Note Summary: Multivariate Calculus

Guoxuan Ma*

August 10, 2022

In this lecture, we quickly review some important concepts in multivariate variable calculus, skipping the proofs of many of the results. we use the Euclidean distance d_2 in $\mathbb{R}^k$ by default when talking about openness, closedness, compactness, limit, and continuity. Also, the product of two vectors in Rk is the dot product, and the norm $||\cdot\|$ of a vector is the Euclidean norm, or L_2 norm.

1 Derivatives

Definition 1. Let $A \subset \mathbb{R}$, and $x \in A \cap A'$[1]. A function $f : A \to \mathbb{R}$ is said to be differentiable at x iff the limit

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

exists. In that case, define the derivative of f at x as the limit above, denoted as $f'(x)$.

A function $f : A \to \mathbb{R}$ is said to be **differentiable** iff $A \subset A'$ and f is differentiable at any $x \in A$.

Let $\hat{A}$ be the set of points in $A \cap A'$ at which f is differentiable. Then the function $f' : \hat{A} \to \mathbb{R}$ is called the derivative (function) of f. Clearly, if a function f is differentiable at x, then it is continuous at x. This is because

$$\begin{aligned}
\lim_{x'\to x} |f(x)' - f(x)| &= \lim_{x'\to x} \left[\frac{f(x') - f(x)}{x' - x} \cdot (x' - x) \right] \\
&= \lim_{x'\to x} \left[\frac{f(x') - f(x)}{x' - x} \right] \cdot \lim_{x'\to x} [x' - x] \\
&= f'(x) \cdot 0 = 0
\end{aligned}$$

*This note mainly borrows from math camp material in Columbia University. https://www.sites.google.com/site/mathcamp2018cu

[1] A' is the set of limit points of A.

Let (X, d) be a metric space, and S a subset of X. A point $x \in X$ is a limit point of S iff $(B_r(x)\backslash\{x\}) \cap S \neq \emptyset \ \forall r > 0$. The set of limit points of S is denoted as S'. The condition $(B_r(x)\backslash\{x\}) \cap S \neq \emptyset \ \forall r > 0$ states that the open ball $B_r(x)$ with the center removed always contains some points in the set S, no matter how small the radius r is. That is, a point x is a limit point of S iff we can use points in S to approximate x arbitrarily well (the point x itself may be a point in S, but we are not allowed to use x to approximate itself).

A function continuous at x may fail to be differentiable at x, since the function may have a kink point. In fact, a function can be continuous everywhere, but not differentiable at a single point (e.g. Weierstrass function).

In Rudin's book, a much simpler way to define the derivative as

Definition 2. Let f be defined (and real-valued) on $[a, b]$. For any $x \in [a, b]$ form the quotient

$$\phi(t) = \frac{f(t) - f(x)}{t - x}, \quad (a < t < b, \ t \neq x)$$

and define

$$f'(x) = \lim_{t \to x} \phi(t),$$

provided this limit exists (we use the epsilon-delta definition here). We thus associate with the function f a function f' whose domain is the set of points x at which the limit exists; f' is called the derivative of f. If f' is defined at a point x, f is differentiable at x.

Derivatives of some commonly used functions:

$$\begin{aligned}
(x^\alpha)' &= \alpha x^{\alpha-1}, \ \alpha > 1 \\
(\ln x)' &= \frac{1}{x} \\
(e^x)' &= e^x \\
(\sin x)' &= \cos x
\end{aligned}$$

Because a derivative is essentially the limit of the slope function $\frac{f(x+h)-f(x)}{h}$ when the deviation h tends to 0, it inherits the properties of limits of functions. Especially, if f and g are both differentiable at x, then $f+g$ is also differentiable at x, and $(f+g)'(x) = f'(x) + g'(x)$. Also, it can be shown that

$$\begin{aligned}
(\lambda f)' &= \lambda f' \\
(fg)' &= f'g + fg' \\
(f/g)' &= \frac{f'g - fg'}{g^2}
\end{aligned}$$

Theorem 1. *(L'Hospital Rule). Let $-\infty < a < b < +\infty$, and $f : (a, b) \to \mathbb{R}$ and $g : (a, b) \to \mathbb{R}\backslash\{0\}$ are differentiable in (a, b). If $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ are both 0 or $\pm\infty$, and $\lim_{x \to a} f'(x)/g'(x)$ has a finite value or is $\pm\infty$, then*

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

The statement is also true for $x \to b$.

L'Hospital rule is particularly useful in obtaining the limit of some particular expression. For example, it might seem difficult to determine the behavior of the function $(\ln x)\sqrt{x}$ when x diverges to $+\infty$, because both the numerator and the denominator diverge to $+\infty$. However, because

$$\frac{(\ln x)'}{(\sqrt{x})'} = \frac{\frac{1}{x}}{\frac{1}{2\sqrt{x}}} = \frac{2}{\sqrt{x}} \to 0$$

as $x \to +\infty$, we $\lim_{x\to+\infty}(\ln x)/\sqrt{x} = 0$ by L'Hospital rule.

According to the theorem, L'Hospital rule applies to functions with the form $0/0$ or ∞/∞, i.e. both the numerator and the denominator converges/diverges to 0 or $\pm\infty$. When a function does not have this form, it must be transformed to this form before L'Hospital rule can be applied. For example, consider the limit $\lim_{x\to+\infty}(1+x^{-1})^x$. It does not have the form $0/0$ or ∞/∞, but its log

$$\ln(1+x^{-1})^x = x\ln(1+x^{-1}) = \frac{\ln(1+x^{-1})}{x^{-1}}$$

takes the form $0/0$, to which L'Hospital rule can be applied.

1.1 Total derivatives

Now we generalize the notion of derivatives to multivariate functions.

Definition 3. Let $A \subset \mathbb{R}^n$ and $x \in int(A)$. A function $f : A \to \mathbb{R}^m$ is said to be differentiable at x iff $\exists$ an $m \times n$ real matrix C s.t.

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

In this case, define the (total) derivative of f at x as the matrix C, denoted as $f'(x)$, or $Df(x)$.

