Lecture 9: Partial derivatives

If $f(x, y)$ is a function of two variables, then the **partial derivative** $\frac{\partial}{\partial x} f(x, y)$ is defined as the derivative of the function $g(x) = f(x, y)$ with respect to x, where y is considered a constant. The partial derivative with respect to y is defined in the same way.

We use the short hand notation $f_x(x, y) = \frac{\partial}{\partial x} f(x, y)$. For iterated derivatives, the notation is similar: we write for example $f_{xy} = \frac{\partial}{\partial x} \frac{\partial}{\partial y} f$. The number $f_x(x_0, y_0)$ gives the slope of the graph sliced at (x_0, y_0) in the x direction. The second derivative f_{xx} is a measure of concavity in that direction. The meaning of f_{xy} is the rate of change of the x-slope if you move the cut along the y-axis.

The notation $\partial_x f, \partial_y f$ was introduced by Carl Gustav Jacobi. Before that, Josef Lagrange used the term "partial differences". For functions of more than two variables, the partial derivatives are defined in a similar way.

1 For $f(x, y) = x^4 - 6x^2 y^2 + y^4$, we have $f_x(x, y) = 4x^3 - 12xy^2$, $f_{xx} = 12x^2 - 12y^2$, $f_y(x, y) = -12x^2 y + 4y^3$, $f_{yy} = -12x^2 + 12y^2$ and see that $\Delta f = f_{xx} + f_{yy} = 0$. A function which satisfies $\Delta f = 0$ is also called **harmonic**. The equation $f_{xx} + f_{yy} = 0$ is an example of a **partial differential equation**: it is an equation for an unknown function $f(x, y)$ which involves partial derivatives with respect to more than one variables.

Clairaut's theorem If f_{xy} and f_{yx} are both continuous, then $f_{xy} = f_{yx}$.

Proof: we look at the equations without taking limits first. We extend the definition and say that the "Planck constant" h is positive, then $f_x(x, y) = [f(x + h, y) - f(x, y)]/h$. For $h = 0$ we define f_x as before. Comparing the two sides of the equation for fixed $h > 0$ shows

$h f_x(x, y) = f(x + h, y) - f(x, y)$ $h f_y(x, y) = f(x, y + h) - f(x, y)$.

$h^2 f_{xy}(x, y) = f(x + h, y + h) - f(x, y + h) - (f(x + h, y) - f(x, y))$ $h^2 f_{yx}(x, y) = f(x + h, y + h) - f(x + h, y) - (f(x, y + h) - f(x, y))$

Without having taken limits we established an identity which holds for all $h > 0$: the discrete derivatives f_x, f_y satisfy the relation $f_{xy} = f_{yx}$ for any $h > 0$. We could fancy it as **"quantum Clairaut"** formula. If the classical derivatives f_{xy}, f_{yx} are both continuous, it is possible to take the limit $h \to 0$. The classical Clairaut's theorem is a "classical limit". The quantum Clairaut holds however for **all** functions $f(x, y)$ of two variables. We do not even need continuity.

2 **Problem:** Find $f_{xxxxxyxxxxx}$ for $f(x) = \sin(x) + x^6 y^{10} \cos(y)$. **Hint:** you do not need to compute. Once you see it, its obvious.

3 The continuity assumption for f_{xy} is necessary. The example

$$f(x, y) = \frac{x^3 y - x y^3}{x^2 + y^2}$$

contradicts the Clairaut theorem:

$f_x(x,y) = (3x^2y - y^3)/(x^2+y^2) - 2x(x^3y - xy^3)/(x^2+y^2)^2$, $f_x(0,y) = -y$, $f_{xy}(0,0) = -1$, $f_y(x,y) = (x^3 - 3xy^2)/(x^2+y^2) - 2y(x^3y - xy^3)/(x^2+y^2)^2$, $f_y(x,0) = x$, $f_{y,x}(0,0) = 1$.

> An equation for an unknown function $f(x,y)$ which involves partial derivatives with respect to at least two different variables is called a **partial differential equation**. We abbreviate PDE. If only the derivative with respect to one variable appears, it is an **ordinary differential equation** abbreviated ODE.

Here are examples of partial differential equations. You **have** to know the first four in the same way than a chemist has to know what $H_2O, CO_2, CH_4, NaCl$ is. You might be surprised how often you encounter them also in domains different from physics, like finance (market prediction) or biology (diffusion). These equations also can be looked at on discrete structures like networks.

4 The **wave equation** $f_{tt}(t,x) = f_{xx}(t,x)$ governs the motion of light or sound. The function $f(t,x) = \sin(x-t) + \sin(x+t)$ satisfies the wave equation.

5 The **heat equation** $f_t(t,x) = f_{xx}(t,x)$ describes diffusion of heat or spread of an epidemic. The function $f(t,x) = \frac{1}{\sqrt{t}} e^{-x^2/(4t)}$ satisfies the heat equation.

6 The **Laplace equation** $f_{xx} + f_{yy} = 0$ determines the shape of a membrane. The function $f(x,y) = x^3 - 3xy^2$ is an example satisfying the Laplace equation.

7 The **advection equation** $f_t = f_x$ is used to model transport in a wire. The function $f(t,x) = e^{-(x+t)^2}$ satisfies the advection equation.

8 The **eiconal equation** $f_x^2 + f_y^2 = 1$ is used to see the evolution of wave fronts in optics. The function $f(x,y) = \cos(x) + \sin(y)$ satisfies the eiconal equation.

