

Lecture 13: Extrema

In many applications we are led to the task of **maximizing** a function f . As in single variable calculus we first search for points where the "derivative" is zero. This is the Fermat principle

A point (a, b) in the plane is called a **critical point** of a function $f(x, y)$ if $\nabla f(a, b) = [0, 0]^T$.

The **Fermat principle** predates the discovery of calculus and says:

If $\nabla f(x, y)$ is not zero, then (x, y) is not a critical point.

Proof. Take the directional derivative in the direction $v = \nabla f / |\nabla f|$. Then $D_{\vec{v}}f = \nabla f \cdot \vec{v} = |\nabla f| > 0$. You can also see it with linear approximation. If $\nabla f \neq 0$, then the linear approximation L is not constant and f is neither maximal nor minimal. QED.

Note that in our definition we do **not** include points, where f or its derivative is not defined. Without stating otherwise, we always assume that a function can be differentiated arbitrarily often. Points where the function has no derivatives are not considered part of the domain and need to be studied separately. For the continuous function $f(x, y) = 1/\log(|xy|)$ for example, we would have to look at the points on the coordinate axes as well as the points $xy = 1$ separately.

- 1 Find the critical points of $f(x, y) = x^4 + y^4 - 4xy + 2$. The gradient is $\nabla f(x, y) = [4(x^3 - y), 4(y^3 - x)]^T$ with critical points $(0, 0), (1, 1), (-1, -1)$.
- 2 $f(x, y) = \sin(x^2 + y) + y$. The gradient is $\nabla f(x, y) = [2x \cos(x^2 + y), \cos(x^2 + y) + 1]^T$. For a critical points, we must have $x = 0$ and $\cos(y) + 1 = 0$ which means $\pi + k2\pi$. The critical points are at $\dots (0, -\pi), (0, \pi), (0, 3\pi), \dots$. There are infinitely many.
- 3 The graph of $f(x, y) = (x^2 + y^2)e^{-x^2 - y^2}$ looks like a volcano. The gradient $\nabla f = [2x - 2x(x^2 + y^2), 2y - 2y(x^2 + y^2)]^T e^{-x^2 - y^2}$ vanishes at $(0, 0)$ and on the circle $x^2 + y^2 = 1$. This function has a continuum of critical points.
- 4 The function $f(x, y) = y^2/2 - g \cos(x)$ is the energy of the pendulum. The variable g is a constant. We have $\nabla f = (y, -g \sin(x)) = [(0, 0)]^T$ for $(x, y) = \dots, (-\pi, 0), (0, 0), (\pi, 0), (2\pi, 0), \dots$. These points are equilibrium points, angles for which the pendulum is at rest.
- 5 The function $f(x, y) = a \log(y) - by + c \log(x) - dx$ is a function which is invariant by the flow of the **Volterra-Lotka** differential equation $\dot{x} = ax - bxy, \dot{y} = -cy + dxy$. The point $(c/d, a/b)$ is a critical point of f and an equilibrium point of the system.
- 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. We have to look at the points on the axes separately. For $y = 0$ we see that $x = 0$ is a minimum. For $x = 0$ we see that $y = 0$ is a minimum. Indeed $(0, 0)$ is a minimum of f . This minimum was not detectable using derivatives.

In one dimension, we needed $f'(x) = 0, f''(x) > 0$ to have a local minimum and $f'(x) = 0, f''(x) < 0$ for a local maximum. If $f'(x) = 0, f''(x) = 0$, the nature of the critical point is 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** of the critical point.

Remark: The discriminant can be remembered better if seen as the determinant of the **Hessian matrix** $H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$.

Theorem: 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 need higher derivatives or ad-hoc methods to determine the nature of the critical point.

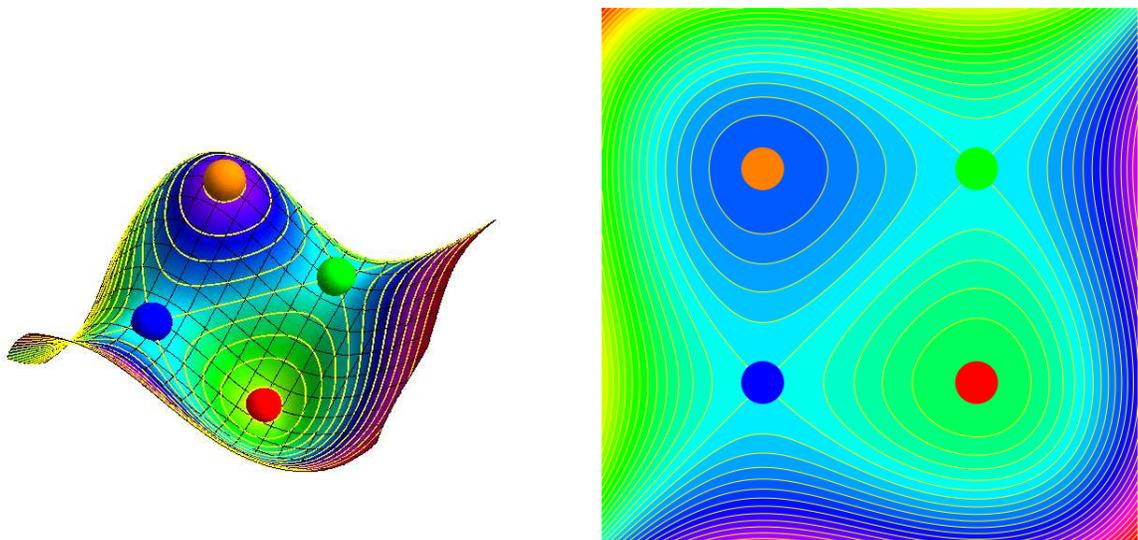
7 The function $f(x, y) = x^3/3 - x - (y^3/3 - y)$ has a graph which looks like a "napkin". It has the gradient $\nabla f(x, y) = [x^2 - 1, -y^2 + 1]^T$. There are 4 critical points $(1, 1), (-1, 1), (1, -1)$ and $(-1, -1)$. The Hessian matrix which includes all partial derivatives is $H = \begin{bmatrix} 2x & 0 \\ 0 & -2y \end{bmatrix}$.

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.

8 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.

Why is it important? 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 $\gamma : t \mapsto \vec{r}(t)$ maximizes the integral $S(\gamma) = \int_a^b m\dot{r}(t)^2/2 - V(r(t)) dt$ among all possible paths. Critical points γ satisfy the Newton equations $m\ddot{r}(t)/2 - \nabla V(r(t)) = 0$.

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

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

with $A = \left(x + \frac{b}{a}y\right), B = b^2/a^2$ 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 $(0, 0)$.

