4 | Topics in calculus

4.1 Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

- **Transformation.** A transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, denoted as $f : \mathbb{R}^n \to \mathbb{R}^m$ is a generalization of a function: it is an operation that takes as an *input* a vector $\boldsymbol{x} \in \mathbb{R}^n$ and gives as an *output* a vector $\boldsymbol{y} = f(\boldsymbol{x}) \in \mathbb{R}^m$. By convention, functions are scalar valued, $m = 1$.

- **Gradient of a function.** Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, that is to say, a transformation taking as input a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}^n$ and rendering a scalar $y = f(\boldsymbol{x})$. Upon differentiating,

$$dy = \frac{\partial f(\boldsymbol{x})}{\partial x_1} dx_1 + \frac{\partial f(\boldsymbol{x})}{\partial x_2} dx_2 + \cdots + \frac{\partial f(\boldsymbol{x})}{\partial x_n} dx_n \,.$$

The differential of y is the sum of the differentials of each argument, x_i, weighted by the partial derivatives $\partial f(\boldsymbol{x})/\partial x_i$. This result can be expressed as the inner product of two vectors,

$$dy = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial f(\boldsymbol{x})}{\partial x_2} & \cdots & \dfrac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}_{1\times n} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}_{n\times 1} \quad \to \quad dy = \boldsymbol{\nabla} f(\boldsymbol{x}) d\boldsymbol{x} \,,$$

where $d\boldsymbol{x} \in \mathbb{R}^n$ is the vector that collects the differentials of the entries of $\boldsymbol{x}$, and $\boldsymbol{\nabla} f(\boldsymbol{x})$ is the *gradient* (by convention a row vector, of dimension $1 \times n$), that collects the partial derivatives.

- **Directional derivative.** Suppose that the function $f(\boldsymbol{x})$ is initially located at point $\boldsymbol{x} = \boldsymbol{a}$. Then, an infinitesimal change from $\boldsymbol{x}$ to $\boldsymbol{x} + \epsilon \boldsymbol{u}$ occurs, where ϵ is an arbitrarily small number, and $\boldsymbol{u}$ is the vector that gives the *direction* of the change. The change produced in $f(\boldsymbol{x})$, *in the direction of* $\boldsymbol{u}$, is known as the *directional derivative*, denoted as

$$D_u f(\boldsymbol{a}) = \boldsymbol{\nabla} f(\boldsymbol{a}) \boldsymbol{u} \,.$$

The similarities with the definition of the differential are rather obvious. The directional derivative simply evaluates the change at $d\boldsymbol{x} = \boldsymbol{u}$. Typically, it is assumed that $\|\boldsymbol{u}\| = 1$, such that $\boldsymbol{u}$ provides indeed the direction of the change, whereas $\boldsymbol{\nabla} f(\boldsymbol{a})$ gives its magnitude.

Suppose that $\boldsymbol{u} = \boldsymbol{e}_i$, a canonical vector. Then, $D_u f(\boldsymbol{a}) = \partial f(\boldsymbol{x})/\partial x_i$. Thus, the directional derivative generalizes the notion of a partial derivative, in which the change is taken along one of the coordinate curves, all other coordinates held constant.

- **Derivative matrix (Jacobian matrix).** Consider a transformation $f : \mathbb{R}^n \to \mathbb{R}^m$. Each element of vector $\boldsymbol{y}$ is the result of evaluating a function $f_i : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x}$. In other words, $y_i = f_i(\boldsymbol{x})$ for $i = 1, 2, \ldots, m$. The transformation $\boldsymbol{y} = f(\boldsymbol{x})$ simply collects the results of all m function involved. Thus,

$$dy_i = \boldsymbol{\nabla} f_i(\boldsymbol{x}) d\boldsymbol{x} \,.$$

Upon stacking the m differentials dy_i in a vector $\boldsymbol{dy} \in \mathbb{R}^m$, we obtain

$$\begin{bmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_m \end{bmatrix}_{m\times 1} = \begin{bmatrix} \boldsymbol{\nabla} f_1(\boldsymbol{x}) \\ \boldsymbol{\nabla} f_2(\boldsymbol{x}) \\ \vdots \\ \boldsymbol{\nabla} f_m(\boldsymbol{x}) \end{bmatrix}_{m\times n} \begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{bmatrix}_{n\times 1} \quad \rightarrow \quad \boldsymbol{dy} = \boldsymbol{D}f(\boldsymbol{x})\boldsymbol{dx}\,.$$

This expression generalizes the notion of the differential for transformations whose range is $\mathbb{R}^m$. The matrix $\boldsymbol{D}f(\boldsymbol{x})$ links the changes in the n entries of $\boldsymbol{x}$, to the changes of the m entries of $\boldsymbol{y}$, and is therefore $m \times n$.