9 The **Burgers equation** $f_t + ff_x = f_{xx}$ describes waves at the beach which break. The function $f(t,x) = \frac{x}{t} \frac{\sqrt{\frac{1}{t}} e^{-x^2/(4t)}}{1 + \sqrt{\frac{1}{t}} e^{-x^2/(4t)}}$ satisfies the Burgers equation.

10 The **KdV equation** $f_t + 6ff_x + f_{xxx} = 0$ models **water waves** in a narrow channel. The function $f(t,x) = \frac{a^2}{2} \cosh^{-2}(\frac{a}{2}(x - a^2t))$ satisfies the KdV equation.

11 The **Schrödinger equation** $f_t = \frac{i\hbar}{2m} f_{xx}$ is used to describe a **quantum particle** of mass m. The function $f(t,x) = e^{i(kx - \frac{\hbar}{2m}k^2 t)}$ solves the Schrödinger equation. [Here $i^2 = -1$ is the imaginary i and $\hbar$ is the **Planck constant** $\hbar \sim 10^{-34} Js$.]

Here are the graphs of the solutions of the equations. Can you match them with the PDE's?
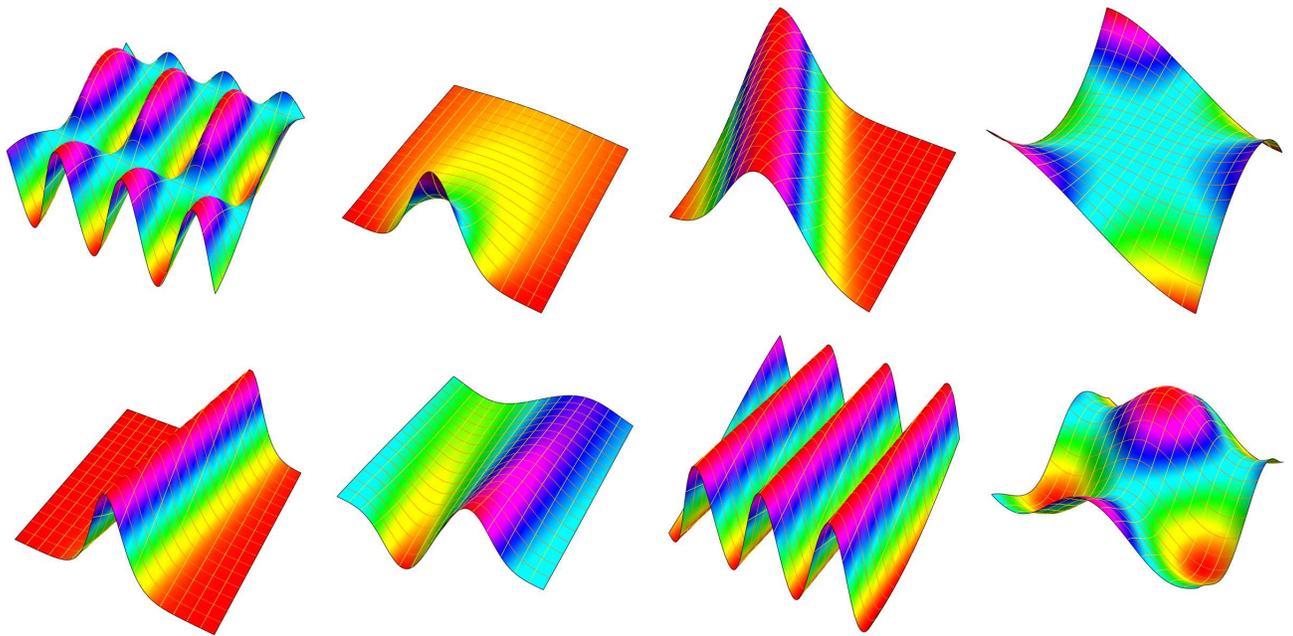
Notice that in all these examples, we have just given one possible solution to the partial differential equation. There are in general many solutions and only additional conditions like initial or boundary conditions determine the solution uniquely. If we know $f(0, x)$ for the Burgers equation, then the solution $f(t, x)$ is determined. A course on partial differential equations would show you how to get the solution.

Paul Dirac once said: "A great deal of my work is just **playing with equations** and seeing what they give. I don't suppose that applies so much to other physicists; I think it's a peculiarity of myself that I like to play about with equations, just **looking for beautiful mathematical relations** which maybe don't have any physical meaning at all. Sometimes they do." Dirac discovered a PDE describing the electron which is consistent both with quantum theory and special relativity. This won him the Nobel Prize in 1933. Dirac's equation could have two solutions, one for an electron with positive energy, and one for an electron with negative energy. Dirac interpreted the later as an **antiparticle**: the existence of antiparticles was later confirmed. We will not learn here to find solutions to partial differential equations. But you should be able to verify that a given function is a solution of the equation.

Homework

1 Verify that $f(t, x) = \sin(t + x) + \cos(\sin(t + x))$ is a solution of the transport equation $f_t(t, x) = f_x(t, x)$.

2 Verify that $f(x, y) = \sin(x)(\cos(3y) + \sin(3y))$ satisfies the **Klein Gordon equation** $u_{xx} - u_{yy} = 8u$. This PDE is useful in quantum mechanics. **Challenge:** Verify that $4 \arctan(e^{(x-t/2)2/\sqrt{3}})$ satisfies the **Sine Gordon equation** $u_{tt} - u_{xx} = -\sin(u)$. If you can do that without technology put it on a separate page and give it to Oliver. It is not for credit.