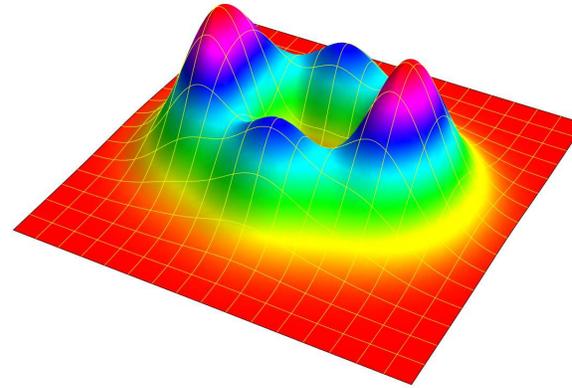
Sometimes, we want to find the overall maximum and not only the local ones.

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) .

9 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 $x^2 = 1, y^2 = 1$.

10

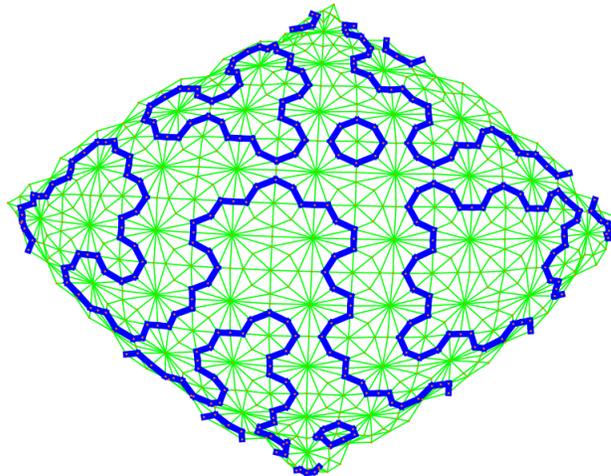
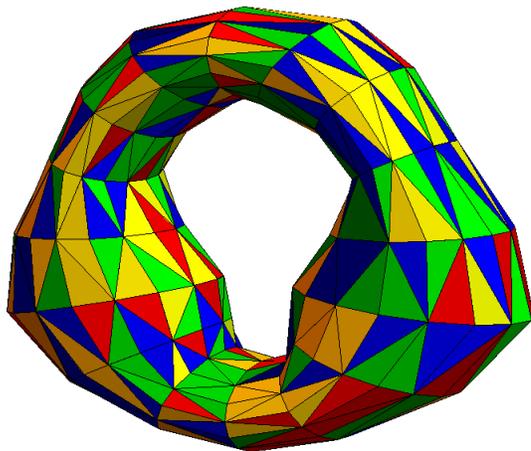
Let $f(x, y)$ be the height of an island for which only finitely many critical points exist and D is not zero there. Label a critical point with $+1$ if it is a maximum or minimum, and with -1 if it is a saddle. If you sum up these “charges” it always gives 1, independent of the function. This Poincaré-Hopf theorem is an example of an “index theorem”, a prototype result in physics and mathematics.



The following remarks can be skipped easily as they are either more theoretical or just point out connections to other fields of mathematics:

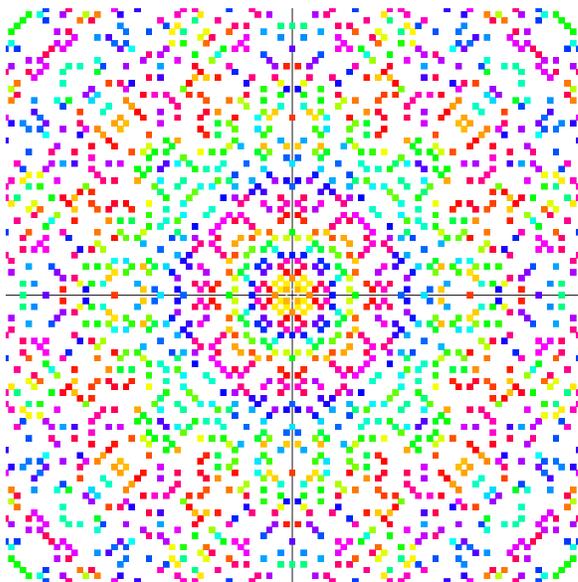
- 1) If you have seen some linear algebra, the discriminant D can be identified with a determinant $\det(H)$ of the matrix $H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$. Besides the determinant, also the trace $f_{xx} + f_{yy}$ is coordinate independent. Linear algebra tells us that the determinant is the product $\lambda_1 \lambda_2$ of the eigenvalues of H and the trace is the sum of the eigenvalues. If the determinant D is positive, then λ_1, λ_2 have the same sign. This sign agrees with the sign of the trace. If the trace and D are positive then both eigenvalues are positive. This means that in the eigen-directions, the graph is concave up. We have a minimum. On the other hand, if the determinant D is negative, then λ_1, λ_2 have different signs and the function is concave up in one eigen-direction and concave down in an other. In any case, if D is not zero, we have an orthonormal eigen-basis of the symmetric matrix A . In that basis, the matrix H is diagonal.
- 2) The discriminant D is also interesting at non-critical point. The number $K = D/(1 + |\nabla f|^2)^2$ is the **Gaussian curvature** of the surface. It is remarkable quantity as it only depends on the intrinsic geometry of the surface and not on the way how the surface is embedded in space. This is the famous **Theorema Egregia** (=great theorem) of Gauss. Note that at a critical point $\nabla f(x) = \vec{0}$, the discriminant agrees with the curvature $D = K$ at that point. Here is an other statement which follows from the **Gauss-Bonnet theorem**: assume you measure the curvature K at each point on the earth. If we integrate the curvature over the entire earth, it always gives 4π . You have seen one-dimensional curvature of a curve earlier. If you take a convex closed curve in the plane (assume the curvature is never zero for example) and integrate up the curvature $\int_0^{2\pi} \kappa(\vec{r}(t)) |\vec{r}'(t)| dt$ then you always get 2π . For a circle of radius r for example, where κ is constant $1/r$ and $|\vec{r}'(t)| = r$ we can confirm this directly. This is a particular case of the **Hopf Umlaufsatz** which can be seen as a particular case of a general Gauss-Bonnet theorem.
- 3) In higher dimensions, the second derivative test definitely requires linear algebra. In three dimensions for example, one can form the second derivative matrix H and look at all the eigenvalues of H . If all eigenvalues are negative, we have a local maximum, if all eigenvalues are positive, we have a local minimum. If the eigenvalues have different signs, we have a saddle point situation where in some directions the function increases and other directions the function decreases. If the Hessian has no zero eigenvalues, then the critical point is called a **Morse point** and the number m of negative eigenvalues is called the **Morse index** of f at the point.
- 4) We have mentioned in the first lecture that some ideas of calculus also go over to the discrete. Extrema problems generalize to discrete geometries like graphs. The analogue of a two dimensional space is a **finite simple graph** G for which every unit circle is a circular graph of length