The i-th row of this matrix contains the gradient of function $f_i(\boldsymbol{x})$. Thus, explicitly

$$\boldsymbol{D}f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_2} & \cdots & \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_n} \\ \dfrac{\partial f_2(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial f_2(\boldsymbol{x})}{\partial x_2} & \cdots & \dfrac{\partial f_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial f_m(\boldsymbol{x})}{\partial x_1} & \dfrac{\partial f_m(\boldsymbol{x})}{\partial x_2} & \cdots & \dfrac{\partial f_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix}_{m\times n}.$$

This is the *Jacobian matrix*, and contains the partial derivatives of all functions $f_i(\boldsymbol{x})$ ($i = 1, 2, \ldots, m$ arrayed in rows), with respect to all the inputs x_j ($j = 1, 2, \ldots, n$ arrayed in columns).

- **Jacobian.** The determinant of $\boldsymbol{D}f(\boldsymbol{x})$, of course when $m = n$.

- **Composition of transformations.** Consider two transformations $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^p \to \mathbb{R}^n$. Transformation $g(\cdot)$ takes a $\boldsymbol{u}$ in $\mathbb{R}^p$ and renders a vector $\boldsymbol{x} = g(\boldsymbol{u})$ from $\mathbb{R}^n$. On the other hand, transformation $f(\cdot)$ takes a vector $\boldsymbol{x}$ in $\mathbb{R}^n$ and gives as a result a vector $\boldsymbol{y} = f(\boldsymbol{x})$ in $\mathbb{R}^m$.

  We could think of a transformation $h : \mathbb{R}^p \to \mathbb{R}^m$ which is the result of the sequential application of the transformations g y f: $\boldsymbol{y} = h(\boldsymbol{u})$ takes a vector $\boldsymbol{u}$ in $\mathbb{R}^p$ and renders a $\boldsymbol{y} = f(g(\boldsymbol{u}))$ de $\mathbb{R}^m$. h is a *composite transformation*. Often, the operation $f(g(\boldsymbol{u}))$ is denoted as $f \circ g(\boldsymbol{u})$.

- **Jacobian matrix of a composition.** We note that $\boldsymbol{dx} = \boldsymbol{D}g(\boldsymbol{u})\boldsymbol{du}$, where $\boldsymbol{D}g(\boldsymbol{u})$ is the $n \times p$ Jacobian matrix of g. Moreover, we have that $\boldsymbol{dy} = \boldsymbol{D}f(\boldsymbol{x})\boldsymbol{dx}$, where $\boldsymbol{D}f(\boldsymbol{x})$ is the $m \times n$ Jacobian matrix of f. Then, for the composite function $\boldsymbol{y} = f \circ g(\boldsymbol{u})$

  $$\boldsymbol{dy} = \boldsymbol{D}f(\boldsymbol{x})\boldsymbol{dx} = \boldsymbol{D}f(\boldsymbol{x})[\,\boldsymbol{D}g(\boldsymbol{u})\boldsymbol{du}\,] = \boldsymbol{D}f(\boldsymbol{x})\boldsymbol{D}g(\boldsymbol{u})\boldsymbol{du} \equiv \boldsymbol{D}f \circ g(\boldsymbol{u})\boldsymbol{du}\,.$$

  The $m \times p$ Jacobian matrix of the composite transformation is the product of the Jacobian matrices of each of the transformations composing $f \circ g$.

