Section 2.4, 2.5

The Derivative and Chain Rule

Learning Objectives:
- Understand the definition and basic properties of the derivative of a vector-valued function of several variables.
- Learn how to use the Chain Rule for functions of several variables.

Keywords: matrix of partial derivatives, the derivative, Chain Rule

The derivative of scalar-valued functions

Let \( f : X \subseteq \mathbb{R}^2 \rightarrow \mathbb{R} \) be a function of two variables. The gradient of \( f \) at \( a \) is the (row) vector

\[
\nabla f(a) = \begin{bmatrix} f_x(a) & f_y(a) \end{bmatrix}
\]

Recall: the linearisation of \( f \) at \( a = (a, b) \in X \) is

\[
L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)
\]

which we may rewrite as

\[
L(x) = f(a) + \nabla f(a)(x - a), \quad x = \begin{bmatrix} x \\ y \end{bmatrix}
\] (1*)

The product here is multiplication of the \( 1 \times 2 \) matrix \( \nabla f(a) \) with the \( 2 \times 1 \) matrix \( x - a \).

Remark:

1. Note the analogy with the equation of a tangent line of the graph of a single variable function:

\[
y = f(a) + f'(a)(x - a).
\]

Thus, the gradient \( \nabla f(x) \) plays a role analogous to the derivative.

2. The above remarks generalise to scalar-valued functions \( f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \), where we define the gradient of \( f \) at \( a \) to be the \( 1 \times n \) row vector

\[
\nabla f(a) = \begin{bmatrix} f_{x_1}(a) & f_{x_2}(a) & \cdots & f_{x_n}(a) \end{bmatrix}
\]

Differentiability of vector-valued functions

Suppose that \( f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m \) is a vector-valued function, \( f(x) = (f_1(x), \ldots, f_m(x)) \), with each \( f_1, \ldots, f_m : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R} \) a scalar-valued function.
Define the matrix of partial derivatives of \( f \) at \( a \in X \), or the Jacobian matrix of \( f \) at \( a \), to be the \( m \times n \) matrix \( Df(a) \) having \( i \)th row \( \nabla f_i(a) \):

\[
Df(a) = \begin{bmatrix}
\frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\
\frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_n}(a) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a)
\end{bmatrix}
\]

We write \( Df(x) \), or simply \( Df \), for the \( m \times n \) matrix whose \( ij \)-entry is \( \frac{\partial f_i}{\partial x_j}(x) \), and call it the Jacobian of \( f \).

**Remark:** If \( f(x) \) is a scalar-valued function then \( Df(x) = \nabla f(x) \); if \( r(t) \) is a path in \( \mathbb{R}^n \) then \( Df_r(t) = r'(t) \) computes the velocity vector of \( r(t) \).

Define the linearisation of \( f \) at \( a \in X \) to be the function

\[
L(x) = f(a) + Df(a)(x - a), \quad x \in \mathbb{R}^n
\]

The product here is multiplication of the \( m \times n \) matrix with the \( n \times 1 \) matrix \( x - a \).

In particular, \( L(x) \in \mathbb{R}^m \).

**Example:** Consider the function

\[
f : \mathbb{R}^2 \to \mathbb{R}^3, \ (x, y) \mapsto (x^2 + y, 2xy, x + y^2)
\]

Then,

\[
Df(x) = \begin{bmatrix}
2x & 1 \\
2y & 2x \\
1 & 2y
\end{bmatrix}
\]

**Differentiability of \( f(x) \)**

Let \( f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m \) be a vector-valued function. We say that \( f \) is differentiable at \( a \in X \) if all partial derivatives \( f_{x_i}(a) \) exist and if

\[
\lim_{x \to a} \frac{f(x) - L(x)}{|x - a|} = 0
\]

If \( f \) is differentiable for every \( a \in X \) then we say that \( f \) is differentiable.

There are analogous results as for the two variable case.

**Sufficient Condition for differentiability**

Let \( f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m, a \in X \). If all partial derivatives \( f_{x_i}(x) \) are continuous nearby to \( a \) then \( f \) is differentiable at \( a \).
Necessary Condition for differentiability

Let \( f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m, \ a \in X \). If \( f \) is differentiable at \( a \) then \( f \) is continuous at \( a \).

Moreover, we can reduce differentiability of vector-valued functions to the differentiability of its component functions

Let \( f : X \subseteq \mathbb{R}^n \to \mathbb{R}^m, f(x) = (f_1(x), \ldots, f_m(x)), \ a \in X. \) If \( f_1, \ldots, f_m \) are differentiable at \( a \) then \( f \) is differentiable at \( a \).

What is the derivative?

Observe the similarity between the linearisation of \( f \) at \( a \)

\[
L(x) = f(a) + Df(a)(x - a)
\]

and function whose graph is the tangent line of a single variable function \( f(x) \):

\[
L(x) = f(a) + f'(a)(x - a)
\]

Define the ‘multiplication by \( Df(a) \)’ linear map

\[
T_{Df(a)} : \mathbb{R}^n \to \mathbb{R}^m, \ x \mapsto Df(a)x
\]

then the linear map \( T_{Df(a)} \) plays the role of the derivative.

The derivative of a vector-valued function \( f \) of several variables is the linear map defined by the Jacobian of \( f \).

Remark: Identifying a linear map with its standard matrix, we will also say that \( Df(a) \) is the derivative of \( f(x) \) at \( x = a \).

The Chain Rule

Recall the Chain Rule for functions of a single variable \( x \): let \( f(x), g(x) \) be differentiable functions defined at \( x = a \). Then,

\[
(f \circ g)'(a) = f'(g(a))g'(a)
\]

In words: the derivative of a composition is an appropriate product of derivatives.