larger than 3. One can now look at a function f on the vertices of a graph. One now looks at the set $S_f^-(x)$ on the unit circle, where $f(y) < f(x)$. If this set $S_f^-(x)$ is the entire circle, we have a maximum. If it is empty, then we have a minimum. And if it is a disconnected set, then we have a saddle point. If $\chi(S_f^-(x))$ is the number of vertices in $S_f^-(x)$ minus the number of edges in $S_f^-(x)$, then the discriminant D is defined as $1 - \chi(S_f^-(x))$. We see that it is equal to 1 at maxima and minima and -1 at saddle points. The sum of the indices is $\chi(G) = v - e + f$, where v is the number of vertices, e the number of edges and f the number of faces in G . It is the Euler characteristic of the graph. For a triangulation of a sphere, it is 2.



Homework

- 1 a) Find the critical points of $f(x, y) = (3x^4 - 8x^3 - 6x^2 + 24x) + (3y^4 - 8y^3 - 6y^2 + 24y)$.
- b) Find all the extrema of the function $f(x, y) = xy + x^2y + xy^2$ and determine whether they are maxima, minima or saddle points.



Remark: b) is the simplest of the **Gaussian Goldbach functions**

$$f_n = \sum_{k+im \text{ prime}, k, m \leq n} x^k y^m.$$

In the problem, we deal with $n = 2$

where we sum over the Gaussian primes $1 + i, 2 + i, 1 + 2i$. (A complex

number $a + ib$ with positive integers a, b is **prime** if $a^2 + b^2$ is a rational

(usual) prime. Oliver's **Goldbach conjecture** claims that all partial

derivatives $g^{(p,q)}(0,0)$ with $1 < p, q \leq n$ of $g(x, y) = f_{2n}^2(x, y)$ at

$(0,0)$ are non-zero if $p+q$ is even. This just means that every Gaussian

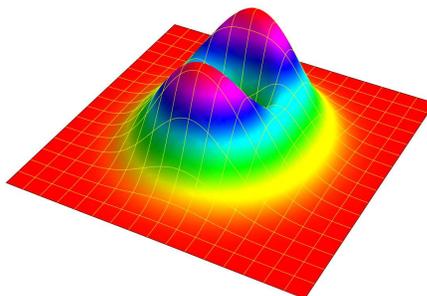
integer $a + ib$ with $a + b$ even and $a > 1, b > 1$ can be written as a sum

of two Gaussian primes in $Q = \{a + ib \mid a > 0, b > 0\}$.

- 2 Where on the parametrized surface $\vec{r}(u, v) = [1 + u^3, v^2, uv]^T$ is the temperature $T(x, y, z) = 2 + x + 12y - 12z$ minimal? To find the minimum, look where the function $f(u, v) = T(\vec{r}(u, v))$ has an extremum. Find all local maxima, local minima or saddle points of f .

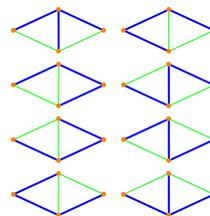
Remark. After you have found the function $f(u, v)$, you could replace the variables u, v again with x, y if you like and look at a function $f(x, y)$.

- 3 Find and classify all the extrema of the function $f(x, y) = e^{-x^2-y^2}(x^2 + 2y^2)$.



- 4 Find all extrema of the function $f(x, y) = 70 + x^3 + y^3 - 3x - 12y$ on the plane and characterize them. Do you find a global maximum or global minimum among them?

Graph theorists look at the **Tutte polynomial** $f(x, y)$ of a network. We work with the Tutte polynomial



- 5 $f(x, y) = x + 2x^2 + x^3 + y + 2xy + y^2$ of the **Kite network**. Classify the two critical using the second derivative test.

Remark. The polynomial is useful: $xf(1-x, 0)$ tells in how many ways one can color the nodes of the network with x colors and $f(1, 1)$ tells how many spanning trees there are. This picture illustrates that the number of spanning trees of the kite graph is $f(1, 1) = 8$ as you see the 8 possible trees.

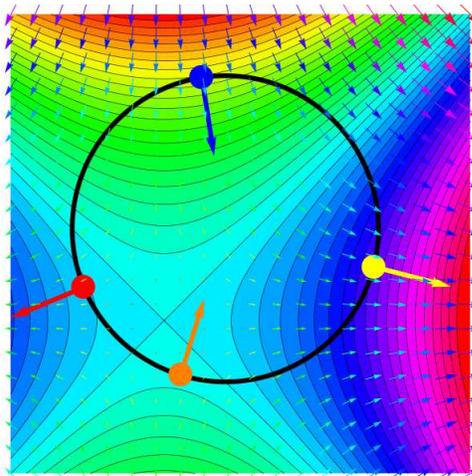
Lecture 14: Lagrange

When looking for maxima and minima of a function $f(x, y)$ in the presence of a **constraint** $g(x, y) = 0$, a necessary condition is that the gradients of f and g are parallel because otherwise, we can move along the curve $g = c$ and increase the value of f : the directional derivative of f in the direction tangent to the level curve is zero if and only if the tangent vector to g is perpendicular to the gradient of f . This can also include the case $\nabla g = 0$.

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

Lagrange theorem: A maximum or minimum of $f(x, y)$ on the curve $g(x, y) = c$ is either a solution of the Lagrange equations or is a critical point of g .

Proof. The condition that ∇f is parallel to ∇g either means $\nabla f = \lambda \nabla g$ or $\nabla f = 0$ or $\nabla g = 0$. The case $\nabla f = 0$ can be included in the Lagrange equation case with $\lambda = 0$. The case $\nabla g = 0$ which would lead to $\lambda = \infty$ has to be included separately. QED.



- 1 Minimize $f(x, y) = x^2 + 2y^2$ under the constraint $g(x, y) = x + y^2 = 1$. **Solution:** The Lagrange equations are $2x = \lambda, 4y = \lambda 2y$. If $y = 0$ then $x = 1$. If $y \neq 0$ we can divide the second equation by y and get $2x = \lambda, 4 = \lambda 2$ again showing $x = 1$. The point $x = 1, y = 0$ is the only solution.
- 2 Find the shortest distance from the origin to the curve $x^6 + 3y^2 = 1$. **Solution:** Minimize the function $f(x, y) = x^2 + y^2$ under the constraint $g(x, y) = x^6 + 3y^2 = 1$. The gradients are $\nabla f = [2x, 2y]^T, \nabla g = [6x^5, 6y]^T$. 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 $g = 1$, 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.