- **Chain rule.** The (i,j)-th entry of $\boldsymbol{D}f \circ g(\boldsymbol{u})$ is the inner product between the i-th row of $\boldsymbol{D}f(\boldsymbol{x})$ and the j-th column of $\boldsymbol{D}g(\boldsymbol{u})$. Hence,

  $$[\boldsymbol{D}f \circ g(\boldsymbol{u})]_{ij} = \frac{\partial f_i(\boldsymbol{x})}{\partial x_1}\frac{\partial g_1(\boldsymbol{u})}{\partial u_j} + \frac{\partial f_i(\boldsymbol{u})}{\partial x_2}\frac{\partial g_2(\boldsymbol{u})}{\partial u_j} + \ldots + \frac{\partial f_i(\boldsymbol{x})}{\partial x_n}\frac{\partial g_n(\boldsymbol{u})}{\partial u_j}\,,$$

  for $i = 1, 2\ldots, m$ and $j = 1, 2, \ldots, p$.

  Recall that $y_i = f_i(\boldsymbol{x})$ and $x_s = g_s(\boldsymbol{u})$. Then, we can write

  $$[\boldsymbol{D}f \circ g(\boldsymbol{u})]_{ij} = \frac{\partial y_i}{\partial x_1}\frac{\partial x_1}{\partial u_j} + \frac{\partial y_i}{\partial x_2}\frac{\partial x_2}{\partial u_j} + \ldots + \frac{\partial y_i}{\partial x_n}\frac{\partial x_n}{\partial u_j} \equiv \frac{\partial y_i}{\partial u_j}\,,$$

  which reveals a generalization of the *chain rule*: how y_i varies with a change in u_j, after taking into account the composition process $f \circ g$.

4.2 Implicit functions

- Consider two vector $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{x} \in \mathbb{R}^n$ and a transformation $F : \mathbb{R}^{m+n} \to \mathbb{R}^m$ such that, for a given point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$, we have that

  $$F(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) = \mathbf{0}\,.$$

  This equality defines an *implicit function*. In particular, the goal is to verify if there exists a transformation $f : \mathbb{R}^n \to \mathbb{R}^m$, such that $\bar{\boldsymbol{y}} = f(\bar{\boldsymbol{x}})$ and

  $$F(\bar{\boldsymbol{x}}, f(\bar{\boldsymbol{x}})) = \mathbf{0}\,.$$

- **Implicit function theorem.** Upon differentiating the equality $F(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) = \mathbf{0}$ we obtain, using partitions,

  $$\begin{bmatrix} DF_x(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) & DF_y(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \end{bmatrix} \begin{bmatrix} d\boldsymbol{x} \\ d\boldsymbol{y} \end{bmatrix} = \mathbf{0} \qquad \to \qquad DF_x(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})d\boldsymbol{x} + DF_y(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})d\boldsymbol{y} = \mathbf{0}\,,$$

  where $DF_x(\cdot)$ is the $m \times n$ Jacobian matrix that collects the partial derivatives of $F(\cdot)$ with respect to $\boldsymbol{x}$, whereas $DF_y(\cdot)$ is the $m \times m$ Jacobian matrix that collects the partial derivatives of $F(\cdot)$ with respect to $\boldsymbol{y}$. Solving for $d\boldsymbol{y}$,

  $$d\boldsymbol{y} = -DF_y(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})^{-1} DF_x(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})d\boldsymbol{x}\,,$$

  which established a relationship between $d\boldsymbol{y}$ and $d\boldsymbol{x}$.

  $d\boldsymbol{y}$ measures the change in $\boldsymbol{y}$ for a change in $\boldsymbol{x}$ in other to keep the equality $F(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) = \mathbf{0}$ and, in doing so, defines $\boldsymbol{y}$ as a function of $\boldsymbol{x}$, around $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$.

- **Explicit functions.** Consider the "explicit function" approach $\boldsymbol{y} = f(\boldsymbol{x})$. Upon differentiating, by the chain rule,

  $$dF(\boldsymbol{x}, f(\boldsymbol{x})) = DF_x(\boldsymbol{x}, \boldsymbol{y})d\boldsymbol{x} + DF_y(\boldsymbol{x}, \boldsymbol{y})Df(\boldsymbol{x})d\boldsymbol{x}\,.$$

  Provided that $dF(\bar{\boldsymbol{x}}, f(\bar{\boldsymbol{x}})) = \mathbf{0}$, we have

  $$Df(\bar{\boldsymbol{x}})d\boldsymbol{x} = -DF_y(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})^{-1} DF_x(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})d\boldsymbol{x} \quad \to \quad Df(\bar{\boldsymbol{x}}) = -DF_y(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})^{-1} DF_x(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})\,.$$

- **Explicit functions, again.** In the case $F(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{y}$, we observe that $DF_y(\boldsymbol{x}, \boldsymbol{y}) = -\boldsymbol{I}_m$ and therefore $Df(\boldsymbol{x}) = DF_x(\boldsymbol{x}, \boldsymbol{y})$.