A function $f : A \to \mathbb{R}$ is said to be differentiable iff A is open and f is differentiable at any $x \in A$. Let $A_1 \subset int(A)$ be the set of points at which f is differentiable. Then the function $f' : A_1 \to \mathbb{R}^{mn}$ is called the derivative (function) of f.

Because the m $\times$ n real matrix C can also be viewed as an mn-dimensional real vector, the codomain of the derivative function f' can be viewed as $\mathbb{R}^{mn}$. For a real-valued function f from$A \subset \mathbb{R}^n$ to $\mathbb{R}$, its derivative $f'(x)$ at $x \in int(A)$ reduces to a $1 \times n$ row vector. In this case, the derivative is also called the gradient of f at x, sometimes denoted as $\nabla f(x)$, which is essentially the same as $f'(x)$ or $Df(x)$.

Clearly, if a function f from $A \subset \mathbb{R}^n$ to $\mathbb{R}^m$ is differentiable at $x \in int(A)$, then it is continuous at x. To see this, by triangle inequality of $||\cdot||$,

$$0 \le ||f(x') - f(x)|| \le ||f(x') - f(x) - f'(x)(x'-x)|| + ||f'(x)(x'-x)||$$

Because the first term

$$\begin{aligned} ||f(x') - f(x) - f'(x'-x)|| &= \frac{||f(x') - f(x) - f'(x)(x'-x)||}{||x'-x||}||x'-x|| \\ &\to 0 \cdot 0 = 0 \end{aligned}$$

as $x' \to x$, and the second term

$$||f'(x)(x'-x)|| \to ||f'(x)(x-x)|| = 0$$

as $x' \to x$, we know that $||f(x') - f(x)|| \to 0$ as $x' \to x$. Therefore, f is continuous at x.

If two functions f and g from $A \subset \mathbb{R}^n$ to $\mathbb{R}^m$ are both differentiable at $x \in int(A)$, then the function $f : A \to \mathbb{R}^m$ is also differentiable at x, and furthermore we have $(f+g)'(x) = f'(x) + g'(x)$.

Similarly, we can show $(\lambda f)' = \lambda f'$. Therefore, taking derivative is a linear operator, i.e.

$$(\lambda_1 f_1 + \lambda_2 f_2)' = \lambda_1 f_1'(x) + \lambda_2 f_2'(2)$$

For a function f from $A \subset \mathbb{R}^n$ to $\mathbb{R}^m$, each coordinate $i \in \{1, ..., m\}$ of f can be regarded as a function f_i from A to R. By definition, it is straightforward to show that f is differentiable at $x \in int(A)$ iff fi is differentiable at x for each i, and furthermore we have

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

We say that f from $A \subset \mathbb{R}^n$ to $\mathbb{R}^m$, is k-th continuously differentiable at x iff $x \in in(A)$ and $f^k(x)$ is continuous at x, where A_k is the set of points at which $f^{(k-1)}$ is differentiable. In this case, f is said to be C^k at x. We say that f is k-th continuously differentiable iff A is open and f is k-th continuously differentiable at all $x \in A$. In this case, f is said to be C^k.

1.2 Partial Derivatives

Definition 4. Let $A \subset \mathbb{R}^n$ and $x \in int(A)$. For a function $f : A \to \mathbb{R}^m$, its **partial derivative of the i-th coordinate w.r.t. the j-th argument at** $x \in A$ is

$$\frac{\partial f_i}{\partial x_j}(x) := \frac{d}{dt} f_i(x + te_j)|_{t=0}$$

if the right-hand side derivative exists.

The vector e_j above is the j-th canonical basis of $\mathbb{R}^n$, i.e. $e_j := (0, ..., 1, ..., 0)$.

In the expression $\frac{d}{dt} f_i(x + te_j)|_{t=0}$, we fix x and consider $f_i(x + te_j)$ as a single variable function in t, then take derivative of this single variable function, and finally evaluate the derivative at $t = 0$.

The vector $x+te_j$ is a deviation from x only in the j-th argument. Therefore, intuitively, the partial derivative $\frac{\partial f_i}{\partial x_j}(x)$ measures the sensitivity of the i-th coordinate f_i of the function f w.r.t. the j-th argument x_j .

Theorem 2. *Let $A \subset \mathbb{R}^n$ and $x \in int(A)$. If function $f : A \to \mathbb{R}^m$ is differentiable at x, then $\frac{\partial f_i}{\partial x_j}(x)$ exists for any $(i, j) \in \{1, ..., m\} \times \{1, ..., n\}$, and furthermore we have*

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

Notice that the theorem above only states that existence of the total derivative implies the existence of all partial derivatives. The reverse is not true, since we can find a function f s.t. $\frac{\partial f_i}{\partial x_j}(x)$ exists for all $(i, j) \in \{1, .., m\} \times \{1, ..., n\}$, but f is not differentiable at x, i.e. its total derivative does not exist. In fact, f may even be discontinuous at x. See the example below.

Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as

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

By definition of partial derivatives, we have

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

and

$$\frac{\partial f}{\partial y}(0, 0) = \frac{d}{dt} f(0, t)|_{t=0} = \lim_{t\to 0} \frac{f(0, t) - f(0, 0)}{t - 0}$$
$$\lim_{t\to 0} \frac{0 - 0}{t - 0} = 0$$

So both of the partial derivatives of f exist. However, f is not differentiable at $(0, 0)$. In fact, f is not even continuous at $(0, 0)$. To see this, notice that f constantly take the value $1/2$ along the path $y = x^2$ except for at the point $(0, 0)$, where the f takes the value 0.

As is shown in the example above, the existence of $\frac{\partial f_i}{\partial x_j}(x)$ for all (i, j) does not imply differentiability of f at x. However, if for each (i, j), the partial $\frac{\partial f_i}{\partial x_j}(x)$ exists not only at x, but also on an open ball around x, and $\frac{\partial f_i}{\partial x_j}(x)$ is continuous at x, then f is differentiable at x. This result is formulated by following theorem.

Let $A \subset \mathbb{R}^n$, $x \in int(A)$, and function $f : A \to \mathbb{R}^m$. Then f is C^1 at x iff $\frac{\partial f_i}{\partial x_j}(x)$ exists on an open ball around x and is continuous at x for any $(i, j) \in \{1, ..., m\} \times \{1, ..., n\}$.