- 3 Which cylindrical soda cans of height h and radius r has minimal surface for fixed volume?
Solution: The volume is $V(r, h) = h\pi r^2 = 1$. The surface area is $A(r, h) = 2\pi rh + 2\pi r^2$. 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. A typical can of "Coca-Cola classic" with 3.7 volumes of CO_2 dissolve has at 75F an internal pressure of 55 psi, where PSI stands for pounds per square inch.
- 4 On the curve $g(x, y) = x^3 - y^2$ the function $f(x, y) = x$ obviously has a minimum $(0, 0)$. The Lagrange equations $\nabla f = \lambda \nabla g$ have no solutions. This is a case where the minimum is a solution to $\nabla g(x, y) = 0$.

Remarks.

- 1) The conditions in the Lagrange theorem are equivalent to $\nabla f \times \nabla g = 0$ in dimensions 2 or 3.
- 2) With $g(x, y) = 0$, the Lagrange equations can also be written as $\nabla F(x, y, \lambda) = 0$, where $F(x, y, \lambda) = f(x, y) - \lambda g(x, y)$.
- 3) The two conditions in the theorem are equivalent to " $\nabla g = \lambda \nabla f$ or f has a critical point".
- 4) Constrained optimization problems work also in higher dimensions. The proof is the same:

Extrema of $f(\vec{x})$ under the constraint $g(\vec{x}) = c$ are either solutions of the Lagrange equations $\nabla f = \lambda \nabla g, g = c$ or points where $\nabla g = \vec{0}$.

- 5 Find the extrema of $f(x, y, z) = z$ on the sphere $g(x, y, z) = x^2 + y^2 + z^2 = 1$. Solution: compute the gradients $\nabla f(x, y, z) = (0, 0, 1), \nabla g(x, y, z) = (2x, 2y, 2z)$ and solve $(0, 0, 1) = \nabla f = \lambda \nabla g = (2\lambda x, 2\lambda y, 2\lambda z), x^2 + y^2 + z^2 = 1$. The case $\lambda = 0$ 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 4th equation gives $z = 1$ or $z = -1$. The minimum is the south pole $(0, 0, -1)$ the maximum the north pole $(0, 0, 1)$.
- 6 A dice shows k eyes with probability p_k with k in $\Omega = \{1, 2, 3, 4, 5, 6\}$. A probability distribution is a non-negative function p on Ω which sums up to 1. It can be written as a vector $(p_1, p_2, p_3, p_4, p_5, p_6)$ with $p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$. The **entropy** of the probability vector \vec{p} is defined as $f(\vec{p}) = -\sum_{i=1}^6 p_i \log(p_i) = -p_1 \log(p_1) - p_2 \log(p_2) - \dots - p_6 \log(p_6)$. Find the distribution p which maximizes entropy under the constrained $g(\vec{p}) = p_1 + p_2 + p_3 + p_4 + p_5 + p_6 = 1$. **Solution:** $\nabla f = (-1 - \log(p_1), \dots, -1 - \log(p_n)), \nabla g = (1, \dots, 1)$. The Lagrange equations are $-1 - \log(p_i) = \lambda, p_1 + \dots + p_6 = 1$, 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**. An unfair dice allows a cheating gambler or casino to gain profit. Cheating through asymmetric weight distributions can be avoided by making the dices transparent.
- 7 Assume that the probability that a physical or chemical system is in a state k is p_k and that the energy of the state k is E_k . Nature tries to minimize the **free energy** $f(p_1, \dots, p_n) = -\sum_i [p_i \log(p_i) - E_i p_i]$ if the energies E_i are fixed. The probability distribution p_i satisfying

$\sum_i p_i = 1$ minimizing the free energy is called a **Gibbs distribution**. Find this distribution in general if E_i are given. **Solution:** $\nabla f = (-1 - \log(p_1) - E_1, \dots, -1 - \log(p_n) - E_n)$, $\nabla g = (1, \dots, 1)$. The Lagrange equation are $\log(p_i) = -1 - \lambda - E_i$, or $p_i = \exp(-E_i)C$, where $C = \exp(-1 - \lambda)$. The constraint $p_1 + \dots + p_n = 1$ gives $C(\sum_i \exp(-E_i)) = 1$ so that $C = 1/(\sum_i e^{-E_i})$. The Gibbs solution is $p_k = \exp(-E_k)/\sum_i \exp(-E_i)$.

Remarks:

1) Can we avoid Lagrange? Sometimes. It is often done in single variable calculus. In order to extremize xy under the constraint $2x + 2y = 4$ for example, we solve for y in the second equation and extremize the single variable problem $f(x, y(x))$. This needs to be done carefully and the boundaries must be considered. To extremize $f(x, y) = y$ on $x^2 + y^2 = 1$ for example we need to extremize $\sqrt{1 - x^2}$. We can differentiate to get the critical points but also have to look at the cases $x = 1$ and $x = -1$, where the actual minima and maxima occur. In general also, we can not do the substitution. To extremize $f(x, y) = x^2 + y^2$ with constraint $g(x, y) = x^4 + 3y^2 - 1 = 0$ for example, we 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 $x = 0$. To find the maximum $(\pm 1, 0)$, we had to maximize $h(x)$ on $[-1, 1]$, which occurs at ± 1 .

To 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 $x = 0, y = 1, \lambda = 1$. We can not solve $g(x, y) = 1$ however for y in an explicit way.

2) How do we determine whether a solution of the Lagrange equations is a maximum or minimum? Instead of using a second derivative test, we make a list of critical points and pick the maximum and minimum. A second derivative test can be designed using second directional derivative in the direction of the tangent.

3) The Lagrange method also works with more constraints. The constraints $g = c, h = d$ define a curve in space. The gradient of f must now be in the plane spanned by the gradients of g and h because otherwise, we could move along the curve and increase f :

Extrema of $f(x, y, z)$ under the constraint $g(x, y, z) = c, h(x, y, z) = d$ are either solutions of the Lagrange equations $\nabla f = \lambda \nabla g + \mu \nabla h, g = c, h = d$ or solutions to $\nabla g = 0, \nabla f(x, y, z) = \mu \nabla h, h = d$ or solutions to $\nabla h = 0, \nabla f = \lambda \nabla g, g = c$ or solutions to $\nabla g = \nabla h = 0$.

