Chapter 4. Extrema and Double integrals

Section 4.1: Extrema, second derivative test

An important problem in multi-variable calculus is to **extremize** functions $f(x,y)$ of two variables. As in one dimensions, in order to look for maxima or minima, we consider points, where the "derivative" is zero.

A point (a,b) in the plane is called a **critical point** of a function $f(x,y)$ if $\nabla f(a,b) = \langle 0,0\rangle$. Critical points are candidates for extrema because at critical points, all directional derivatives $D_{\vec{v}} f = \nabla f \cdot \vec{v}$ are zero.

Note that that we do not include points, where f or its derivative is not defined in the set of critical points. We usually assume that a function is arbitrarily often differentiable. Points where the function is not smooth are part of the boundary of the region and would have to be studied seperatly.

Examples:
1) $f(x,y) = x^4 + y^4 - 4xy + 2$. The gradient is $\nabla f(x,y) = \langle 4(x^3 - y), 4(y^3 - x)\rangle$ with critical points $(0,0), (1,1), (-1,-1)$.

2) $f(x,y) = \sin(x^2 + y) + y$. The gradient is $\nabla f(x,y) = \langle 2x\cos(x^2+y), \cos(x^2+y)+1\rangle$. For a critical points, we must have $x = 0$ and $\cos(y) + 1 = 0$ which means $\pi + k2\pi$. The critical points are at $(0,\pi), (0,3\pi), \ldots$.

3) The graph of $f(x,y) = (x^2+y^2)e^{-x^2-y^2}$ looks like a volcano. The gradient $\nabla f = \langle 2x - 2x(x^2+y^2), 2y - 2y(x^2+y^2)\rangle e^{-x^2-y^2}$ vanishes at $(0,0)$ and on the circle $x^2+y^2 = 1$. There is a continuum of critical points.

4) The function $f(x,y) = y^2/2 - g\cos(x)$ is the energy of the pendulum. We have $\nabla f = (y, -g\sin(x)) = \langle(0,0\rangle$ for $x = 0, \pi, 2\pi, \ldots, y = 0$. These points are equilibrium points, where the pendulum is at rest.

5) The function $f(x,y) = a\log(y) - by + c\log(x) - dx$ is left invariant by the flow of the Volterra Lodka differential equation $\dot{x} = ax - bxy, \dot{y} = -cy + dxy$. The point $(c/d, a/b)$ is a critical point.

6) The function $f(x,y) = |x| + |y|$ is smooth on the first quadrant. It does not have critical points there. The function has a minimum at $(0,0)$ but it is not in the domain, where f and ∇f are defined.

For functions in one dimension, we needed $f'(x) = 0, f''(x) > 0$ to have a local minimum, $f'(x) = 0, f''(x) < 0$ for a local maximum. If $f'(x) = 0, f''(x) = 0$, then the critical point was undetermined and could be a maximum like for $f(x) = -x^4$, or a minimum like for $f(x) = x^4$

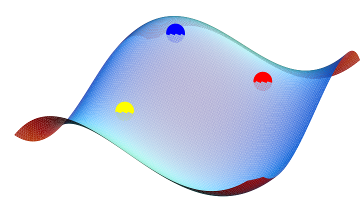
or a flat inflection point like for $f(x) = x^3$.

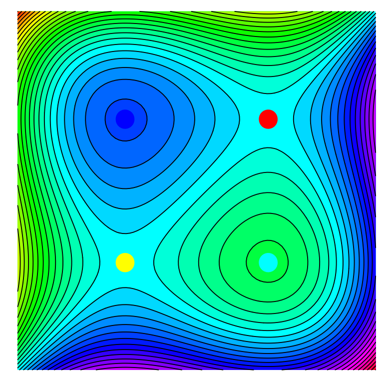
Let now $f(x,y)$ be a function of two variables with a critical point (a,b). Define $D = f_{xx}f_{yy} - f_{xy}^2$. It is called the **discriminant**. You might want to remember it with the help of the **Hessian matrix** $H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$ because the discriminant is the **determinant** of H. Here is the important **second derivative test**.

Assume (a,b) is a critical point for $f(x,y)$.
If $D > 0$ and $f_{xx}(a,b) > 0$ then (a,b) is a local minimum.
If $D > 0$ and $f_{xx}(a,b) < 0$ then (a,b) is a local maximum.
If $D < 0$ then (a,b) is a saddle point.

In the case $D = 0$, we would need higher derivatives to determine the nature of the critical point.

