# LINEAR ALGEBRA AND VECTOR ANALYSIS

#### $\mathrm{MATH}\ 22\mathrm{B}$

# Unit 16: Chain rule

#### INTRODUCTION

**16.1.** In calculus, we can build from basic functions more general functions. One possibility is to add functions like  $f(x) + g(x) = x^2 + \sin(x)$ . An other possibility is to multiply functions like  $f(x)g(x) = x^2\sin(x)$ . A third possibility is to combine functions like  $f \circ g(x) = f(g(x)) = \sin^2(x)$ . The composition of functions is non-commutative:  $f \circ g \neq g \circ f$ . Indeed, we have  $g \circ f(x) = \sin(x^2)$  which is completely different from  $f \circ g(x) = \sin^2(x)$ .



FIGURE 1.  $f : \mathbb{R}^p \to \mathbb{R}^n$  and  $g : \mathbb{R}^m \to \mathbb{R}^p$  can be combined to  $f(g) : \mathbb{R}^m \to \mathbb{R}^n$ .

16.2. How can we express the rate of change of a composite function in terms of the basic functions it is built of? For the sum of two functions, we have the **addition rule** (f + g)'(x) = f'(x) + g'(x), for multiplication we have the **product rule** (fg)'(x) = f'(x)g(x) + f(x)g(x). We usually just write (f + g)' = f' + g' or (fg)' = f'g + fg' and do not always write the argument. As you know from single variable calculus, the derivative of the composite function is given by **chain rule**. This is  $(f \circ g)' = f'(g)g'$ . Written out in more details with argument, we can write

 $\frac{d}{dx}f(g(x)) = \frac{d}{dx}f'(g(x))g'(x)$ . We generalize this here to higher dimensions. Instead of  $\frac{d}{dx}f$  we just write df. This is the Jacobean matrix we know. Now, the same rule holds as before df(g(x)) = df(g(x))dg(x) and this is called the chain rule in higher dimensions. On the right hand side, we have the matrix product of two matrices.

**16.3.** Let us see why this makes sense in terms of dimensions:  $g : \mathbb{R}^m \to \mathbb{R}^p$  and  $f : \mathbb{R}^p \to \mathbb{R}^n$ , then  $dg(x) \in M(p,m)$  and  $df(g(x)) \in M(n,p)$  and  $df(g(x))dg(x) \in M(n,m)$  which is the same type of matrix than  $d(f \circ g)$  because  $f \circ g(x)$  maps  $\mathbb{R}^m \to \mathbb{R}^n$  so that also  $d(f \circ g)(x) \in M(n,m)$ . The name **chain rule** comes because it deals with functions that are chained together.

## LECTURE

**16.4.** Given a differentiable function  $r : \mathbb{R}^m \to \mathbb{R}^p$ , its derivative at x is the Jacobian matrix  $dr(x) \in M(p,m)$ . If  $f : \mathbb{R}^p \to \mathbb{R}^n$  is another function with  $df(y) \in M(n,p)$ , we can combine them and form  $f \circ r(x) = f(r(x)) : \mathbb{R}^m \to \mathbb{R}^n$ . The matrices  $df(y) \in M(n,p)$  and  $dr(x) \in M(p,m)$  combine to the matrix product df dr at a point. This matrix is in M(n,m). The **multi-variable chain rule** is:

**Theorem:**  $d(f \circ r)(x) = df(r(x))dr(x)$ 

16.5. For m = n = p = 1, the single variable calculus case, we have df(x) = f'(x)and  $(f \circ r)'(x) = f'(r(x))r'(x)$ . In general, df is now a matrix rather than a number. By checking a single matrix entry, we reduce to the case n = m = 1. In that case,  $f : \mathbb{R}^p \to \mathbb{R}$  is a scalar function. While df is a row vector, we define the column vector  $\nabla f = df^T = [f_{x_1}, f_{x_2}, \dots, f_{x_p}]^T$ . If  $r : \mathbb{R} \to \mathbb{R}^p$  is a curve, we write  $r'(t) = [x'_1(t), \dots, x'_p(t)]^T$  instead of dr(t). The symbol  $\nabla$  is addressed also as "nabla". <sup>1</sup> The special case n = m = 1 is:

