Contact: `gmelvin@middlebury.edu`

April 2 Lecture

Textbook Reference:
- *Vector Calculus*, Colley, 4th Edition: §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 $f : X \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables. The **gradient of** f **at** $\underline{a}$ is the (row) vector

$$\nabla f(\underline{a}) = \begin{bmatrix} f_x(\underline{a}) & f_y(\underline{a}) \end{bmatrix}$$

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

$$L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

which we may rewrite as

$$L(\underline{x}) = f(\underline{a}) + \nabla f(\underline{a})(\underline{x} - \underline{a}), \qquad \underline{x} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{1*}$$

The product here is muliplication of the 1×2 matrix $\nabla f(\underline{a})$ with the 2×1 matrix $\underline{x} - \underline{a}$.

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 $\nabla f(\underline{x})$ plays a role analogous to the derivative.

2. The above remarks generalise to scalar-valed functions $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$, where we define **the gradient of** f **at** $\underline{a}$ to be the $1 \times n$ row vector
$$\nabla f(\underline{a}) = \begin{bmatrix} f_{x_1}(\underline{a}) & f_{x_2}(\underline{a}) & \cdots & f_{x_n}(\underline{a}) \end{bmatrix}$$

Differentiability of vector-valued functions

Suppose that $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is a **vector-valued** function, $\mathbf{f}(\underline{x}) = (f_1(\underline{x}), \ldots, f_m(\underline{x}))$, with each $f_1, \ldots, f_m : X \subseteq \mathbb{R}^n \to \mathbb{R}$ a scalar-valued function.

Goal: we want to define what it means for $\mathbf{f}(\underline{x})$ to be differentiable at a point.

Define **the matrix of partial derivatives of f at** $\underline{a} \in X$, or **the Jacobian matrix of f at** $\underline{a}$, to be the $m \times n$ matrix $D\mathbf{f}(\underline{a})$ having i^{th} row $\nabla f_i(\underline{a})$:

$$D\mathbf{f}(\underline{a}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\underline{a}) & \frac{\partial f_1}{\partial x_2}(\underline{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\underline{a}) \\ \frac{\partial f_2}{\partial x_1}(\underline{a}) & \frac{\partial f_2}{\partial x_2}(\underline{a}) & \cdots & \frac{\partial f_2}{\partial x_n}(\underline{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\underline{a}) & \frac{\partial f_m}{\partial x_2}(\underline{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\underline{a}) \end{bmatrix}$$

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

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

Define the **linearisation of f at** $\underline{a} \in X$ to be the function

$$\mathbf{L}(\underline{x}) = \mathbf{f}(\underline{a}) + D\mathbf{f}(\underline{a})(\underline{x} - \underline{a}), \quad \underline{x} \in \mathbb{R}^n$$

The product here is multiplication of the $m \times n$ matrix with the $n \times 1$ matrix $\underline{x} - \underline{a}$. In particular, $\mathbf{L}(\underline{x}) \in \mathbb{R}^m$.

Example: Consider the function

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

Then,

$$D\mathbf{f}(\underline{x}) = \begin{bmatrix} 2x & 1 \\ 2y & 2x \\ 1 & 2y \end{bmatrix}$$

Differentiability of $\mathbf{f}(\underline{x})$

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

$$\lim_{\underline{x} \to \underline{a}} \frac{\mathbf{f}(\underline{x}) - \mathbf{L}(\underline{x})}{|\underline{x} - \underline{a}|} = 0$$

If **f** is differentiable for every $\underline{a} \in X$ then we say that **f is differentiable**.

There are analogous results as for the two variable case.

Sufficient Condition for differentiability

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

Necessary Condition for differentiability

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

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

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

What is the derivative?

Observe the similarity between the linearisation of $\mathbf{f}$ at $\underline{a}$

$$\mathbf{L}(\underline{x}) = \mathbf{f}(\underline{a}) + D\mathbf{f}(\underline{a})(\underline{x} - \underline{a})$$

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* $D\mathbf{f}(\underline{a})$' linear map

$$T_{D\mathbf{f}(\underline{a})} : \mathbb{R}^n \to \mathbb{R}^m \ , \ \underline{x} \mapsto D\mathbf{f}(\underline{a})\underline{x}$$

then the linear map $T_{D\mathbf{f}(\underline{a})}$ 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 $D\mathbf{f}(\underline{a})$ is the derivative of $\mathbf{f}(\underline{x})$ at $\underline{x} = \underline{a}$.

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,

$$(f \circ g)'(a) = f'(g(a))g'(a) \qquad (*)$$

In words: **the derivative of a composition is an <u>appropriate</u> product of derivatives**.

