - (f \circ g)(a)}{x - a} = F(g(x))\frac{\Delta(x)}{x - a} \to f'(g(a))g'(a) \quad \text{as} \quad x \to a,$$

which shows that $f \circ g$ is differentiable at a and proves the chain rule.

1.1.8 Inverse Function Theorem (for derivatives)

Here we will derive the formula for the derivative of the inverse function. Similar to the setting of 1.1.5, let $f : [a, b] \to [c, d]$ be a strictly increasing function, continuous on $[a, b]$, such that $c = f(a)$ and $d = f(b)$. From 1.1.5 we know that its inverse $g : [c, d] \to [a, b]$ is also strictly increasing and continuous. A natural question now is *whether the differentiability of f at some $\alpha \in (a, b)$ implies the differentiability of g at $\beta = f(\alpha) \in (c, d)$.* Unfortunately, there is an obstruction to such a statement. Indeed, supposing that this was true, let us differentiate the identity $x = (g \circ f)(x)$ at α. Then by the chain rule we get

$$1 = (g \circ f)'(\alpha) = g'(\beta) f'(\alpha).$$

However, this produces a contradiction if $f'(\alpha) = 0$, since the right hand side is then zero. Fortunately, this is the only obstruction for us to be able to differentiate the inverse function.

Theorem. *Let $f : [a, b] \to [c, d]$ be a strictly increasing function, continuous on $[a, b]$, such that $c = f(a)$ and $d = f(b)$. Let $g : [c, d] \to [a, b]$ be the inverse of f. Let $\alpha \in (a, b)$ and suppose that f is differentiable at α and that $f'(\alpha) \neq 0$. Then g is differentiable at $\beta = f(\alpha)$ and*

$$g'(\beta) = \frac{1}{f'(\alpha)}.$$

Expressing everything in terms of β, i.e. using that $\alpha = g(\beta)$, we can write the formula for the derivative as

$$\boxed{g'(\beta) = \frac{1}{f'(g(\beta))}}$$

We can also note that it is in fact enough to only show that g is differentiable at β. Indeed, in this case the required formula follows from the established above identity $1 = (g \circ f)'(\alpha) = g'(\beta) f'(\alpha)$. However, below we give the prove which also yields the formula for the derivative.

Proof. We have to show that

$$\frac{g(y) - g(\beta)}{y - \beta} \to \frac{1}{f'(\alpha)} \quad \text{as} \quad y \to \beta.$$

Let us now rewrite this in terms of expressions involving f so that we could use that f is differentiable at α. Let us define

$$\Delta(y) = g(y) - g(\beta).$$

Since $g(\beta) = \alpha$, we can write this as $g(y) = \alpha + \Delta(y)$, which means that $y = f(\alpha + \Delta(y))$. Thus, we have to prove that

$$\frac{\Delta(y)}{f(\alpha + \Delta(y)) - f(\alpha)} \to \frac{1}{f'(\alpha)} \quad \text{as} \quad y \to \beta,$$

where we used that $\beta = f(\alpha)$. We know that this is equivalent to proving that

$$\frac{f(\alpha + \Delta(y)) - f(\alpha)}{\Delta(y)} \to f'(\alpha) \quad \text{as} \quad y \to \beta,$$

provided that we never divided by zero. This is indeed the case if we notice that $\Delta(y) = 0$ if and only if $y = \beta$ since g is strictly increasing. Also, $f(\alpha + \Delta(y)) = f(\alpha)$ if and only if $y = \beta$ since f is strictly increasing. We also know that we need not look at the limit point $y = \beta$ when taking limits, so all the expressions on the left hand sides are non-zero. Now, let us introduce function

$$F(x) = \begin{cases} \frac{f(x)-f(\alpha)}{x-\alpha}, & \text{if} \quad x \neq \alpha \\ f'(\alpha), & \text{if} \quad x = \alpha. \end{cases}$$