Homework

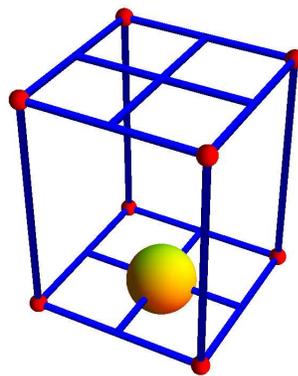
1) Find the cylindrical basket which is open on the top has the largest volume for fixed area π . If x is the radius and y is the height, we have to extremize $f(x, y) = \pi x^2 y$ under the constraint $g(x, y) = 2\pi xy + \pi x^2 = \pi$. Use the method of Lagrange multipliers.

2) Find the extrema of the same function

$$f(x, y) = e^{-x^2 - y^2} (x^2 + 2y^2)$$

you have seen in the last homework set but now on the entire disc $\{x^2 + y^2 \leq 4\}$ of radius 2. Besides the already found extrema **inside** the disk, now find also the extrema on the boundary.

- 3 Motivated by the Disney movie “Tangled”, we want to build a hot air balloon with a cuboid mesh of dimension x, y, z which together with the top and bottom fortifications uses wires of total length $g(x, y, z) = 6x + 6y + 4z = 32$. Find the balloon with maximal volume $f(x, y, z) = xyz$.



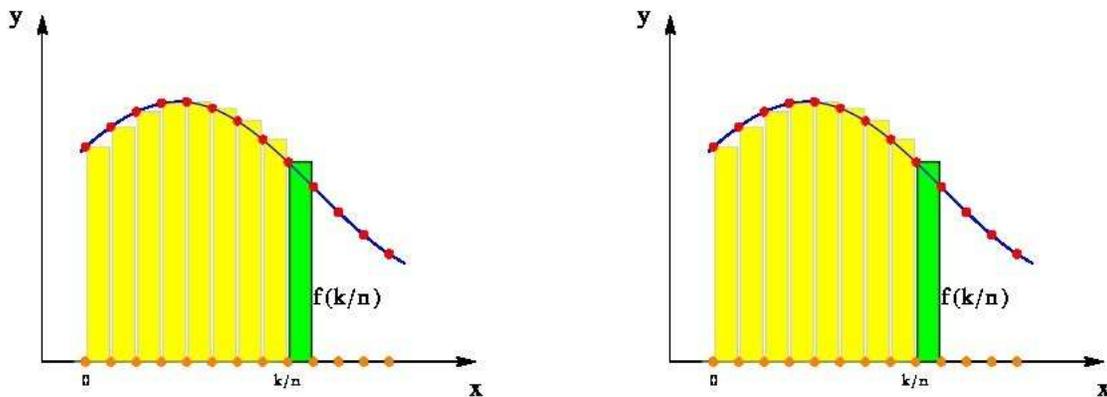
- 4 A **solid bullet** made of a half sphere and a cylinder has the volume $V = 2\pi r^3/3 + \pi r^2 h$ and surface area $A = 2\pi r^2 + 2\pi r h + \pi r^2$. Doctor Manhattan designs a bullet with fixed volume and minimal area. With $g = 3V/\pi = 1$ and $f = A/\pi$ he therefore minimizes $f(h, r) = 3r^2 + 2rh$ under the constraint $g(h, r) = 2r^3 + 3r^2 h = 1$. Use the Lagrange method to find a local minimum of f under the constraint $g = 1$.
- 5 Which pyramid of height h over a square $[-a, a] \times [-a, a]$ with surface area is $4a\sqrt{h^2 + a^2} + 4a^2 = 4$ has maximal volume $V(h, a) = 4ha^2/3$? By using new variables (x, y) and multiplying V with a constant, we get to the equivalent problem to maximize $f(x, y) = yx^2$ over the constraint $g(x, y) = x\sqrt{y^2 + x^2} + x^2 = 1$. Use the later variables.

Lecture 15: Double integrals

If $f(x)$ is a continuous function, then the **Riemann integral** $\int_a^b f(x) dx$ is defined as the limit of the **Riemann sums** $S_n f(x) = \frac{1}{n} \sum_{k/n \in [a,b]} f(k/n)$ for $n \rightarrow \infty$. The **derivative** of f is the limit of **difference quotients** $D_n f(x) = n[f(x + 1/n) - f(x)]$ as $n \rightarrow \infty$. The integral $\int_a^b f(x) dx$ is the **signed area** under the graph of f and above the x -axes, where "signed" indicates that parts below have a negative sign. The function $F(x) = \int_0^x f(y) dy$ is called an **anti-derivative** of f . It is determined up a constant. The **fundamental theorem of calculus** states

$$F'(x) = f(x), \quad \int_0^x f(x) = F(x) - F(0).$$

It allows to compute integrals by inverting differentiation so that **differentiation rules** become **integration rules**: the product rule leads to integration by parts, the chain rule becomes partial integration. For a 20 × 20 seconds review of 1D calculus, see www.math.harvard.edu/~knill/pedagogy/pechakucha, For a 140 character fundamental theorem: <https://twitter.com/oliverknill/status/320289197653106688>.

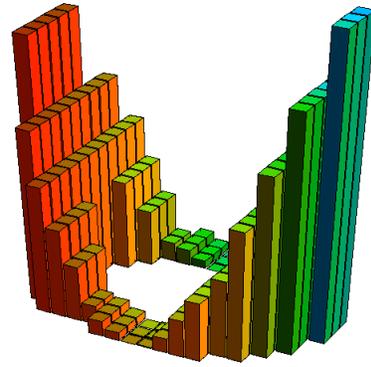
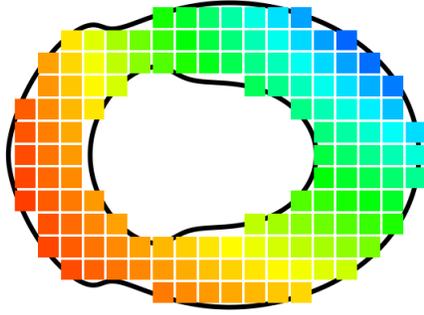


If $f(x, y)$ is continuous on a region R , the integral $\int_R f(x, y) dx dy$ is defined as the limit of the Riemann sum

$$\frac{1}{n^2} \sum_{\left(\frac{i}{n}, \frac{j}{n}\right) \in R} f\left(\frac{i}{n}, \frac{j}{n}\right)$$

when $n \rightarrow \infty$. We write also $\int_R f(x, y) dA$, where $dA = dx dy$ is a notation standing for "an area element".