3 Verify that for any constant b, the function $f(x,t) = e^{-bt}\sin(x + t)$ satisfies the driven transport equation $f_t(x,t) = f_x(x,t) - bf(x,t)$ It is sometimes also called the **advection equation** with damping b.

4 The differential equation

$$f_t = f - xf_x - x^2 f_{xx}$$

is a version of the **infamous Black-Scholes equation**. Here $f(x,t)$ is the prize of a **call option** and x the stock prize ad t is time. Find a function $f(x,t)$ solving it which depends both on x and t. Hint: look first for functions only involve one variable.

5 The partial differential equation $f_t + ff_x = f_{xx}$ is called **Burgers equation** and describes waves at the beach. In higher dimensions, it leads to the Navier-Stokes equation which are used to describe the weather. Verify that

$$f(t,x)\frac{\left(\frac{1}{t}\right)^{3/2}xe^{-\frac{x^2}{4t}}}{\sqrt{\frac{1}{t}e^{-\frac{x^2}{4t}}+1}}$$

solves the Burgers equation.

Remark. You better use technology. Here is an example on how to check that a function is a solution of the heat equation in Mathematica:

```
f[t_,x_]:=(1/Sqrt[t])*Exp[-x^2/(4t)];
Simplify[D[f[t,x],t] == D[f[t,x],{x,2}]]
```

And here is the function

```
(1/t)^(3/2)*x*Exp[-x^2/(4t)]/((1/t)^(1/2)*Exp[-x^2/(4t)]+1);
```

Lecture 10: Linearization

In single variable calculus, you have seen the following definition:

> The **linear approximation** of $f(x)$ at a point a is the linear function
>
> $$L(x) = f(a) + f'(a)(x - a) \ .$$

If you have seen **Taylor series**, this is part of the series $f(x) = \sum_{k=0}^{\infty} f^{(k)}(a)(x-a)^k/k!$ but only considering the $k = 0$ and $k = 1$ term. Think about the linear approximation L as a function and not as a graph because we will also look at this also for functions of three variables, where we can not draw graphs.
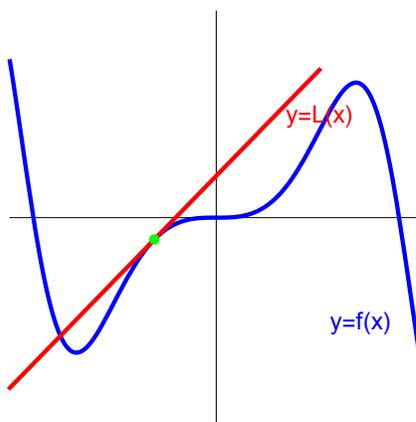
The graph of the function L is close to the graph of f at a. What about higher dimensions?

> The **linear approximation** of $f(x, y)$ at (a, b) is the linear function
>
> $$L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) \ .$$
>
> The **linear approximation** of a function $f(x, y, z)$ at (a, b, c) is
>
> $$L(x, y, z) = f(a, b, c) + f_x(a, b, c)(x - a) + f_y(a, b, c)(y - b) + f_z(a, b, c)(z - c) \ .$$

Using the **gradient**

$$\nabla f(x, y) = \langle f_x, f_y \rangle, \qquad \nabla f(x, y, z) = \langle f_x, f_y, f_z \rangle \ ,$$

the linearization can be written more compactly as

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

How do we justify the linearization? If the second variable $y = b$ is fixed, we have a one-dimensional situation, where the only variable is x. Now $f(x, b) = f(a, b) + f_x(a, b)(x - a)$ is the linear approximation. Similarly, if $x = x_0$ is fixed y is the single variable, then $f(x_0, y) = f(x_0, y_0) + f_y(x_0, y_0)(y - y_0)$. Knowing the linear approximations in both the x and y variables, we can get the general linear approximation by $f(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$.

1 What is the linear approximation of the function $f(x, y) = \sin(\pi x y^2)$ at the point $(1, 1)$? We have $\langle f_x(x, y), f_y(x, y) \rangle = \langle \pi y^2 \cos(\pi x y^2), 2xy\pi \cos(\pi x y^2) \rangle$ which is at the point $(1, 1)$ equal to $\nabla f(1, 1) = \langle \pi \cos(\pi), 2\pi \cos(\pi) \rangle = \langle -\pi, -2\pi \rangle$. The function is $L(x, y) = 0 + (-\pi)(x - 1) - 2\pi(y - 1)$.

2 Linearization can be used to estimate functions near a point. In the previous example,

$$-0.0095 = f(1+0.01, 1+0.01) \sim L(1+0.01, 1+0.01) = -\pi 0.01 - 2\pi 0.01 = -3\pi/100 = -0.00942 \ .$$

3 Here is an example in three dimensions: find the linear approximation to $f(x, y, z) = xy + yz + zx$ at the point $(1, 1, 1)$. Since $f(1, 1, 1) = 3$, and $\nabla f(x, y, z) = (y + z, x + z, y + x)$, $\nabla f(1, 1, 1) = (2, 2, 2)$. we have $L(x, y, z) = f(1, 1, 1) + (2, 2, 2) \cdot (x - 1, y - 1, z - 1) = 3 + 2(x - 1) + 2(y - 1) + 2(z - 1) = 2x + 2y + 2z - 3$.

4 Estimate $f(0.01, 24.8, 1.02)$ for $f(x, y, z) = e^x \sqrt{y} z$.
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.