**Theorem:**  $\frac{d}{dt}f(r(t)) = \nabla f(r(t)) \cdot r'(t).$ 

**16.6.** Proof.  $d/dt f(x_1(t), x_2(t), \ldots, x_p(t))$  is the limit  $h \to 0$  of

$$\begin{aligned} & [f(x_1(t+h), x_2(t+h), \dots, x_p(t+h)) - f(x_1(t), x_2(t), \dots, x_p(t))]/h = \\ & = [f(x_1(t+h), x_2(t+h), \dots, x_p(t+h)) - f(x_1(t), x_2(t+h), \dots, x_p(t+h)]/h \\ & + [f(x_1(t), x_2(t+h), \dots, x_p(t+h)) - f(x_1(t), x_2(t), \dots, x_p(t+h)]/h + \cdots \\ & + [f(x_1(t), x_2(t), \dots, x_p(t+h)) - f(x_1(t), x_2(t), \dots, x_p(t))]/h \end{aligned}$$

which is (1D chain rule) in the limit  $h \to 0$  the sum  $f_{x_1}(x)x'_1(t) + \cdots + f_{x_p}(x)x'_p(t)$ .

**16.7.** Proof of the general case: Let  $h = f \circ r$ . The entry ij of the Jacobian matrix dh(x) is  $dh_{ij}(x) = \partial_{x_j}h_i(x) = \partial_{x_j}f_i(r(x))$ . The case of the entry ij reduces with  $t = x_j$  and  $h_i = f$  to the case when r(t) is a curve and f(x) is a scalar function. This is the case we have proven already.

<sup>&</sup>lt;sup>1</sup>Etymology tells that the symbol is inspired by a Egyptian or Phoenician harp.

## EXAMPLE

**16.8.** Assume a ladybug walks on a circle  $r(t) = \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix}$  and  $f(x, y) = x^2 - y^2$  is the temperature at the position (x, y), then f(r(t)) is the rate of change of the temperature. We can write  $f(r(t)) = \cos^2(t) - \sin^2(t) = \cos(2t)$ . Now,  $d/dtf(r(t)) = -2\sin(2t)$ . The gradient of f and the velocity are  $\nabla f(x, y) = \begin{bmatrix} 2x \\ -2y \end{bmatrix}$ ,  $r'(t) = \begin{bmatrix} -\sin(t) \\ \cos(t) \end{bmatrix}$ . Now

$$\nabla f(r(t)) \cdot r'(t) = \begin{bmatrix} 2\cos(t) \\ -2\sin(t) \end{bmatrix} \cdot \begin{bmatrix} -\sin(t) \\ \cos(t) \end{bmatrix} = -4\cos(t)\sin(t) = -2\sin(2t) \cdot \frac{1}{2} + \frac$$



FIGURE 2. If f(x, y) is a height, the rate of change d/dtf(r(t)) is the gain of height the bug climbs in unit time. It depends on how fast the bug walks and in which direction relative to the gradient  $\nabla f$  it walks.

#### Illustrations

**16.9.** The case n = m = 1 is extremely important. The chain rule  $d/dtf(r(t)) = \nabla f(r(t)) \cdot r'(t)$  tells that the rate of change of the **potential energy** f(r(t)) at the position r(t) is the dot product of the **force**  $F = \nabla f(r(t))$  at the point and the **velocity** with which we move. The right hand side is **power** = **force** times **velocity**. We will use this later in the fundamental theorem of line integrals.

**16.10.** If  $f, g: \mathbb{R}^m \to \mathbb{R}^m$ , then  $f \circ g$  is again a map from  $\mathbb{R}^m$  to  $\mathbb{R}^n$ . We can also iterate a map like  $x \to f(x) \to f(f(x)) \to f(f(f(x))) \dots$ . The derivative  $df^n(x)$  is by the chain rule the product  $df(f^{n-1}(x)) \cdots df(f(x)) df(x)$  of Jacobian matrices. The number  $\lambda(x) = \limsup_{n \to \infty} (1/n) \log(|df^n(x)|)$  is called the **Lyapunov exponent** of the map fat the point x. It measures the amount of **chaos**, the "sensitive dependence on initial conditions" of f. These numbers are hard to estimate mathematically. Already for simple examples like the **Chirikov map**  $f([x,y]) = [2x - y + c\sin(x), x]$ , one can measure **positive entropy** S(c). A conjecture of Sinai tells that that the **entropy** of the map is positive for large c. Measurements show that this entropy  $S(c) = \int_0^{2\pi} \int_0^{2\pi} \lambda(x,y) dx dy/(4\pi^2)$  satisfies  $S(x) \ge \log(c/2)$ . The conjecture is still open.<sup>2</sup>

**16.11.** If H(x, y) is a function called the **Hamiltonian** and  $x'(t) = H_y(x, y), y'(t) = -H_x(x, y)$ , then d/dtH(x(t), y(t)) = 0. This can be interpreted as **energy conservation**. We see that a Hamiltonian differential equation always preserves the energy. For the **pendulum**,  $H(x, y) = y^2/2 - \cos(x)$ , we have  $x' = y, y' = -\sin(x)$  or  $x'' = -\sin(x)$ .