- 1 If we integrate $f(x, y) = xy$ over the unit square we can sum up the Riemann sum for fixed $y = j/n$ and get $y/2$. Now perform the integral over y to get $1/4$. This example shows how to reduce double integrals to single variable integrals.
- 2 If $f(x, y) = 1$, then the integral is the **area** of the region R . The integral is the limit $L(n)/n^2$, where $L(n)$ is the number of lattice points $(i/n, j/n)$ contained in R .
- 3 The integral $\iint_R f(x, y) dA$ divided by the area of R is the **average** value of f on R .



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

Fubini's theorem allows to switch the order of integration over a rectangle if the function f is continuous: $\int_a^b \int_c^d f(x, y) dx dy = \int_c^d \int_a^b f(x, y) dy dx$.

Proof. For every n , there is the "quantum Fubini identity"

$$\sum_{\frac{i}{n} \in [a, b]} \sum_{\frac{j}{n} \in [c, d]} f\left(\frac{i}{n}, \frac{j}{n}\right) = \sum_{\frac{j}{n} \in [c, d]} \sum_{\frac{i}{n} \in [a, b]} f\left(\frac{i}{n}, \frac{j}{n}\right)$$

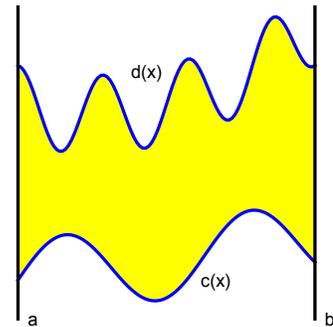
holding for all functions. Now divide both sides by n^2 and take the limit $n \rightarrow \infty$. This is possible for continuous functions. Fubini's theorem only holds for rectangles. We extend the class of regions now to so called Type I and Type II regions:

A **type I region** is of the form

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

An integral over a type I region is called a **type I integral**

$$\iint_R f dA = \int_a^b \int_{c(x)}^{d(x)} f(x, y) dy dx .$$

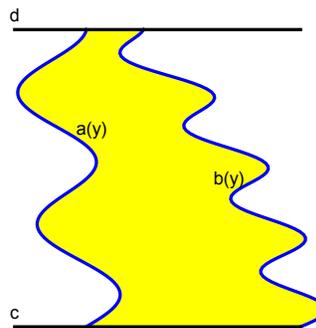


A **type II region** is of the form

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

An integral over such a region is called a **type II integral**

$$\iint_R f \, dA = \int_c^d \int_{a(y)}^{b(y)} f(x, y) \, dx \, dy .$$



- 5 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 called **moment of inertia**.

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

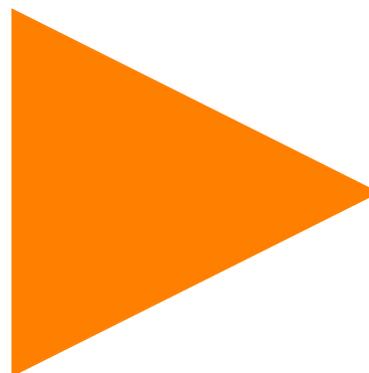
We have now an integral, which we can solve by substitution

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



- 6 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 $x = 1 - y$

$$\int_0^1 \int_{x-1}^{1-x} y^2 \, dy \, dx = 1/6 .$$



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

$$\int \int_R e^{-x^2} \, dx \, dy ?$$

The type II 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 I integral $\int_0^1 [\int_0^x e^{-x^2} \, dy] dx$ however can be solved:

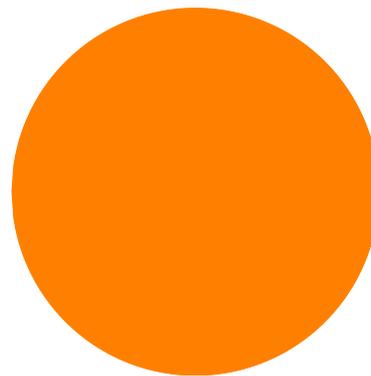
$$= \int_0^1 x e^{-x^2} \, dx = -\frac{e^{-x^2}}{2} \Big|_0^1 = \frac{(1 - e^{-1})}{2} = 0.316... .$$



8 The area of a disc of radius R is

$$\int_{-R}^R \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} 1 \, dy dx = \int_{-R}^R 2\sqrt{R^2-x^2} \, dx .$$

Substitute $x = R \sin(u)$, $dx = R \cos(u)$, to
get $\int_{-\pi/2}^{\pi/2} 2\sqrt{R^2 - R^2 \sin^2(u)} R \cos(u) \, du =$
 $\int_{-\pi/2}^{\pi/2} 2R^2 \cos^2(u) \, du$. Using a double angle formula.
this gives $R^2 \int_{-\pi/2}^{\pi/2} 2 \frac{(1+\cos(2u))}{2} \, du = R^2 \pi$.



Remark: The Riemann integral just defined works for more general functions. But it has limitations. In other branches of mathematics like in **probability theory**, a better integral is needed. It is the **Lebesgue integral** which fits the bill. Its definition is close to the Riemann integral which we have given as the limit $n^{-2} \int_{(x_k, y_l) \in R} f(x_k, y_l)$ where $x_k = k/n, y_l = l/n$. The Lebesgue integral replaces the regularly spaced (x_k, y_l) grid with random points x_k, y_l and uses the same formula. The following Mathematica code computes the integral $\int_0^1 \int_0^1 x^2 y$ using this **Monte Carlo definition** of the Lebesgue integral.

```
M=10000; R:=Random[]; f[x_, y_]:=x^2 y; Sum[f[R,R], {M}]/M
```

It is as elegant than the numerical Riemann sum computation

```
M=100; f[x_, y_]:=x^2 y; Sum[f[k/M, l/M], {k,M}, {l,M}]/M^2
```

but the Lebesgue integral is usually closer to the actual answer $1/6$ than the Riemann integral. Note that for all continuous functions, the Lebesgue integral gives the same results than the Riemann integral. It does not change calculus. But it is useful for example to compute nasty integrals like the area of the Mandelbrot set.

Homework

1 a) (4 points) Find the iterated integral

$$\int_0^1 \int_0^2 3xy / \sqrt{x^2 + (y^2/2)} \, dy \, dx .$$

b) (4 points) Now compute

$$\int_0^1 \int_0^2 3xy / \sqrt{x^2 + y^2/2} \, dx \, dy .$$

c) (2 points) Wouldn't Fubini assure that a) and b) are the same? What change would be needed in b) to make the results agree?

2 Find the area of the region

$$R = \{(x, y) \mid 0 \leq x \leq 2\pi, \sin(x) - 1 \leq y \leq \cos(x) + 2\}$$

and use it to compute the average value $\int \int_R f(x, y) \, dx dy / \text{area}(R)$ of $f(x, y) = y$ over that region.