We immediately notice that F is continuous at α because f is differentiable at α (similar to the argument in the proof of the chain rule), and that we have the formula

$$(x - \alpha)F(x) = f(x) - f(\alpha)$$

for all x. In particular, taking $x = \alpha + \Delta(y)$, we get

$$\Delta(y)F(\alpha + \Delta(y)) = f(\alpha + \Delta(y)) - f(\alpha).$$

Now, for $y \neq \beta$, we can write

$$\frac{f(\alpha + \Delta(y)) - f(\alpha)}{\Delta(y)} = \frac{\Delta(y)F(\alpha + \Delta(y))}{\Delta(y)} = F(\alpha + \Delta(y)),$$

where we divided by non-zero $\Delta(y)$. Since Δ is continuous at β (because g is continuous at β), $\Delta(\beta) = 0$, and since F is continuous at $\alpha = \alpha + \Delta(\beta)$, we get that the composition $F(\alpha + \Delta(y))$ is continuous at β. Hence its limit as $y \to \beta$ is equal to its value at β which is $F(\alpha + \Delta(\beta)) = F(\alpha) = f'(\alpha)$, which is what was required to prove.

1.1.9 Derivative at an extremal point is zero

It turns out that the extremal points discussed in 1.1.3 have a useful characterisation in terms of derivatives.

Theorem. *Let f be differentiable at a and assume that f has a local maximum or a local minimum at a. Then $f'(a) = 0$.*

Proof. Let us deal with the case when f has a maximum at a. For the case of a local minimum we simply observe that f has a local maximum at a point if and only if $-f$ has a local minimum at the same point, so an application of the maximum case to $-f$ would immediately imply the statement for the minimum.

Thus, without loss of generality we may assume that f has a local maximum at a, i.e. there is some $\delta > 0$ such that $|x - a| < \delta$ implies $f(x) \leq f(a)$. We may restrict our attention only to x with $|x - a| < \delta$ since we do not need to look at other x to deal with the derivative of f at a. So, if $x < a$, we must have $f(x) - f(a) \leq 0$ and $x - a < 0$, and hence $\frac{f(x)-f(a)}{x-a} \geq 0$. Since f is differentiable at a, it is also left differentiable at a, so this expression must have a limit as $x \to a-$. By the sandwich theorem we get

$$f'_-(a) = \lim_{x \to a-} \frac{f(x) - f(a)}{x - a} \geq 0.$$

On the other hand, if $x > a$, we have $f(x) - f(a) \leq 0$ and $x - a > 0$, and hence $\frac{f(x)-f(a)}{x-a} \leq 0$. So by the sandwich theorem again we get

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

Since f is differentiable at a, the values of left and right derivatives at a should coincide, so we must have $f'(a) = f'_-(a) \geq 0$ and $f'(a) = f'_+(a) \leq 0$, which implies that $f'(a) = 0$.

1.1.10 Rolle's Theorem

Now we can argue that if we have a function which is differentiable in a closed interval, it should attain a maximum (and a minimum) somewhere by the Extreme Value Theorem. But then at such point the derivative should be equal to zero by Theorem 1.1.9. This is the statement of the following theorem where we may actually assume slightly less, namely that the function is differentiable only up to the end-points of our interval. At the end-points it is sufficient to assume only continuity.

Rolle's Theorem. *Let f be continuous on $[a, b]$ and differentiable on (a, b). Assume that $f(a) = f(b)$. Then there is some point $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. The proof is essentially already given above before the formulation, we only have to take care of what may happen at the end-points a and b. Thus, since f is continuous on $[a, b]$, the extreme value theorem EVT implies that there are some points $c_1, c_2 \in [a, b]$ such that f attains its maximum at c_1 and its minimum at c_2.

If any of points c_1 or c_2 lies in the open interval (a, b), let us call this point c. Since $c \in (a, b)$ and f is differentiable on (a, b), function f must be differentiable in particular at c. But then by Theorem 1.1.9 we must have $f'(c) = 0$, which proves the theorem.