It is typically difficult to find the total derivative of a function f, since we need to find a $m \times n$ matrix that satisfies the limit condition specified by the definition. However, the mn partial derivatives are much easier to find, since they are essentially derivatives of single variable derivatives. Therefore, to find the total derivative of a function at x, we usually don't directly work with the definition of total derivatives. Instead, we look at all partial derivatives of f and see if all of them exist in an open ball around x and are continuous at x. If yes, then by the theorem above we know that the total derivative exists at x, and is exactly the matrix of all partial derivatives at x.

1.3 Directional Derivatives

Definition 5. Let $A \subset \mathbb{R}^n$ and $x \in int(A)$. For a function $f : A \to \mathbb{R}^m$, and a vector $z \in \mathbb{R}^n$ with $||z|| = 1$, the directional derivative of f along the vector $z \in \mathbb{R}^n$at $x \in A$ is

$$f'_z(x) \coloneqq \frac{d}{dt} f(x+tz)|_{t=0} = \begin{bmatrix} \frac{d}{dt} f_1(x+tz)|_{t=0} \\ \frac{d}{dt} f_2(x+tz)|_{t=0} \\ \vdots \\ \frac{d}{dt} f_m(x+tz)|_{t=0} \end{bmatrix}$$

if the right-hand side derivative exists.

Consider a function f from $A \subset \mathbb{R}^n$ to $\mathbb{R}$ and $x \in int(A)$, the gradient $\nabla f(x)$ can be interpreted as the direction in which f increases the fastest at x. This is formulated in the proposition below.

Proposition 1. *Let f be a function from $A \subset \mathbb{R}^n$ to $\mathbb{R}$ that is differentiable at $x \in int(A)$, and $\nabla f(x) \neq 0$. Then the directional derivative $f'_z(x)$ is maximized when $z = \frac{\nabla f(x)}{||\nabla f(x)||}$, and the maximized directional derivative is $||\nabla f(x)||$.*

1.4 Chain Rule

Proposition 2. *(Chain Rule). Let S be a subset of R, and $f : S \to \mathbb{R}$. Let T be a set s.t. $f(S) \subset T \subset \mathbb{R}$, and $g : T \to \mathbb{R}$. If f is differentiable at x, and g is differentiable at f ($\mathbb{R}$ is differentiable at x, and we have*

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$$

The chain rule for single variable functions to can be generalized to multivariate functions.

Proposition 3. *Let $S \in \mathbb{R}^n$, $x \in int(S)$, and $f : S \to \mathbb{R}^m$. Let T be s.t. $f(S) \subset T \subset \mathbb{R}^m$ and $f(x) \in int(T)$, and let $g : T \to \mathbb{R}^k$. If f is differentiable at x, and g is differentiable at $f(x)$, then g is differentiable at $f(x)$, then $g \circ f : S \to \mathbb{R}^k$ is differentiable at x. Furthermore, we have*

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$$

In the equation $(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$, the equality between the (i, j)-th entries of the matrices on two sides is

$$\frac{\partial (g \circ f)_i}{\partial x_j}(x) = \sum_{l=1}^{m} \left[\frac{\partial g_i}{\partial y_l}(f(x)) \cdot \frac{\partial f_l}{\partial x_j}(x) \right]$$

2 Mean Value Theorem

Theorem 3. *(Mean Value Theorem). Let $f : [a, b] \to \mathbb{R}$, differentiable on (a, b), and continuous on $[a, b]$. Then there exists $x \in (a, b)$ s.t.*

$$f'(x) = \frac{f(b) - f(a)}{b - a}$$

One implication of mean value theorem is: if $f' > (\geq)0$ on (a, b), then f is strictly (weakly) increasing on (a, b). If we have $f' < (\leq)$, then f is strictly (weakly) decreasing on (a, b).

Theorem 4. *Let $f : A \subset \mathbb{R}^n$ is C^1 in an open set in A which contains $[x, y]$ $(x_i < y_i, \forall i = 1, 2, ..., n)$. Then there exists a point $\mathbf{w}$ in $(\mathbf{x}, \mathbf{y})$ (i.e. $x_i < w_i < y_i,\ \forall i = 1, 2, ..., n$) s.t.*

$$f(\mathbf{x}) - f(\mathbf{y}) = \nabla f(\mathbf{w}) \cdot (\mathbf{x} - \mathbf{y})$$

Finally, here is a similar theorem that does not require differentiability.

Theorem 5. *(Intermediate Value Theorem). Let $f : [a, b] \to \mathbb{R}$ continuous and u is a number between $f(a)$ and $f(b)$, then there exists $c \in [a, b]$ s.t. $u = f(c)$.*

3 Higher Order Derivatives and Taylor Expansion

3.1 Second Order Derivatives of $f : A \subset \mathbb{R}^n \to \mathbb{R}$

For a function f from $f : A \subset \mathbb{R}^n \to \mathbb{R}$, we know that its gradient at $x \in int(A)$ is equal to the vector of partial derivatives, i.e.

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), ..., \frac{\partial f}{\partial x_n}(x) \right)$$

The second derivative of the real-valued function f at x is also known as the Hessian matrix of f at x, denoted as $H_f(x)$:

$$H_f(x) \coloneqq f''(x) = (\nabla f)'(x) = \begin{bmatrix} \left(\nabla \frac{\partial f}{\partial x_1}\right)(x) \\ \left(\nabla \frac{\partial f}{\partial x_1}\right)(x) \\ \vdots \\ \left(\nabla \frac{\partial f}{\partial x_1}\right)(x) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial\left(\frac{\partial f}{\partial x_1}\right)}{\partial(x_1)}(x) & \frac{\partial\left(\frac{\partial f}{\partial x_1}\right)}{\partial(x_2)}(x) & \cdots & \frac{\partial\left(\frac{\partial f}{\partial x_1}\right)}{\partial(x_n)}(x) \\ \frac{\partial\left(\frac{\partial f}{\partial x_2}\right)}{\partial(x_1)}(x) & \frac{\partial\left(\frac{\partial f}{\partial x_2}\right)}{\partial(x_2)}(x) & \cdots & \frac{\partial\left(\frac{\partial f}{\partial x_2}\right)}{\partial(x_n)}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial\left(\frac{\partial f}{\partial x_n}\right)}{\partial(x_1)}(x) & \frac{\partial\left(\frac{\partial f}{\partial x_n}\right)}{\partial(x_2)}(x) & \cdots & \frac{\partial\left(\frac{\partial f}{\partial x_n}\right)}{\partial(x_n)}(x) \end{bmatrix}$$