Example: The function $f(x,y) = x^3/3 - x - (y^3/3 - y)$ has a graph which looks like a "napkin". It has the gradient $\nabla f(x,y) = \langle x^2 - 1, -y^2 + 1\rangle$. There are 4 critical points $(1,1)$,$(-1,1)$,$(1,-1)$ and $(-1,-1)$. The Hessian matrix which includes all partial derivatives is $H = \begin{bmatrix} 2x & 0 \\ 0 & -2y \end{bmatrix}$.
For $(1,1)$ we have $D = -4$ and so a saddle point,
For $(-1,1)$ we have $D = 4, f_{xx} = -2$ and so a local maximum,
For $(1,-1)$ we have $D = 4, f_{xx} = 2$ and so a local minimum.
For $(-1,-1)$ we have $D = -4$ and so a saddle point. The function has a local maximum, a local minimum as well as 2 saddle points.





To determine the maximum or minimum of $f(x,y)$ on a domain, determine all critical points **in the interior the domain**, and compare their values with maxima or minima **at the boundary**. We will see next time how to get extrema on the boundary.

Example: Find the maximum of $f(x,y) = 2x^2 - x^3 - y^2$ on $y \geq -1$. With $\nabla f(x,y) = 4x - 3x^2, -2y)$, the critical points are $(4/3, 0)$ and $(0,0)$. The Hessian is $H(x,y) = \begin{bmatrix} 4-6x & 0 \\ 0 & -2 \end{bmatrix}$. At $(0,0)$, the discriminant is -8 so that this is a saddle point. At $(4/3,0)$, the discriminant is 8 and $H_{11} = 4/3$, so that $(4/3,0)$ is a local maximum. We have now also to look at the boundary $y = -1$ where the function is $g(x) = f(x,-1) = 2x^2 - x^3 - 1$. Since $g'(x) = 0$ at $x = 0, 4/3$, where 0 is a local minimum, and 4/3 is a local maximum on the line $y = -1$. Comparing $f(4/3,0), f(4/3,-1)$ shows that $(4/3,0)$ is the global maximum.

As in one dimensions, knowing the critical points helps to understand the function. Critical points are also physically relevant. Examples are configurations with lowest energy. Many physical laws are based on the principle that the equations are critical points. Newton equations in Classical mechanics are an example: a particle of mass m moving in a field V along a path $\gamma : t \mapsto \vec{r}(t)$ extremizes the integral $S(\gamma) = \int_a^b mr'(t)^2/2 - V(r(t))\ dt$ among all possible paths. Critical points γ satisfy the Newton equations $mr''(t)/2 - \nabla V(r(t)) = 0$.

Why is the second derivative test true? Assume $f(x,y)$ has the critical point $(0,0)$ and is a quadratic function satisfying $f(0,0) = 0$. Then

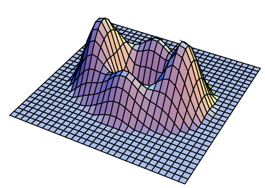
$$ax^2 + 2bxy + cy^2 = a(x + \frac{b}{a}y)^2 + (c - \frac{b^2}{a})y^2 = a(A^2 + DB^2)$$

with $A = (x + \frac{b}{a}y), B = b^2/a^2$ and discriminant D. You see that if $a = f_{xx} > 0$ and $D > 0$ then $c - b^2/a > 0$ and the function has positive values for all $(x,y) \neq (0,0)$. The point $(0,0)$ is a minimum. If $a = f_{xx} < 0$ and $D > 0$, then $c - b^2/a < 0$ and the function has negative values for all $(x,y) \neq (0,0)$ and the point (x,y) is a local maximum. If $D < 0$, then the function can take both negative and positive values. A general smooth function can be approximated by a quadratic function near $(0,0)$.

Sometimes, we want to find the overall maximum and not only the local ones. A point (a,b) in the plane is called a **global maximum** of $f(x,y)$ if $f(x,y) \leq f(a,b)$ for all (x,y). For example, the point $(0,0)$ is a global maximum of the function $f(x,y) = 1 - x^2 - y^2$. Similarly, we call (a,b) a **global minimum**, if $f(x,y) \geq f(a,b)$ for all (x,y).

Problem: Does the function $f(x,y) = x^4 + y^4 - 2x^2 - 2y^2$ have a global maximum or a global minimum? If yes, find them.
Solution: the function has no global maximum. This can be seen by restricting the function to the x-axis, where $f(x,0) = x^4 - 2x^2$ is a function without maximum. The function has four global minima however. They are located on the 4 points $(\pm 1, \pm 1)$. The best way to see this is to note that $f(x,y) = (x^2-1)^2 + (y-1)^2 - 2$ which is minimal when $x^2 = 1, y^2 = 1$.

Here is a curious remark: let $f(x,y)$ be the height of an island. Assume there are only finitely many critical points on the island and all of them have nonzero determinant. Label each critical point with a $+1$ if it is a maximum or minimum, and with -1 if it is a saddle point. Sum up all these number and you will get 1, independent of the function. This property is an example of an "index theorem", a prototype for important theorems in physics and mathematics.