Now, if none of c_1, c_2 lie in (a, b), they must be the end-points. Since $f(a) = f(b)$ by assumption, we must have $f(c_1) = f(c_2)$. But this means that the maximal and the minimal values of f on $[a, b]$ are the same. The only possibility is that f is constant on $[a, b]$. In turn, this means that $f'(c) = 0$ for all points $c \in (a, b)$, proving the theorem also in this case.

Remark. Note that in the formulation of Rolle's theorem we have not assumed f to be differentiable at the end-points a and b. This means, in particular, that we may apply Rolle's theorem to the function

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

on any interval $[0, b_k]$, with any $b_k = \frac{1}{2\pi k}$, for $k \in \mathbb{N}$. Indeed, although f is not differentiable at 0, it satisfies the assumptions on Rolle's theorem on the closed interval $[0, b_k]$ since $f(0) = f(b_k) = 0$, and it is differentiable on $(0, b_k)$ and continuous on $[0, b_k]$. Thus, there must be some $0 < c < b_k$ such that $f'(c) = 0$. Applying this argument for all $k \in \mathbb{N}$, we can easily observe that for any positive $b > 0$, there are infinitely many points $c_k \in [0, b]$ where the derivative f' vanishes. Of course, we could reach the same conclusion by simply looking at the graph of f or by writing out explicit formulae for all c_k's, but Rolle's theorem provides a rigorous (although quite inefficient in this case) proof for this observation.

1.1.11 Test for extremal points

It is very useful to be able to determine properties of function using properties of derivatives. Indeed, in many situations it may be difficult to compare different values of a function while it may be much easier to determine the sign of its first and higher order derivatives.

Theorem. *Let us assume that $f'(a) = 0$ and that f is twice differentiable at a. Then*

(i) *if $f''(a) > 0$, then f has a local minimum at a;*

(ii) *if $f''(a) < 0$, then f has a local maximum at a;*

(iii) *if $f''(a) = 0$, we can not make any conclusion.*

Proof. (i) Since f is twice differentiable at a and $f'(a) = 0$, we have

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

Since $f''(a) > 0$, the expression under the limit must be positive for x close to a. In other words, there is some $\delta > 0$ such that $\frac{f'(x)}{x-a} > 0$ for all $0 < |x - a| < \delta$.

Now, if we are to the left of a, we have $x - a < 0$, and hence $f'(x) < 0$. Since f is continuous on $[a - \delta, a]$, differentiable on $(a - \delta, a)$, and $f'(x) < 0$ for all $x \in (a - \delta, a)$, it follows that f is strictly decreasing in $[a - \delta, a]$. In particular it means that $f(a) < f(x)$ for all $a - \delta \le x < a$.

If we are to the right of a, we have $x - a > 0$, and hence $f'(x) > 0$. By an argument similar to the one just given, we may conclude that f is strictly increasing on $[a, a + \delta]$, which means, in particular, that $f(a) < f(x)$ for all $a < x \le a + \delta$. Thus, we proved that $f(a) < f(x)$ for all $a - \delta \le x \le a + \delta$, which means that f has a local minimum at a.

To prove (ii) we may notice that we can apply part (i) to $-f$ instead of f to conclude that $-f$ has a local minimum at a. But this means that f has a local maximum at a.

Finally, to prove (iii) we just need to produce some examples. So, if we take $f_\pm(x) = \pm x^4$, then $f''_\pm(0) = 0$, and f_+ has a local minimum at 0 while f_- has a local maximum. In addition, function $f(x) = x^3$ also has $f''(0) = 0$, and neither minimum nor maximum at 0.