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

$$D(\mathbf{f} \circ \mathbf{g})(\underline{a}) = D\mathbf{f}(\mathbf{g}(\underline{a}))D\mathbf{g}(\underline{a}) \qquad (**)$$

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 $D\mathbf{f}(\mathbf{g}(\underline{a}))$ with the $m \times n$ matrix $D\mathbf{g}(\underline{a})$.

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 $\frac{\partial(f_i \circ \mathbf{g})}{\partial u_j}(\underline{a})$ and the ij entry on the RHS is $\nabla f_i(\mathbf{g}(\underline{a}))(D\mathbf{g}(\underline{a}))_j$, where $(D\mathbf{g}(\underline{a}))_j$ is the j^{th} column of $D\mathbf{g}(\underline{a})$. 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 $f(x, y) = x^2 + 3y^2$, $\underline{r}(t) = (2t, t^2)$. Then,

$$f \circ \underline{r} : \mathbb{R} \to \mathbb{R} \ , \ t \mapsto (2t)^2 + 3(t^2)^2 = 4t^2 + 3t^4$$

   In this case, $D(f \circ \underline{r})(t)$ is precisely the derivative $(f \circ \underline{r})'(t) = 8t + 12t^3$.

   Let's compute the right-hand side of the Chain Rule: we have

$$D\underline{r}(t) = \underline{r}'(t) = \begin{bmatrix} 2 \\ 2t \end{bmatrix}$$

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

   Hence,

$$Df(\underline{r}(t))D\underline{r}(t) = \begin{bmatrix} 4t & 6t^2 \end{bmatrix} \begin{bmatrix} 2 \\ 2t \end{bmatrix} = 8t + 12t^3$$

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

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

   Then,

$$\begin{aligned} (h \circ \mathbf{f})(x, y, z) &= h(x^2 + y, 2xy, x + y^2) \\ &= (x^2 + y) + (2xy)(x + y^2) \\ &= x^2 + y + 2x^2 y + 2xy^3 \end{aligned}$$

   Hence,

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

   Computing the right-hand side of (**):

$$Dh(\underline{x}) = \nabla h(\underline{x})) = \begin{bmatrix} 1 & z & y \end{bmatrix} \quad \Longrightarrow \quad Dh(\mathbf{f}(\underline{x})) = \begin{bmatrix} 1 & x + y^2 & 2xy \end{bmatrix}$$

   and

$$D\mathbf{f}(\underline{x}) = \begin{bmatrix} 2x & 1 \\ 2y & 2x \\ 1 & 2y \end{bmatrix}$$

   Then,

$$Dh(\mathbf{f}(\underline{x}))D\mathbf{f}(\underline{x}) = \begin{bmatrix} 1 & x + y^2 & 2xy \end{bmatrix} \begin{bmatrix} 2x & 1 \\ 2y & 2x \\ 1 & 2y \end{bmatrix} = \begin{bmatrix} 2x + 4xy + 2y^3 & 1 + 2x^2 + 6xy^2 \end{bmatrix}$$

3. Let

$$\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3 \ , \ (x, y) \mapsto (x^2 + y, 2xy, x + y^2)$$
$$\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^2 \ , \ (u, v, w) \mapsto (u^2 + v, 3w - u)$$

Then,

$$\mathbf{f} \circ \mathbf{g}(\underline{u}) = ((u^2 + v)^2 + 3w - u, 2(u^2 + v)(3w - u), u^2 + v + (3w - u)^2)$$

and

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

Computing the righ-hand side of (∗∗):

$$D\mathbf{f}(\underline{x}) = \begin{bmatrix} 2x & 1 \\ 2y & 2x \\ 1 & 2y \end{bmatrix} \quad \Longrightarrow \quad D\mathbf{f}(\mathbf{g}(\underline{u})) = \begin{bmatrix} 2(u^2 + v) & 1 \\ 2(3w - u) & 2(u^2 + v) \\ 1 & 2(3w - u) \end{bmatrix}$$

$$\text{and} \qquad D\mathbf{g}(\underline{u}) = \begin{bmatrix} 2u & 1 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

Hence,

$$D\mathbf{f}(\mathbf{g}(\underline{u}))D\mathbf{g}(\underline{u}) = \begin{bmatrix} 2(u^2 + v) & 1 \\ 2(3w - u) & 2(u^2 + v) \\ 1 & 2(3w - u) \end{bmatrix} \begin{bmatrix} 2u & 1 & 0 \\ -1 & 0 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 4u^3 + 4uv - 1 & 2u^2 + 2v & 3 \\ 12uw - 6u^2 - 2v & 6w - 2u & 6v + 6u^2 \\ 4u - 6w & 1 & 18w - 6u \end{bmatrix}$$