<sup>&</sup>lt;sup>2</sup>To generate orbits, see http://www.math.harvard.edu/~knill/technology/chirikov/.



FIGURE 3. The map  $f([x, y]) = [x^2 - x/2 - y, x]$  is a **Henon map**. We see some orbits. The map  $f([x, y]) = [2x - y + 4\sin(x), x]$  on the right appeared in the first hourly. The torus  $\mathbb{T}^2 = \mathbb{R}^2/(2\pi\mathbb{Z})^2$  is filled with a blue "stochastic sea" containing red "stable islands".

16.12. The chain rule is useful to get derivatives of inverse functions. Like

$$1 = \frac{d}{dx}x = \frac{d}{dx}\sin(\arcsin(x)) = \cos(\arcsin(x))\arcsin(x)$$

which then gives  $\arcsin'(x) = 1/\sqrt{1 - \sin^2(\arcsin(x))} = 1/\sqrt{1 - x^2}$ .

**16.13.** Assume  $f(x, y) = x^3y + x^5y^4 - 2 - \sin(x - y) = 0$  is a curve. We can not solve for y. Still, we can assume f(x, y(x)) = 0. Differentiation using the chain rule gives  $f_x(x, y(x)) + f_y(x, y(x))y'(x) = 0$ . Therefore

$$y'(x) = -\frac{f_x(x, y(x))}{f_y(x, y(x))}$$

In the above example, the point (x, y) = (1, 1) is on the curve. Now  $g_x(x, y) = 3 + 5 - 1 = 7$  and  $g_y(x, y) = 1 + 4 + 1 = 6$ . So, g'(1) = -7/6. This is called **implicit differentiation**. We could compute with it the derivative of a function which was not known.

16.14. The implicit function theorem assures that a differentiable implicit function g(x) exists near a root (a, b) of a differentiable function f(x, y).

**Theorem:** If f(a, b) = 0,  $f_y(a, b) \neq 0$  there exists c > 0 and a function  $g \in C^1([b-c, b+c])$  with f(x, g(x)) = 0.

Proof. Let c be so small that for fixed  $x \in [a-c, a+c]$ , the function  $y \in [b-c, b+c] \rightarrow h(y) = f(x, y)$  has the property h(b-c) < 0 and h(b+c) > 0 and  $h'(y) \neq 0$  in [b-c, b+c]. The **intermediate value theorem** for h now assures a unique root z = g(x) of h near b. The chain rule formula above then assures that for a - c < x < a + c, the differential quotient [g(x+h)-g(x)]/h written down for g has a limit  $-f_x(x,g(x))/f_y(x,g(x))$ .

P.S. We can get the root of h by applying **Newton steps** T(y) = y - h(y)/h'(y). Taylor (seen in the next class) shows the error is squared in every step. The Newton step  $T(y) = y - dh(y)^{-1}h(y)$  works also in arbitrary dimensions. One can prove the implicit function theorem by just establishing that  $\text{Id} - T = dh^{-1}h$  is a contraction and then use the **Banach fixed point theorem** to get a fixed point of Id - T which is a root of h.



FIGURE 4. The Newton step.



FIGURE 5. If we apply the map  $f([x, y]) = [x^2 - x^4 - y, x]$  again and again and plot points, we get an **orbit**. Such simple dynamical systems are largely not understood. Which points do not escape to infinity? What is the boundary of this set. Proving that there are regions which stay bounded is hard and needs "hard implicit function theorems". The Newton method allows to get a grip on proving this, where the Newton step is applied on spaces of functions. Some of the hardest analysis which humans have invented for tackling mathematical problems come to play in this seemingly simple map  $f : \mathbb{R}^2 \to \mathbb{R}^2$ .

Units 16 and 17 are together taught on Wednesday. Homework is all in unit 17. OLIVER KNILL, KNILL@MATH.HARVARD.EDU, MATH 22B, HARVARD COLLEGE, SPRING 2022