Differentiation in Linear Spaces - II

Prof. Dan A. Simovici

UMB

Theorem

(Chain Rule Theorem) *Let* S, T, U *be three normed spaces,* X *be an open subset of* S *and* Y *an open subset of* T*.*
If $f : X \longrightarrow T$ *is a Gâteaux differentiable function at* $x_0 \in X$ *and* $g : Y \longrightarrow U$ *is a Gâteaux differentiable function at* $y_0 = f(x_0) \in Y$*, then* gf *is differentiable at* x_0 *and*

$$D_x(gf)(x_0) = (D_y g)(f(x_0))(D_x f)(x_0).$$

Proof

Since $f(x_0+h)-f(x_0)-(D_xf)(x_0)(h)=o_1(h)$ and $g(y_0+k)-g(y_0)-(D_yg)(y_0)(k)=o_2(k)$, we have

$$
\begin{aligned}
& g(f(x_0+p))-g(f(x_0)) \\
= \;& g\left(f(x_0)+(D_xf)(x_0)(p)+o_1(p)\right)-g(f(x_0)) \\
= \;& (D_yg)(y_0)\left((D_xf)(x_0)(p)+o_1(p)\right)+o_2((D_xf)(x_0)(p)+o_1(p)) \\
= \;& (D_yg)(y_0)((D_xf)(x_0)(p)) \\
& +(D_yg)(y_0)(o_1(p))+o_2((D_xf)(x_0)(p)+o_1(p)).
\end{aligned}
$$

Observe that

$$\lim_{h\to 0}\frac{(D_yg)(y_0)(o_1(p))+o_2((D_xf)(x_0)(p)+o_1(p))}{\| p \|}\}=0,$$

because $\| (D_yg)(y_0)(o_1(p)) \| \leqslant \| (D_yg)(y_0) \| \| o_1(p) \|$.

Proof (cont'd)

Also,

$$\| o_2((D_x f)(x_0)(p) + o_1(p)) \| \leqslant \| o_2((D_x f)(x_0)(p)) \| + \| o_2(o_1(p)) \|$$

which shows that $D_x(gf)(x_0) = (D_y g)(f(x_0))(D_x f)(x_0)$.
Important: the equality above is the composition of two linear operators, which corresponds, in the finite dimensional case to a matrix product (see next example).

Example (cont'd)

Example

Let $\mathbf{f} : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ and $\mathbf{g} : \mathbb{R}^4 \longrightarrow \mathbb{R}^3$ be the functions defined by

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_1 + x_2 \\ x_1^2 - x_2^2 \\ x_1 - x_2 \end{pmatrix} \text{ and } \mathbf{g}(\mathbf{y}) = \begin{pmatrix} y_1 + y_2 \\ y_2 + y_3 \\ y_1 + y_3 \\ y_1^2 \end{pmatrix}$$

for $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{y} \in \mathbb{R}^3$.
We have

$$(D\mathbf{f})(\mathbf{x}) = \begin{pmatrix} 1 & 1 \\ 2x_1 & -2x_2 \\ 1 & -1 \end{pmatrix} \text{ and } (D\mathbf{g})(\mathbf{y}) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2y_1 & 0 & 0 \end{pmatrix}.$$

Example (cont'd)

We have

$$(D\mathbf{f})(\mathbf{x}) = \begin{pmatrix} 1 & 1 \\ 2x_1 & -2x_2 \\ 1 & -1 \end{pmatrix} \text{ and } (D\mathbf{g})(\mathbf{y}) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2y_1 & 0 & 0 \end{pmatrix}.$$

If $\mathbf{x}_0 = \begin{pmatrix} a \\ b \end{pmatrix}$, then

$$\mathbf{f}(\mathbf{x}_0) = \begin{pmatrix} a+b \\ a^2-b^2 \\ a-b \end{pmatrix}$$

Example (cont'd)

By applying the Chain Rule we can write:

$$\begin{aligned} D(\mathbf{gf})(\mathbf{x}_0) &= (D\mathbf{g})(\mathbf{f}(\mathbf{x}_0))(D\mathbf{f})(\mathbf{x}_0) \\ &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 2(a+b) & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2a & -2b \\ 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1+2a & 1-2b \\ 2a+1 & -2b-1 \\ 2 & 0 \\ 2(a+b) & 2(a+b) \end{pmatrix}. \end{aligned}$$