2 Basics of analysis in $\mathbb{R}^n$

2.1 Continuity in $\mathbb{R}^n$

2.1.1 Properties of vectors in $\mathbb{R}^n$

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Let also $\alpha \in \mathbb{R}$ and $y \in \mathbb{R}^n$. The space $\mathbb{R}^n$ is a linear space with operations

$$\alpha x = (\alpha x_1, \dots, \alpha x_n), \quad x + y = (x_1 + y_1, \dots, x_n + y_n).$$

It has the zero element $0 = (0, \dots, 0) \in \mathbb{R}^n$, the standard inner product $x \cdot y = x_1 y_1 + \dots + x_n y_n$, the norm (length)

$$|x| = \sqrt{x \cdot x} = \sqrt{x_1^2 + \cdots + x_n^2},$$

and the distance $|x-y| = \sqrt{(x-y)\cdot(x-y)}$. The norm satisfies the following properties

(i) $|x| \geq 0$ for all $x \in \mathbb{R}^n$, and $|x| = 0$ is equivalent to $x = 0$;

(ii) $|\alpha x| = |\alpha| \cdot |x|$ for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$;

(iii) the triangle inequality $|x+y| \leq |x| + |y|$ holds for all $x, y \in \mathbb{R}^n$.

The triangle inequality follows from the following

Cauchy–Schwartz inequality. *Let $x, y \in \mathbb{R}^n$. Then $|x \cdot y| \leq |x||y|$.*

Proof. For $\epsilon > 0$, we have $0 \leq |x \pm \epsilon y|^2 = |x|^2 \pm 2\epsilon x \cdot y + \epsilon^2 |y|^2$. This implies $\pm x \cdot y \leq \frac{1}{2\epsilon}|x|^2 + \frac{\epsilon}{2}|y|^2$. Setting $\epsilon = \frac{|x|}{|y|}$, we obtain the required inequality, provided $y \neq 0$ (if $y = 0$ it is trivial).

An alternative proof may be given as follows. We can observe that the inequality $0 \leq |x + \epsilon y|^2 = |x|^2 + 2\epsilon x \cdot y + \epsilon^2 |y|^2$ implies that the discriminant of the quadratic (in ϵ) polynomial on the right hand side must be non-positive, which means $|x \cdot y|^2 - |x|^2 |y|^2 \leq 0$.

The Cauchy–Schwartz inequality implies the triangle inequality because we have

$$\begin{aligned} |x+y|^2 &= (x+y)\cdot(x+y) = x \cdot x + 2x \cdot y + y \cdot y \\ &\leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2, \end{aligned}$$

implying the triangle inequality.

Linear spaces with a norm are called *normed linear spaces.* So $\mathbb{R}^n$ is an example of a normed linear space.

2.1.2 Continuity

Recall that all Theorems of Chapter 6 of M1P1 continue to hold for functions $f : \mathbb{R}^n \to \mathbb{R}$. In particular, we (re)define the notion of continuity in the following way. We say that f *is continuous at* $a \in \mathbb{R}^n$ if

for every $\epsilon > 0$ there is some $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$.

We will also say that f *has limit* ℓ *at* a, and will write $\lim_{x \to a} f(x) = \ell$, if

for every $\epsilon > 0$ there is some $\delta > 0$ such that $0 < |x - a| < \delta$ implies $|f(x) - \ell| < \epsilon$.

One has to pay more attention to continuity in $\mathbb{R}^n$ compared to $\mathbb{R}$. In particular, continuity coordinate-wise is not equivalent to continuity for $n \geq 2$.

Theorem. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous at $a \in \mathbb{R}^n$. Let $a = (a_1, \dots, a_n)$. Then the function $x_k \mapsto f(a_1, \dots, a_{k-1}, x_k, a_{k+1}, \dots, a_n)$ is continuous at $a_k \in \mathbb{R}$, for all $k = 1, \dots, n$.*

Proof. Let $\epsilon > 0$. Since f is continuous at a, there is some $\delta > 0$ such that $|x - a| < \delta$ implies $|f(x) - f(a)| < \epsilon$. If we observe that $|x_k - a_k| \leq |x - a|$, we see that $|x_k - a_k| < \delta$ implies that $|f(a_1, \ldots, a_{k-1}, x_k, a_{k+1}, \ldots, a_n) - f(a)| < \epsilon$, which means that the function in the theorem is continuous at a_k.