5 Find the tangent line to the graph of the function $g(x) = x^2$ at the point $(2, 4)$.
Solution: the level curve $f(x, y) = y - x^2 = 0$ is the graph of a function $g(x) = x^2$ and the tangent at a point $(2, g(2)) = (2, 4)$ is obtained by computing the gradient $\langle a, b \rangle = \nabla f(2, 4) = \langle -g'(2), 1 \rangle = \langle -4, 1 \rangle$ and forming $-4x + y = d$, where $d = -4 \cdot 2 + 1 \cdot 4 = -4$. The answer is $\boxed{-4x + y = -4}$ which is the line $y = 4x - 4$ of slope 4.

6 The **Barth surface** is defined as the level surface $f = 0$ of

$$\begin{aligned} f(x, y, z) &= (3 + 5t)(-1 + x^2 + y^2 + z^2)^2(-2 + t + x^2 + y^2 + z^2)^2 \\ &+ 8(x^2 - t^4 y^2)(-(t^4 x^2) + z^2)(y^2 - t^4 z^2)(x^4 - 2x^2 y^2 + y^4 - 2x^2 z^2 - 2y^2 z^2 + z^4) \ , \end{aligned}$$

where $t = (\sqrt{5} + 1)/2$ is a constant called the **golden ratio**. If we replace t with $1/t = (\sqrt{5} - 1)/2$ we see the surface to the middle. For $t = 1$, we see to the right the surface $f(x, y, z) = 8$. Find the tangent plane of the later surface at the point $(1, 1, 0)$. **Answer:** We have $\nabla f(1, 1, 0) = \langle 64, 64, 0 \rangle$. The surface is $x + y = d$ for some constant d. By plugging in $(1, 1, 0)$ we see that $x + y = 2$.
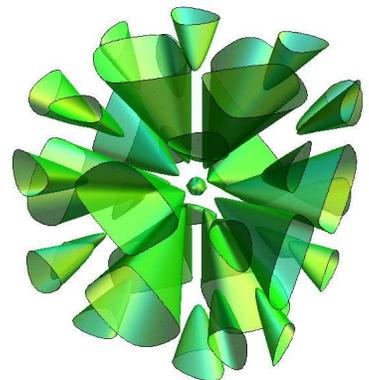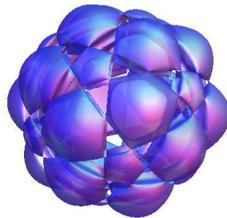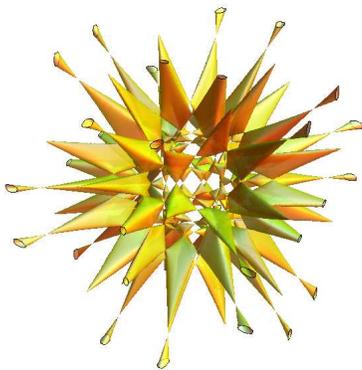
7 The quartic surface

$$f(x, y, z) = x^4 - x^3 + y^2 + z^2 = 0$$

is called the **piriform**. What is the equation for the tangent plane at the point $P = (2, 2, 2)$ of this pair shaped surface? We get $\langle a, b, c \rangle = \langle 20, 4, 4 \rangle$ and so the equation of the plane $20x + 4y + 4z = 56$, where we have obtained the constant to the right by plugging in the point $(x, y, z) = (2, 2, 2)$.
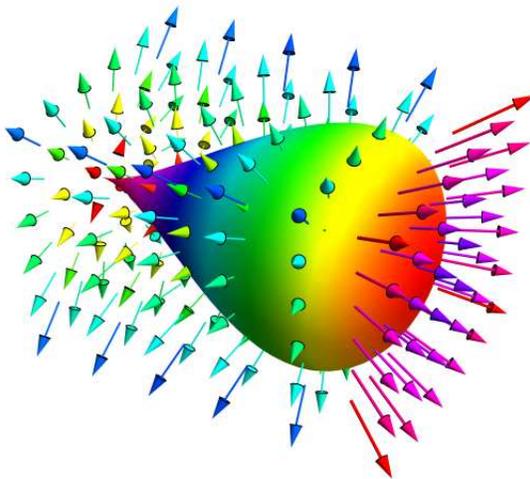
Remark: Sometimes, **differentials** are used to describe linearizations. Try to avoid it or at least use it for intuition only. Like Newtons "fluxions", the Leibniz "differentials" are outdated. Its is a good intuitive notion but its also easy to abuse it. The linearlization of a function f at a point is just a **linear function** L in the same number of variables. There is a modern notion of **differential forms** which is well defined, but which needs some multi-linear algebra for definition. The notion of **infinitesimal small quantities** has been clarified within non-standard analysis but that theory needs a solid logic background. The notion of "differentials" comes from a time, when calculus was in the early development. To see why the notion "differential" is a bit murky, try to find out what the definition of "differential" is: you find notions like "change in the linearization of a function" or "infinitesimals". Both are "foggy terminology". To add to the confusion, expressions like dx are also called a "differentials". They appear in **total differentials** $df = f_x dx + f_y dy$ which can make sense as an intuitive shortcut for the chain rule $df(x(t), y(t))/dt = f_x(x(t), y(t))dx(t)/dt + f_y(x(t), y(t))dy(t)/dt$ as the later can be multiplied by dt to get the differential expression. (The chain rule will come later). The expression dx also appears in integrals $\int_0^\pi \sin(x) \, dx$ but there, it is just **notation** to indicate with respect to which variable we integrate. Mathematica for example writes $Integrate[Sin[x], x, 0, Pi]$. Leibniz used $\int_a^b f(x) \, dx$ because it is close to the Riemann sum $\sum_i f(x_i)dx_i$ notation, in which $dx_i = x_{i+1} - x_i$ are **differences**. Leibniz notation stands for a limit of Riemann sums which is well defined if f is continuous. But expressions like dx, dt alone are not defined without considerably more theory. As a **notation** it can be useful in the **separation of variable technique**: when solving $f' = f$ we write $df/dx = f$ leading to $df/f = dx$. Integration produces $\log(f) = x + c$, leading to $f = e^{x+c}$. This separation of variable technique can be backed up with theory.