Alternatively, the derivative of **gf** can be computed by composing first the functions **f** and **g**. This yields

$$\mathbf{gf}(\mathbf{x}) = \begin{pmatrix} x_1 + x_2 + x_1^2 - x_2^2 \\ x_1^2 - x_2^2 + x_1 - x_2 \\ 2x_1 \\ (x_1 + x_2)^2 \end{pmatrix}.$$

This implies

$$(D\mathbf{gf})(\mathbf{x}) = \begin{pmatrix} 1 + 2x_1 & 1 - 2x_2 \\ 2x_1 + 1 & -2x_2 - 1 \\ 2 & 0 \\ 2(x_1 + x_2) & 2(x_1 + x_2) \end{pmatrix}.$$

Substituting a for x_1 and b for x_2 we retrieve the first expression.

The next definition introduces the notation "∇f" (read *"nabla f"*).
Let $f : X \longrightarrow \mathbb{R}$, where $X \subseteq \mathbb{R}^n$, and let $\mathbf{z} \in X$. The gradient of f in $\mathbf{z}$ is the vector

$$(\nabla f)(\mathbf{z}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{z}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{z}) \end{pmatrix} \in \mathbb{R}^n.$$

If $\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ then $\mathbf{f}$ is Gâteaux differentiable at $\mathbf{x}_0$ if there exists a matrix $A \in \mathbb{R}^{m \times n}$ such that

$$\lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\mathbf{u}) - \mathbf{f}(\mathbf{x}_0)}{t} = A\mathbf{u}$$

for every $\mathbf{u} \in \mathbb{R}^n$. Namely, A is the matrix of the linear transformation $(D\mathbf{f})(x_0) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

If $\mathbf{a}_i$ is the i^{th} row of A, where $1 \leqslant i \leqslant m$, the previous equality can be written componentwise as

$$\lim_{t \to 0} \frac{f_i(\mathbf{x}_0 + t\mathbf{u}) - f_i(\mathbf{x}_0)}{t} = \mathbf{a}_i\mathbf{u}$$

for $1 \leqslant i \leqslant m$. Thus, we have $\mathbf{a}_i = (\nabla f_i)(\mathbf{x}_0)'$ for $1 \leqslant i \leqslant m$.

$$A = \begin{pmatrix} (\nabla f_1)(\mathbf{x}_0)' \\ \vdots \\ (\nabla f_m)(\mathbf{x}_0)' \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \cdots & \vdots \\ \frac{\partial f_i}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_i}{\partial x_n}(\mathbf{x}_0) \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}_0) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}_0) \end{pmatrix}$$

The matrix $A \in \mathbb{R}^{m \times n}$ is referred to as the *Jacobian of* $\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *at* $\mathbf{x}_0$. As before, if $\mathbf{x}$ is understood from context we may omit occasionally the subscript $\mathbf{x}$ and write A simply as $(D\mathbf{f})(\mathbf{x}_0)$. The rows of Jacobian of $\mathbf{f}$ at $\mathbf{x}_0$ consist of the transposes of the gradients of its component functions $f_1, \ldots, f_m$ at $\mathbf{x}_0$.

Example

Let $\mathbf{f} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ be the function that maps spherical coordinates into Cartesian coordinates. Its components are:

$$\begin{aligned} f_1(r,\theta,\phi) &= r\sin\theta\cos\phi, \\ f_2(r,\theta,\phi) &= r\sin\theta\sin\phi, \\ f_3(r,\theta,\phi) &= r\cos\theta. \end{aligned}$$

The Jacobian matrix in (r,θ,ϕ) is:

$$(D\mathbf{f})(r,\theta,\phi) = \begin{pmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}.$$

Note that $\det(D\mathbf{f})(r,\theta,\phi) = r^2\sin\theta$.

Example

The function $\mathbf{g} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ that maps polar coordinates into Cartesian coordinates, given by

$$\begin{aligned} g_1(r,\phi) &= r\cos\phi, \\ g_2(r,\phi) &= r\sin\phi \end{aligned}$$

has the Jacobian matrix

$$(D\mathbf{g})(r,\phi) = \begin{pmatrix} \cos\phi & -r\sin\phi, \\ \sin\phi & r\cos\phi \end{pmatrix}$$

and its determinant is $\det(D\mathbf{g})(r,\phi) = r$.

