MATH11007 NOTES 17: THE CHAIN RULE AND THE GRADIENT.

1. The chain rule

Let $A \subseteq \mathbb{R}^d$ and $f : A \mapsto \mathbb{R}$. Consider a curve in $\mathbb{R}^d$ given in the parametric form $\mathbf{x}(t)$, $t_0 \leq t \leq t_1$, and suppose that $\mathbf{x}(t) \in A$ for every $t \in [t_0, t_1]$. How does one work out

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}(t))\,?$$

To give an illustration, take $d = 3$, $A = \mathbb{R}^3$, and let (x, y, z) denote the usual spatial coordinates. It is customary to make the x-axis point eastward, the y-axis point northward and the z axis point upward. Imagine you take a trip on a hot air balloon! Your position then changes with time, and as you get higher, you might become interested in the temperature at your current position. So let f be the temperature, viewed as a function of the spatial coordinates x, y and z. If the coordinates of the balloon at time t are $x(t)$, $y(t)$ and $z(t)$, then the temperature you feel in the baloon becomes a function of time alone, say

$$T(t) := f\left(x(t), y(t), z(t)\right) .$$

The rate of change of the temparure T is of obvious interest if you happen to be underdressed...

Notation . It will be convenient in what follows to use a "dot" to indicate differentiation with respect to the variable t. Thus, for example, we shall sometimes write

$$\dot{T}(t)$$

instead of

$$\frac{\mathrm{d}T}{\mathrm{d}t}(t)\,.$$

Suppose that $d = 1$. Then, by the chain rule we know so well,

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x(t)) = \frac{\mathrm{d}f}{\mathrm{d}x}(x(t))\,\dot{x}(t)\,.$$

This formula generalises to several variables:

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}(t)) = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(\mathbf{x}(t))\,\dot{x}_i(t)\,. \tag{1.1}$$

Example 1.1. *Let* $f(x, y) = \sqrt{x^2 + y^2}$ *and*

$$x(t) = \cos t\,, \quad y(t) = \sin t\,.$$

Then

$$f(x(t),\, y(t)) = 1$$

and so, by direct calculation,

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x(t), y(t)) = 0\,.$$

On the other hand, the chain rule gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} f(x(t), y(t)) &= f_x(x(t), y(t))\dot{x}(t) + f_y(x(t), y(t))\dot{y}(t) \\ &= \frac{x(t)\dot{x}(t)}{\sqrt{x^2(t) + y^2(t)}} + \frac{y(t)\dot{y}(t)}{\sqrt{x^2(t) + y^2(t)}} \\ &= -\cos t \sin t + \sin t \cos t = 0\,.\end{aligned}$$

Example 1.2. *Consider again the hot air balloon and suppose that its position at time t is given by*

$$(x(t),\, y(t),\, z(t)) = \left(at\,, at\,, \frac{bt}{1+t}\right)\,.$$

For $a > 0$ and $b > 0$, this corresponds to a steady drift northeastward. The height increases monotonically with time but remains bounded. One might assume that the temperature, as a function of position, increases with y and z, say

$$f(x, y, z) = \alpha y + \beta z$$

for some positive constants α and β. Then

$$T(t) = f(x(t),\, y(t),\, z(t)) = \alpha y(t) + \beta z(t) = \alpha a t + \beta \frac{bt}{1+t}\,.$$

Hence, by direct calculation,

$$\dot{T}(t) = \alpha a + \frac{\beta b}{(1+t)^2}\,.$$

Alternatively, by the chain rule

$$\begin{aligned}\dot{T}(t) &= f_x(x(t),\, y(t),\, z(t))\,\dot{x}(t) \\ &\quad + f_y(x(t),\, y(t),\, z(t))\,\dot{y}(t) + f_z(x(t),\, y(t),\, z(t))\,\dot{z}(t) \\ &= 0a + \alpha a + \beta \frac{b}{(1+t)^2} = \alpha a + \frac{\beta b}{(1+t)^2}\,.\end{aligned}$$

Let us give a justification of the chain rule in the case $d = 2$. The case $d > 2$ is no more difficult— only, the notation gets in the way. By definition of the derivative,

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x(t), y(t)) = \lim_{h \to 0} \frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h}\,.$$

Now,

$$\begin{aligned}&\frac{f(x(t+h), y(t+h)) - f(x(t), y(t))}{h} \\ &= \underbrace{\frac{f(x(t+h), y(t+h)) - f(x(t+h), y(t))}{h}}_{R} + \underbrace{\frac{f(x(t+h), y(t)) - f(x(t), y(t))}{h}}_{S}\end{aligned}$$

and we have, by the one-variable chain rule,

$$S \xrightarrow[h \to 0]{} f_x(x(t), y(t))\,\dot{x}(t)\,.$$

Also

$$R \overset{\tau=t+h}{\overset{\downarrow}{=}} \frac{f(x(\tau),y(\tau)) - f(x(\tau),y(\tau-h))}{h}$$

$$\overset{H=-h}{\overset{\downarrow}{=}} \frac{-f(x(\tau),y(\tau)) + f(x(\tau),y(\tau+H))}{H}$$

$$\xrightarrow[h\to 0]{} f_y(x(t),y(t))\,\dot{y}(t)\,.$$

2. The gradient

Definition 2.1. Let $A \subseteq \mathbb{R}^d$ and suppose that the function $f : A \mapsto \mathbb{R}$ has first partial derivatives at every point of A. The map $\nabla f : A \to \mathbb{R}^d$ defined by

$$\nabla f(\mathbf{x}) = (f_{x_1}(\mathbf{x}),\, \ldots,\, f_{x_d}(\mathbf{x}))$$

is called the *gradient* of f. The gradient of f at a point $\mathbf{x} \in A$ is the value of this map at $\mathbf{x}$.

