10/29/2002, CHAIN RULE — Math 21a, O. Knill

Homework: 11.5: 6, 17, 18, 24, 28

REMEMBER? If f and g are functions of one variable t, then $d/dt f(g(t)) = f'(g(t))g'(t)$. For example, $d/dt \sin(\log(t)) = \cos(\log(t))/t$.

THE CHAIN RULE If r is curve (a function of one variable which has values in space) and f is a function of three variables, we get a function of one variables $t \mapsto f(r(t))$. The chain rule looks the same now, we just have to replace the derivative f' with the gradient ∇f. The derivative is $\boxed{d/dt f(r(t)) = \nabla f(r(t)) \cdot r'(t)}$

INTUITIVE EXAMPLE. If f is the temperature distribution in a room and $r(t)$ is the path of a spider (it is soon Halloween), then $f(r(t))$ is the temperature, the spider experiences at time t. The rate of change depends on the velocity $r'(t)$ of the spider as well as the temperature gradient ∇f and the angle between gradient and velocity. For example, if the spider moves perpendicular to the gradient, it moves tangent to a level curve and the rate of change is 0.

WRITING IT OUT. Writing the dot product out gives

$$\frac{d}{dt} f(r(t)) = f_x(x(t), y(t), z(t))x'(t) + f_y(x(t), y(t), z(t))y'(t) + f_z(x(t), y(t), z(t))z'(t) \; .$$

SPIDER. Assume a spider moves along a circle $r(t) = (\cos(t), \sin(t))$ on a table with temperature distribution $T(x, y) = x^2 - y^3$. Find the rate of change of the temperature, the spider experiences in time.

SOLUTION. $\nabla T(x, y) = (2x, -3y^2)$, $r'(t) = (-\sin(t), \cos(t))$ $d/dt T(r(t)) = \nabla T(r(t)) \cdot r'(t) = (2\cos(t), -3\sin(t)^2) \cdot (-\sin(t), \cos(t)) = -2\cos(t)\sin(t) - 3\sin^2(t)\cos(t)$.

APPLICATION: IMPLICIT DIFFERENTIATION.

1) $\boxed{\text{2D}}$ From $f(x, y) = 0$ we can express y is a function of x. From $d/df(x, y(x)) = \nabla f \cdot (1, y'(x)) = f_x + f_y y' = 0$, we obtain $\boxed{y' = -f_x/f_y}$.

2) $\boxed{\text{3D}}$ If $z = g(x, y)$ is given by $f(x, y, z) = 0$, then $f_x + f_z g_x = 0$ and $f_y + f_z g_y = 0$, so that $\boxed{g_x = -f_x/f_z}$ and $\boxed{g_y = -f_y/f_z}$.

EXAMPLE. $f(x, y) = x^4 + x\sin(xy) = 0$ defines $y = g(x)$. If $f(x, g(x)) = 0$, then $g_x(x) = -f_x/f_y = -(4x^3 + \sin(xy) + xy\cos(xy))/(x^2\cos(xy))$.

APPLICATION: If $f(x, y, z) = 0$, then $x = x(y, z), y = y(x, z)$ and $z = z(x, y)$. From $y_x = -f_x/f_y, x_z = -f_z/f_x, z_y = -f_y/f_z$ we get the relation $\boxed{y_x x_z z_y = -1}$. This formula appears in thermodynamics.

DIETERICI EQUATION. In thermodynamics, where temperature T, pressure p and volume V are related. One refinement of the ideal gas law $pV = RT$ is the Dieterici equation $f(p, V, T) = p(V - b)e^{a/RVT} - RT = 0$. The constant b depends on the volume of the molecules and a depends on the interaction of the molecules. A different variation is van der Waals law. Problem: compute V_T.

If $V = V(T, p)$, the chain rule says $f_V V_T + f_T = 0$, so that $\boxed{V_T = -f_T/f_V} = -(-ap(V - b)e^{a/RVT}/(RVT^2) - R)/(pVe^{a/RVT} - p(V - b)e^{a/RVT}/(RV^2T))$. (This could be simplified to $(R + a/TV)/(RT/(V - b) - a/V^2)$).

MORE GENERAL CHAIN RULE. If $\vec{f}$ is a vector valued function, we can apply the chain rule for each of the components. The derivative $\vec{f}'$ of $\vec{f}$ is then a vector too: $d/dt f_i(r(t)) = \nabla f_i(r(t)) \cdot r'(t)$.

DERIVATIVE. If $f : \mathbf{R}^n \to \mathbf{R}^m$ is a map, its **derivative** f' is the $m \times n$ matrix $[f']_{ij} = \frac{\partial}{\partial x_j} f_i$.

EXAMPLES OF DERIVATIVES.

$f : \mathbf{R} \to \mathbf{R}^3$	$f : \mathbf{R}^3 \to \mathbf{R}$	$f : \mathbf{R}^2 \to \mathbf{R}^3$	$f : \mathbf{R}^2 \to \mathbf{R}^2$	$f : \mathbf{R}^3 \to \mathbf{R}^3$
curve	scalar function	surface	coordinate change	gradient field
f' velocity vector.	f' gradient vector.	f' tangent matrix	f' Jacobean matrix.	f' Hessian matrix.