The converse to this theorem is not true. Let us consider an example. Let us define $g : \mathbb{R}^2 \to \mathbb{R}$ by

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

Let use denote $g_x(y) = g(x, y)$ and $g^y(x) = g(x, y)$. These are functions obtained from looking at g coordinate-wise. Then for all $x, y \in \mathbb{R}$, functions $g_x(\cdot)$ and $g^y(\cdot)$ are continuous. This can be formulated as saying that

Statement 1. For any fixed $x \in \mathbb{R}$, the function $y \mapsto g(x, y)$ is continuous on $\mathbb{R}$, and for any fixed $y \in \mathbb{R}$, the function $x \mapsto g(x, y)$ is continuous on $\mathbb{R}$.

Proof. Indeed, let us show, for example, that $g_x : \mathbb{R} \to \mathbb{R}$ is continuous. If $x = 0$, we have $g_x(y) = 0$ for all $y \in \mathbb{R}$, so g_x is continuous in this case. Let us now assume $x \neq 0$. Then functions $y \mapsto xy$ and $y \mapsto x^2 + y^2$ are continuous, while the second function is never zero. Hence their quotient is also continuous. Since this quotient is precisely g_x, we obtain the claim.

So, if we look at g coordinate-wise, it is everywhere continuous. On the other hand, we have

Statement 2. Function g is continuous at any point except $(0, 0)$. It is not continuous at $(0, 0)$.

Proof. Let us show first that g is not continuous at $(x, y) = (0, 0)$. Indeed, if g were continuous at $(0, 0)$, by the characterisation of continuity in terms of sequences, we would have $g(x_k, y_k) \to g(0, 0) = 0$ for all sequences $(x_k, y_k) \to (0, 0)$. To get a contradiction, let us pick a sequence for which this is not true. Let $(x_k, y_k) = (1/k, 1/k)$. Then $g(x_k, y_k) = 1/2$ for all k, yielding the contradiction.

Finally, let us show that g is continuous at all other points. Indeed, if $a \in \mathbb{R}^2$ is such that $a \neq 0$, both functions $(x, y) \mapsto xy$ and $(x, y) \mapsto x^2 + y^2$ are continuous at a. The second function is non-zero for (x, y) close to a, hence their quotient is also continuous at a.

Conclusion: a function g may be continuous separately in x and y, without being continuous in (x, y).

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

2.2.1 Differentiation: gradient

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called *differentiable at a point* $a \in \mathbb{R}^n$ if there exists $v \in \mathbb{R}^n$ such that

$$\boxed{f(a + h) - f(a) = v \cdot h + o(h)}$$

where $o(h)$ denotes any function such that $\frac{o(h)}{|h|} \to 0$ as $h \to 0$. In other words, if $[f(a+h) - f(a) - v \cdot h]/|h| \to 0$ as $h \to 0$. The vector v is called *the gradient* of f at a, and is denoted by $\nabla f(a)$ or by $\frac{df}{dx}(a)$.

Note that in the case $n = 1$ we could simply divide by h to recover the definition in Chapter 3 with $v = f'(a)$. Here we can not simply divide by h, because is a vector, or even by $|h|$, since the inner product has no division.

Alternatively, we can say that f is differentiable at a if

$$f(x) = f(a) + v \cdot (x - a) + o(x - a),$$

where $\frac{o(x-a)}{|x-a|} \to 0$ as $x \to a$.