3 Find the volume of the solid lying under the paraboloid $z = x^2 + y^2$ and above the rectangle $R = [-2, 2] \times [-3, 6] = \{(x, y) \mid -2 \leq x \leq 2, -3 \leq y \leq 6\}$.

4 Calculate the iterated integral $\int_0^1 \int_x^{2-x} (x^2 - y) \, dy dx$. Sketch the corresponding type I region. Write this integral as integral over a type II region and compute the integral again.

5 There is only one way to identify zombies: throw two difficult integrals at them and see whether they can solve them. Prove that you are not a zombie!

a) (6 points) Find the integral

$$\int_0^1 \int_{\sqrt{y}}^{y^2} \frac{3x^7}{\sqrt{x} - x^2} \, dx \, dy .$$

b) (4 points) Integrate

$$\int_0^1 \int_0^{\sqrt{1-y^2}} 11(x^2 + y^2)^{10} \, dx \, dy .$$

You might want to “time travel” one lecture forward, where polar coordinates are known to solve this problem.

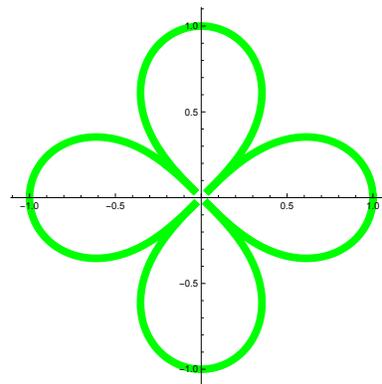
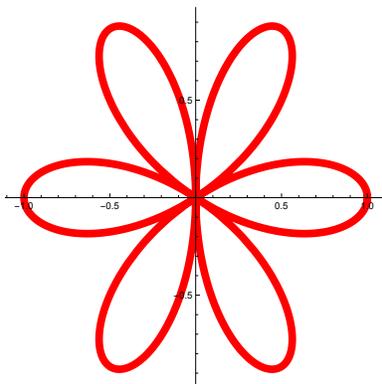
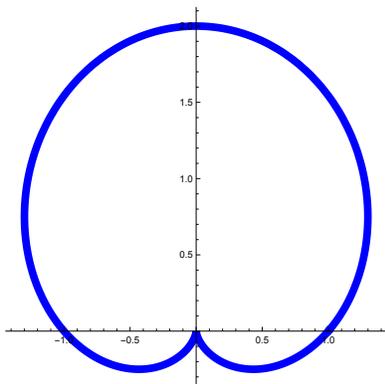


Lecture 16: Surface area

A **polar region** is a planar region bound by a simple closed curve given in polar coordinates as the curve $(r(t), \theta(t))$.

The most common case is $\theta(t) = t$. In Cartesian coordinates the parametrization of the boundary of a polar region is $\vec{r}(t) = [r(t) \cos(\theta(t)), r(t) \sin(\theta(t))]^T$, a **polar graph** like the spiral with $\theta(t) = t$.

- 1 The polar graph defined by $r(\theta) = |\cos(3\theta)|$ belongs to the class of **roses** $r(\theta) = |\cos(n\theta)|$. Regions enclosed by this graph are also called **rhododenea**.
- 2 The polar curve $r(\theta) = 1 + \sin(\theta)$ is called a **cardioid**. It looks like a heart. It belongs to the class of **limaçon** curves $r(\theta) = 1 + b \sin(\theta)$.
- 3 The polar curve $r(\theta) = |\sqrt{\cos(2t)}|$ is called a **lemniscate**. It looks like an infinity sign.



To integrate in polar coordinates, we evaluate the integral

$$\iint_R f(x, y) \, dx dy = \iint_R f(r \cos(\theta), r \sin(\theta)) r \, dr d\theta$$

- 4 Integrate

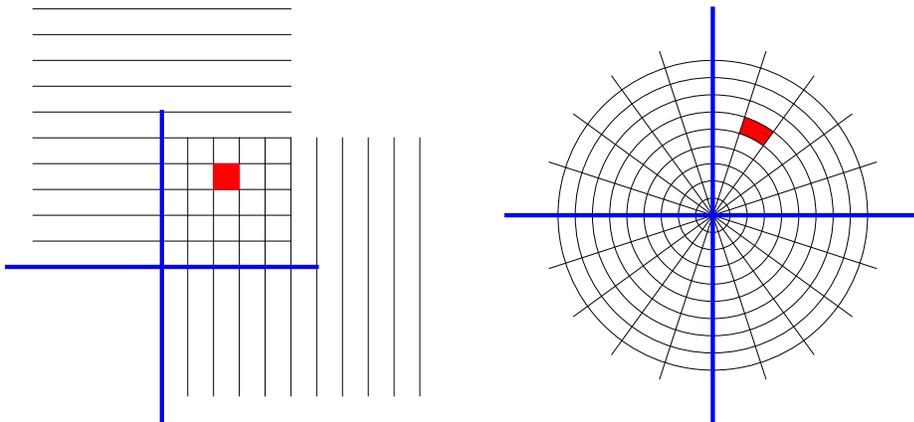
$$f(x, y) = x^2 + y^2 + xy,$$

over the unit disc. We have $f(x, y) = f(r \cos(\theta), r \sin(\theta)) = r^2 + r^2 \cos(\theta) \sin(\theta)$ so that $\iint_R f(x, y) \, dx dy = \int_0^1 \int_0^{2\pi} (r^2 + r^2 \cos(\theta) \sin(\theta)) r \, d\theta dr = 2\pi/4$.

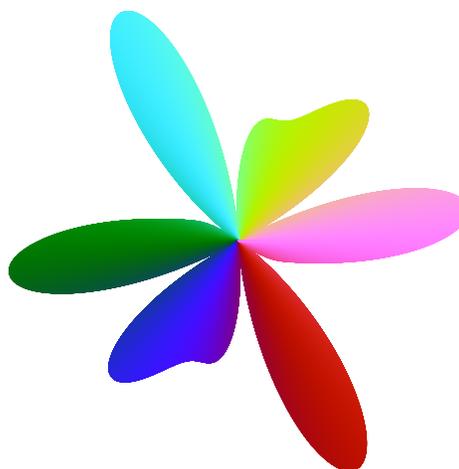
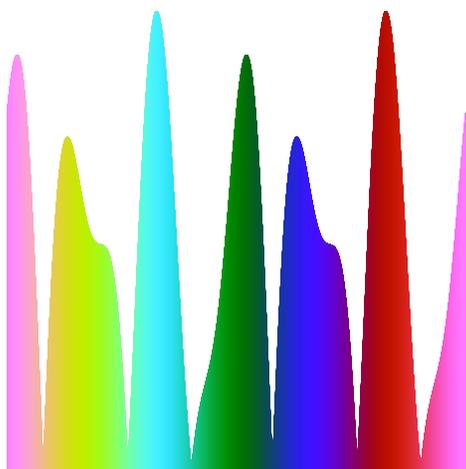
- 5 We have earlier computed area of the disc $\{x^2 + y^2 \leq 1\}$ using substitution. It is more elegant to do this integral in polar coordinates:

$$\int_0^{2\pi} \int_0^1 r \, dr d\theta = 2\pi r^2/2|_0^1 = \pi.$$

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 the area $r d\theta dr$. If you have seen some linear algebra, note that the Jacobean matrix dT has the determinant r .