Section 4.2: Extrema with constraints

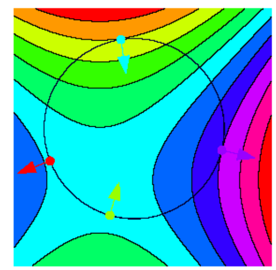
We aim now to find maxima and minima of a function $f(x,y)$ in the presence of a **constraint** $g(x,y) = 0$. You can see that a necessary condition for a critical point is that the gradients of f and g are parallel because otherwise, we can go along the level curve of g and get a nonzero directional derivatives. The condition of having parallel gradients, leads to a system of equations

which are called the **Lagrange equations** $\nabla f(x,y) = \lambda \nabla g(x,y), g(x,y) = 0$. These are three equations for the three unknowns x, y, λ. The variable λ is called the **Lagrange multiplier**.

Given a function $f(x,y)$ of two variables and a level curve $g(x,y) = c$. Find the extrema of f on the curve. You see that at places, where the gradient of f is not parallel to the gradient of g, the function f changes when we change position on the curve $g = c$. Therefore, we must have a solution of three equations

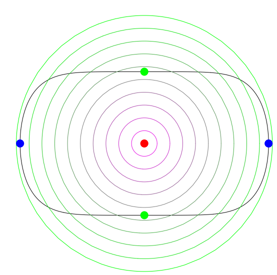
$$\boxed{\nabla f(x,y) = \lambda \nabla g(x,y), g(x,y) = c}$$

to the three unknowns (x,y,λ). Additionally we must check points with $\nabla g(x,y) = (0,0)$ because also then the gradients are parallel.



Example: To find the shortest distance from the origin to the curve $x^6 + 3y^2 = 1$, we extremize $f(x,y) = x^2 + y^2$ under the constraint $g(x,y) = x^6 + 3y^2 - 1 = 0$.

Solution: $\nabla f = \langle 2x, 2y \rangle, \nabla g = \langle 6x^5, 6y \rangle$. The Lagrange equations $\nabla f = \lambda \nabla g$ lead to the system $2x = \lambda 6x^5, 2y = \lambda 6y, x^6 + 3y^2 - 1 = 0$. We get $\lambda = 1/3, x = x^5$, so that either $x = 0$ or 1 or -1. From the constraint equation, we obtain $y = \sqrt{(1-x^6)/3}$. So, we have the solutions $(0, \pm\sqrt{1/3})$ and $(1,0), (-1,0)$. To see which is the minimum, just evaluate f on each of the points. We see that $(0, \pm\sqrt{1/3})$ are the minima.



Constrained optimization problems work also in more than 2 dimensions. For example, if $f(x,y,z)$ is a function of three variables and $g(x,y,z) = c$ is a surface, we solve the system of 4 equations

$$\nabla f(x,y,z) = \lambda \nabla g(x,y,z), g(x,y,z) = c$$

to the 4 unknowns (x,y,z,λ). If we want to extremize $f(x,y,z)$ under two constraints $g(x,y,z) = c$ and $h(x,y,z) = d$, we have a system of 5 equations for 5 unknowns:

$$\boxed{\nabla f(x,y,z) = \lambda \nabla g(x,y,z) + \mu \nabla h(x,y,z), g(x,y,z) = c, h(x,y,z) = d}$$

Example: Extrema of $f(x,y,z) = z$ on the sphere $g(x,y,z) = x^2 + y^2 + z^2 = 1$ are obtained by calculating the gradients $\nabla f(x,y,z) = (0,0,1)$, $\nabla g(x,y,z) = (2x, 2y, 2z)$ and solving $(0,0,1) = \nabla f = \lambda \nabla g = (2\lambda x, 2\lambda y, 2\lambda z), x^2 + y^2 + z^2 = 1$. $\lambda = 0$ is excluded by the third equation $1 = 2\lambda z$ so that the first two equations $2\lambda x = 0, 2\lambda y = 0$ give $x = 0, y = 0$. The 4'th equation gives $z = 1$ or $z = -1$. The extrema are the north pole $(0,0,1)$ (maximum) and the south pole $(0,0,-1)$ (minimum).

Example: Consider a dice showing i with probability p_i, where $i = 1, ..., 6$. The **entropy** of the probability distribution is defined as $S(\vec{p}) = -\sum_{i=1}^6 p_i \log(p_i)$. Find the distribution p which maximizes entropy under the constrained $g(\vec{p}) = \sum_{i=1}^6 p_i = 1$.
Solution: $\nabla f = (-1 - \log(p_1), \dots, -1 - \log(p_n))$, $\nabla g = (1, \dots, 1)$. The Lagrange equations are $-1 - \log(p_i) = \lambda, p_1 + ... + p_6 = 1$, from which we get $p_i = e^{-(\lambda+1)}$. The last equation $1 = \sum_i \exp(-(\lambda + 1)) = 6 \exp(-(\lambda + 1))$ fixes $\lambda = -\log(1/6) - 1$ so that $p_i = 1/6$. The distribution, where each event has the same probability is the distribution of maximal entropy. Maximal entropy means **least information content**. A dice which is fixed (asymmetric weight distribution for example) allows a cheating gambler to gain profit. Cheating through asymmetric weight distributions can be avoided by making the dices transparent.

