§12.5: Chain Rule

Recall the chain rule for single-variable functions:

$$\frac{d}{dt}f(x(t)) = f'(x(t))x'(t), \quad \text{i.e. } \frac{df}{dt} = \frac{df}{dx}\frac{dx}{dt}.$$

Here's an informal way to understand the chain rule.

The linearisation of f says:
$$f(x+\delta x) \approx f(x) + f'(x)\delta x. \qquad (*)$$

Write $x+\delta x$ for $x(t+\delta t)$. Using the linearisation of x:

$$\begin{aligned} x+\delta x = x(t+\delta t) &\approx x(t) + x'(t)\delta t \\ \delta x &\approx x'(t)\delta t \end{aligned}$$

Substituting into $(*)$:

$$f(x(t+\delta t)) \approx f(x(t)) + \boxed{f'(x(t))x'(t)}\,\delta t.$$

Compare the above to the linearisation of the composite function $f(x(t))$:

$$f(x(t+\delta t)) \approx f(x(t)) + \boxed{\frac{d}{dt}f(x(t))}\,\delta t.$$

So the quantities in the blue rectangles should be the same.

Now we derive a simple example of a multivariate chain rule in the same way.

Imagine that you are walking on $\mathbb{R}^2$, and your position at time t is $(x(t), y(t))$. The temperature at the point (x, y) is $f(x, y)$. So the temperature that you feel at time t is the composite function $f(x(t), y(t))$. What is $\frac{d}{dt}f((x(t), y(t))$, the rate of change of temperature that you feel?

Now we derive a simple example of a multivariate chain rule in the same way.

Imagine that you are walking on $\mathbb{R}^2$, and your position at time t is $(x(t), y(t))$. The temperature at the point (x, y) is $f(x, y)$. So the temperature that you feel at time t is the composite function $f(x(t), y(t))$. What is $\frac{d}{dt} f((x(t), y(t))$, the rate of change of temperature that you feel?

The linearisation of the temperature function is

$$f(x + \delta x, y + \delta y) \approx f(x, y) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y.$$

And the linearisations of x and y tell us that

$$\delta x \approx \frac{df}{dx}\delta t; \quad \delta y \approx \frac{df}{dy}\delta t.$$

Substituting into $(*)$

$$f(x(t + \delta t), y(t + \delta t)) \approx f(x, y) + \frac{\partial f}{\partial x}\frac{df}{dx}\delta t + \frac{\partial f}{\partial y}\frac{df}{dy}\delta t.$$

Now we derive a simple example of a multivariate chain rule in the same way.

Imagine that you are walking on $\mathbb{R}^2$, and your position at time t is $(x(t), y(t))$. The temperature at the point (x, y) is $f(x, y)$. So the temperature that you feel at time t is the composite function $f(x(t), y(t))$. What is $\frac{d}{dt} f((x(t), y(t))$, the rate of change of temperature that you feel?

The linearisation of the temperature function is

$$f(x + \delta x, y + \delta y) \approx f(x, y) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y.$$

And the linearisations of x and y tell us that

$$\delta x \approx \frac{df}{dx}\delta t; \quad \delta y \approx \frac{df}{dy}\delta t.$$

Substituting into $(*)$

$$f(x(t + \delta t), y(t + \delta t)) \approx f(x, y) + \frac{\partial f}{\partial x}\frac{df}{dx}\delta t + \frac{\partial f}{\partial y}\frac{df}{dy}\delta t.$$

Comparing with the linearisation of $f(x(t), y(t))$:

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Now we derive a simple example of a multivariate chain rule in the same way.

Imagine that you are walking on $\mathbb{R}^2$, and your position at time t is $(x(t), y(t))$. The temperature at the point (x, y) is $f(x, y)$. So the temperature that you feel at time t is the composite function $f(x(t), y(t))$. What is $\frac{d}{dt} f((x(t), y(t))$, the rate of change of temperature that you feel?