Homework

1 Given $f(x, y) = \sin(x) - 3yx/\pi$. Estimate $f(\pi + 0.01, \pi - 0.03)$ using linearization

2 Estimate $2'000'000^{1/7}$ using linear approximation of $f(x) = x^{1/7}$ near $x_0 = 8^7$.

Source: http://www.cdnpapermoney.com

3 Estimate $f(0.001, 0.9999)$ for $f(x, y) = \cos(\pi y) + \sin(x + \pi y)$ using linearization.

4 Find the linear approximation $L(x, y)$ of the function

$$f(x, y) = \sqrt{10 - x^2 - 5y^2}$$

at $(2, 1)$ and use it to estimate $f(1.95, 1.04)$.

5 Estimate $(99^3 * 101^2)$ by linearizing the function $f(x, y) = x^3 y^2$ at $(100, 100)$. What is the difference between $L(100, 100)$ and $f(100, 100)$?

Lecture 11: Chain rule

If f and g are functions of a variable t, the **single variable chain rule** tells us that $d/dt f(g(t)) = f'(g(t))g'(t)$. For example, $d/dt \sin(\log(t)) = \cos(\log(t))/t$.
It can be proven by linearizing the functions f and g and verifying the chain rule in the linear case. The **chain rule** is also useful:

To find $\arccos'(x)$ for example, we differentiate $x = \cos(\arccos(x))$ to get $1 = d/dx \cos(\arccos(x)) = -\sin(\arccos(x)) \arccos'(x) = -\sqrt{1 - \cos^2(\arccos(x))} \arccos'(x) = -\sqrt{1 - x^2} \arccos'(x)$ so that $\arccos'(x) = -1/\sqrt{1 - x^2}$.

> Define the **gradient** $\nabla f(x, y) = \langle f_x(x, y), f_y(x, y) \rangle$ or $\nabla f(x, y, z) = \langle f_x(x, y, z), f_y(x, y, z), f_z(x, y, z) \rangle$.

If $\vec{r}(t)$ is curve and f is a function of several variables we get a function $t \mapsto f(\vec{r}(t))$ of one variable. Similarly, if $\vec{r}(t)$ is a parametrization of a planar curve f is a function of two variables, then $t \mapsto f(\vec{r}(t))$ is a function of one variable.

> The **multivariable chain rule** is $\frac{d}{dt} f(\vec{r}(t)) = \nabla f(\vec{r}(t)) \cdot \vec{r}'(t)$.

Proof. When written out in two dimensions, it is

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

The identity

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

holds for every $h > 0$. The left hand side converges to $\frac{d}{dt} f(x(t), y(t))$ in the limit $h \to 0$ and the right hand side to $f_x(x(t), y(t))x'(t) + f_y(x(t), y(t))y'(t)$ using the single variable chain rule twice. Here is the proof of the later, when we differentiate f with respect to t and y is treated as a constant:

$$\frac{f(\boxed{x(t+h)}) - f(x(t))}{h} = \frac{[f(\boxed{x(t) + \boxed{(x(t+h)-x(t))}}) - f(x(t))]}{\boxed{[x(t+h)-x(t)]}} \cdot \frac{\boxed{[x(t+h)-x(t)]}}{h} .$$

Write $H(t) = \boxed{x(t+h)-x(t)}$ in the first part on the right hand side.

$$\frac{f(x(t+h)) - f(x(t))}{h} = \frac{[f(x(t) + H) - f(x(t))]}{H} \cdot \frac{x(t+h) - x(t)}{h} .$$

As $h \to 0$, we also have $H \to 0$ and the first part goes to $f'(x(t))$ and the second factor to $x'(t)$.

> **1** We move on a circle $\vec{r}(t) = \langle \cos(t), \sin(t) \rangle$ on a table with temperature distribution $f(x, y) = x^2 - y^3$. Find the rate of change of the temperature $\nabla f(x, y) = (2x, -3y^2)$, $\vec{r}'(t) = (-\sin(t), \cos(t))$ $d/dt f(\vec{r}(t)) = \nabla T(\vec{r}(t)) \cdot \vec{r}'(t) = (2\cos(t), -3\sin(t)^2) \cdot (-\sin(t), \cos(t)) = -2\cos(t)\sin(t) - 3\sin^2(t)\cos(t)$.

From $f(x, y) = 0$, one can express y as a function of x, at least near a point where f_y is not zero. From $d/df(x, y(x)) = \nabla f \cdot (1, y'(x)) = f_x + f_y y' = 0$, we obtain $y' = -f_x/f_y$. Even so, we do not know $y(x)$, we can compute its derivative! Implicit differentiation works also in three variables. The equation $f(x, y, z) = c$ defines a surface. Near a point where f_z is not zero, the surface can be described as a graph $z = z(x, y)$. We can compute the derivative z_x without actually knowing the function $z(x, y)$. To do so, we consider y a fixed parameter and compute, using the chain rule

$$f_x(x, y, z(x, y))1 + f_z(x, y)z_x(x, y) = 0$$

so that $z_x(x, y) = -f_x(x, y, z)/f_z(x, y, z)$. This works at points where f_z is not zero.