Example: You manufacture cylindrical soda cans of height h and radius r. You want for a fixed volume $V(r,h) = h\pi r^2 = 1$ a minimal surface area $A(r,h) = 2\pi r h + 2\pi r^2$. With $x = h\pi, y = r$, you need to optimize $f(x,y) = 2xy + 2\pi y^2$ under the constrained $g(x,y) = xy^2 = 1$. Calculate $\nabla f(x,y) = (2y, 2x + 4\pi y), \nabla g(x,y) = (y^2, 2xy)$. The task is to solve $2y = \lambda y^2, 2x + 4\pi y = \lambda 2xy, xy^2 = 1$. The first equation gives $y\lambda = 2$. Putting that in the second one gives $2x + 4\pi y = 4x$ or $2\pi y = x$. The third equation finally reveals $2\pi y^3 = 1$ or $y = 1/(2\pi)^{1/3}, x = 2\pi(2\pi)^{1/3}$. This means $h = 0.54.., r = 2h = 1.08$.

Remark. Other factors can influence the shape. For example, the can has to withstand a pressure up to 100 psi.
The Lagrange equations $\nabla f(x,y) = \lambda \nabla g(x,y), g(x,y) = c$ do not give all the extrema. We can also have $\nabla g(x,y) = (0,0)$. The parallel condition also could have been written as $\lambda \nabla f(x,y) = \nabla g(x,y)$, $f(x,y) = x^2 + (y-1)^2, g(x,y) = x^2 - y^3$. The function f has a local maximum 1 at $(0,0)$ under the constraint $g(x,y) = 0$ but the Lagrange equations do not find it. The problem is that the gradient of g vanishes. ∇g is technically parallel to ∇f but there is no λ such that $\nabla f = \lambda \nabla g$ at this point. The reason for this "mistake" (which is present in virtually all calculus text books), is that parallel of the two gradient is not equivalent to $\nabla f = \lambda \nabla g$ but can also mean $\lambda \nabla f = \nabla g$ with $\lambda = 0$.

Can we avoid Lagrange? We could extremize $f(x,y)$ under the constraint $g(x,y) = 0$ by finding $y = y(x)$ from the later and extremizing the 1D problem $f(x, y(x))$.

Example: To extremize $f(x,y) = x^2 + y^2$ with constraint $g(x,y) = x^4 + 3y^2 - 1 = 0$, solve $y^2 = (1 - x^4)/3$ and minimize $h(x) = f(x, y(x)) = x^2 + (1 - x^4)/3$. $h'(x) = 0$ gives $x = 0$. The find the maximum $(\pm 1, 0)$, we had to maximize $h(x)$ on $[-1, 1]$, which occurs at ± 1.

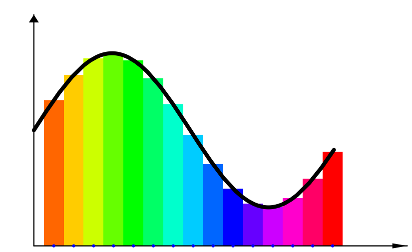
Sometimes, Lagrange could be avoided:
Extremize $f(x,y) = x^2 + y^2$ under the constraint $g(x,y) = p(x) + p(y) = 1$, where p is a complicated function in x which satisfies $p(0) = 0, p'(1) = 2$ The Lagrange equations

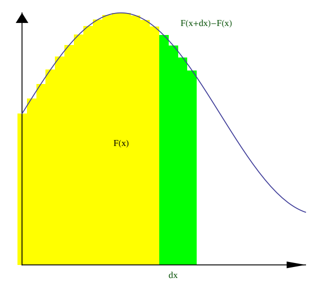
$2x = \lambda p'(x), 2y = \lambda p'(y), p(x) + p(y) = 1$ can be solved: with $x = 0, y = 1, \lambda = 1$, however, we can not solve $g(x,y) = 1$ for y. Substitution fails:

In general, the Lagrange method is more powerful.