Example

Let $\mathbf{f} : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ be the function defined by:

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_1^2 + x_2^2 + x_3^2 \\ e^{x_1} + e^{x_2} + e^{x_3} \end{pmatrix}$$

for $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$. The Jacobian of $\mathbf{f}$ is the matrix

$$(D\mathbf{f})(\mathbf{x}) = \begin{pmatrix} 2x_1 & 2x_2 & 2x_3 \\ e^{x_1} & e^{x_2} & e^{x_3} \end{pmatrix} \in \mathbb{R}^{2\times 3}.$$

Definition

Let $\mathbf{f} : X \longrightarrow \mathbb{R}^m$, where $X \subseteq \mathbb{R}^n$ is an open set and $\mathbf{x}_0 \in X$. If the limit

$$\lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\mathbf{u}) - \mathbf{f}(\mathbf{x}_0)}{t}$$

exists, then we refer to it as the directional derivative of $\mathbf{f}$ in $\mathbf{x}_0$ with respect to $\mathbf{u}$ and we write

$$\frac{\partial \mathbf{f}}{\partial \mathbf{u}}(\mathbf{x}_0) = \lim_{t \to 0} \frac{\mathbf{f}(\mathbf{x}_0 + t\mathbf{u}) - \mathbf{f}(\mathbf{x}_0)}{t}.$$

If $\mathbf{u} = \mathbf{e}_i$, then $\frac{\partial \mathbf{f}}{\partial \mathbf{e}_i}(\mathbf{x})$ is called the *partial derivative with respect to* x_i and is denoted by $\frac{\partial \mathbf{f}}{\partial x_i}$.

If $\mathbf{f}$ is Gâteaux differentiable in $\mathbf{x}_0$, then the directional derivative of $\mathbf{f}$ exists with respect to every direction $\mathbf{u}$ and we have

$$\frac{\partial \mathbf{f}}{\partial \mathbf{u}}(\mathbf{x}_0) = (D\mathbf{f})(\mathbf{x}_0)(\mathbf{u}).$$

Unlike the case of single-argument functions, where the existence of a derivative at a point $\mathbf{x}_0$ implies the continuity in $\mathbf{x}_0$, the existence of partial derivatives of a multi-argument function does not imply its continuity.

Example

Let $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be the function defined by

$$f(\mathbf{x}) = \begin{cases} \cos \frac{\pi}{2} \frac{x_1^2 - x_2^2}{x_1^2 + x_2^2} & \text{if } \mathbf{x} \neq \mathbf{0}_2, \\ 0 & \text{if } \mathbf{x} = \mathbf{0}_2. \end{cases}$$

Note that $\lim_{n\to\infty} f\left(\frac{1}{n}, \frac{1}{n}\right) = 1$, while $f(\mathbf{0}_2) = 0$, so f is not continuous in $\mathbf{0}_2$. However, partial derivatives do exist in $\mathbf{0}_2$ because

$$\frac{\partial f}{\partial x_1}(\mathbf{0}_2) = \lim_{x_1 \to 0} \frac{f(x_1, 0)}{x_1} = 0 \text{ and } \frac{\partial f}{\partial x_2}(\mathbf{0}_2) = \lim_{x_2 \to 0} \frac{f(0, x_2)}{x_2} = 0.$$

If $\frac{\partial \mathbf{f}}{\partial \mathbf{u}}$ exists relative to every direction $\mathbf{u}$, then $\frac{\partial \mathbf{f}}{\partial x_i}$ exists for every i, $1 \leqslant i \leqslant n$. The converse is not true, as the next example shows.

Example

Consider the function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$f(\mathbf{x}) = \begin{cases} a & \text{if } \mathbf{x} = a\mathbf{e}_i \text{ for some } i \\ 1 & \text{otherwise.} \end{cases}$$