> **2** The surface $f(x, y, z) = x^2 + y^2/4 + z^2/9 = 6$ is an ellipsoid. Compute $z_x(x, y)$ at the point $(x, y, z) = (2, 1, 1)$.
> **Solution:** $z_x(x, y) = -f_x(2, 1, 1)/f_z(2, 1, 1) = -4/(2/9) = -18$.

The chain rule is powerful because it implies other differentiation rules like the addition, product and quotient rule in one dimensions: $f(x, y) = x + y, x = u(t), y = v(t), d/dt(x+y) = f_x u' + f_y v' = u' + v'$.
$f(x, y) = xy, x = u(t), y = v(t), d/dt(xy) = f_x u' + f_y v' = vu' + uv'$.
$f(x, y) = x/y, x = u(t), y = v(t), d/dt(x/y) = f_x u' + f_y v' = u'/y - v'u/v^2$.
As in one dimensions, the chain rule follows from linearization. If f is a linear function $f(x, y) = ax + by - c$ and if the curve $\vec{r}(t) = \langle x_0 + tu, y_0 + tv \rangle$ parametrizes 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 $\vec{r}\,'(t) = (u, v)$. Since the chain rule only refers to the derivatives of the functions which agree at the point, the chain rule is also true for general functions.

Homework

1 You know that $d/dt f(\vec{r}(t)) = 5$ at $t = 0$ if $\vec{r}(t) = \langle t, t \rangle$ and $d/dt f(\vec{r}(t)) = 9$ at $t = 0$. $\vec{r}(t) = \langle t, -t \rangle$. Find the gradient of f at $(0, 0)$.

2 The pressure in the space at the position (x, y, z) is $p(x, y, z) = x^2 + y^2 - z^3$ and the trajectory of an observer is the curve $\vec{r}(t) = \langle t, t, 1/t \rangle$. Using the chain rule, compute the rate of change of the pressure the observer measures at time $t = 2$.

3 The chain rule is closely related to linearization as it could be proven by linearlization. Lets get back to linearlization a bit: A farm costs $f(x, y)$, where x is the number of cows and y is the number of ducks. There are 10 cows and 20 ducks and $f(10, 20) = 1000000$. We know that $f_x(x, y) = 2x$ and $f_y(x, y) = y^2$ for all x, y. Estimate $f(12, 19)$.
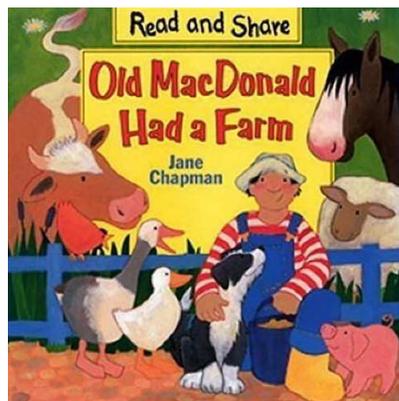
P.S. In the fall of 2013, Oliver made a song out of this:

"Old MacDonald had a million dollar farm, E-I-E-I-O,
and on that farm he had $x = 10$ cows, E-I-E-I-O,
and on that farm he had $y = 20$ ducks, E-I-E-I-O,
with $f_x = 2x$ here and $f_y = y^2$ there,
and here two cows more, and there a duck less,
how much does the farm cost now, E-I-E-I-O?"

4 Find, using implicit differentiation the derivative $d/dx \ \text{arctanh}(x)$, where

$$\tanh(x) = \sinh(x)/\cosh(x) \, .$$

The **hyperbolic sine** and **hyperbolic cosine** are defined as are $\sinh(x) = (e^x - e^{-x})/2$ and $\cosh(x) = (e^x + e^{-x})/2$. We have $\sinh' = \cosh$ and $\cosh' = \sinh$ and $\cosh^2(x) - \sinh^2(x) = 1$.

5 The equation $f(x, y, z) = e^{xyz} + z = 1 + e$ implicitly defines z as a function $z = g(x, y)$ of x and y. Find formulas (in terms of x,y and z) for $g_x(x, y)$ and $g_y(x, y)$. Estimate $g(1.01, 0.99)$ using linear approximation.

Lecture 12: Gradient

The **gradient** of a function $f(x, y)$ is defined as

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y) \rangle .$$

For functions of three variables, define

$$\nabla f(x, y, z) = \langle f_x(x, y, z), f_y(x, y, z), f_z(x, y, z) \rangle .$$

The symbol ∇ is spelled "Nabla" and named after an Egyptian or Assyrian harp. Early on, the name "Atled" was suggested. But the textbook of 1901 of Gibbs in Nabla was used, was too persuasive. Here is a very important fact:

Gradient Theorem: Gradients are orthogonal to level curves and level surfaces.

Proof. Every curve $\vec{r}(t)$ on the level curve or level surface satisfies $\frac{d}{dt} f(\vec{r}(t)) = 0$. By the chain rule, $\nabla f(\vec{r}(t))$ is perpendicular to the tangent vector $\vec{r}'(t)$.