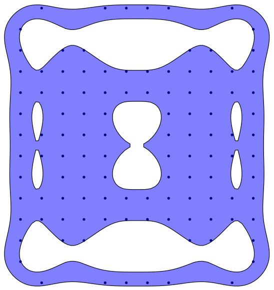
Section 4.3: Double integrals

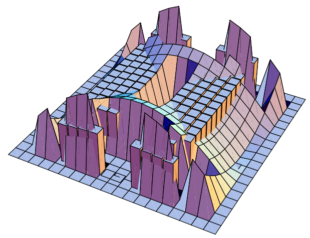
Lets first review integration in one dimensions. If $f(x)$ is a continuous function, then $\int_a^b f(x)\,dx$ is defined as a limit of the **Riemann sum** $f_n(x) = \sum_{x_k \in [a,b]} f(x_k)\Delta x$ for $n \to \infty$ with $x_k = k/n$ and $\Delta x = 1/n$. This Riemann integral divided by $|b - a|$ is the **average** of f on $[a,b]$. The integral $\int_a^b f(x)\,dx$ can be interpreted as an **signed area** under the graph of f, which can be negative too. If $f(x) = 1$, the integral is the **length** of the interval. The function $F(x) = \int_a^x f(y)\,dy$ is called an **anti-derivative** of f. The **fundamental theorem of calculus** states $F'(x) = f(x)$. This allows to compute integrals by inverting differentiation. Differentiation rules like the Leibniz rule become integration rules like integration by part, the chain rule becomes partial integration. Note that unlike the derivative, anti-derivatives can not always be expressed in terms of known functions. An example is: $F(x) = \int_0^x e^{-t^2}\,dt$. Often, the anti-derivative can be found: Example: $f(x) = \cos^2(x) = (\cos(2x) + 1)/2, F(x) = x/2 - \sin(2x)/4$.





If $f(x,y)$ is a continuous function of two variables on a region R, the integral $\int_R f(x,y)\,dxdy$ can be defined as the limit $\sum_{i,j,x_{i,j} \in R} f(x_i, y_j)\Delta x \Delta y$ with $x_{i,j} = (i/n, j/n)$ when n goes to infinity. If $f(x,y) = 1$, then the integral is the **area** of the region R. The integral divided by the area of R is the **average** value of f on R. For many regions, the integral can be calculated as a **double integral** $\int_a^b [\int_{c(x)}^{d(x)} f(x,y)\,dy]\,dx$.





One can interpret $\int\int_R f(x,y)\ dydx$ as the **signed volume** of solid below the graph of f and above R in the $x-y$ plane. As in 1D integration, the volume of the solid below the xy-plane is counted negatively.

Example: Calculate $\int\int_R f(x,y)\ dxdy$, where $f(x,y) = 4x^2y^3$ and where R is the rectangle $[0,1] \times [0,2]$.

$$\int_0^1 [\int_0^2 4x^2y^3\ dy]\ dx = \int_0^1 [x^2y^4|_0^2]\ dx = \int_0^1 x^2(16-0)\ dx = 16x^3/3|_0^1 = \frac{16}{3}\ .$$

Fubini's theorem states that we can interchange the order of integration if we integrate over a rectangle: $\boxed{\int_a^b \int_c^d f(x,y)\ dxdy = \int_c^d \int_a^b f(x,y)\ dydx.}$

This can be seen by approximate both sides with a Riemann sum, where $\Delta x = \Delta y = 1/n$. We have the identity

$$\sum_{x_i \in [a,b]} \sum_{y_j \in [c,d]} f(x_i,y_j)\ \Delta y \Delta y = \sum_{y_j \in [c,d]} \sum_{x_i \in [a,b]} f(x_i,y_j)\ \Delta x \Delta x\ .$$

Now take the limit $n \to \infty$.

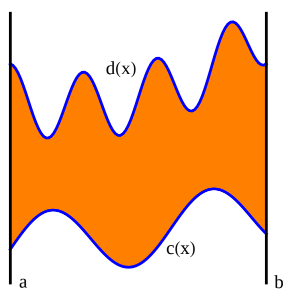
Fubini's theorem only holds for rectangles. We extend the class of regions now to so called Type I and Type II regions:

If the region satisfies $a \leq x \leq b$ and is bounded by the graphs of two functions $c(x)$ and $d(x)$, it is called of **type I**. One can write the region as

$$R = \{(x,y) \mid a \leq x \leq b\ c(x) \leq y \leq d(x)\ \}\ .$$

An integral over such a region is an iterated integral which is:

$$\int\int_R f\ dA = \int_a^b \int_{c(x)}^{d(x)} f(x,y)\ dydx\ .$$

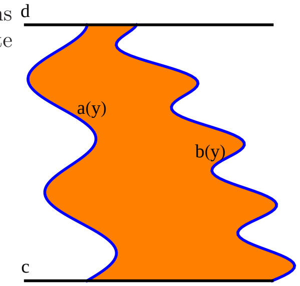


If the region is bound between the graphs of the functions $a(y)$ and $b(y)$, the region is called of **type II**. One can write the region as