We have

$$\frac{\partial f}{\partial x_i}(\mathbf{0}) = \lim_{t \to 0} \frac{f(t\mathbf{e}_i) - f(\mathbf{0})}{t} = 1.$$

for $1 \leqslant i \leqslant n$. However, if $\mathbf{u} \not\in \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, then

$$\frac{f(t\mathbf{u}) - f(\mathbf{0})}{t} = \frac{1}{t}$$

and $\lim_{t \to 0} \frac{1}{t}$ does not exist, which shows that $\frac{\partial f}{\partial \mathbf{u}}(\mathbf{0})$ does not exists if $\mathbf{u} \not\in \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Starting with a vector $\mathbf{h} \in \mathbb{R}^n$ and the symbolic vector ∇ we consider a symbolic "scalar product" $\mathbf{h}'\nabla$ defined by

$$\mathbf{h}'\nabla = h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n},$$

which is a differential operator that can be applied to a functional $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. The definition of the differential can now be written in a condensed form as

$$\delta f(\mathbf{x}_0; \mathbf{h}) = ((\mathbf{h}'\nabla) f)(\mathbf{x}_0).$$

For a function $\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ that is Gâteaux differentiable at $\mathbf{x}_0$, the vector $(D_{\mathbf{x}}\mathbf{f})(\mathbf{x}_0)(\mathbf{u})$ is the product of the Jacobian matrix of $\mathbf{f}$ computed at $\mathbf{x}_0$ and $\mathbf{u}$.

Example

Let $\mathbf{h} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be the linear operator defined by $\mathbf{h}(\mathbf{x}) = A\mathbf{x}$, where $A \in \mathbb{R}^{m \times n}$. By the definition of the Gâteaux derivative of $\mathbf{h}$ we can write

$$\begin{aligned}(D\mathbf{h})(\mathbf{x}_0)(u) &= \lim_{t \to 0} \frac{\mathbf{h}(\mathbf{x}_0 + t\mathbf{u}) - \mathbf{h}(\mathbf{x}_0)}{t} \\ &= \lim_{t \to 0} \frac{A(\mathbf{x}_0 + t\mathbf{u}) - Ax_0}{t} \\ &= \lim_{t \to 0} \frac{tA\mathbf{u}}{t} = A\mathbf{u},\end{aligned}$$

hence $(D\mathbf{h})(\mathbf{x}_0) = A$ for every $\mathbf{x}_0 \in \mathbb{R}^n$.

Example (cont'd)

Example

Suppose now that, as before, $\mathbf{h}(\mathbf{x}) = A\mathbf{x}$, and $\mathbf{x} = \mathbf{g}(\mathbf{z})$, where $\mathbf{g} : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ is a Gâteaux differentiable function. An application of the chain rule yields the formula

$$(D_{\mathbf{z}}\mathbf{h})(\mathbf{z}_0) = A(D_{\mathbf{z}}\mathbf{g})(\mathbf{z}_0).$$

Example

Let $f : \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}$ be the functional defined by $f(\mathbf{x}, \mathbf{y}) = \mathbf{y}' A \mathbf{x}$, where $\mathbf{y} \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n}$, and $\mathbf{x} \in \mathbb{R}^n$. We have:

$$(D_{\mathbf{x}} f)(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{y}_0' A \text{ and } (D_{\mathbf{y}} f)(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{x}_0' A'.$$

By the definition of the Gâteaux differential, $(D_{\mathbf{x}} f)(\mathbf{x}_0, \mathbf{y}_0)$ is given by

$$\begin{aligned}
(D_{\mathbf{x}} f)(\mathbf{x}_0, \mathbf{y}_0) &= \lim_{t \to 0} \frac{f(\mathbf{x}_0 + t\mathbf{u}, \mathbf{y}_0) - f(\mathbf{x}_0, \mathbf{y}_0)}{t} \\
&= \lim_{t \to 0} \frac{\mathbf{y}_0' A(\mathbf{x}_0 + t\mathbf{u}) - \mathbf{y}_0' A \mathbf{x}_0}{t} \\
&= \lim_{t \to 0} \frac{t\mathbf{y}_0' A \mathbf{u}}{t} \\
&= \mathbf{y}_0' A \mathbf{u}
\end{aligned}$$

for $\mathbf{u} \in \mathbb{R}^n$. Therefore, since the limit of the fraction when $t \to 0$ is 0, we have $(D_{\mathbf{x}} f)(\mathbf{x}_0) = \mathbf{y}_0' A$.

