Unit 16: Chain rule

Lecture

16.1. 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: $d(f \circ r)(x) = df(r(x))dr(x)$

16.2. 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 $r'(t) = [x_1'(t), \cdots, x_p'(t)]^T$ instead of $dr(t)$. The symbol ∇ is addressed also as "nabla". [1] The special case $n = m = 1$ is:

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

16.3. Proof. $d/dt f(x_1(t), x_2(t), \dots, x_p(t))$ 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 $f_{x_1}(x)x_1'(t) + \cdots + f_{x_p}(x)x_p'(t)$.

16.4. 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.

[1] Etymology tells that the symbol is inspired by a Egyptian or Phoenician harp.

Example

16.5. 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, $d/dt f(r(t)) = -2\sin(2t)$. 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

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

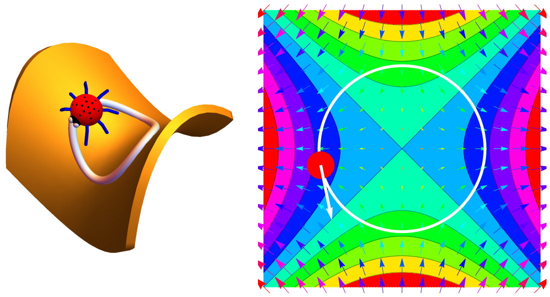

FIGURE 1. If $f(x,y)$ is a height, the rate of change $d/dt f(r(t))$ is the gain of height the bug climbs in unit time. It depends on how fast the bug walks and in which direction relative to the gradient ∇f it walks.

Illustrations

16.6. The case $n = m = 1$ is extremely important. The chain rule $d/dt f(r(t)) = \nabla f(r(t)) \cdot r'(t)$ 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.

16.7. 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 f at 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 $S(c) = \int_0^{2\pi}\int_0^{2\pi} \lambda(x,y)\, dxdy/(4\pi^2)$ satisfies $S(x) \geq \log(c/2)$. The conjecture is still open. [2]

16.8. If $H(x,y)$ is a function called the **Hamiltonian** and $x'(t) = H_y(x,y), y'(t) = -H_x(x,y)$, then $d/dt H(x(t), y(t)) = 0$. 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 $x'' = -\sin(x)$.

[2] To generate orbits, see http://www.math.harvard.edu/~knill/technology/chirikov/.

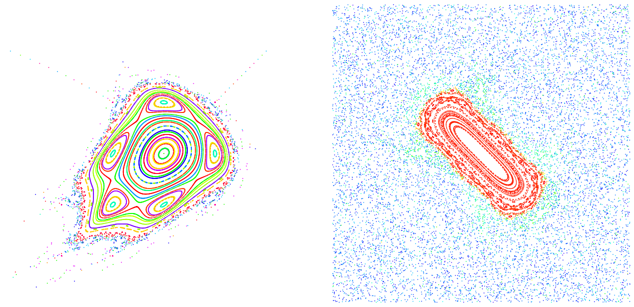

FIGURE 2. The map $f([x,y]) = [x^2 - x/2 - y, x]$ is a **Henon map**. We see some orbits. The map $f([x,y]) = [2x - y + 4\sin(x), x]$ on the right appeared in the first hourly. The torus $\mathbb{T}^2 = \mathbb{R}^2/(2\pi\mathbb{Z})^2$ is filled with a blue "stochastic sea" containing red "stable islands".

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

16.10. 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.

16.11. 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 $y \in [b-c,b+c] \to h(y) = f(x,y)$ 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 $\mathrm{Id} - T = dh^{-1}h$ is a contraction and then use the **Banach fixed point theorem** to get a fixed point of $\mathrm{Id} - T$ which is a root of h.

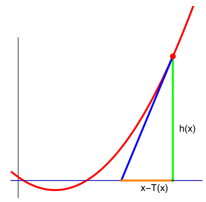

FIGURE 3. The Newton step.

HOMEWORK

Problem 16.1: Let $r(t) = [3t + \cos(t), t + 4\sin(t)]^T$ be a curve and $f([x, y]^T) = [x^3 + y, x + 2y + y^3]^T$ be a coordinate change. a) Compute $v = r'(0)$ at $t = 0$, then $df(x, y)$ and $A = df(r(0))$ and $df(r(0))r'(0) = Av$. b) Compute $R(t) = f(r(t))$ first, then find $w = R'(0)$. It should agree with a).

Problem 16.2: a) Define the function $f(x, y) = x \cdot y$ from $\mathbb{R}^2$ to $\mathbb{R}$. If both x and y are functions of t we get the curve $r(t) = [x(t), y(t)]^T$. What does the chain rule tell for $t \to f(r(t))$ from $\mathbb{R}$ to $\mathbb{R}$?
b) Do the same for the function $f(x, y) = x/y$. What rule do you get now?

Problem 16.3: The surface $f(x, y, z) = x^2 + y^2/4 + z^2/9 = 4 + 1/4 + 1/9$ is an ellipsoid. Compute $z_x(x, y)$ at the point $(x, y, z) = (2, 1, 1)$.

Problem 16.4: Consider the Hénon map $f([x, y]^T) = [x^2 - x^4 - y, x]^T$. Compute either $d(f \circ f)([1, 1]^T)$ or $df(f([1, 1]^T))df([1, 1]^T)$. The chain rule tells it is the same matrix.

Problem 16.5: Apply the Newton step 3 times starting with $x = 2$ to solve the equation $x^2 - 2 = 0$.

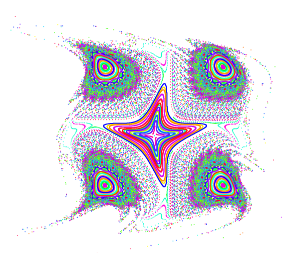

FIGURE 4. Some orbits of the Henon map $f([x, y]) = [x^2 - x^4 - y, x]$.

OLIVER KNILL, KNILL@MATH.HARVARD.EDU, MATH 22A, HARVARD COLLEGE, FALL 2018