The linearisation of the temperature function is

$$f(x + \delta x, y + \delta y) \approx f(x, y) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y.$$

And the linearisations of x and y tell us that

$$\delta x \approx \frac{df}{dx}\delta t; \quad \delta y \approx \frac{df}{dy}\delta t.$$

Substituting into $(*)$

$$f(x(t + \delta t), y(t + \delta t)) \approx f(x, y) + \frac{\partial f}{\partial x}\frac{df}{dx}\delta t + \frac{\partial f}{\partial y}\frac{df}{dy}\delta t.$$

Comparing with the linearisation of $f(x(t), y(t))$:

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

(This is not a rigorous proof because we haven't checked that the errors are small enough. We sketch a rigorous and more general version of this argument on p10. For a different rigorous proof, see the first page of §12.5 in the textbook.)

Example: Let $f(x, y) = xy^2$, and $x = \ln t, y = 3t^2$.
Find $\dfrac{df}{dt}$.

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

We showed that, if $f(x, y)$ is a 2-variable function, and x and y are functions of t, then

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Now suppose x, y are multivariate functions, e.g. $x(s, t), y(s, t)$.

We showed that, if $f(x,y)$ is a 2-variable function, and x and y are functions of t, then

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Now suppose x, y are multivariate functions, e.g. $x(s,t), y(s,t)$.

To find $\frac{\partial f}{\partial t}$, we treat s as a constant throughout, so

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t};$$

And similarly:

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}.$$

Example: Let $f(x,y) = xy^2$, and $x = \ln(s+t), y = 3t^2 \cos s$. Find $\dfrac{\partial f}{\partial s}$ and $\dfrac{\partial f}{\partial s}(0,1)$.

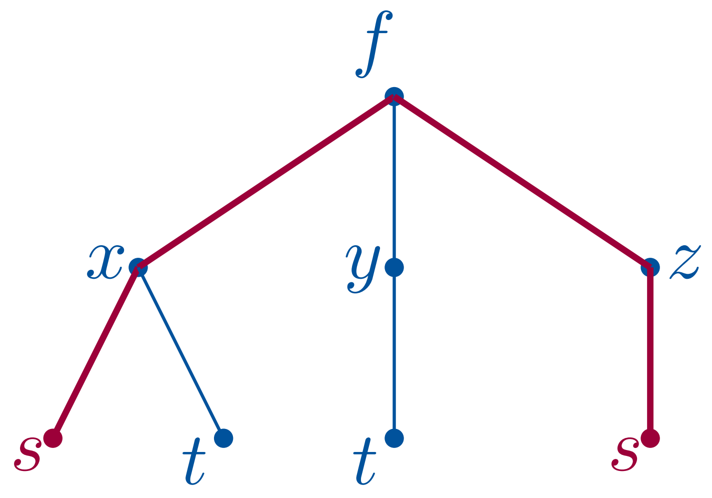
$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}.$$

In ex. sheet #15 Q1, we are given $f(x, y, z)$ and $x(s, t) = e^{st}$, $y(s, t) = t^2$, $z(s, t) = s^2 + 1$. The chain rule says

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial s},$$

but the second term (in y) is unnecessary because y does not depend on s. To simplify things in such cases, we can draw a dependency chart showing which functions depend on which variables. Then the terms in the chain rule for $\frac{\partial f}{\partial s}$ correspond to all the paths from s to f.

Dependency charts can be really useful when there are many variables, or when dealing with a triple composition (e.g. if s and t here are functions of u, v, w).



As in the 1D case, we can compute higher order derivatives of composite functions by applying the chain rule repeatedly.

Example: Let $f(x, y)$ be a two variable function, and $x = 2s + 3t, y = st$. Find an expression for $\frac{\partial^2}{\partial s \partial t} f(x(s, t), y(s, t))$ in terms of the partial derivatives of f.