2.2.2 Differentiation: partial derivatives

By fixing all coordinates but x_k, one can consider a function

$$x_k \mapsto f(a_1, \ldots, a_{k-1}, x_k, a_{k+1}, \ldots, a_n),$$

as above. This is a function on $\mathbb{R}$ and can be differentiated in the usual way. Its derivative is denoted by $\frac{\partial f}{\partial x_k}$, and we have

$$\frac{\partial f}{\partial x_k}(a) = \lim_{t \to 0} \frac{f(a_1, \ldots, a_k + t, \ldots, a_n) - f(a_1, \ldots, a_n)}{t}.$$

2.2.3 Properties of derivatives

Theorem. *If function f is differentiable at $a \in \mathbb{R}^n$, then f is continuous at a.*

Proof. We will use the property that $f(a+h) - f(a) = v \cdot h + o(h)$ with some $v \in \mathbb{R}^n$ and function $o(h)$ is such that $\frac{o(h)}{|h|} \to 0$ as $h \to 0$. Indeed, it follows that $o(h) = \frac{o(h)}{|h|}|h| \to 0$ as $h \to 0$. Therefore, we have $f(a + h) = f(a) + v \cdot h + o(h) \to f(a)$ as $h \to 0$, which means that f is continuous at a.

Theorem. *If function f is differentiable at $a \in \mathbb{R}^n$, then all partial derivatives exist at a and we have*

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a) \right).$$

Proof. Let us take $h = (0, \ldots, 0, t, 0, \ldots, 0)$, with all zeros except for t at the k^{th} position. Then the differentiability of f implies that

$$f(a + h) - f(a) = f(a_1, \ldots, a_{k-1}, a_k + t, a_{k+1}, \ldots, a_n) - f(a) = v_k t + o(t).$$

Dividing by t and taking the limit as $t \to 0$, we obtain $\frac{\partial f}{\partial x_k}(a) = v_k$, which implies the statement.

Theorem. *Conversely, the existence of partial derivatives does not imply differentiability, not even continuity. Function g in 2.1.2 provides such an example.*

Indeed, let us look at $g_x(y) = g(x, y)$. If $x = 0$, this function is constant (equal to zero), hence differentiable. If $x \neq 0$, g_x is a quotient of two differentiable functions, with the one in the denominator non-zero, hence it is also differentiable. The similar argument shows that g^y is also everywhere differentiable. On the other hand, g is not even continuous at $(0, 0)$, hence also not differentiable, since differentiability would imply continuity.

However, we still have some equivalence provided that partial derivatives are continuous:

Theorem. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that $\frac{\partial f}{\partial x_k}$ exist for all $k = 1, \dots, n$, at all points, and they are all continuous on $\mathbb{R}^n$. Then f is differentiable at any point and*

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_n}(x)\right).$$

Proof. For simplicity, let us give the proof in the case of two variables $(x_1, x_2) \in \mathbb{R}^2$. We can write

$$\begin{aligned} f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) &= f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) \\ &\quad + f(x_1 + h_1, x_2) - f(x_1, x_2) \\ &= \frac{\partial f}{\partial x_2}(x_1 + h_1, x_2 + \theta h_2)h_2 + \frac{\partial f}{\partial x_1}(x_1 + \tau h_1, x_2)h_1, \end{aligned}$$

for some $\tau, \theta \in (0, 1)$, by the mean value theorem applied to f coodrinate-wise. Note that τ and θ may depend on x and h, but what is important is that they are in the interval $(0, 1)$. Now, we have

$$\alpha_1(h) := \frac{\partial f}{\partial x_1}(x_1 + \tau h_1, x_2) - \frac{\partial f}{\partial x_1}(x_1, x_2) \to 0 \text{ as } h \to 0$$

because $h \to 0$ implies $h_1 \to 0$, and because $\frac{\partial f}{\partial x_1}$ is continuous at (x_1, x_2). Similarly, we have

$$\alpha_2(h) := \frac{\partial f}{\partial x_2}(x_1 + h_1, x_2 + \theta h_2) - \frac{\partial f}{\partial x_2}(x_1, x_2) \to 0 \text{ as } h \to 0.$$