Example cont'd

Example

Similarly, the fraction that enters in the definition of $(D_{\mathbf{x}}f)(\mathbf{x}_0, \mathbf{y}_0)$ is

$$\begin{aligned}(D_{\mathbf{y}}f)(\mathbf{x}_0, \mathbf{y}_0)(\mathbf{v}) &= \lim_{t\to 0} \frac{f(\mathbf{x}_0, \mathbf{y}_0 + t\mathbf{v}) - f(\mathbf{x}_0, \mathbf{y}_0)}{t} \\ &= \lim_{t\to 0} \frac{(\mathbf{y}_0' + t\mathbf{v}')A\mathbf{x}_0 - \mathbf{y}_0' A\mathbf{x}_0}{t} \\ &= \lim_{t\to 0} \frac{t\mathbf{v}' A\mathbf{x}_0}{t} = \mathbf{v}' A\mathbf{x}_0.\end{aligned}$$

Since $\mathbf{v}' A\mathbf{x}_0$ is a scalar, it is equal to its own transpose, that is $\mathbf{v}' A\mathbf{x}_0 = \mathbf{x}_0' A' \mathbf{v}$ and we have:

$$(D_{\mathbf{y}}f)(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{x}_0' A'.$$

Theorem

Let $\mathbf{h} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *and* $\mathbf{g} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *be two differentiable operators in* $\mathbf{z}_0 \in \mathbb{R}^n$*. We have*

$$D_{\mathbf{z}}(\mathbf{g}(\mathbf{z})'\mathbf{h}(\mathbf{z}))(\mathbf{z}_0) = \mathbf{h}(\mathbf{z}_0)'(D_{\mathbf{z}}\mathbf{g})(\mathbf{z}_0) + \mathbf{g}(\mathbf{z}_0)'(D_{\mathbf{z}}\mathbf{h})(\mathbf{z}_0).$$

Under certain conditions of differentiability of a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, Taylor's formula shows that there exists a polynomial in the variables of the function that provides an approximation of f.
We use symbolic powers of the operator $\mathbf{h}'\nabla$ that act as operators on real-valued functions of n variables. These powers are defined inductively by $(\mathbf{h}'\nabla)^0 f = f$, and

$$(\mathbf{h}'\nabla)^{k+1} f = (\mathbf{h}'\nabla)((\mathbf{h}'\nabla)^k f)$$

for $k \in \mathbb{N}$.

Example

For

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{pmatrix} \text{ and } \mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

we have

$$\mathbf{h}'\nabla = h_1 \frac{\partial}{\partial x_1} + h_2 \frac{\partial}{\partial x_2},$$

and $(\mathbf{h}'\nabla)f = h_1 \frac{\partial f}{\partial x_1} + h_2 \frac{\partial f}{\partial x_2}$.

Example cont'd

Example

Note that

$$\begin{aligned}
(\mathbf{h}'\nabla)^2 f &= (\mathbf{h}'\nabla)\left(h_1\frac{\partial f}{\partial x_1} + h_2\frac{\partial f}{\partial x_2}\right) \\
&= h_1\frac{\partial}{\partial x_1}\left(h_1\frac{\partial f}{\partial x_1} + h_2\frac{\partial f}{\partial x_2}\right) + h_2\frac{\partial}{\partial x_2}\left(h_1\frac{\partial f}{\partial x_1} + h_2\frac{\partial f}{\partial x_2}\right) \\
&= h_1^2\frac{\partial^2 f}{\partial {x_1}^2} + 2h_1 h_2\frac{\partial^2 f}{\partial x_1\,\partial x_2} + h_2^2\frac{\partial^2 f}{\partial {x_2}^2}.
\end{aligned}$$

The m^{th} iteration of the operator $\mathbf{h}'\nabla$ can be computed using the multinomial formula

$$(\mathbf{h}'\nabla)^m = \sum \left\{ \binom{m}{p_1 \; p_2 \; \cdots \; p_k} h_1^{p_1} h_2^{p_2} \cdots h_k^{p_k} \frac{\partial^m}{\partial x_1^{p_1} x_2^{p_2} \partial \cdots x_k^{p_k}} \right. \left| p_1, p_2, \cdots, p_k \in \mathbb{N}, \sum_{i=1}^{k} p_i = m \right\}.$$

Let X be an open subset of $\mathbb{R}^n$. The function $f : X \longrightarrow \mathbb{R}$ belongs to the class $C^k(X)$ if it has continuous partial derivatives of order at most k.