Example 2.1. *The gradient of the function*

$$f(x,y,z) = xy^2z^3$$

is

$$\nabla f(x,y,z) = yz^2(yz,\, 2xz,\, 3xy)\,.$$

The gradient of the function

$$f(\mathbf{x}) = \|\mathbf{x}\| = \left[\sum_{j=1}^{d} x_j^2\right]^{\frac{1}{2}}$$

is

$$\nabla f(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|}\mathbf{x}\,.$$

Using the gradient and the dot product between two vectors in $\mathbb{R}^d$, the chain rule can be expressed neatly as

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}) = \nabla f(\mathbf{x}(t)) \cdot \dot{\mathbf{x}}(t)\,.$$

For $d = 1$, the gradient is, of course, just the familiar derivative. The derivative of a function $f(x)$ has a well-known interpretation as the slope of the tangent to the curve of equation $y = f(x)$. Let us now look at the geometrical meaning of the gradient when $d = 2$.

Consider a function $f(x,y)$. Pick a point (x_0, y_0) in the domain of f. The level curve going through this point is the curve of equation

$$f(x,y) = c := f(x_0,y_0)\,.$$

Look at a neighbourhood of the point (x_0, y_0). In this neighbourhood, along the level curve, we may view y as a function of x, and we have

$$f(x, y(x)) = c\,.$$

Differentiate this identity with respect to x: the chain rule gives

$$\nabla f(x,y(x)) \cdot (1, y'(x)) = 0\,.$$

Evaluating at $x = x_0$, we obtain

$$\nabla f(x_0, y_0) \cdot (1, y'(x_0)) = 0 \,. \tag{2.1}$$

Now, the vector $(1, y'(x_0))$ is parallel to the line tangent to the level curve $f(x_0, y_0)$; see see Figure 1. So Equation (2.1) expresses the geometrical fact that the gradient at (x_0, y_0) is normal (perpendicular, orthogonal) to the line tangent to the level curve through (x_0, y_0).

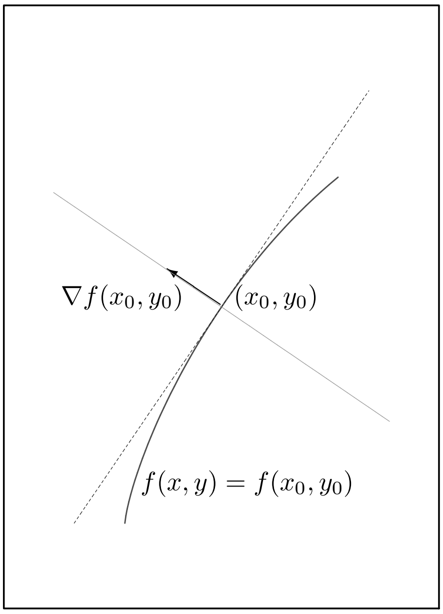

FIGURE 1. The gradient of f at (x_0, y_0) is orthogonal to the level curve through $f(x, y) = f(x_0, y_0)$.

Example 2.2. *Find the normal and tangent lines to the curve of equation $y^3 = 4x^2$ at $(\sqrt{2}, 2)$.*

Solution: *Take*

$$f(x, y) = 4x^2 - y^3 \,.$$

The point $(\sqrt{2}, 2)$ lies on the level curve $f(x, y) = 0$. The gradient

$$\nabla f(x, y) = \big(8x,\, -3y^2\big)_{|_{x=\sqrt{2}, y=2}} = \Big(8\sqrt{2},\, -12\Big)$$

is a vector normal to the level curve. Hence the equation of the normal line is

$$x(t) = \sqrt{2} + 8\sqrt{2}t, \quad y(t) = 2 - 12t \,, \quad t \in \mathbb{R} \,.$$

This expresses the line in parametric form. To obtain the familiar cartesian form, eliminate t.

The vector $\big(12,\, 8\sqrt{2}\big)$ is perpendicular to the gradient, and so the equation of the tangent line is

$$x(t) = \sqrt{2} + 12t \,, \quad y(t) = 2 + 8\sqrt{2}t \,, \quad t \in \mathbb{R} \,.$$

In the case $d = 3$, the equation

$$f(x, y, z) = f(x_0, y_0, z_0)$$

defines a *surface*. It may be shown that the gradient $\nabla f(x_0, y_0, z_0)$ is normal to the plane tangent to the surface at (x_0, y_0, z_0). The equation of the tangent plane (in cartesian form) is

$$\nabla f(x_0, y_0, z_0) \cdot (x - x_0, y - y_0, z - z_0) = 0\,.$$

The equation of the line normal to the surface (in parametric form) is therefore

$$x = x_0 + t f_x(x_0, y_0, z_0)\,, \quad y = y_0 + t f_y(x_0, y_0, z_0)\,, \quad z = z_0 + t f_z(x_0, y_0, z_0)\,, \quad t \in \mathbb{R}\,.$$

References

1. Frank Ayres, Jr. and Elliott Mendelson, *Schaum's Outline of Calculus, Fourth Edition*, McGraw–Hill, 1999.
2. E. Hairer and G. Wanner, *Analysis by its History*, Springer-Verlag, New-York, 1996.