If \( f = Y \subseteq \mathbb{R}^m \to \mathbb{R}^p, \ g : X \subseteq \mathbb{R}^n \to \mathbb{R}^m \) are functions of several variables for which the composition \( f \circ g \) makes sense (i.e. \( g(u) \subseteq Y \), for any \( u \in X \)) then it’s reasonable to expect the following analog of \((*)\):

\[
D(f \circ g)(a) = Df(g(a))Dg(a)
\]

This is the Chain Rule for functions of several variables.

Remark:
1. The product on the right-hand side of (**) is the product of the $p \times m$ matrix $Df(g(a))$ with the $m \times n$ matrix $Dg(a)$.

2. To prove the Chain Rule you need to show an equality of matrices: this means you must show that the $ij$ entry on the LHS equals the $ij$ entry on the RHS. The $ij$ entry on the LHS is $\frac{\partial (f \circ g)}{\partial u_j}(a)$ and the $ij$ entry on the RHS is $\nabla f_i(g(a))(Dg(a))_j$, where $(Dg(a))_j$ is the $j^{th}$ column of $Dg(a)$. That these two quantities are equal now follows from the Chain Rule for single variable functions and the definition of partial derivatives.

Example:

1. Let $f(x, y) = x^2 + 3y^2$, $r(t) = (2t, t^2)$. Then, $f \circ r : \mathbb{R} \to \mathbb{R} : t \mapsto 4t^2 + 3(t^2)^2 = 4t^2 + 3t^4$. In this case, $D(f \circ r)(t)$ is precisely the derivative $(f \circ r)'(t) = 8t + 12t^3$.

   Let’s compute the right-hand side of the Chain Rule: we have
   
   $$Dr(t) = r'(t) = \begin{bmatrix} 2 \\ 2t \end{bmatrix}$$

   $$Df(x) = \nabla f(x) = \begin{bmatrix} 2x \\ 6y \end{bmatrix} \implies Df(r(t)) = \begin{bmatrix} 4t \\ 6t^2 \end{bmatrix}$$

   Hence,
   
   $$Df(r(t))Dr(t) = \begin{bmatrix} 4t \\ 6t^2 \end{bmatrix} \begin{bmatrix} 2 \\ 2t \end{bmatrix} = 8t + 12t^3$$

2. Let $h(x, y, z) = x + yz$ and

   $$f : \mathbb{R}^2 \to \mathbb{R}^3 : (x, y) \mapsto (x^2 + y, 2xy, x + y^2)$$

   Then,
   
   $$(h \circ f)(x, y, z) = h(x^2 + y, 2xy, x + y^2)$$
   $$= (x^2 + y) + (2xy)(x + y^2)$$
   $$= x^2 + y + 2x^2y + 2xy^3$$

   Hence,
   
   $$D(h \circ f)(x) = \nabla(h \circ f)(x) = \begin{bmatrix} 2x + 4xy + 2y^3 \\ 1 + 2x^2 + 6xy^2 \end{bmatrix}$$

   Computing the right-hand side of (**):
   
   $$Dh(x) = \nabla h(x) = \begin{bmatrix} 1 \\ z \\ y \end{bmatrix} \implies Dh(f(x)) = \begin{bmatrix} 1 \\ x + y^2 \\ 2xy \end{bmatrix}$$

   and
   
   $$Df(x) = \begin{bmatrix} 2x \\ 2y \\ 2x \\ 1 \\ 2y \end{bmatrix}$$

   Then,
   
   $$Dh(f(x))Df(x) = \begin{bmatrix} 1 \\ x + y^2 \\ 2xy \end{bmatrix} \begin{bmatrix} 2x \\ 2y \\ 2x \\ 1 \\ 2y \end{bmatrix} = \begin{bmatrix} 2x + 4xy + 2y^3 \\ 1 + 2x^2 + 6xy^2 \end{bmatrix}$$
3. Let
\[ f : \mathbb{R}^2 \to \mathbb{R}^3, \quad (x, y) \mapsto (x^2 + y, 2xy, x + y^2) \]
\[ g : \mathbb{R}^3 \to \mathbb{R}^2, \quad (u, v, w) \mapsto (u^2 + v, 3w - u) \]

Then,
\[ f \circ g(u) = ((u^2 + v)^2 + 3w - u, 2(u^2 + v)(3w - u), u^2 + v + (3w - u)^2) \]

and
\[ D(f \circ g)(u) = \begin{bmatrix} 4u^3 + 4uv - 1 & 2u^2 + 2v & 3 \\ 12uw - 6u^2 - 2v & 6w - 2u & 6v + 6u^2 \\ 4u - 6w & 1 & 18w - 6u \end{bmatrix} \]

Computing the righthand side of (\(**\)):
\[ Df(x) = \begin{bmatrix} 2x & 1 \\ 2y & 2x \\ 1 & 2y \end{bmatrix} \quad \Rightarrow \quad Df(g(u)) = \begin{bmatrix} 2(u^2 + v) & 1 \\ 2(3w - u) & 2(u^2 + v) \\ 1 & 2(3w - u) \end{bmatrix} \]

and
\[ Dg(u) = \begin{bmatrix} 2u & 1 & 0 \\ -1 & 0 & 3 \end{bmatrix} \]

Hence,
\[ Df(g(u))Dg(u) = \begin{bmatrix} 2(u^2 + v) & 1 \\ 2(3w - u) & 2(u^2 + v) \\ 1 & 2(3w - u) \end{bmatrix} \begin{bmatrix} 2u & 1 & 0 \\ -1 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 4u^3 + 4uv - 1 & 2u^2 + 2v & 3 \\ 12uw - 6u^2 - 2v & 6w - 2u & 6v + 6u^2 \\ 4u - 6w & 1 & 18w - 6u \end{bmatrix} \]