Lecture 4: Linear approximation and the chain rule

Learning goals:

1. How is the linear approximation used?
2. How do derivation calculation rules generalise to multivariable functions?
3. In particular, how does the chain rule work for multivariable functions?

Where to find the material?

Guichard et friends 14.4 (does not contain linear approximation)
Active Calculus 10.4, 10.5
Adams-Essex 13.5, 13.6

Before we move on few remarks about differentiability and partial derivatives.

The existence of partial derivatives does not even guarantee continuity

The existence of partial derivatives is a weaker condition than differentiability:

1. $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = \begin{cases} 0, & \text{when } x = 0 \quad \text{or} \quad y = 0 \\ 1, & \text{elsewhere} \end{cases}$$

   both partial derivatives are at the origin, but the function is not continuous at the origin (This function is in Round 2 exercises)

2. $f : \mathbb{R}^2 \to \mathbb{R}$

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

   In lecture 2 it was calculated that this does not have limit at the origin, so it is not continuous at the origin, but still both partial derivatives exists at the origin.

If the partial derivatives are continuous, then...

- If all partial derivatives are continuous around the point x_0, then the function itself is differentiable at the point x_0 and the Jacobian matrix can be calculated using the partial derivatives
- almost all functions discussed in this course are of this type

Linear approximation when $f\colon \mathbb{R}^2 \to \mathbb{R}$

- Consider a function $f\colon \mathbb{R}^2 \to \mathbb{R}$, all partial derivatives of which are continuous.
- The Jacobian matrix of this is

$$Df(\mathbf{x}) = \begin{bmatrix} \partial_1 f(\mathbf{x}) & \partial_2 f(\mathbf{x}) \end{bmatrix}$$

- Thus the linear approximation near the point $\mathbf{x}$ (i.e. $\mathbf{h}$ is small = near the origin)

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{h}) + \phi(\mathbf{h})$$

- $\Rightarrow$

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) &\approx f(\mathbf{x}) + Df(\mathbf{x})\mathbf{h} \\ &= f(\mathbf{x}) + \begin{bmatrix} \partial_1 f(\mathbf{x}) & \partial_2 f(\mathbf{x}) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \\ &= f(\mathbf{x}) + \partial_1 f(\mathbf{x}) h_1 + \partial_2 f(\mathbf{x}) h_2 \end{aligned}$$

Example of linear approximation when $f\colon \mathbb{R}^2 \to \mathbb{R}$

Task: Find an approximation for the value of the function $f(x, y) = \sqrt{2x^2 + e^{2y}}$ at the point $(2.2, -0.2)$ without using calculator.
Solution: Now $f(2, 0) = 3$ and $\mathbf{h} = (0.2, -0.2)$. The partial derivatives of the function are

$$\partial_1 f(x, y) = \frac{2x}{\sqrt{2x^2 + e^{2y}}}, \quad \partial_1 f(2, 0) = \frac{4}{3}.$$

$$\partial_2 f(x, y) = \frac{e^{2y}}{\sqrt{2x^2 + e^{2y}}}, \quad \partial_2 f(2, 0) = \frac{1}{3}.$$

Hence the desired approximation using linear approximation is

$$f(2.2, -0.2) \approx 3 + \frac{4}{3}0.2 + \frac{1}{3}(-0.2) = 3.2.$$

Comparison: Using calculator we find that the value of the function at the point $(2.2, -0.2)$ is about 3.2172.

Linear approximation when $f\colon \mathbb{R}^3 \to \mathbb{R}$

Similarly if $f\colon \mathbb{R}^3 \to \mathbb{R}$, all partial derivatives of which are continuous. The Jacobian matrix is

$$Df(\mathbf{x}) = \begin{bmatrix} \partial_1 f(\mathbf{x}) & \partial_2 f(\mathbf{x}) & \partial_3 f(\mathbf{x}) \end{bmatrix}$$

Linear approximation near the point $\mathbf{x}$ (i.e. $\mathbf{h}$ is small = near the origin)

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{h}) + \phi(\mathbf{h})$$

$\Rightarrow$

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) &\approx f(\mathbf{x}) + Df(\mathbf{x})\mathbf{h} \\ &= f(\mathbf{x}) + \begin{bmatrix} \partial_1 f(\mathbf{x}) & \partial_2 f(\mathbf{x}) & \partial_3 f(\mathbf{x}) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \\ &= f(\mathbf{x}) + \partial_1 f(\mathbf{x}) h_1 + \partial_2 f(\mathbf{x}) h_2 + \partial_3 f(\mathbf{x}) h_3 \end{aligned}$$

Another look to the linear approximation - differentials

- The quantity $Df(\mathbf{x})\mathbf{h}$ is called the **differential** of f, and denoted by df
- It approximates the change of the function $\Delta f = f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})$
- **Common notation:** if $f : \mathbb{R}^2 \to \mathbb{R}$, then $df = f_x dx + f_y dy$

Example

Estimate the percentage change in the period $T = 2\pi\sqrt{\frac{L}{g}}$ of a simple pendulum if the length, L, of the pendulum increases by 2% and the acceleration of gravity, g, decreases by 0.6%.

- So we need to estimate $\frac{\Delta T}{T}$
- With differential $\Delta T \approx dT = \frac{\partial T}{\partial L} dL + \frac{\partial T}{\partial g} dg$
- $dL = 0.02L$ and $dg = -0.006g$
- Do all the needed calculations
- Answer: the period T of the pendulum increases approximately 1.3%

Calculation rules for derivatives

Just as for the derivative of a single-variable function, so for a Jacobian matrix:

Basic claculation rules

$f, g\colon \mathbb{R}^n \to \mathbb{R}^m$.