It follows now that

$$\begin{aligned} f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) &= \frac{\partial f}{\partial x_2}(x_1 + h_1, x_2 + \theta h_2)h_2 + \frac{\partial f}{\partial x_1}(x_1 + \tau h_1, x_2)h_1 \\ &= \left(\frac{\partial f}{\partial x_2}(x) + \alpha_2(h)\right)h_2 + \left(\frac{\partial f}{\partial x_1}(x) + \alpha_1(h)\right)h_1 \\ &= v \cdot h + o(h), \end{aligned}$$

with $v = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x)\right)$ and with $o(h) = \alpha_1(h)h_1 + \alpha_2(h)h_2$. Finally, we have

$$\frac{|o(h)|}{|h|} \leq |\alpha_1(h)|\frac{|h_1|}{|h|} + |\alpha_2(h)|\frac{|h_2|}{|h|} \leq |\alpha_1(h)| + |\alpha_2(h)| \to 0 \text{ as } h \to 0,$$

implying that f is differentiable at x and that $\nabla f(x) = v$.

2.2.4 Rules for differentiation

Theorem. *If f, g are differentiable at a, so are $f+g$, fg; we also have $\nabla(f+g) = \nabla f + \nabla g$ and $\nabla(fg) = f\nabla g + g\nabla f$. Moreover, f/g is differentiable at a provided $g(a) \neq 0$, and $\nabla(f/g) = \frac{g\nabla f - f\nabla g}{g^2}$.*

Proof. The proof of this theorem is straightforward from the definition. For example, since f and g are differentiable at a, we can write

$$f(a+h) = f(a) + \nabla f(a) \cdot h + o_1(h)$$

and

$$g(a+h) = g(a) + \nabla g(a) \cdot h + o_2(h),$$

where $\frac{o_1(h)}{|h|} \to 0$ and $\frac{o_2(h)}{|h|} \to 0$ as $h \to 0$. Consequently, adding these equalities, we get

$$(f+g)(a+h) = (f+g)(a) + (\nabla f(a) + \nabla g(a)) \cdot h + o(h),$$

where $o(h) = o_1(h) + o_2(h)$ satisfies $\frac{o(h)}{|h|} \to 0$ as $h \to 0$. Therefore, $f+g$ is differentiable at a and $\nabla(f+g)(a) = \nabla f(a) + \nabla g(a)$. Similarly, multiplying formulae for $f(a+h)$ and $g(a+h)$ above, we get

$$f(a+h)g(a+h) = f(a)g(a) + (f(a)\nabla g(a) + g(a)\nabla f(a)) \cdot h + o(h),$$

where

$$o(h) = (\nabla f(a) \cdot h)(\nabla g(a) \cdot h) + (f(a) + \nabla f(a) \cdot h + o_1(h))o_2(h) + (g(a) + \nabla g(a) \cdot h + o_2(h))o_1(h).$$

It follows from properties of $o_1(h)$ and $o_2(h)$ that also $\frac{o(h)}{|h|} \to 0$ as $h \to 0$. Therefore, fg is differentiable at a and $\nabla(fg)(a) = f(a)\nabla g(a) + g(a)\nabla f(a)$.

The calculation for the differentiability of f/g and its gradient is similar to those above and we omit it here (but recall that two proofs were given in the lectures).

The following chain rule is more interesting and less straightforward. First, we need a definition. We will say that $\varphi : \mathbb{R} \to \mathbb{R}$ is continuously differentiable at $\alpha \in \mathbb{R}$ if φ is differentiable in some open interval containing α and if φ' is continuous at α.

Chain rule. *If f is differentiable at a and $\varphi : \mathbb{R} \to \mathbb{R}$ is continuously differentiable at $f(a)$, then $\varphi \circ f$ is differentiable at a and $\nabla(\varphi \circ f)(a) = \varphi'(f(a))\nabla f(a)$.*