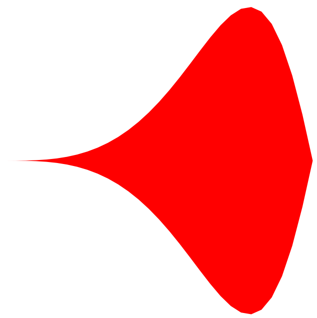
$$R = \{(x,y) \mid c \leq y \leq d, a(y) \leq x \leq b(y)\ \}\ .$$

An integral over such a region is an iterated integral:

$$\int\int_R f\ dA = \int_c^d \int_{a(y)}^{b(y)} f(x,y)\ dxdy\ .$$



Example: Integrate $f(x,y) = x^2$ over the region bounded above by $\sin(x^3)$ and bounded below by the graph of $-\sin(x^3)$ for $0 \leq x \leq \pi$. The value of this integral has a physical meaning. It is a moment of inertia. We will come back to that next week.



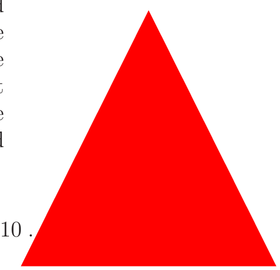
$$\int_0^{\pi^{1/3}} \int_{-\sin(x^3)}^{\sin(x^3)} x^2\ dydx = 2\int_0^{\pi^{1/3}} \sin(x^3)x^2\ dx$$

We have now an integral, which we can solve by substitution

$$= -\frac{2}{3}\cos(x^3)|_0^{\pi^{1/3}} = \frac{4}{3}$$

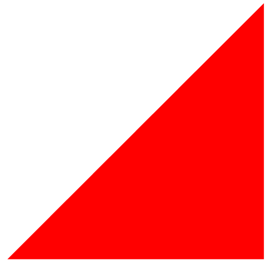
Example: Integrate $f(x,y) = y^2$ over the region bound by the x-axes, the lines $y = x+1$ and $y = 1-x$. The problem is best solved as a type I integral. As you can see from the picture, we would have to compute 2 different integrals as a type I integral. To do so, we have to write the bounds as a function of y: they are $x = y-1$ and $x = 1-y$

$$\int_0^1 \int_{y-1}^{1-y} y^3\ dx\ dy = 2\int_0^1 y^3(1-y)\ dy = 2(1/4-1/3) = 1/10\ .$$



Example: Let R be the triangle $1 \geq x \geq 0, 0 \leq y \leq x$. What is

$$\int \int_R e^{-x^2} \; dxdy \; ?$$

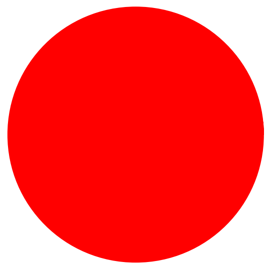


The type I integral $\int_0^1 [\int_y^1 e^{-x^2} \; dx] dy$ can not be solved because e^{-x^2} has no anti-derivative in terms of elementary functions.
The type II integral $\int_0^1 [\int_0^x e^{-x^2} \; dy] \; dx$ however can be solved:

$$= \int_0^1 x e^{-x^2} \; dx = -\frac{e^{-x^2}}{2}|_0^1 = \frac{(1-e^{-1})}{2} = 0.316... \; .$$

Example: The area of a disc of radius R is

$$\int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} 1 \; dydx = \int_{-R}^{R} \sqrt{R^2-x^2} \; dx \; .$$



This integral can be solved with the substitution $x = R\sin(u), dx = R\cos(u)$

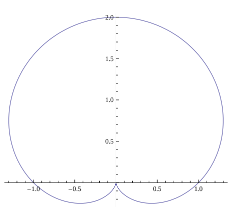
$$\int_{-\pi/2}^{\pi/2} \sqrt{R^2 - R^2\sin^2(u)} R\cos(u) \; du = \int_{-\pi/2}^{\pi/2} R^2 \cos^2(u) \; du$$

Now continue with a trigonometc identity to get $R^2 \int_{-\pi/2}^{\pi/2} \frac{(1+\cos(2u))}{2} \; du = R^2\pi$. This is too complicated. We will see how to do that better in polar coordinates.

Section 4.4: Polar coordinates and surface area

A **polar region** is bound by a curve given in polar coordinates as the curve $(r(t), \theta(t))$. In cartesian coordinates the parametrization is $\vec{r}(t) = \langle r(t)\cos(\theta(t), r(t)\sin(\theta(t)) \rangle$. We are especially interested in regions which are bound by **polar graphs**, where $\theta(t) = t$.
Example: The polar graph defined by $r(\theta) = \cos(3\theta)$ belongs to the class of **roses** $r(t) = |\cos(nt)|$. Regions enclosed by this graph are also called **rhododenea**.