Because $\vec{n} = \nabla f(p, q) = \langle a, b \rangle$ is perpendicular to the level curve $f(x, y) = c$ through (p, q), the equation for the tangent line is $ax + by = d$, $a = f_x(p, q)$, $b = f_y(p, q)$, $d = ap + bq$. Compactly written, this is

$$\nabla f(\vec{x}_0) \cdot (\vec{x} - \vec{x}_0) = 0$$

and means that the gradient of f is perpendicular to any vector $(\vec{x} - \vec{x}_0)$ in the plane. It is one of the most important statements in multivariable calculus as it gives a crucial link between calculus and geometry. The just mentioned gradient theorem is also useful. We can immediately compute tangent planes and tangent lines, without linearization!
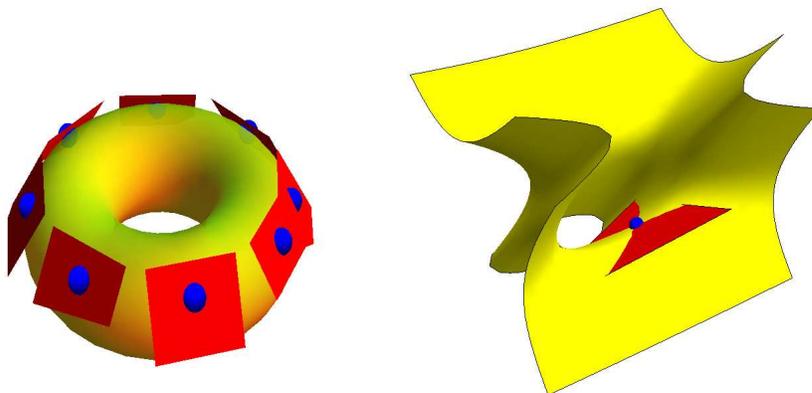

1 Compute the tangent plane to the surface $3x^2y + z^2 - 4 = 0$ at the point $(1, 1, 1)$. **Solution:** $\nabla f(x, y, z) = \langle 6xy, 3x^2, 2z \rangle$. And $\nabla f(1, 1, 1) = \langle 6, 3, 2 \rangle$. The plane is $6x + 3y + 2z = d$ where d is a constant. We can find the constant d by plugging in a point and get $6x + 3y + 2z = 11$.

2 **Problem:** reflect the ray $\vec{r}(t) = \langle 1 - t, -t, 1 \rangle$ at the surface

$$x^4 + y^2 + z^6 = 6 .$$

Solution: $\vec{r}(t)$ hits the surface at the time $t = 2$ in the point $(-1, -2, 1)$. The velocity vector in that ray is $\vec{v} = \langle -1, -1, 0 \rangle$. The normal vector at this point is $\nabla f(-1, -2, 1) = \langle -4, -4, 6 \rangle = \vec{n}$. The reflected vector is $R(\vec{v}) = 2\text{Proj}_{\vec{n}}(\vec{v}) - \vec{v}$. We have $\text{Proj}_{\vec{n}}(\vec{v}) = 8/68 \langle -4, -4, 6 \rangle$. Therefore, the reflected ray is $\vec{w} = (4/17)\langle -4, -4, 6 \rangle - \langle -1, -1, 0 \rangle$.
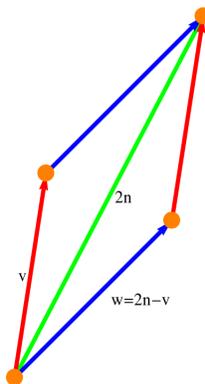
> If f is a function of several variables and $\vec{v}$ is a unit vector then $D_{\vec{v}}f = \nabla f \cdot \vec{v}$ is called the **directional derivative** of f in the direction $\vec{v}$.

The name directional derivative is related to the fact that every unit vector gives a direction. If $\vec{v}$ is a unit vector, then the chain rule tells us $\frac{d}{dt}D_{\vec{v}}f = \frac{d}{dt}f(x + t\vec{v})$.

The directional derivative tells us how the function changes when we move in a given direction. Assume for example that $T(x, y, z)$ is the temperature at position (x, y, z). If we move with velocity $\vec{v}$ through space, then $D_{\vec{v}}T$ tells us at which rate the temperature changes for us. If we move with velocity $\vec{v}$ on a hilly surface of height $h(x, y)$, then $D_{\vec{v}}h(x, y)$ gives us the slope we drive on.

3 If $\vec{r}(t)$ is a curve with velocity $\vec{r}\,'(t)$ and the speed is 1, then $D_{\vec{r}\,'(t)}f = \nabla f(\vec{r}(t)) \cdot \vec{r}\,'(t)$ is the temperature change, one measures at $\vec{r}(t)$. The chain rule told us that this is $d/dt f(\vec{r}(t))$.

4 For $\vec{v} = (1, 0, 0)$, then $D_{\vec{v}}f = \nabla f \cdot v = f_x$, the directional derivative is a generalization of the partial derivatives. It measures the rate of change of f, if we walk with unit speed into that direction. But as with partial derivatives, it is a **scalar**.

> The directional derivative satisfies $\boxed{|D_{\vec{v}}f| \leq |\nabla f||\vec{v}|}$ because $\nabla f \cdot \vec{v} = |\nabla f||\vec{v}||\cos(\phi)| \leq |\nabla f||\vec{v}|$.

> The direction $\vec{v} = \nabla f/|\nabla f|$ is the direction, where f **increases** most. It is the direction of **steepest ascent**.