Proof. We will use the mean value theorem for φ. Indeed, if $v = \nabla f(a)$, we can write

$$\begin{aligned}(\varphi \circ f)(a+h) - (\varphi \circ f)(a) &= \varphi(f(a+h)) - \varphi(f(a)) \\ &= \varphi(f(a) + v \cdot h + o_1(h)) - \varphi(f(a)) \\ &= \varphi'(f(a) + \theta(h))[v \cdot h + o_1(h)] \\ &= [\varphi'(f(a))v] \cdot h + o_2(h),\end{aligned}$$

where $|\theta(h)| \leq |v \cdot h + o_1(h)|$ by the mean value theorem, and where

$$o_2(h) = \varphi'(f(a) + \theta(h))o_1(h) + [\varphi'(f(a) + \theta(h)) - \varphi'(f(a))]v \cdot h.$$

It follows that $\theta(h) \to 0$ as $h \to 0$. Moreover, since φ' is continuous at $f(a)$, we also have $\frac{o_2(h)}{|h|} \to 0$ as $h \to 0$. This calculation means that $\varphi \circ f$ is differentiable at a and that $\nabla(\varphi \circ f)(a) = \varphi'(f(a))v = \varphi'(f(a))\nabla f(a)$.

2.2.5 Directional derivative

Let $\nu \in \mathbb{R}^n$ be a vector. Then we define *the derivative of f in the direction ν* by

$$\frac{\partial f}{\partial \nu}(a) = \lim_{t \to 0} \frac{f(a + t\nu) - f(a)}{t},$$

if such limit exists. In particular, if $\nu = e_k$ is the standard k^{th} coordinate vector in $\mathbb{R}^n$, we have $\frac{\partial f}{\partial \nu} = \frac{\partial f}{\partial x_k}$. We also note that $\frac{\partial f}{\partial \nu}(a)$ can be viewed as the derivative of the function $t \mapsto f(a + t\nu)$ at $t = 0$. As with derivatives in the directions of coordinate axis, we have that the differentiability implies the existence of derivative in any direction ν:

Theorem. *If f is differentiable at a, then $\frac{\partial f}{\partial \nu}(a)$ exists for all vectors $\nu \in \mathbb{R}^n$, and $\frac{\partial f}{\partial \nu}(a) = \nabla f(a) \cdot \nu$.*

Proof. We easily have from the definition that

$$f(a + t\nu) - f(a) = \nabla f(a) \cdot t\nu + o(t\nu) = t\nabla f(a) \cdot \nu + o(t),$$

which immediately implies the statement.

2.2.6 Mean value theorem in $\mathbb{R}^n$

There are different versions of the mean value theorem in $\mathbb{R}^n$. Let us give an example of one of the simplest formulations.

Theorem (MVT in $\mathbb{R}^n$). *Let $x, y \in \mathbb{R}^n$ and let f be differentiable in some set containing the segment joining points x and y. Then there exists some $\theta \in (0, 1)$ such that*

$$f(y) - f(x) = \nabla f(x + \theta(y - x)) \cdot (y - x).$$

The point $x + \theta(y - x)$ lies on the segment joining points x and y, so this theorem is a generalisation of the one dimensional MVT to $\mathbb{R}^n$.

Proof. If $x = y$, the statement is trivial, so let us assume $x \neq y$, and let us denote $\nu = y - x$. Let us consider function $h(t) = f(x + t(y - x))$. Then h is differentiable on $(0, 1)$ and continuous on $[0, 1]$. Hence, by the MVT, there is some $\theta \in (0, 1)$ such that $h(1) - h(0) = h'(\theta)$. We have $h(1) = f(y)$, $h(0) = f(x)$, and $h'(t) = \frac{\partial f}{\partial \nu}(x + t\nu)$ by the definition of the partial derivative. By Theorem 2.2.5 we obtain $h'(t) = \frac{\partial f}{\partial \nu}(x + t\nu) = \nabla f(x + t\nu) \cdot \nu$, finally implying

$$f(y) - f(x) = h(1) - h(0) = h'(\theta) = \nabla f(x + \theta\nu) \cdot \nu,$$

which is the statement of the theorem, if we recall that $\nu = y - x$.

Applying the Taylor's theorem to the function h in the proof, we immediately obtain the Taylor's theorem in $\mathbb{R}^n$. We leave the formulation of this theorem as an exercise.