LINEARIZATION AND CHAIN RULE

Maths21a, O. Knill

LINEAR APPROXIMATION.

1D: The **linear approximation** of a function $f(x)$ at a point x_0 is the linear function

$$L(x) = f(x_0) + f'(x_0)(x - x_0)$$

The graph of L is tangent to the graph of f.

2D: The **linear approximation** of a function $f(x,y)$ at (x_0,y_0) is

$$L(x,y) = f(x_0,y_0) + f_x(x_0,y_0)(x - x_0) + f_y(x_0,y_0)(y - y_0)$$

The level curve of g is tangent to the level curve of f at (x_0,y_0). The graph of L is tangent to the graph of f.

3D: The **linear approximation** of a function $f(x,y,z)$ at (x_0,y_0,z_0) by

$$L(x,y,z) = f(x_0,y_0,z_0) + f_x(x_0,y_0,z_0)(x - x_0) + f_y(x_0,y_0,z_0)(y - y_0) + f_z(x_0,y_0,z_0)(z - z_0).$$

The level surface of L is tangent to the level surface of f at (x_0,y_0,z_0).

GRADIENT. Define $\nabla f(x,y) = \langle f_x(x,y), f_y(x,y)\rangle$. It is called the **gradient** of f. The symbol ∇ is calle "nabla".

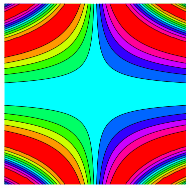
Using the gradient, one can write the linearization in a short convenient way which ressembles the situation in one dimensions:

$$L(\vec{x}) = f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0)$$

JUSTIFYING THE LINEAR APPROXIMATION

The higher dimensional case can be reduced to the one dimensional case: if $y = y_0$ is fixed and x, then $f(x,y_0) = f(x_0,y_0) + f_x(x_0,y_0)(x - x_0)$ is the linear approximation of the function. Similarly, if $x = x_0$ is fixed y is the single variable, then $f(x_0,y) = f(x_0,y_0) + f_y(x_0,y_0)(y - y_0)$. So along two directions, the linear approximations are the best. Together we get the approximation for $f(x,y)$.

EXAMPLE. Find the linear approximation of the function $f(x,y) = \sin(\pi x y^2)$ at the point $(1,1)$. The gradient is $\nabla f(x,y) = (\pi y^2 \cos(\pi x y^2), 2y\pi\cos(\pi x y^2))$. At the point $(1,1)$, we have $\nabla f(1,1) = (\pi\cos(\pi), 2\pi\cos(\pi)) = (-\pi, 2\pi)$. The linear function approximating f is $L(x,y) = f(1,1) + \nabla f(1,1)\cdot(x-1, y-1)$ $= 0 - \pi(x-1) - 2\pi(y-1) = -\pi x - 2\pi y + 3\pi$. The level curves of G are the lines $x + 2y = const$. The line which passes through $(1,1)$ satisfies $x + 2y = 3$.



Application: $-0.00943407 = f(1+0.01, 1+0.01) \sim g(1+0.01, 1+0.01) = -\pi 0.01 - 2\pi 0.01 + 3\pi = -0.00942478$.

EXAMPLE. Find the linear approximation to $f(x,y,z) = xy + yz + zx$ at the point $(1,1,1)$.

We have $f(1,1,1) = 3$, $\nabla f(x,y,z) = (y+z, x+z, y+x), \nabla f(1,1,1) = (2,2,2)$. Therefore $L(x,y,z) = f(1,1,1) + (2,2,2)\cdot(x-1, y-1, z-1) = 3 + 2(x-1) + 2(y-1) + 2(z-1) = 2x + 2y + 2z - 3$.

EXAMPLE. Use the best linear approximation to $f(x,y,z) = e^x\sqrt{y}z$ to estimate the value of f at the point $(0.01, 24.8, 1.02)$.