Notice that in the expressions above, the notation $(\nabla \frac{\partial f}{\partial x_i})(x)$ stands for the gradient of the function $\frac{\partial f}{\partial x_i}$ at x. The notation $\frac{\partial(\frac{\partial f}{\partial x_i})}{\partial x_j}(x)$ stands for the partial derivative of the function $\frac{\partial f}{\partial x_i}$ at x w.r.t. the j-th argument, which is usually referred to as a cross partial at x. The notation for the cross partial $\frac{\partial(\frac{\partial f}{\partial x_i})}{\partial x_j}$ is usually simplified as $\frac{\partial^2 f}{\partial x_i \partial x_j}$.

Notice that the cross partial

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) := \frac{\partial(\frac{\partial f}{\partial x_i})}{\partial x_j}(x)$$

and the cross partial

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) := \frac{\partial(\frac{\partial f}{\partial x_j})}{\partial x_i}(x)$$

are conceptually very different when $i \neq j$. However, they are equal if f is twice-differentiable at x, and this result is usually known as Young's theorem or Schwarz's theorem.

Let $A \subset \mathbb{R}^n$ and $x \in int(A)$. If function $f : A \to \mathbb{R}$ is C^2 at x, then for any $i, j \in \{1, ..., n\}$ both $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}(x)$ exists and

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

By the theorem above, when f is twice-differentiable at x, the Hessian matrix of f at x

$$H_f(x) = \begin{bmatrix} \frac{\partial^2 f}{(\partial x_1)^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{bmatrix}$$

is a symmetric matrix.

3.2 Taylor's Theorem

Definition 6. Let $f : [a, b] \to \mathbb{R}$ be C^{n-1} and $f^{(n)}(t)$ exists at every $t \in (a, b)$. Let a and b be distinct points in $[a, b]$, and define

$$P_{n-1}(t) := f(\alpha) + f'(\alpha)(t - \alpha) + \frac{f''(\alpha)}{2}(t - \alpha)^2 + ... + \frac{f^{(n-1)}(\alpha)}{(n-1)!}(t - \alpha)^{n-1}$$

Then there exists x strictly between α and β s.t.

$$f(\beta) = P_{n-1}(\beta) + \frac{f^{(n)}(x)}{n!}$$

In the theorem, β is allowed to be greater or less than α. Notice that Taylor's theorem reduces to the mean value theorem when $n = 1$, and so Taylor's theorem can be viewed as a generalization of the mean value theorem.

This theorem states that under some differentiability and continuity conditions, $f(\beta)$ can be approximated by the polynomial

$$\begin{aligned} P_{n-1}(\beta) &\coloneqq f(\alpha) + f'(\alpha)(\beta - \alpha) + \frac{f''(\alpha)}{2}(\beta - \alpha)^2 \\ &+ ... + \frac{f^{(n-1)}(\alpha)}{(n-1)!}(\beta - \alpha)^{n-1} \end{aligned}$$

and the error is $\frac{f^{(n)}(x)}{n!}(\beta - \alpha)^n$. if we rewrite β as $\alpha + h$, then $f(\alpha + h)$ can be approximated by the polynomial

$$f(\alpha) + f'(\alpha)h + \frac{f''(\alpha)}{(n-1)!}h^{n-1}$$

and the error is $\frac{f^{(n)}(x)}{n!}h^n$, where x is some point between x and $x + h$.

If we further assume that $f \in C^n$, then $f^{(n)}$ is continuous at α, and thus

$$\frac{\frac{f^{(n)}(x)}{n!}h^n}{h^{n-1}} = \frac{f^{(n)}(x)}{n!}h \to \frac{f^{(n)}(\alpha)}{n!}0 = 0$$

as $h \to 0$, which means that the error is small compared to h^{n-1} as h tends to 0. Conventionally, the notation $o(f(t))$ is used to denote any function $g(t)$ s.t. $\lim_{t\to 0} g(t)/f(t) = 0$. So the error term is $o(h^{n-1})$. Therefore, Taylor's theorem can be rewritten as

$$f(a + h) = f(\alpha) + f'(\alpha)h + \frac{f''(\alpha)}{2}h^2 + ... + \frac{f(n-1)(\alpha)}{(n-1)!}h^{n-1} + o(h^{n-1})$$

when f is C^n, and this is sometimes known as the $(n-1)$-th order Taylor expansion of f at α. Notice that the correct interpretation of the equality above is

$$\lim_{h\to 0} \frac{f(a + h) - \left[f(\alpha) + f'(a)h + \frac{f''(\alpha)*h^2}{2} + ... + \frac{f^{(n-1)}(\alpha)}{(n-1)!}h^{n-1}\right]}{h^{n-1}} = 0$$

we can also write the $(n-1)$-th order Taylor approximation of f at α:

$$f(a + h) \approx f(\alpha) + f'(\alpha)h + \frac{f''(\alpha)}{2}h^2 + ... + \frac{f^{(n-1)}(\alpha)}{(n-1)!}h^{n-1}$$

The following Theorem states the first and second order Taylor expansion

Let f be a function from $A \subset \mathbb{R}^n$ to $\mathbb{R}$, and f is C^2 at $x \in int(A)$. Then we have

$$f(x+h) = f(x) + \nabla f(x)h + o(||h||)$$

If f is C^3 at x, we have

$$f(x+h) = f(x) + \nabla f(x)h + \frac{1}{2}h^T H_f(x) + o(||h||^2)$$

Recall that the correct interpretation of the two equations above is

$$\lim_{h \to 0} \frac{f(x+h) - [f(x) + \nabla f(x)h]}{||h||} = 0$$

and

$$\lim_{h \to 0} \frac{f(x+h) - [f(x) + \nabla f(x)h + \frac{1}{2}h^T H_f(x)h]}{||h||^2} = 0$$

4 Log-linearization

In dynamic macro economics models, we sometimes use log-linearization to approximate a non-linear dynamic system using a linear dynamic system. This invokes Taylor's theorem, which tells us know to construct linear approximations of (non-linear) functions, at least near some point x (which is usually the steady state point of the system).