- $D(f+g)(\mathbf{x}) = Df(\mathbf{x}) + Dg(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$
- $D(cf)(\mathbf{x}) = cDf(\mathbf{x})$ for all $c \in \mathbb{R}$ and all $\mathbf{x} \in \mathbb{R}^n$

Thus the derivation is a linear operation.

Chain rule

The chain rule also holds: $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^m \to \mathbb{R}^k$

$$D(g \circ f)(\mathbf{x}) = Dg(f(\mathbf{x}))Df(\mathbf{x}) \qquad \text{Note! Matrix multiplication}$$

The idea of the proof for the chain rule 1/2

$f\colon \mathbb{R}^n \to \mathbb{R}^m$ and $g\colon \mathbb{R}^m \to \mathbb{R}^k$ such that f is differentiable at the point $\mathbf{x}$ and g is differentiable at the point $y = f(\mathbf{x})$

- By the definition of differentiability, we find a linear map $T_f\colon \mathbb{R}^n \to \mathbb{R}^m$ and a function $\phi_f\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + T_f(\mathbf{h}) + \phi_f(\mathbf{h}) \quad \text{and} \quad \frac{\phi_f(\mathbf{h})}{\|\mathbf{h}\|} to \mathbf{0} \text{ when } \mathbf{h} \to \mathbf{0}.$$

- Similarly, we find a linear map $T_g\colon \mathbb{R}^m \to \mathbb{R}^k$ and a function $\phi_g\colon \mathbb{R}^m \to \mathbb{R}^k$ such that

$$g(\mathbf{y}+\mathbf{h}) = g(\mathbf{y}) + T_g(\mathbf{h}) + \phi_g(\mathbf{h}) \quad \text{ja} \quad \frac{\phi_g(\mathbf{h})}{\|\mathbf{h}\|} \to \mathbf{0} \text{ kun } \mathbf{h} \to \mathbf{0}.$$

The idea of the proof for the chain rule 2/2

$$
\begin{aligned}
g \circ f(\mathbf{x}+\mathbf{h}) &= g(f(\mathbf{x}+\mathbf{h})) = g(f(\mathbf{x}) + T_f(\mathbf{h}) + \phi_f(\mathbf{h})) \\
&= g(f(\mathbf{x}) + \mathbf{h}_1) = g(f(\mathbf{x})) + T_g(\mathbf{h}_1) + \phi_g(\mathbf{h}_1) \\
&= g \circ f(\mathbf{x}) + T_g(T_f(\mathbf{h}) + \phi_f(\mathbf{h})) + \phi_g(\mathbf{h}_1) \\
&= g \circ f(\mathbf{x}) + T_g(T_f(\mathbf{h})) + T_g(\phi_f(\mathbf{h})) + \phi_g(\mathbf{h}_1) \\
&= g \circ f(\mathbf{x}) + T_g(T_f(\mathbf{h})) + \phi_{g\circ f}(\mathbf{h})
\end{aligned}
$$

It can be shown that $\frac{\phi_{g\circ f}(\mathbf{h})}{\|\mathbf{h}\|} \to 0$ when $\mathbf{h} \to \mathbf{0}$.

Thus, the derivative of the composition function is obtained by taking the composition of linear maps $T_g \circ T_f$. As matrices $T_g = Dg(f(\mathbf{x}))$ and $T_f = Df(\mathbf{x})$ Written in matrices (and remembering that the multiplication of matrices gives the matrix of the composition of linear maps) we get

$$D(f \circ g)(\mathbf{x}) = Dg(f(\mathbf{x}))Df(\mathbf{x})$$

Chain rule - example 1

Let $f\colon \mathbb{R} \to \mathbb{R}^2$, $f(t) = (2t, t^2)$ and $g\colon \mathbb{R}^2 \to \mathbb{R}$, $g(x, y) = x^2 + y^2$.

Calculate what is $D(g \circ f)(t) \quad (= (g \circ f)'(t))$

a) by substitutions

b) by the chain rule

$$\begin{aligned} D(f \circ g)(t) &= Dg(f(x))Df(t) \\ &= \begin{bmatrix} \partial_1 g(f(t)) & \partial_2 g(f(t)) \end{bmatrix} \begin{bmatrix} f_1'(t) \\ f_2'(t) \end{bmatrix} \\ &= \begin{bmatrix} 2 \cdot (2t) & 2 \cdot (t^2) \end{bmatrix} \begin{bmatrix} 2 \\ 2t \end{bmatrix} = 8t + 4t^3 \end{aligned}$$

Chain rule - example 1 (other notations)

Let $f\colon \mathbb{R} \to \mathbb{R}^2$, $f(t) = (2t, t^2)$ and $g\colon \mathbb{R}^2 \to \mathbb{R}$, $g(x, y) = x^2 + y^2$.

Let's write $z = g(x, y) = x^2 + y^2$, where $x(t) = 2t$ and $y(t) = t^2$.
Calculate what is $\frac{dz}{dt}$.
Calculated above

$$\begin{aligned} D(f \circ g)(t) &= Dg(f(t))Df(t) \\ &= \begin{bmatrix} \partial_1 g(f(t)) & \partial_2 g(f(t)) \end{bmatrix} \begin{bmatrix} f_1'(t) \\ f_2'(t) \end{bmatrix} \\ &= \partial_1 g(f(t)) f_1'(t) + \partial_2 g(f(x)) f_2'(t) \end{aligned}$$

Using the other notations we obtain:

$$\frac{dz}{dt} = \frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt}$$

Chain rule - example 2

If $z = \sin(x^2 y)$, where $x = st^2$ and $y = s^2 + \frac{1}{t}$, find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$ using the chain rule.