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



A polar region shown in polar coordinates. It is a type I region.

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

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

$$\int \int_R 1 \, dx dy = \int_0^{2\pi} \int_0^{\cos(3\theta)} r \, dr \, d\theta = \int_0^{2\pi} \frac{\cos(3\theta)^2}{2} \, d\theta = \pi/2 .$$

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

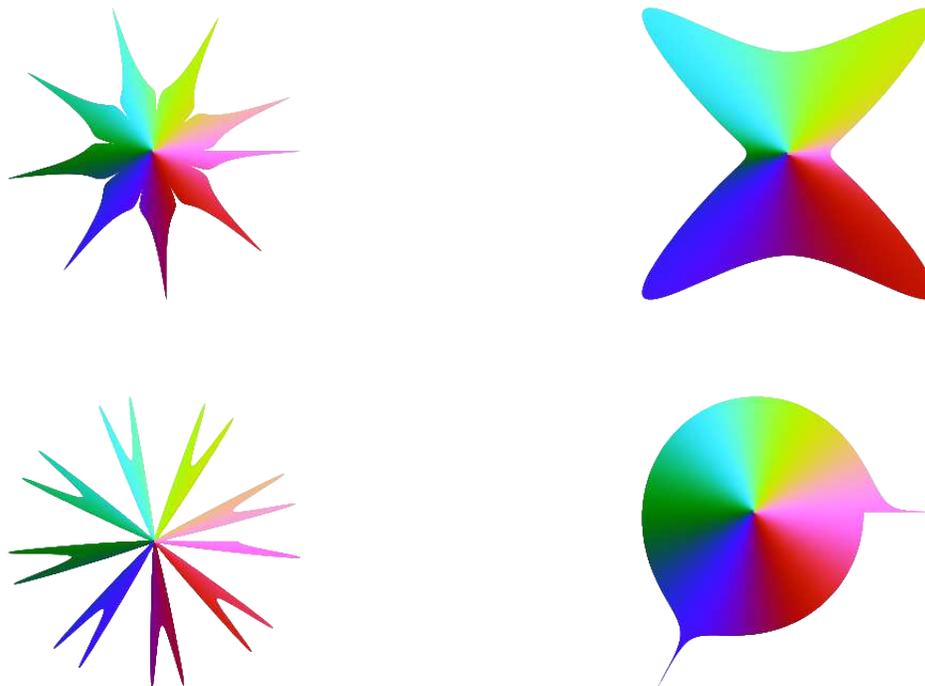
$$\int_1^2 \int_0^\pi r \sin(\theta) r \, d\theta dr = \int_1^2 r^3 \int_0^\pi \sin(\theta) \, d\theta dr = \frac{(2^4 - 1^4)}{4} \int_0^\pi \sin(\theta) \, d\theta = 15/2$$

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 $x = r \cos(\theta), y = r \sin(\theta)$.

- 8 The Belgian Biologist **Johan Gielis** defined in 1997 with the family of curves given in polar coordinates as

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

This so called **super-curve** can produce a variety of shapes like circles, square, triangle, stars. It can also be used to produce "super-shapes". The super-curve generalizes the **super-ellipse** which had been discussed in 1818 by Lamé and helps to **describe forms** in biology. ¹



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

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

Proof. The vector \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 $|\vec{r}_u(u, v) \times \vec{r}_v(u, v)| \, dudv$.

- 9 The parametrized surface $\vec{r}(u, v) = [2u, 3v, 0]^T$ 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. $\vec{r}_u \times \vec{r}_v = [0, 0, 6]^T$ and we see for example that the area of $\vec{r}(G)$ is 6 times the area of G .
- 10 The map $\vec{r}(u, v) = [L \cos(u) \sin(v), L \sin(u) \sin(v), L \cos(v)]^T$ 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 $\int \int_R 1 \, dS = \int_0^{2\pi} \int_0^\pi L^2 \sin(v) \, dvdu = 4\pi L^2$

¹Gielis, J. A 'generic geometric transformation that unifies a wide range of natural and abstract shapes'. American Journal of Botany, 90, 333 - 338, (2003).

- 11 For graphs $(u, v) \mapsto [u, v, f(u, v)]^T$, 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 $\int \int_G \sqrt{1 + f_u^2 + f_v^2} \, dudv$.
- 12 Lets take a surface of revolution $\vec{r}(u, v) = [v, f(v) \cos(u), f(v) \sin(u)]^T$ 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 $\int \int |\vec{r}_u \times \vec{r}_v| \, dudv = 2\pi \int_a^b |f(v)| \sqrt{1 + f'(v)^2} \, dv$.

Homework

- 1 a) A city near the sea is modeled by a half disk $D = \{(x, y) \mid x^2 + y^2 \leq 49, x \geq 0\}$ with center the origin and radius 7. What is the average distance of a point in D to the origin? in other words, what is the integral $\int \int_D \sqrt{x^2 + y^2} \, dxdy / \int \int_D 1 \, dxdy$.
- b) The distance to the beach is x . Find the average distance $\int \int_D x \, dxdy / \int \int_D 1 \, dxdy$ to the beach.
- 2 Find $\int \int_R (x^2 + y^2)^{44} \, dA$, where R is the part of the unit disc $\{x^2 + y^2 \leq 1\}$ for which $y > x$.
- 3 What is the area of the region which is bounded by the following three curves, first by the polar curve $r(\theta) = \theta$ with $\theta \in [0, 2\pi]$, second by the polar curve $r(\theta) = 2\theta$ with $\theta \in [0, 2\pi]$ and third by the positive x -axis?
- 4 The average of a function f on a region is defined as

$$\frac{\int_R f \, dxdy}{\int_R 1 \, dxdy}.$$

Find the average value of $f(x, y) = 2(x^2 + y^2)$ on the annular region $R : 1 \leq |(x, y)| \leq 2$.

- 5 Find the surface area of the part of the paraboloid $x = y^2 + z^2$ which is inside the cylinder $y^2 + z^2 \leq 16$.