Solution. Take $(x_0,y_0,z_0) = (0,25,1)$, where $f(x_0,y_0,z_0) = 5$. The gradient is $\nabla f(x,y,z) = (e^x\sqrt{y}z, e^x z/(2\sqrt{y}), e^x\sqrt{y})$. At the point $(x_0,y_0,z_0) = (0,25,1)$ the gradient is the vector $(5, 1/10, 5)$. The linear approximation is $L(x,y,z) = f(x_0,y_0,z_0) + \nabla f(x_0,y_0,z_0)(x - x_0, y - y_0, z - z_0) = 5 + (5, 1/10, 5)(x - 0, y - 25, z - 1) = 5x + y/10 + 5z - 2.5$. We can approximate $f(0.01, 24.8, 1.02)$ by $5 + (5, 1/10, 5)\cdot(0.01, -0.2, 0.02) = 5 + 0.05 - 0.02 + 0.10 = 5.13$. The actual value is $f(0.01, 24.8, 1.02) = 5.1306$, very close to the estimate.

1D CHAIN RULE 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 $\vec{r}(t)$ is curve in space and f is a function of three variables, we get a function of one variables $t \mapsto f(\vec{r}(t))$. The **chain rule** is

$$d/dt f(\vec{r}(t)) = \nabla f(\vec{r}(t)) \cdot \vec{r}\,'(t)$$

WRITING IT OUT. Writing the dot product out gives

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

EXAMPLE. Let $z = \sin(x + 2y)$, where x and y are functions of t: $x = e^t, y = \cos(t)$. What is $\frac{dz}{dt}$?

Here, $z = f(x,y) = \sin(x + 2y)$, $z_x = \cos(x + 2y)$, and $z_y = 2\cos(x + 2y)$ and $\frac{dx}{dt} = e^t$, $\frac{dy}{dt} = -\sin(t)$ and $\frac{dz}{dt} = \cos(x + 2y)e^t - 2\cos(x + 2y)\sin(t)$.

EXAMPLE. If f is the temperature distribution in a room and $\vec{r}(t)$ is the path of the spider Shelob, then $f(\vec{r}(t))$ is the temperature, Shelob experiences at time t. The rate of change depends on the velocity $\vec{r}\,'(t)$ 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, its velocity is tangent to a level curve and the rate of change is zero.

EXAMPLE. A nicer spider called "Nabla" moves along a circle $\vec{r}(t) = (\cos(t), \sin(t))$ on a table with temperature distribution $T(x,y) = x^2 - y^3$. Find the rate of change of the temperature, "Nabla" experiences.

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

APPLICATION ENERGY CONSERVATION. If $H(x,y)$ is the energy of a system, the system moves satisfies the equations $x'(t) = H_y, y'(t) = -H_x$. For example, if $H(x,y) = y^2/2 + V(x)$ is a sum of kinetic and potential energy, then $x'(t) = y, y'(t) = V'(x)$ is equivalent to $x''(t) = -V'(x)$. In the case of the Kepler problem, we had $V(x) = Gm/|x|$. The energy H is conserved. Proof. The chain rule shows that $d/dt H(x(t), y(t)) = H_x(x,y)x'(t) + H_y(x,y)y'(t) = H_x(x,y)H_y(x,y) - H_y(x,y)H_x(x,y) = 0$.

APPLICATION: IMPLICIT DIFFERENTIATION.

From $f(x,y) = 0$ one can express y as a function of x. From $d/df(x, y(x)) = \nabla f \cdot (1, y'(x)) = f_x + f_y y' = 0$ we obtain

$$y' = -f_x/f_y.$$

If $z = g(x,y)$ is implicitely defined y $f(x,y,z) = 0$ then $f_x + f_z g_x = 0$ and $f_y + f_z g_y = 0$ so that

$$g_x = -f_x/f_z \text{ and } g_y = -f_y/f_z$$

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

PROOF OF THE CHAIN RULE.

Near any point, we can approximate f by a linear function L. It is enough to check the chain rule for linear functions $f(x,y) = ax + by - c$ and if $\vec{r}(t) = (x_0 + tu, y_0 + tv)$ is a line. Then $\frac{d}{dt} f(\vec{r}(t)) = \frac{d}{dt}(a(x_0 + tu) + b(y_0 + tv)) = au + bv$ and this is the dot product of $\nabla f = (a,b)$ with $\vec{r}\,'(t) = (u,v)$.

WHERE IS THE CHAIN RULE NEEDED?

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.

It is used to see the fact that **gradients are orthogonal to level surfaces**, appears in **change of variable** formulas and will later be used in the **fundamental theorem for line integrals**