- **Existence.** From our previous analysis, we can conclude that an implicit function exists around point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ if $DF_y(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ is a nonsingular matrix. Hence, if the *Jacobian* satisfies

  $$|\, DF_y(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \,| \neq 0\,.$$

4.3 Taylor polynomials and Taylor series

- **Basics.** The purpose is to derive polynomial approximations of a continuous function $f : \mathbb{R} \to \mathbb{R}$, around (i.e., *in the neighborhood of*) a point a.

  To this end, consider the polynomial

  $$p_n(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^3 + \cdots + c_n(x-a)^n\,,$$

  in x of degree n.

  Also, denote the i-th derivative of $f(x)$ with respect to x as

  $$f^{(i)}(x) = \frac{d^i f(x)}{d\,x^i} \qquad \text{and} \qquad f^{(i)}(a) = \left.\frac{d^i f(x)}{d\,x^i}\right|_{x=a}\,.$$

- **Linear approximation.** From the definition of the derivative, we know that the linear function (i.e., a polynomial of degree 1) $p_1(x) = c_0 + c_1(x-a)$ that best approximates function $f(x)$ around $x = a$, is the tangent curve that passes through $f(a)$. At the tangency point, $p_1(a) = f(a)$ and $p_1^{(1)}(a) = f^{(1)}(a)$. Then,

$$\begin{array}{lclcl} p_1(x) = c_0 + c_1(x-a) & \to & p_1(a) = c_0 = f(a) & \to & c_0 = f(a) \\ p_1^{(1)}(x) = c_1 & \to & p_1^{(1)}(a) = c_1 = f^{(1)}(a) & \to & c_1 = f^{(1)}(a) \,. \end{array}$$

  Upon replacing these findings on the definition of $p_1(x)$, we obtain the equation of the tangent curve

$$p_1(x) = f(a) + f^{(1)}(a)(x-a) \,.$$

- **Quadratic approximation.** Let us find now the quadratic function $p_2(x) = c_0 + c_1(x-a) + c_2(x-a)^2$ that best approximates $f(x)$ around $x = a$. At a we have that $p_1(a) = f(a)$, $p_1^{(1)}(a) = f^{(1)}(a)$ *and* $p_1^{(2)}(a) = f^{(2)}(a)$. That is, the value, slope and curvature of both $f(x)$ and $p_2(x)$ coincide at $x = a$. Then,

$$\begin{array}{lclcl} p_2(x) = c_0 + c_1(x-a) + c_2(x-a)^2 & \to & p_2(a) = c_0 = f(a) & \to & c_0 = f(a) \\ p_2^{(1)}(x) = c_1 + 2c_2(x-a) & \to & p_2^{(1)}(a) = c_1 = f^{(1)}(a) & \to & c_1 = f^{(1)}(a) \\ p_2^{(2)}(x) = 2c_2 & \to & p_2^{(2)}(a) = 2c_2 = f^{(2)}(a) & \to & c_2 = \frac{1}{2}f^{(2)}(a) \,. \end{array}$$

  Upon replacing these findings on the definition of $p_2(x)$, we obtain

$$p_2(x) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2}(x-a)^2 \,.$$

  Note that $p_2(x)$ equals the linear approximation $p_1(x)$ *plus* the quadratic term $\frac{1}{2}f^{(2)}(a)(x-a)^2$.

- **Cubic approximation.** Now we get the cubic polynomial $p_3(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^2$ that best approximates $f(x)$ around $x = a$. We now require that at point a the value of $p_3(x)$ and its first three derivatives to be the same as those of $f(x)$. Thus,

$$\begin{array}{lclcl} p_3(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + c_3(x-a)^n & \to & p_3(a) = c_0 = f(a) & \to & c_0 = f(a) \\ p_3^{(1)}(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 & \to & p_n^{(1)}(a) = c_1 = f^{(1)}(a) & \to & c_1 = f^{(1)}(a) \\ p_3^{(2)}(x) = 2c_2 + (3\cdot 2)c_3(x-a) & \to & p_3^{(2)}(a) = 2c_2 = f^{(2)}(a) & \to & c_2 = \frac{1}{2}f^{(2)}(a) \\ p_3^{(3)}(x) = (3\cdot 2)c_3(x-a) & \to & p_n^{(3)}(a) = 3!c_3 = f^{(3)}(a) & \to & c_3 = \frac{1}{3!}f^{(3)}(a) \,. \end{array}$$