If $\vec{v} = \nabla f/|\nabla f|$, then the directional derivative is $\nabla f \cdot \nabla f/|\nabla f| = |\nabla f|$. This means f **increases**, if we move into the direction of the gradient. The slope in that direction is $|\nabla f|$.

5 You are on a trip in a air-ship over Cambridge at $(1, 2)$ and you want to avoid a thunderstorm, a region of low pressure. The pressure is given by a function $p(x, y) = x^2 + 2y^2$. In which direction do you have to fly so that the pressure change is largest? **Solution:** The gradient $\nabla p(x, y) = \langle 2x, 4y \rangle$ at the point $(1, 2)$ is $\langle 2, 8 \rangle$. Normalize to get the direction $\langle 1, 4 \rangle/\sqrt{17}$.

The directional derivative has the same properties than any derivative: $D_v(\lambda f) = \lambda D_v(f)$, $D_v(f + g) = D_v(f) + D_v(g)$ and $D_v(fg) = D_v(f)g + fD_v(g)$.

We will see later that points with $\nabla f = \vec{0}$ are candidates for **local maxima** or **minima** of f. Points (x, y), where $\nabla f(x, y) = (0, 0)$ are called **critical points** and help to understand the function f.

6 The Matterhorn is a mountain in Switzerland with an altitude of 4'478 meters. Even so there are quite many climbing accidents at the Matterhorn, it does not stop you from trying an ascent. In suitable units on the ground, the height $f(x, y)$ of the Matterhorn is approximated by the function $f(x, y) = 4000 - x^2 - y^2$. At height $f(-10, 10) = 3800$, at the point $(-10, 10, 3800)$, you rest. The climbing route continues into the south-east direction $v = \langle 1, -1 \rangle/\sqrt{2}$. Calculate the rate of change in that direction. We have $\nabla f(x, y) = \langle -2x, -2y \rangle$, so that $\langle 20, -20 \rangle \cdot \langle 1, -1 \rangle/\sqrt{2} = 40/\sqrt{2}$. This is a place, with a ladder, where you climb $40/\sqrt{2}$ meters up when advancing 1m forward.
The rate of change in all directions is zero if and only if $\nabla f(x, y) = 0$: if $\nabla f \neq \vec{0}$, we can choose $\vec{v} = \nabla f/|\nabla f|$ and get $D_{\nabla f}f = |\nabla f|$.

7 Assume we know $D_v f(1, 1) = 3/\sqrt{5}$ and $D_w f(1, 1) = 5/\sqrt{5}$, where $v = \langle 1, 2 \rangle/\sqrt{5}$ and $w = \langle 2, 1 \rangle/\sqrt{5}$. Find the gradient of f. Note that we do not know anything else about the function f.
Solution: Let $\nabla f(1, 1) = \langle a, b \rangle$. We know $a + 2b = 3$ and $2a + b = 5$. This allows us to get $a = 7/3, b = 1/3$.

Homework

1 Find the directional derivative $D_{\vec{v}}f(2, 1) = \nabla f(2, 1) \cdot \vec{v}$ into the direction $\vec{v} = \langle 3, 4 \rangle/5$ for the function $f(x, y) = 2 + x^5y + y^3 + y$.

2 A surface $x^2 + y^2 - z = 1$ radiates light away. It can be parametrized as $\vec{r}(x, y) = \langle x, y, x^2 + y^2 - 1 \rangle$. Find the parametrization of the wave front $\vec{r}(x, y) + \vec{n}(x, y)$, which is distance 1 from the surface. Here $\vec{n}$ is a unit vector normal to the surface.

3 Assume $f(x, y) = 1 - x^2 + y^2$. Compute the directional derivative $D_{\vec{v}}(x, y)$ at $(0, 0)$, where $\vec{v} = \langle \cos(t), \sin(t) \rangle$ is a unit vector. Now compute

$$D_v D_v f(x, y)$$

at $(0, 0)$, for any unit vector. For which directions is this **second directional derivative** positive?

4 The **Kitchen-Rosenberg formula** gives the curvature of a level curve $f(x, y) = c$ as

$$\kappa = \frac{f_{xx} f_y^2 - 2 f_{xy} f_x f_y + f_{yy} f_x^2}{(f_x^2 + f_y^2)^{3/2}}$$

Use this formula to find the curvature of the ellipsoid $f(x, y) = x^2 + 2y^2 = 1$ at the point $(1, 0)$.

This formula is useful in computer vision. If you want to derive the formula, you can check that the angle

$$g(x, y) = \arctan(f_y / f_x)$$

of the gradient vector has κ as the directional derivative in the direction $\vec{v} = \langle -f_y, f_x \rangle / \sqrt{f_x^2 + f_y^2}$ tangent to the curve.

5 One can fin the maximum of a function numerically by moving in the direction of the gradient. This is called the **steepest ascent method**. You start at a point (x_0, y_0) then move in the direction of the gradient for some time c to be at $(x_1, y_1) = (x_0, y_0) + c\nabla f(x_0, y_0)$. Repeat to $(x_2, y_2) = (x_1, y_1) + c\nabla f(x_1, y_1)$ etc. It can be a bit difficult if the function has a flat ridge like in the **Rosenbrock function**

$$f(x, y) = 1 - (1 - x)^2 - 100(y - x^2)^2 .$$

Plot the contour map of this function on $-0.6 \leq x \leq 1, -0.1 \leq y \leq 1.1$, then and find the directional derivative at $(1/5, 0)$ in the direction $(1, 1)/\sqrt{2}$.