Theorem

(Taylor's Formula) *Let* $f : B(\mathbf{x}, r) \longrightarrow \mathbb{R}$ *be a function that belongs to the class* $C^n(B(\mathbf{x}, r))$*, where* $B(\mathbf{x}, r) \subseteq \mathbb{R}^k$*. If* $\mathbf{h} \in \mathbb{R}^k$ *is such that* $\| \mathbf{h} \| < r$ *then there exists* $\theta \in (0, 1)$ *such that:*

$$f(\mathbf{x} + \mathbf{h}) = \sum_{m=0}^{n-1} \frac{1}{m!}((\mathbf{h}'\nabla)^m f)(\mathbf{x}) + \frac{1}{n!}((\mathbf{h}'\nabla)^n f)(\mathbf{x} + \theta\mathbf{h}).$$

Proof

Define the function $g : (0,1) \longrightarrow \mathbb{R}$ by $g(a) = f(\mathbf{x} + a\mathbf{h})$. Since $f \in C^n(B(\mathbf{x}, r))$ it follows that g belongs to the class $C^n([0,1])$. Thus, by the standard Taylor formula to g there exists $\theta \in (0,1)$ such that

$$g(1) = \sum_{m=0}^{n-1} \frac{1}{m!} g^{(m)}(0) + \frac{1}{n!} g^{(n)}(\theta).$$

Proof (cont'd)

Note that differentiating the function $g(a) = f(\mathbf{x} + a\mathbf{h})$ with respect to a is the same thing as applying the differential operator $\mathbf{h}'\nabla$ to f. Indeed, we have

$$\begin{aligned} g'(a) &= (f(x_1 + ah_1, \ldots, x_n + ah_n)' \\ &= h_1 \frac{\partial f}{\partial x_1}(\mathbf{x} + a\mathbf{h}) + \cdots h_n \frac{\partial f}{\partial x_n}(\mathbf{x} + a\mathbf{h}) \\ &= ((\mathbf{h}'\nabla)f)(\mathbf{x} + a\mathbf{h}). \end{aligned}$$

It follows immediately that

$$g^{(m)} = ((\mathbf{h}'\nabla)^m f)(\mathbf{x} + a\mathbf{h})$$

for $m \in \mathbb{N}$, so there exists $\theta \in (0, 1)$ such that

$$f(\mathbf{x} + \mathbf{h}) = \sum_{m=0}^{n-1} \frac{1}{m!}((\mathbf{h}'\nabla)^m f)(\mathbf{x}) + \frac{1}{n!}((\mathbf{h}'\nabla)^n f)(\mathbf{x} + \theta\mathbf{h}).$$

Example

For the real-valued function $f \in C^1(B(\mathbf{x}, r))$, there is $\theta \in (0, 1)$ such that

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) &= f(\mathbf{x}) + ((\mathbf{h}'\nabla)f)(\mathbf{x}+\theta\mathbf{h}) \\ &= f(\mathbf{x}) + \sum_{i=1}^{k} h_i \frac{\partial f}{\partial x_i}(\mathbf{x}+\theta\mathbf{h}), \end{aligned}$$

where $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}$ and $\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_k \end{pmatrix}$.

Example

For $f \in C^2(B(\mathbf{x}, r))$, the same theorem implies the existence of $\theta \in (0, 1)$ such that

$$
\begin{aligned}
f(\mathbf{x}+\mathbf{h}) &= f(\mathbf{x}) + ((\mathbf{h}'\nabla)f)(\mathbf{x}) + ((\mathbf{h}'\nabla)^2 f)(\mathbf{x}+\theta\mathbf{h}) \\
&= f(\mathbf{x}) + \sum_{i=1}^{k} h_i \frac{\partial f}{\partial x_i}(\mathbf{x}) \\
&\quad + \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} h_{i_1} h_{i_2} \frac{\partial^2 f}{\partial x_{i_1}\, \partial x_{i_2}}(\mathbf{x}+\theta\mathbf{h}).
\end{aligned}
$$

Example cont'd

The matrix-valued function H_f defined by

$$H_f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_{i_1} \, \partial x_{i_2}} \right)$$

is the Hessian matrix of f. Using the Hessian matrix of $f \in C^2(B(\mathbf{x}, r))$ we can write

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + ((\mathbf{h}'\nabla)f)(\mathbf{x}) + \mathbf{h}' H_f(\mathbf{x} + \theta\mathbf{h})\mathbf{h}$$

for some $\theta \in (0, 1)$.