  Upon replacing these findings on the definition of $p_3(x)$, we obtain

$$p_3(x) = f(a) + f^{(1)}(a)(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 \,.$$

  Note that $p_3(x)$ equals the quadratic approximation $p_2(x)$ *plus* the cubic term $\frac{1}{3!}f^{(3)}(a)(x-a)^3$.

- **General case.** Consider now a polynomial of degree n. To determine its $n+1$ coefficients, we impose the $n+1$ restrictions $p_n^{(k)}(a) = f^{(k)}(a)$ for $k = 0, 1, 2, \ldots, n$. That is to say, at point a the value of $f(x)$ and $p_n(x)$ coincides, as well as their first n derivatives. Thus,

$$\begin{array}{lcl} p_n(x) = c_0 + c_1(x-a) + c_2(x-a)^2 + \cdots + c_n(x-a)^n & \to & p_n(a) = c_0 \\ p_n^{(1)}(x) = c_1 + 2c_2(x-a) + 3c_3(x-a)^2 + \cdots + nc_n(x-a)^{n-1} & \to & p_n^{(1)}(a) = c_1 \\ p_n^{(2)}(x) = 2c_2 + (3\cdot 2)c_3(x-a) + (4\cdot 3)c_4(x-a)^2 + \cdots + n(n-1)c_n(x-a)^{n-2} & \to & p_n^{(2)}(a) = 2c_2 \\ p_n^{(3)}(x) = (3\cdot 2)c_3(x-a) + (4\cdot 3\cdot 2)c_4(x-a) + \cdots + n(n-1)(n-2)c_n(x-a)^{n-3} & \to & p_n^{(3)}(a) = 3!c_3 \,. \\ \quad\vdots & & \\ p_n^{(n)}(x) = n\cdot(n-1)\cdot(n-2)\cdots 4\cdot 3\cdot 2c_n & \to & p_n^{(n)}(a) = n!c_n \end{array}$$

from which we deduce $p_n^{(k)}(a) = k!c_k$ for $k = 0, 1, 2, \ldots, n$ (recall that $0! = 1$).

Then, from the equality $p_n^{(k)}(a) = f^{(k)}(a)$ we obtain that $c_k = f^{(k)}(a)/!k$, rendering the so-called *Taylor polynomial of degree n*

$$p_n(x) = \frac{f(a)}{0!} + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}(x-a)^k + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n .$$

Note that

$$p_n(x) = p_{n-1}(x) + \frac{f^{(n)}(a)}{n!}(x-a)^n .$$

Thus, Taylor polynomials are formed by sequentially adding terms to the approximation in order to match higher-order derivatives of the original function $f(x)$. For this reason, we can expect the degree of accuracy of the approximation (*in the neighborhood of* a) to be increasing in n.

- **Remainder (Optional).** When approximating a function by a Taylor polynomial of degree n, we generate an approximation error, a *remainder*, which is equal to

  $$R_n(x) = f(x) - p_n(x) .$$

  An explicit, useful form for $R_n(x)$ can be found through a simple application of the *Mean Value Theorem*. This important theorem says that if we have a function $g(x)$ that is continuous on the interval $[a, b]$, then there exists a value (the mean value) $c \in [a, b]$ such that

  $$g(b) = g(a) + g'(c)(b-a) .$$

  Note that the theorem *does not refer to an approximation*, it involves an *equality* (with no error). The theorem states that c exists, it does not tell us how much it is, only that it belongs to $[a, b]$.

  Back to the remainder of the Taylor approximation, we then have that

  $$R_n(x) = \frac{f^{(n+1)}(z)}{(n+1)!}(x-a)^{n+1} ,$$

  where z lies between x and a. This looks quite similar to the extra term we would add when we move from $p_n(x)$ to $p_{n+1}(x)$. The important difference is that in the definition of $R_n(x)$, $f^{(n+1)}(\cdot)$ is evaluated at z, whereas it is evaluted at a in $p_{n+1}(x) - p_n(x)$.