Consider multivariate function $f : A \subset \mathbb{R}^n \to \mathbb{R}$, we want to approximate it around point $x^* = (x_1^*, x_2^*, ..., x_n^*)$ s.t. $x_i^* \neq 0$, $\forall i$, for each variable x_i, we define $\hat{x}_i := \ln(x_i/x_i^*)$ to be its log-deviation when x_i and x_i^* have the same sign (which is reasonable when x is "near" x^*).

Since $x_i = x_i^* e^{\hat{e}_i}$, we can rewrite $f(\mathbf{x}) = f(x_1, x_2, ..., x_n)$ as a function h of $\hat{x}_1, \hat{x}_2, ..., \hat{x}_n$:

$$h(\hat{x}_1, \hat{x}_2, ..., \hat{x}_n) = f(x_1^* e^{\hat{x}_1}, x_2^* e^{\hat{x}_2}, ..., x_n^* e^{\hat{x}_n}) = f(\mathbf{x})$$

Note that $h(\mathbf{0}) = f(\mathbf{x}^*)$ and $h_i'(\mathbf{0}) = f_i'(\mathbf{x}^*)x_i^*, \forall i = 1, 2, ..., n$.

We then take a first order Taylor expansion of h around the point $\mathbf{0}$(replace $\approx$ with $=$ here)

$$\begin{aligned} f(\mathbf{x}) = h(\hat{x}_1, \hat{x}_2, ..., \hat{x}_n) &= h(0) + h_1'(\mathbf{0})\hat{x}_1 + h'(\mathbf{0})\hat{x}_2 + ... + h_n'(\mathbf{0})\hat{x}_n \\ &= f(\mathbf{x}^*) + f_1'(\mathbf{x}^*)x_1^*\hat{x}_1 + f_2'(\mathbf{x}^*)x_2^*\hat{x}_2 + ... + f_n'(\mathbf{x}^*)x_n^*\hat{x}_n \end{aligned}$$

The approximation above, in the form of $f(\mathbf{x}) = a_0 + \sum_{i=1}^n a_i \hat{x}_i$, is called the log-linear approximation of function f around point $\mathbf{x}^*$.

Often, instead of log-linearizing a function, we want to log-linearize an equation (which is (a part of) the characterization of a system at its steady state):

$$f(\mathbf{x}) = f(x_1, x_2, ..., x_n) = 0$$

around root $\mathbf{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ satisfying $f(\mathbf{x}^*) = 0$. In this case, we first write a log-linear approximation of the LHS, $f(\mathbf{x})$, then we set this log-linear approximation equal to zero. So we have

$$f_1'(\mathbf{x}^*)x_1^*\hat{x}_1 + f_2'(\mathbf{x}^*)x_2^*\hat{x}_2 + ... + f_n'(\mathbf{x}^*)x_n^*\hat{x}_n = 0$$

which, in the form of $\sum_{i=1}^n b_i\hat{x}_i = 0$, is called the log-linearization of equation $f(\mathbf{x}) = 0$ around x s.t. $f(\mathbf{x}^*\ ) = 0$. The discussion below is devoted to showing how to perform log-linearization of equations (faster).

If $f(\mathbf{x}^*) \neq 0$, define $\eta_i \coloneqq \frac{f_i'(\mathbf{x}^*)x_i^*}{f(\mathbf{x}^*)}$ $(i = 1, 2, ..., n)$ the elasticity of f w.r.t x_i at x , we can also write:

$$f(\mathbf{x}) = f(\mathbf{x}^*)[1 + \eta_1\hat{x}_1 + \eta_2\hat{x}_2 + ... + \eta_n\hat{x}_n]$$

therefore

$$\frac{f(\mathbf{x}) - f(\mathbf{x}^*)}{f(\mathbf{x}^*)} = \eta_1\hat{x}_1 + \eta_2\hat{x}_2 + ... + \eta_n\hat{x}_n$$

Now we define the log-deviation of function f around some point $\mathbf{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ s.t. $f(\mathbf{x}^*) \neq 0$:

$$f(\hat{x}) \coloneqq \ln(f(\mathbf{x})/f(\mathbf{x}^*))$$

(when $f(\mathbf{x})$ and $f(\mathbf{x})^*$ have the same sign). Notice that $\ln(f(\mathbf{x})/f(\mathbf{x}^*)) \approx \frac{f(\mathbf{x})-f(\mathbf{x}^*)}{f(\mathbf{x}^*)}$, we then have

$$f(\hat{x}) \coloneqq \eta_1\hat{x}_1 + \eta_2\hat{x}_2 + ... + \eta_n\hat{x}_n$$

The following are the log-deviations of some simple functions (please verify by yourselves), which are "shortcuts" you might want to memorize (note here $x, x1, x2$ are scalars):

- $\widehat{\alpha x} = \hat{x}$
- $\widehat{x_1 + x_2} = \frac{x_1^*}{x_1^*+x_2^*}\hat{x}_1 + \frac{x_2^*}{x_1^*+x_1^*}\hat{x}_2$
- $\widehat{x_1 x_2} = \hat{x_1} + \hat{x}_2$
- $\widehat{x^\alpha} = \alpha\hat{x}$
- $\hat{c} = 0$ where c is a constant

Sometimes the function $f(\mathbf{x})$ can be written in the form of $f(x) = g(\mathbf{x}) - l(\mathbf{x})$. Then the equation $f(\mathbf{x}) = 0$ can be written as $g(\mathbf{x}) = l(\mathbf{x})$. To log-linearize equation $g(\mathbf{x}) = h(\mathbf{x})$ around some $\mathbf{x}^*$ satisfying $g(\mathbf{x}^*) = l(\mathbf{x}^*)$, we can just derive the log-deviation $\widehat{g(\mathrm{x}^*)}$ and $\widehat{h(\mathbf{x}^*)}$ around $\mathbf{x}^*$, and set them equal to one another.