The chain rule in terms of Jacobian matrices and the derivative linear transformation

Remember from p4 that, for $f(x, y)$, $x(s, t)$, $y(s, t)$, we have

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}; \qquad \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

The chain rule in terms of Jacobian matrices and the derivative linear transformation

Remember from p4 that, for $f(x, y)$, $x(s, t)$, $y(s, t)$, we have

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}; \qquad \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}.$$

In the notation of Jacobian matrices, we have

$$Df(s,t) = \begin{pmatrix} \frac{\partial f}{\partial s} & \frac{\partial f}{\partial t} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{pmatrix} = Df(x(s,t), y(s,t))D\mathbf{g}(s,t),$$

writing $\mathbf{g}(s, t)$ for $(x(s, t), y(s, t))$ (i.e. $g_1 = x$ and $g_2 = y$).

In general, the Jacobian matrix of a composite function is the matrix product of the Jacobian matrices

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{t}) = D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t}).$$

Because the product of matrices correspond to the composition of linear transformations, this says that the derivative of a composition is a composition of the derivatives.

Example: Let $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^3$ be a function such that

$\mathbf{g}(1,2) = (1,2,1)$ and $D\mathbf{g}(1,2) = \begin{pmatrix} 1/2 & 1/2 \\ 2 & 1 \\ 1 & 0 \end{pmatrix}$.

Let $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $\mathbf{f}(x,y) = (x^2 e^y, y^2 z)$.
Find $D(\mathbf{f} \circ \mathbf{g})(1,2)$.

$$D(\mathbf{f}\circ\mathbf{g})(\mathbf{t}) = D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t}).$$

Non-examinable: the proof of the chain rule

The main idea is the linearisation argument on pp1-2. We will show carefully that the errors in the linearisation are small compared to $|\delta\mathbf{t}|$, as required in the definition of the derivative.

We wish to show that $D(\mathbf{f}\circ\mathbf{g})(\mathbf{t}) = D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t})$. So we need to show that $D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t})$ satisfies the definition of the derivative $D(\mathbf{f}\circ\mathbf{g})$, i.e.

$$\frac{(\mathbf{f}\circ\mathbf{g})(\mathbf{t}+\delta\mathbf{t}) - (\mathbf{f}\circ\mathbf{g})(\mathbf{t}) - [D\mathbf{f}(\mathbf{g}(\mathbf{t}))][D\mathbf{g}(\mathbf{t})]\delta\mathbf{t}}{|\delta\mathbf{t}|} \to 0 \text{ as } \delta\mathbf{t} \to \mathbf{0}.$$

Non-examinable: the proof of the chain rule

The main idea is the linearisation argument on pp1-2. We will show carefully that the errors in the linearisation are small compared to $|\delta\mathbf{t}|$, as required in the definition of the derivative.

We wish to show that $D(\mathbf{f}\circ\mathbf{g})(\mathbf{t}) = D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t})$. So we need to show that $D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t})$ satisfies the definition of the derivative $D(\mathbf{f}\circ\mathbf{g})$, i.e.

$$\frac{(\mathbf{f}\circ\mathbf{g})(\mathbf{t}+\delta\mathbf{t}) - (\mathbf{f}\circ\mathbf{g})(\mathbf{t}) - [D\mathbf{f}(\mathbf{g}(\mathbf{t}))][D\mathbf{g}(\mathbf{t})]\delta\mathbf{t}}{|\delta\mathbf{t}|} \to 0 \text{ as } \delta\mathbf{t}\to\mathbf{0}.$$