Example. The polar curve $r(\theta) = 1 + \sin(\theta)$ is called a **cardioid**. It looks like a heart. It is a special case of **limacon** curves $r(\theta) = 1 + b\sin(\theta)$. We call the inside of the region a limacon region.



Integration in polar coordinates is

$$\boxed{\int \int_R f(x,y) \; dxdy = \int \int_R f(r\cos(\theta), r\sin(\theta) r \; drd\theta}$$

Example: If

$$f(x,y) = x^2 + x^2 + xy \; ,$$

then

$$f(r\cos(\theta), r\sin(\theta)) = r^2 + r^2\cos(\theta)\sin(\theta) \; .$$

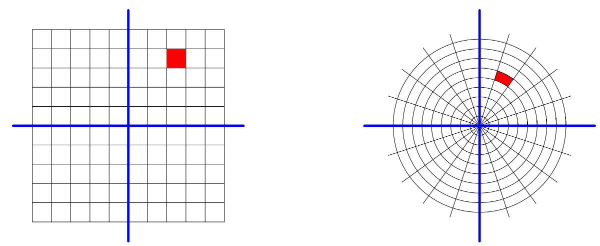
Example: We have earlier computed area of the disc $\{x^2 + y^2 \leq 1\}$ using substitution. It is more elegant to do this integral in polar coordinates:

$$\int_0^{2\pi} \int_0^1 r \; drd\theta = 2\pi r^2/2|_0^1 = \pi \; .$$

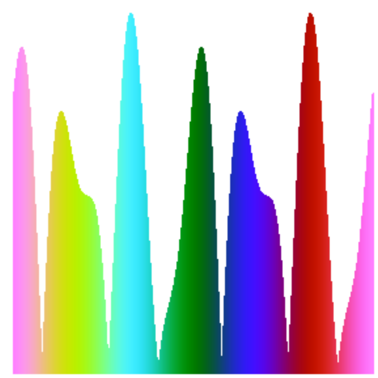
Why do we have to include the factor r, when we move to polar coordinates? The reason is that a small rectangle R with dimensions $d\theta dr$ in the (r,θ) plane is mapped by

$$T : (r,\theta) \mapsto (r\cos(\theta), r\sin(\theta))$$

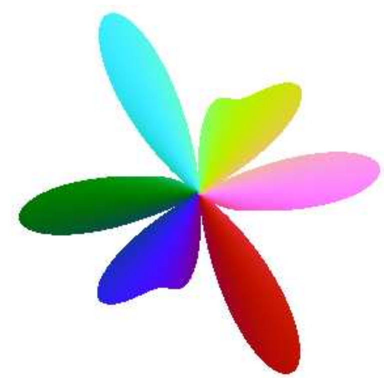
to a sector segment S in the (x,y) plane. It has approximately the area $r \; d\theta dr$.



We can now integrate over type I or type II regions in the (θ, r) plane. like **flowers**: $\{(\theta, r) \,|\, 0 \leq r \leq f(\theta)\}$ where $f(\theta)$ is a periodic function of θ.



A region R $\theta - r$ coordinates is a type I region



The same region in the xy coordinate system is not type I or II.

Example: Integrate the function $f(x,y) = 1 \ \{(\theta, r(\theta)) \mid r(\theta) \leq \ \mid \cos(3\theta)| \ \}$.

$$\int\int_R 1 \; dxdy = \int_0^{2\pi} \int_0^{\cos(3\theta)} r \; dr \; d\theta = \int_0^{2\pi} \frac{\cos(3\theta)^2}{2} \; d\theta = \pi/2$$

Example: Integrate $f(x,y) = y\sqrt{x^2+y^2}$ over the region $R = \{(x,y) \mid 1 < x^2+y^2 < 4, y > 0 \ \}$.

Solution.

$$\int_1^2 \int_0^{\pi} r\sin(\theta) r \; r \; d\theta dr = \int_1^2 r^3 \int_0^{\pi} \sin(\theta) \; d\theta dr = \frac{(2^4-1^4)}{4} \int_0^{\pi} \sin(\theta) \; d\theta = 15/2$$