5 Implicit Function Theorem and Inverse Function Theorem

For a function f from $A \subset \mathbb{R}^n$ to $\mathbb{R}^k$ and a point $(x_0, y_0) \in A$, the Jacobian matrix $f'_x(x_0, y_0)$ at (x_0, y_0) is a $k \times n$ matrix defined as the derivative of $f(x, y_0)$ viewed as a function of x, evaluated at $x = x_0$. Similarly, the Jacobian matrix $f'_y(x_0, y_0)$ at (x_0, y_0) is a $k \times m$ matrix defined as the derivative of $f(x_0, y)$ viewed as a function of y, evaluated at $y = y0$.

Theorem 6. *(Implicit Function). Let f be a function from $A \subset \mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^m$. Let $(x_0, y_0) \in int(A)$ s.t. $f(x_0, y_0) = 0$. If f is C^1 at (x_0, y_0) and the $m \times m$ Jacobian matrix $f'_y(x_0, y_0)$ is invertible, then there exist an open ball B_x around x_0 and an open ball B_y around y_0 s.t. $\forall x \in B_x$ there exists a unique $y \in B_y$ s.t. $f(x, y) = 0$. Therefore, the equation $f(x, y) = 0$ implicitly defines a function $g : B_x \to B_y$ with the property*

$$f(x, g(x)) = 0$$

for any $x \in B_x$. Furthermore, we know that the function g is differentiable at any $x \in B_x$, and

$$g'(x) = -\left[f'_y(x, g(x))\right]^{-1} f'_x(x, g(x))$$

Here let's admit that the implicit function g is well-defined and is differentiable, and provide some intuitions only for the derivative formula $g'(x) = -\left[f'_y(x, g(x))\right]^{-1} f'_x(x, g(x))$ using chain rule.

The next theorem, often known as the inverse function theorem, is just a special case of the implicit function theorem.

Theorem 7. *(Inverse Function). Let f be a function from $A \subset \mathbb{R}^n$ to $\mathbb{R}^n$. Let $x_0 \in int(A)$ and $y_0 := f(x_0)$. If f is C^1 at (x_0, y_0) and the derivative $f'(x_))$ is invertible, then there exists an open ball B_y around y_0 and an open ball B_x s.t. $\forall y \in B_y$ there exists a unique $x \in B_x$ s.t. $f(x) = y$. Therefore, the equation $f(x) = y$ implicitly defines a function $g : B_y \to B_x$ with the property*

$$f(g(y)) = y$$

for any $y \in B_y$. Furthermore, the function g is differentiable at any $y \in B_y$, and we have

$$g'(y) = f'(g(y))^{-1}$$

To see why the inverse function theorem is a special case of the implicit function theorem, define

$$F(y, x) := y - f(x)$$

for any $(y, x) \in \mathbb{R}^n \times A$. Clearly, $(y_0, x_0) \in int(\mathbb{R}^n \times A)$ and $F(y_0, x_0) = 0$, and F is C^1. Furthermore $F'_x(y_0, x_0) = -f'(x_0)$ is invertible by assumption. Invoke the implicit function theorem for function F, and we know that x is implicitly defined as a function g of y on an open ball B_y around y_0, with the property

$F(y, g(y)) = 0$ for any $y \in B_y$. Furthermore, the function g is differentiable at any $y \in B_y$, and

$$g'(y) = -\left[F_x'(y, g(y))\right]^{-1} F_y'(y, g(y)) = -\left[-f'(g(y))\right]^{-1} \cdot I_n = f'(g(y))^{-1}$$

So we have the implicit function theorem. Again, we can obtain some intuitions of this result using chain rule. Think of both sides of the equation $f(g(y)) = y$ as a function in y and take derivative:

$$\begin{aligned} \frac{d}{dy} f(g(y)) &= \frac{d}{dy} y \\ f'(g(y)) \cdot g'(y) &= I_n \end{aligned}$$

Because $f'(g(y))$ is invertible when $y = y_0$, and so we can set the open ball B to be small enough s.t. $f'(g(y))$ is invertible for any $y \in B_y$. Left multiplying the equation above by $f'(g(y))^{-1}$, and we have $g'(y) = f'(g(y))^{-1}$.

6 Integrals

6.1 Riemann Integrability

In $\mathbb{R}^k$, a finite partition P of the cell $C = [a_1, b_1] \times ... \times [a_k, b_k]$ is a finite set of points $\left\{ \{x_i^n\}_{n=0}^{N_i} \right\}_{i=1}^{k}$ s.t. $\{x_i^n\}_{n=0}^{N_i}$ is a partition of $[a_i, b_i]$ for each dimension $i \in \{1, ..., k\}$, i.e.,

$$a_I = x_i^0 \leq x_i^1 \leq ... \leq x_i^{N_i} = b_i$$

Let $\mathcal{P}$ be the set of all finite partitions of C.

We can interpret a partition $\left\{ \{x_i^n\}_{n=0}^{N_i} \right\}_{i=1}^{k}$ as partitioning the cell $[a_1, b_1] \times \cdots \times [a_k, b_k]$ into $\prod_{i=1}^{k} N_i$ sub-cells. Denote the set of sub-cells of $[a_1, b_1] \times \cdots \times [a_k, b_k]$ under P as $\mathcal{C}(P)$. For each subcell $c \in \mathcal{C}(P)$, let the "volume", or measure, of it be defined as the product of the lengths of the cell in each dimension, and we denote it as $\mu(c)$. For any bounded function $f : C \to \mathbb{R}$ and a partition P of the cell $C = [a_1, b_1] \times \cdots \times [a_k, b_k]$, for each sub-cell $c \in \mathcal{C}(P)$ let

$$\begin{aligned} M_c &\coloneqq \sup\{f(x) : x \in c\} \\ m_c &\coloneqq \inf\{f(x) : x \in c\} \end{aligned}$$

and then define

$$U(P,f) := \sum_{c \in \mathcal{C}(P)} M_c \mu(c)$$
$$L(P,f) := \sum_{c \in \mathcal{C}(P)} m_c \mu(c)$$
$$\overline{\int}_c f dx := \inf\{U(P,f) : P \in \mathcal{P}\}$$
$$\underline{\int}_c f dx := \sup\{L(P,f) : P \in \mathcal{P}\}$$