THE GENERAL CHAIN RULE.
If $f : \mathbf{R}^n \to \mathbf{R}^m$ and $g : \mathbf{R}^k \to \mathbf{R}^n$, we can compose $f \circ g$, which is a map from $\mathbf{R}^k$ to $\mathbf{R}^m$. The chain rule expresses the derivative of $f \circ g(x) = f(g(x))$ in terms of the derivatives of f and g.

$\frac{\partial}{\partial x_j} f(g(x))_i = \sum_k \frac{\partial}{\partial x_k} f_i(g(x)) \frac{\partial}{\partial x_j} g_k(x)$ or short $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$

Here $f'(g(x))$ and $g'(x)$ are matrices and $\cdot$ is the matrix multiplication. The chain rule in higher dimensions looks like the chain rule in one dimension, only that the objects are matrices and the multiplication is matrix multiplication.

REMARK. An other notation is used used especially in general relativity: if f^i_j denotes the derivative of f^i with respect to x_j, where $f = (f^1, \dots, f^n)$. The chain rule is $(f \circ g)^i_j(x) = \sum_k f^i_k(g(x)) g^k_j(x) = f^i_k(g(x)) g^k_j(x)$. The expression to the right uses automatic summation over the index n which appears in upper and lower places (Einstein convention).

EXAMPLE. GRADIENT IN POLAR COORDINATES. In Polar coordinates, the gradient is defined as $\nabla f = (f_r, f_\theta / r)$. Using the chain rule, we can relate this to the usual gradient: $d/dr f(x(r,\theta), y(r,\theta)) = f_x(x,y)\cos(\theta) + f_y(x,y)\sin(\theta)$ and $d/(rd\theta) f(x(r,\theta), y(r,\theta)) = -f_x(x,y)\sin(\theta) + f_y(x,y)\cos(\theta)$ means that the length of ∇f is the same in both coordinate systems.

WHERE DO WE NEED THE CHAIN RULE.
While the chain rule is useful in calculations using the composition of functions, the iteration of maps or in doing change of variables, it is also useful for **understanding** some theoretical aspects. Examples (both practical and theoretical):

- The chain rule is used in **change of variable** formulas. For example, if f is a function in the plane and $x = r\cos(\theta), y = r\sin(\theta)$ is a change of coordinates into polar coordinates, then calculating the gradient of f in the new coordinates requires the chain rule.

- The chain rule will be used in the **fundamental theorem for line integrals**: $\int_a^b \nabla U(r(t)) \cdot r'(t)\ dt = U(b) - U(a)$. The chain rule said that inside the integral, we have $d/dt U(r(t))$.

- Gradients are orthogonal to level surfaces: assume we have a curve $r(t)$ on a surface $g(x,y,z) = c$. Because we move on the surface where g is constant, we have $d/dt g(r(t)) = 0$. The chain rule says $\nabla g(r(t)) \cdot r'(t) = 0$. In other words, the gradient of g is orthogonal to the surface $g = const$. (We had argued earlier using the linear approximation of g.)

- The chain rule illustrates also the **Lagrange multiplier** method which we will see later. If we want to extremize $f(x,y)$ on the constraint $g(x,y)$ and $r(t)$ is a curve on $g(x,y) = c$, then at a critical point, we must have $d/dt f(r(t)) = 0$. This means with the chain rule that the velocity vector $r'(t)$ is orthogonal to $\nabla f(r(t))$. Because the velocity vector is orthogonal to ∇g The vectors ∇g and ∇f have to be parallel.

- In **fluid dynamics**, you often see partial differential equations starting with $u_t + u\nabla u$, where u is the velocity of the fluid. For example, the conservation of momentum of an ideal fluid in the presence of an external force f is $u_t + u \cdot \nabla u = (1/\rho)\nabla p + f$, where p is the pressure and ρ is the density. What does this mean? The term $u_t + u \cdot \nabla u$ is the change of velocity in the coordinate frame of a particle in the fluid. $(u(t,x,y,z) = (u_1(t,x,y,z), u_2(x,y,z), u_3(x,y,z)$ is a vector and ∇u means applying the gradient to each coordinate). For any quantity like pressure, vorticity, density etc. , $Df/Dt = f_t + u \cdot \nabla f$ means the time derivative of f **moving with the fluid**.

- In **chaos theory**, one wants to understand what happens after iterating a map f. If $f^{(n)}(x) = f f^{(n-1)}(x)$, then $(f^{(n)})'(x) = f'(f^{(n-1)}x) f'(f^{(n-2)}(x)) ... f'(x)$ is a product of matrices. Chaos means that $(f^n)'(x)$ grows exponentially for a large set of x. A measure of chaos is the Lyapunov exponent $\lambda = n^{-1}\log|(f^n)'(x)|$ in the limit $n \to \infty$. If this number is positive, one has sensitive dependence on initial conditions. The map $x \mapsto 4x(1-x)$ on the interval $[0,1]$ for example has this property. Small chaos of order $\lambda = 10^{-15}$ is present in the solar system. Large chaos $\lambda = 10^{15}$ is present in a small volume of argon gas at room temperature.

PROOF OF THE CHAIN RULE. The chain rule can be plugging in the definitions of the derivatives. An other proof: by linear approximation it is enough to check the chain rule for linear functions.