  This expression for $R_n(x)$ is known as the **Lagrange remainder**.

- **Bounding the approximation error (Optional).** We can use the Lagrange remainder to compute the magnitude of the maximum error (in absolute value) that the Taylor polynominal $p_n(x)$, centered at a, produces when approximating $f(x)$.

  Let $M_n = \max_{z \in [x,a]} | f^{(n+1)}(z) |$. That is, M_n is the maximum absolute value of $f^{(n+1)}(\cdot)$ on the interval between x and a. Then,

  $$|R_n(x)| = \left| \frac{f^{(n+1)}(z)}{(n+1)!}(x-a)^{n+1} \right| \leq M_n \frac{|x-a|^{n+1}}{(n+1)!} = \text{Error bound} .$$

- **Taylor's theorem.** Suppose that for all n, $M_n \leq M$ which is a finite number. Then,

  $$\lim_{n\to\infty} |R_n(x)| \leq M \lim_{n\to\infty} \frac{|x-a|^{n+1}}{(n+1)!} = 0 .$$

  This happens because the factorial function grows faster than the power function.

  The implication of Taylor's theorem is that since $R_n(x) \to 0$ as $n \to \infty$, then

  $$\lim_{n\to\infty} p_n(x) = \lim_{n\to\infty} \left[f(x) - R_n(x) \right] = f(x) - \lim_{n\to\infty} R_n(x) = f(x) .$$

  As n increases, our approximation becomes progressively more accurate to the extent that the Taylor polynominal *converges* to $f(x)$.

- **Taylor series.** Also known as the *power series representation* of $f(x)$:

$$f(x) = \lim_{n\to\infty} p_n(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k \,.$$

This is a Taylor polynomial of infinite degree.

- **Ratio test.** Consider an infinite series

$$S = \sum_{k=0}^{\infty} s_k \,.$$

The ratio tests says that this infinite sum converges to a finite value (that we actually do not always know) if

$$\lim_{n\to\infty}\left|\frac{s_{n+1}}{s_n}\right| = L < 1 \,.$$

- **Range of convergence.** For a Taylor series, $s_k = f^{(k)}(a)(x-a)^k/k!$. Then, taking the limit of the ratio test,

$$L = \lim_{n\to\infty}\left|\frac{s_{n+1}}{s_n}\right| = \lim_{n\to\infty}\left|\frac{f^{(n+1)}(a)}{f^{(n)}(a)} \cdot \frac{(x-a)^{n+1}}{(x-a)^n} \cdot \frac{n!}{(n+1)!}\right| = |x-a| \lim_{n\to\infty}\left|\frac{f^{(n+1)}(a)}{f^{(n)}(a)} \cdot \frac{(x-a)^{n+1}}{(x-a)^n} \cdot \frac{1}{n+1}\right| \,.$$

This will be less than one if

$$|x-a| < \lim_{n\to\infty}\left|\frac{f^{(n)}(a)}{f^{(n+1)}(a)}(n+1)\right| \equiv \text{RC} \,.$$

The quantity RC is the *range of convergence*. This means that Taylor's theorem holds (so $p_n(x)$ converges to $f(x)$ as n increases) for all values of x such that $|x-a| < \text{RC}$.

The interval $\text{IC}_a = (\text{RC} - a, \text{RC} + a)$ is the *interval of convergence*. Taylor's theorem holds if $x \in \text{IC}_a$.

- **McLaurin polynominals and series.** Taylor polynomials and series for $a = 0$.

- **Multivariate functions (Optional).** If $f : \mathbb{R}^n \to \mathbb{R}$, the second order Taylor polynominal of $f(\boldsymbol{x})$ around point $\boldsymbol{a} \in \mathbb{R}^n$ is

$$p_2(\boldsymbol{x}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a})(\boldsymbol{x}-\boldsymbol{a}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{H}(\boldsymbol{a})(\boldsymbol{x}-\boldsymbol{a}) \,,$$

where $\nabla f(\boldsymbol{a})$ is the gradient of f evaluated at point $\boldsymbol{a}$, and $\boldsymbol{H}(\boldsymbol{a})$ is the *Hessian* (i.e., the matrix of second order partial derivatives) also evaluated at point $\boldsymbol{a}$.

This is a straightforward generalization of the scalar case. The linear term comes from the inner product of two vectors, whereas the quadratic term comes from a quadratic form.