If $\overline{\int}_C f dx$ and $\underline{\int}_c f dx$ give us the same value, we define this value as the Riemann integral of f over the cell C, denoted as $\int_C f dx$, or $\int_C f(x)dx$. In this case, we say that f is Riemann integrable over the cell C. We can define the generalized Riemann integral over an "infinite cell" $[a_1, +\infty) \times ... \times [a_k, +\infty)$ as the limit of the integral over $[a_1, b_1] \times \cdots \times [a_k, b_k]$ when $b_i \to +\infty$ for each i. We can also generalize the notion of Riemann integral to allow for integration over a general set S that may not be a cell. For a function $f : S \to \mathbb{R}$, we can find a potentially infinite cell C s.t. $C \supset S$, and extend the domain of f to C by defining a new function $f_C : C \to \mathbb{R}$ as

$$f_C(x) := \begin{cases} f(x) & if\ x \in S \\ 0 & if\ x \in C \backslash S \end{cases}$$

and then we define $\int_S f(x)dx := \int_C f_C(x)dx$.

Note that not all functions are Riemann integrable. For example, consider the function $f : [0,1] \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

f is not Riemann integrable over $[0,1]$.

The following theorem provides a sufficient condition for Riemann integrability of single-variate functions.

Theorem 8. *Let f be a bounded real function on $[a,b]$, and it is discontinuous only at finitely many points on $[a,b]$. Then f is Riemann-Stieltjes integrable over $[a,b]$.*

6.2 Fundamental Theorem of Calculus

The next theorem relates integration to differentiation. It provides us the most fundamental tool to calculate the value of a particular integral.

Theorem 9. *(Fundamental Theorem of Calculus). If f is (Riemann) integrable w.r.t. x on $[a, b]$, and if there is a differentiable function F on $[a, b]$ s.t. $F^0 = f$, then*

$$\int_a^b f dx = F(b) - F(a)$$

F is called the antiderivative (or indefinite integral) of f on $[a, b]$, noted $\int f(x)dx$.

The next three properties are especially useful in calculation of single integrals, and we state them below in an informal way.

1. Differentiation of α:

$$\int_a^b f(x)d\alpha(x) = \int_a^b f(x)\alpha'(x)dx$$

2. Change of variable:

$$\int_a^b f(\phi(x))d\alpha(\phi(x)) = \int_{\phi(a)}^{\phi(b)} f(y)d\alpha(y)$$

3. Integration by part:

$$\int_a^b f(x)dg(x) = f(b)g(b) - f(a)g(a) - \int_a^b g(x)df(x)$$

6.3 Multiple Integrals over Product Domains

Theorem 10. *(Fubini). Let $C_X = [a_1, b_1] \times ... \times [a_k, b_k]$ and $C_Y = [a_{k+1}, b_{k+1}] \times ... \times [a_{k+m}, b_{k+m}]$. Consider a continuous function $f : C_X \times C_Y \to \mathbb{R}$. We have*

$$\int_{C_x \times C_Y} f(x, y)d(x, y) = \int_{C_Y} \left(\int_{C_X} f(x, y)dx \right) dy = \int_{C_X} \left(\int_{C_y} f(x, y)dy \right) dx$$

Fubini's theorem allows us to rewrite a double integral as an iterated integral, and the order of integration does not matter. In the theorem above, I assume f to be continuous in order to make sure that all integrals are well-defined. However, this continuity assumption is not necessary and can be relaxed if we work with Lebesgue integrals.

If $C = [a_1, b_1] \times ... \times [a_k, b_k]$ and $f : C \to \mathbb{R}$ is continuous, then we can repeatedly apply Fubini's theorem and calculate $\int_C f(x)dx$ as k nested single variable integrals. To see this,

$$\begin{aligned}
\int_C f(x)dx &= \int_{[a_1,b_1]\times\ldots\times[a_k,b_k]} f(x)d(x_1,(x_2,...,x_k)) \\
&= \int_{[a_2,b_2]\times\ldots\times[a_k,b_k]} \left(\int_{a_1}^{b_1} f(x)dx_1\right) d(x_2,...,x_k) \\
&= \int_{[a_2,b_2]\times\ldots\times[a_k,b_k]} \left(\int_{a_1}^{b_1} f(x)dx_1\right) f(x)d(x_2,(x_3,...,x_k)) \\
&\ldots \\
&= \int_{a_k}^{b_k} \left(\cdots\left(\int_{a_2}^{b_2}\left(\int_{a_1}^{b_1} f(x)dx_1\right)dx_2\right)\cdots\right)dx_k
\end{aligned}$$

Double Integrals Over General Regions

We use double integral as an example to illustrate integration over general domains. Suppose we want to integrate continuous $f(x,y)$ over a set $A := \{(x,y)\in\mathbb{R}^2 : x\in[a,b], y\in[u(x),v(x)]\}$.

$$\int_A f(x,y)dxdy = \int_a^b \left(\int_{u(x)}^{v(x)} f(x,y)dy\right)dx$$

6.4 Change of Variables

Next we show change of variables in multiple integrals, still using double integral as an example. Consider double integral $\int_A f(x,y)dxdy$. Suppose that

$$x = g(u,v),\ y = h(u,v)$$

defines a one-to-one C^1 transformation from an open and bounded set A' in the uv-plane onto an open and bounded set A in the xy-plane, and assume the Jacobian determinant

$$\frac{\partial(g,h)}{\partial(u,v)} := det\left(\begin{bmatrix} \partial g/\partial u & \partial g/\partial v \\ \partial h/\partial u & \partial h/\partial v \end{bmatrix}\right)$$

is bounded on A'. Assume f is bounded and continuous on A. Then

$$\int_A f(x,y)dxdy = \int_{A'} f(g(u,v),h(u,v))d|\frac{\partial(g,h)}{\partial(u,v)}$$

where $|\frac{\partial(g,h)}{\partial(u,v)}|$ is the absolute value of the Jacobian determinant.