Let $\mathbf{x}=\mathbf{g}(\mathbf{t})$ and $\mathbf{x}+\delta\mathbf{x}=\mathbf{g}(\mathbf{t}+\delta\mathbf{t})$, and rewrite the expression above as

$$\frac{\mathbf{f}(\mathbf{g}(\mathbf{t}+\delta\mathbf{t})) - \mathbf{f}(\mathbf{g}(\mathbf{t})) - [D\mathbf{f}(\mathbf{g}(\mathbf{t}))]\delta\mathbf{x}}{|\delta\mathbf{t}|} + \frac{[D\mathbf{f}(\mathbf{g}(\mathbf{t}))]\delta\mathbf{x} - [D\mathbf{f}(\mathbf{g}(\mathbf{t}))][D\mathbf{g}(\mathbf{t})]\delta\mathbf{t}}{|\delta\mathbf{t}|}$$

$$=\frac{\mathbf{f}(\mathbf{x}+\delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - [D\mathbf{f}(\mathbf{x})]\delta\mathbf{x}}{|\delta\mathbf{x}|}\frac{|\delta\mathbf{x}|}{|\delta\mathbf{t}|} + [D\mathbf{f}(\mathbf{g}(\mathbf{t}))]\left(\frac{\delta\mathbf{x} - [D\mathbf{g}(\mathbf{t})]\delta\mathbf{t}}{|\delta\mathbf{t}|}\right)$$

Non-examinable: the proof of the chain rule

The main idea is the linearisation argument on pp1-2. We will show carefully that the errors in the linearisation are small compared to $|\delta\mathbf{t}|$, as required in the definition of the derivative.

We wish to show that $D(\mathbf{f}\circ\mathbf{g})(\mathbf{t}) = D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t})$. So we need to show that $D\mathbf{f}(\mathbf{g}(\mathbf{t}))D\mathbf{g}(\mathbf{t})$ satisfies the definition of the derivative $D(\mathbf{f}\circ\mathbf{g})$, i.e.

$$\frac{(\mathbf{f}\circ\mathbf{g})(\mathbf{t}+\delta\mathbf{t}) - (\mathbf{f}\circ\mathbf{g})(\mathbf{t}) - [D\mathbf{f}(\mathbf{g}(\mathbf{t}))][D\mathbf{g}(\mathbf{t})]\delta\mathbf{t}}{|\delta\mathbf{t}|} \to 0 \text{ as } \delta\mathbf{t}\to\mathbf{0}.$$

Let $\mathbf{x} = \mathbf{g}(\mathbf{t})$ and $\mathbf{x}+\delta\mathbf{x} = \mathbf{g}(\mathbf{t}+\delta\mathbf{t})$, and rewrite the expression above as

$$= \frac{\mathbf{f}(\mathbf{g}(\mathbf{t}+\delta\mathbf{t})) - \mathbf{f}(\mathbf{g}(\mathbf{t})) - [D\mathbf{f}(\mathbf{g}(\mathbf{t}))]\delta\mathbf{x}}{|\delta\mathbf{t}|} + \frac{[D\mathbf{f}(\mathbf{g}(\mathbf{t}))]\delta\mathbf{x} - [D\mathbf{f}(\mathbf{g}(\mathbf{t}))][D\mathbf{g}(\mathbf{t})]\delta\mathbf{t}}{|\delta\mathbf{t}|}$$

$$= \underbrace{\frac{\mathbf{f}(\mathbf{x}+\delta\mathbf{x}) - \mathbf{f}(\mathbf{x}) - [D\mathbf{f}(\mathbf{x})]\delta\mathbf{x}}{|\delta\mathbf{x}|}}_{\text{goes to } 0 \text{ because } D\mathbf{f} \text{ is the derivative of } \mathbf{f}.} \underbrace{\frac{|\delta\mathbf{x}|}{|\delta\mathbf{t}|}}_{\text{is finite because } \mathbf{x}=\mathbf{g} \text{ is differentiable.}} + [D\mathbf{f}(\mathbf{g}(\mathbf{t}))] \underbrace{\left(\frac{\delta\mathbf{x} - [D\mathbf{g}(\mathbf{t})]\delta\mathbf{t}}{|\delta\mathbf{t}|}\right)}_{\text{goes to } 0 \text{ because } D\mathbf{g} \text{ is the derivative of } \mathbf{g}=\mathbf{x}.}$$