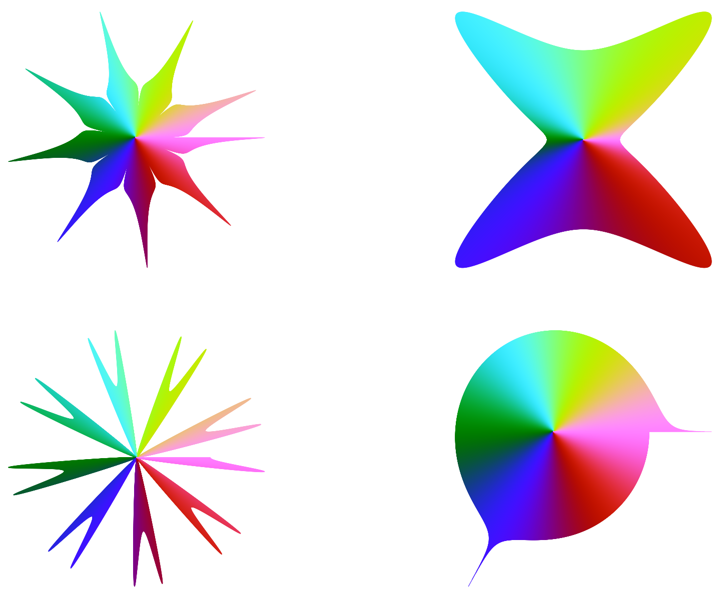
For integration problems, where the region is part of an annular region, or if you see function with terms x^2+y^2 try to use polar coordinates $x = r\cos(\theta), y = r\sin(\theta)$.

Example: The Belgian Biologist **Johan Gielis** came up in 1997 with the family of curves given in polar coordinates as

$$r(\phi) = (\frac{|\cos(\frac{m\phi}{4})|^{n_1}}{a} + \frac{|\sin(\frac{m\phi}{4})|^{n_2}}{b})^{-1/n_3}$$

It is called the **super-curve**, because it can produce a variety of shapes like circles, square, triangle, stars. It can also be used to produce "super-shapes" (see later).

The super-curve generalizes the **super-ellipse** which had been discussed in 1818 by Lamé and helps to tackle one of the more intractable problems in biology: describing form. A twist: Gielis has patented his discovery!



A surface $\vec{r}(u,v)$ parametrized on a parameter domain R has the **surface area**

$$\int\int_R |\vec{r}_u(u,v) \times \vec{r}_v(u,v)| \; dudv \; .$$

Note that $\vec{r}_u$ is tangent to the grid curve $u \mapsto \vec{r}(u,v)$ and $\vec{r}_v$ is tangent to $v \mapsto \vec{r}(u,v)$, the two vectors span a parallelogram with area $|\vec{r}_u \times \vec{r}_v|$. A small rectangle $[u, u+du] \times [v, v+dv]$ is mapped by $\vec{r}$ to a parallelogram spanned by $[\vec{r}, \vec{r}+\vec{r}_u]$ and $[\vec{r}, \vec{r}+\vec{r}_v]$ which has the area $|\vec{r}_u(u,v) \times \vec{r}_v(u,v)| \; dudv$.

Example: consider the parametrized surface $\vec{r}(u,v) = \langle 2u, 3v, 0 \rangle$. This surface is part of the xy-plane. The parameter region G just gets stretched by a factor 2 in the x coordinate and by a factor 3 in the y coordinate. $\vec{r}_u \times \vec{r}_v = \langle 0,0,6 \rangle$ and we see for example that the area of $\vec{r}(G)$ is 6 times the area of G.

Example: The map $\vec{r}(u,v) = \langle L\cos(u)\sin(v), L\sin(u)\sin(v), L\cos(v) \rangle$ maps the rectangle $G = [0,2\pi] \times [0,\pi]$ onto the sphere of radius L. We compute $\vec{r}_u \times \vec{r}_v = L\sin(v)\vec{r}(u,v)$. So, $|\vec{r}_u \times \vec{r}_v| = L^2|\sin(v)|$ and $\int\int_R 1 \; dS = \int_0^{2\pi} \int_0^{\pi} L^2 \sin(v) \; dvdu = 4\pi L^2$.

Example: For graphs $(u,v) \mapsto \langle u, v, f(u,v) \rangle$, we have $\vec{r}_u = (1, 0, f_u(u,v))$ and $\vec{r}_v = (0, 1, f_v(u,v))$. The cross product $\vec{r}_u \times \vec{r}_v = (-f_u, -f_v, 1)$ has the length $\sqrt{1+f_u^2+f_v^2}$. The area of the surface above a region G is $\int\int_G \sqrt{1+f_u^2+f_v^2} \; dudv$.

Example: Lets take a surface of revolution $\vec{r}(u,v) = \langle v, f(v)\cos(u), f(v)\sin(u) \rangle$ on $R = [0,2\pi] \times [a,b]$. We have $\vec{r}_u = (0, -f(v)\sin(u), f(v)\cos(u)), \vec{r}_v = (1, f'(v)\cos(u), f'(v)\sin(u))$ and $\vec{r}_u \times \vec{r}_v = (-f(v)f'(v), f(v)\cos(u), f(v)\sin(u)) = f(v)(-f'(v), \cos(u), \sin(u))$. The surface area is $\int\int |\vec{r}_u \times \vec{r}_v| \; dudv = 2\pi \int_a^b |f(v)|\sqrt{1+f'(v)^2} \; dv$.