Notice that we sometimes do not need to solve for $\frac{\partial(g,h)}{\partial(u,v)}$ explicityly. We have the following identity

$$\frac{\partial(g,h)}{\partial(u,v)}\frac{\partial(u,v)}{\partial(x,y)} = 1$$

The right-hand Jacobian is easy to calculate if you know $u(x,y)$ and $v(x,y)$; then the left-hand one - the one needed - will be its reciprocal. This result can be generalized to n-dimensional multiple integrals.

6.5 Derivatives of Integrals

Consider a parameterized integral

$$\int_{u(x)}^{v(x)} f(x,t)dt$$

The variable of integration is t, and x is a real-valued parameter. We allow both the function $f(x,t)$ of t and the interval of integration $[u(x), v(x)]$ to depend on the parameter x. As a consequence, the value of this integral also depends on the parameter x. Define the value of the integral as $I(x) \coloneqq \int_{u(x)}^{v(x)} f(x,t)dt$, and then $I(x)$ can be viewed as a function of the parameter x. The next theorem provides a sufficient condition for $I(x)$ to be differentiable, and also a formula for calculating the derivative of $I(x)$.

Theorem 11. *(Leibniz's Formula). Let f be a function from a subset A of $\mathbb{R}^2$ to $\mathbb{R}$. Let rectangle $E \coloneqq [a,b] \times [c,d] \subset A$ with $a < b$ and $c < d$. Let u and v be two C^1 functions from $[a,b]$ to $[c,d]$. If $\frac{\partial f}{\partial x}(x,t)$ exists for any $(x,t) \in E$ and $\frac{\partial f}{\partial x}$ is continuous on E, then $I(x) \coloneqq \int_{u(x)}^{v(x)} f(x,t)dt$ is differentiable on $[a,b]$, and*

$$I'(x) = f(x, v(x))v'(x) - f(x, u(x))u'(x) + \int_{u(x)}^{v(x)} \frac{\partial f}{\partial x}(x,t)dt$$

The next theorem states that under some conditions on function f, Leibniz's Formula can be applied to cases where the region of integration is unbounded.

Theorem 12. *Let f be a function from a subset A of $\mathbb{R}^2$ to $\mathbb{R}$. Let infinite rectangle $E \coloneqq [a,b] \times [c,+\infty) \subset A$. with $a < b$. Let u be a C^1 function from $[a,b]$ to $[c,\infty)$. If*

1. $\frac{\partial f}{\partial x}(x,t)$ exists for any $(x,t) \in E$ and $\frac{\partial f}{\partial x}$ is continuous on E, and
2. $\frac{\partial f}{\partial x}(x,t)$ is integrably bounded, that is, there exists a function $p : [c,\infty) \to \mathbb{R}_+$ s.t. $|\frac{\partial f}{\partial x}(x,t)| \leq p(t)$ for any $t \in [c,\infty)$, $x \in [a,b]$ and $\int_c^{+\infty} p(t)dt < \infty$

then $I(x) \coloneqq \int_{u(x}^{\infty} f(x,t)dt$ is differentiable on $[a,b]$, and

$$I'(x) = -f(x, u(x))u'(x) + \int_{u(x)}^{\infty} \frac{\partial f}{\partial x}(x,t)dt$$

7 Homogeneous Functions

Definition 7. A set C in real vector space V is said to be a **cone**, iff $\lambda v \in C$ for any $\lambda \in \mathbb{R}_{++}$ and $v \in C$.

Definition 8. Let C be a cone in real vector space V, and let W be another real vector space. For $k \in \mathbb{R}$, a function $f : C \to W$ is said to be homogeneous of degree k iff $f(\lambda v) = \lambda^k f(v)$ for any $\lambda \in \mathbb{R}_{++}$ and $v \in C$.

In the definition above, because C is a cone, we know that $f(\lambda v)$ is defined whenever $\lambda \in \mathbb{R}_{++}$ and $v \in C$.

In most applications, C is a cone in $\mathbb{R}^n$ (usually $C = \mathbb{R}^n_{++}$ or $\mathbb{R}^n_+$), $W = \mathbb{R}$, and k is a non-negative integer.

Definition 9. Let C be a cone in $\mathbb{R}^n$, and $f : C \to \mathbb{R}$ homogeneous of degree k. Let $x \in int(C)$ and $\lambda > 0$. If $\frac{\partial f}{\partial x_i}$ exists at x, then $\frac{\partial f}{\partial x_i}$ exists at λx, and we have

$$\frac{\partial f}{\partial x_i}(\lambda x) = \lambda^{k-1} \frac{\partial f}{\partial x_i}(x)$$

Shortly put, the theorem says that a partial of a function homogeneous of degree k is homogeneous of degree k $-$ 1, if the partial exists.

The next theorem is known as Euler's equation for homogeneous functions.

Theorem 13. *(Euler's Equation). Let C be a cone in $\mathbb{R}^n$, and $f : C \to \mathbb{R}$ homogeneous of degree k and differentiable at $x \in int(x)$, and then we have*

$$\nabla f(x) \cdot x = kf(x)$$

Definition 10. Let C be a cone in real vector space V. A function $f : C \to \mathbb{R}$ is said to be homothetic iff there exists $h : C \to \mathbb{R}$ homogeneous of some degree k and $g \coloneqq \mathbb{R} \to \mathbb{R}$ strictly increasing s.t. $f = g \circ h$.

Clearly, if f is homothetic, we have $f(\lambda v) = f(\lambda w)$ for any $\lambda \in \mathbb{R}_{++}$ and $v, w \in C$ s.t. $f(v) = f(w)$.

Intuitively, it says that the "marginal rate of substitution" of a homothetic function is preserved under scalar multiplication

Let C be a cone in $\mathbb{R}^n$, and $f : C \to \mathbb{R}$ homothetic. Let $x \in int(C)$ and $\lambda > 0$. If f is differentiable at x and λx, and $\frac{\partial f}{\partial x_j}(\lambda x)$ and $\frac{\partial f}{\partial x_j}(x)$ are not zero, then we have

$$\frac{\frac{\partial f}{\partial x_i}(\lambda x)}{\frac{\partial f}{\partial x_j}(\lambda x)} = \frac{\frac{\partial f}{\partial x_j}(x)}{\frac{\partial f}{\partial x_j}(x)}$$