# Lecture 14: The Chain Rule

## 14.1 The gradient

**Definition** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and all first-order partial derivatives of $f$ exist at $a$. We call

$$\nabla f(\mathbf{a}) = \left( \frac{\partial}{\partial x_1} f(\mathbf{a}), \frac{\partial}{\partial x_2} f(\mathbf{a}), \ldots, \frac{\partial}{\partial x_n} f(\mathbf{a}) \right)$$

the *gradient* of $f$ at $\mathbf{a}$.

**Example** If $f(x, y, z) = xyz - 10x^2$, then

$$\nabla f(x, y, z) = (yz - 20x, xz, xy)$$

and, for example,

$$\nabla f(1, -1, 2) = (-22, 2, -1).$$

Note that if $f : \mathbb{R}^n \to \mathbb{R}$, then $\nabla : \mathbb{R}^n \to \mathbb{R}^n$. Such functions are called *vector fields*. Moreover, note that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{a}$, then

$$df_{\mathbf{a}}(\mathbf{x}) = \nabla f(\mathbf{a}) \cdot \mathbf{x}.$$

## 14.2 The chain rule

Recall: If $\varphi : \mathbb{R} \to \mathbb{R}$ is differentiable at $a$ and $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $\varphi(a)$, then $f \circ \varphi$ is differentiable at $a$ and

$$(f \circ \varphi)'(a) = f'(\varphi(a))\varphi'(a).$$

**Chain Rule** If $\varphi : \mathbb{R} \to \mathbb{R}^n$ is differentiable at $a$ and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\varphi(a)$, then $f \circ \varphi$ is differentiable at $a$ and

$$(f \circ \varphi)'(a) = \nabla f(\varphi(a)) \cdot \varphi'(a).$$

**Sketch of proof** We need to show that

$$\lim_{h \to 0} \frac{f(\varphi(a + h)) - f(\varphi(a)) - (\nabla f(\varphi(a)) \cdot \varphi'(a))h}{h} = 0.$$

Let $\mathbf{s} = \varphi(a+h) - \varphi(a)$. Then

$$\frac{f(\varphi(a+h)) - f(\varphi(a)) - (\nabla f(\varphi(a)) \cdot \varphi'(a))h}{h} = \frac{f(\varphi(a)+\mathbf{s}) - f(\varphi(a)) - (\nabla f(\varphi(a)) \cdot \varphi'(a))h}{|\mathbf{s}|} \frac{|\mathbf{s}|}{h}.$$

Now

$$\varphi'(a)h \approx \varphi(a+h) - \varphi(a) = \mathbf{s},$$

so the first factor in the above becomes, approximately,

$$\frac{f(\varphi(a)+\mathbf{s}) - f(\varphi(a)) - \nabla f(\varphi(a)) \cdot \mathbf{s}}{|\mathbf{s}|},$$

which goes to 0 as $\mathbf{s}$ goes to $\mathbf{0}$ by the differentiability of $f$. Moreover, $\mathbf{s}$ goes to $\mathbf{0}$ as $h$ goes to 0 by the continuity of $\varphi$. Finally,

$$\frac{|\mathbf{s}|}{h}$$

approaches $|\varphi'(a)|$ as $h$ approaches 0 through positive values and $-|\varphi'(a)|$ as $h$ approaches 0 through negative values. Hence our original limit approaches 0 as $h$ goes to 0.

**Example** Suppose a metal plate is heated in such a way that the temperature at point $(x, y)$ is given by

$$T(x,y) = 100xye^{-\frac{1}{20}(x^2+y^2)}.$$

Moreover, suppose a bug moves along the plate with position at time $t$ given by

$$\varphi(t) = (2\cos(t), \sin(t)).$$

Now

$$\frac{\partial}{\partial x}T(x,y) = -10x^2ye^{-\frac{1}{20}(x^2+y^2)} + 100ye^{-\frac{1}{20}(x^2+y^2)} = 10(10y - x^2y)e^{-\frac{1}{20}(x^2+y^2)},$$

$$\frac{\partial}{\partial x}T(x,y) = -10xy^2e^{-\frac{1}{20}(x^2+y^2)} + 100xe^{-\frac{1}{20}(x^2+y^2)} = 10(10x - xy^2)e^{-\frac{1}{20}(x^2+y^2)},$$

$$\nabla T(x,y) = 10e^{-\frac{1}{20}(x^2+y^2)}(10y - x^2y, 10x - xy^2),$$

and

$$\varphi'(t) = (-2\sin(t), \cos(t)).$$

Hence, for example, at time $t = \frac{\pi}{4}$,

$$\varphi\left(\frac{\pi}{4}\right) = \left(\sqrt{2}, \frac{1}{\sqrt{2}}\right),$$

$$\varphi'\left(\frac{\pi}{4}\right) = \left(-\sqrt{2}, \frac{1}{\sqrt{2}}\right),$$

$$\nabla T\left(\sqrt{2}, \frac{1}{\sqrt{2}}\right) = 10e^{-\frac{1}{8}}\left(\frac{10}{\sqrt{2}} - \sqrt{2}, 10\sqrt{2} - \frac{1}{\sqrt{2}}\right),$$

and so if $g(t) = T(\varphi(t))$, then

$$g'\left(\frac{\pi}{4}\right) = 10e^{-\frac{1}{8}}\left(-10 + 2 + 10 - \frac{1}{2}\right) = 15e^{-\frac{1}{8}} \approx 13.24.$$

Put another way,

$$\left.\frac{dT}{dt}\right|_{t=\frac{\pi}{4}} \approx 13.24.$$

Note that in the previous example, if we let $x = 2\cos(t)$ and $y = \sin(t)$, then we have

$$\frac{dT}{dt} = \frac{\partial T}{\partial x}\frac{dx}{dt} + \frac{\partial T}{\partial y}\frac{dy}{dt}.$$

More generally, if $z = f(x_1, x_2, \ldots, x_n)$ and $x_1 = f_1(t)$, $x_2 = f_2(t)$, $\ldots$, $x_n = f_n(t)$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial z}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial z}{\partial x_n}\frac{dx_n}{dt}.$$

**Example** Suppose $w = x^2 - y^2 + z^2$, $x = 3t + 4$, $y = t^2 + 1$, and $z = 4t^3$. Then

$$\begin{aligned}\frac{dw}{dt} &= \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt} \\ &= (2x)(3) - (2y)(2t) + (2z)(12t^2) \\ &= 6x - 4yt + 24zt^2.\end{aligned}$$

For example, when $t = 1$, $x = 7$, $y = 2$, and $z = 4$, so

$$\left.\frac{dw}{dt}\right|_{t=1} = 42 - 8 + 96 = 130.$$

More generally, if $u$ is a function of $x_1$, $x_2$, $\ldots$, $x_n$ and, for $k = 1, 2, \ldots, n$, $x_k$ is a function of $t_1$, $t_2$, $\ldots$, $t_m$, then, for $i = 1, 2, \ldots, m$,

$$\frac{\partial u}{\partial t_i} = \frac{\partial u}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial u}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial u}{\partial x_n}\frac{\partial x_n}{\partial t_i}.$$

**Example** Suppose $u = 4x^2y^2$, $x = t^2 - s^2$, and $y = 20ts$. Then

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t} = (8xy^2)(2t) + (8x^2y)(20s) = 16xy^2t + 160x^2ys.$$

and

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} = (8xy^2)(-2s) + (8x^2y)(20t) = -16xy^2s + 160x^2yt.$$