SHORTEST PATHS IN THE PLANE — Math118, O. Knill

ABSTRACT. The minimization of the arc-length while connecting two points in the plane has been studied by Archimedes already. It can also be solved, if the arc length is generalized. It leads to differential equations.

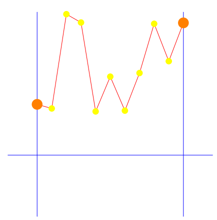
PLANE. Given two points P, Q in the plane. What is the path connecting P with Q which minimizes the length? While everybody knows that the straight line solves this problem, how does one prove this?

CONNECTING POINTS IN THE PLANE. Let $f(x)$ be a graph over the interval $[a, b]$ such that $P = (a, f(a))$ and $Q = (b, f(b))$. The length of this graph is

$$I(f) = \int_a^b \sqrt{1 + f'(x)^2}\, dx \ .$$

Which function f minimizes that? We could look at paths connecting points (x_i, y_i) with $(x_0, y_0) = P$ and $(x_n, y_n) = Q$ using $f_i(x) = f(a)(x - x_i) + (x - x_i)(y_{i+1} - y_i)/(x_{i+1} - x_i)$, to connect neighboring points. The length of such a graph is by Pythagoras $I(y_1, ..., y_{n-1}) = \sum_{i=0}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} = \sum_{i=0}^{n-1} l_i$. To minimize this, the gradient of I must vanish. Because the partial derivative with respect to y_i is $(y_i - y_{i-1})/l_i - (y_{i+1} - y_i)/l_i = \sin(\alpha_i) - \sin(\alpha_{i+1})$, all the slopes of the polygonal graph must agree and the line has to be a straight line. We have verified

> LEMMA. Among all polygonal graphs connecting P and Q, the straight line has minimal length.



One can also see by the triangle inequality that any corner in the graph can be shortened. A polygon which is not a straight line can be shortened by a definite amount. For any given differentiable function f, we can approximate the graph of f by piecewise linear graphs of g_n so that the length differences ϵ_n of the f and g graphs goes to zero. If there was a f for which the length were by an amount $\delta > 0$ smaller than the length of the straight line, we could approximate that function f with a polygon g_n for which $\epsilon_n < \delta$ and have a polygon with smaller length contradicting the lemma. We have now shown:

> THEOREM (Archimedes). Among all differentiable functions whose graph connects two points P and Q in the plane, the straight line minimizes the length.

Remark: this proof seems oblivious, since it can be shot down with mathematical cannon called "calculus of variations". Besides the fact that it is always nice to avoid heavy artillery, if not needed, the Archimedes proof has an advantage: it goes through also in a larger class of rectifiable functions which do not need to be differentiable like Snells refraction example below. The discretization approach also generalizes to inhomogeneous media, where it gives a numerical method. Remarkably, the proof does not need the notion of "derivative" at all, if one defines "rectifiable curves", as curves for which the lengths of the polygonal approximations converges and replaces $\nabla I = 0$ with the triangle inequality.

INHOMOGENEOUS MEDIUM. Lets assume that we are in a medium, where it is difficult to travel at some places and hard at others. If we replace the length by the work

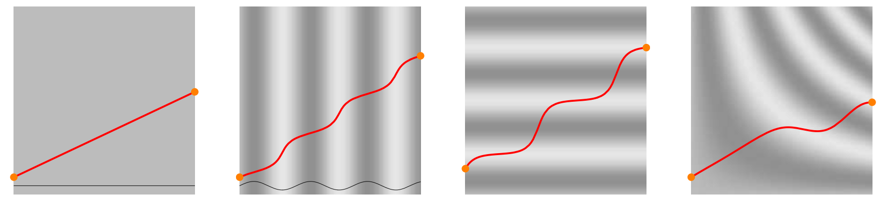
$$I_g(f) = \int_a^b g(x, f(x)) \sqrt{1 + f'(x)^2}\, dx$$

and again ask for the problem to find the most efficient path connecting two points P and Q, the result will critically depend on the function $g(x, y)$. There will be no more unique solutions. Lets discretize the problem again: we have to minimize $I(y_1, ..., y_{n-1}) = \sum_{i=0}^{n-1} g(x_i, y_i) l_i$. Setting the partial derivatives with respect to y_i equal to zero shows that $g_y(x_i, y_i) l_i + g(x_i, y_i)(y_{i+1} - y_i)/l_i = C$ is constant. This allows to compute recursively the slope

$$\sin(\alpha_i) = (C - g_y(x_i, y_i))/g(x_i, y_i) \ .$$

The constant C is obtained by the requirement that P and Q are connected.

EXAMPLES. The following examples were obtained by numerically solving for the shortest path connecting two given points.



Flat medium. The shortest connection between two points is a line.

Rippled medium. The path prefers to stay in the bright regions, where traveling is easy.

The inhomogeneity is vertical. Again, the particle prefers to stay in the bright regions.

A more general optimization problem: again the path tries to avoid staying too long in the dark area.

EULER-LAGRANGE EQUATIONS. Let $F(t, x, p)$ be a function of three variables. We look at the **variational problem** to extremize

$$I(\gamma) = \int_a^b F(t, x(t), \dot{x}(t))\, dt$$

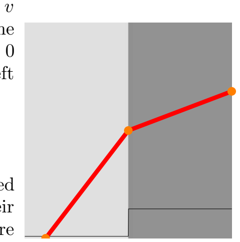
among all smooth paths γ connecting $x(a)$ with $x(b)$. If $t \mapsto h(t)$ is an other path, then $(I(\gamma + h) - I(\gamma)) = D_h I h + O(h^2)$ for $h \to 0$ defines a "directional derivative" $D_h I$ called here the **first variation**. By linearizing F, we know that $I(\gamma + h) - I(\gamma) = \int_a^b F_x(t, x, \dot{x}) + F_{\dot{x}}(t, x, \dot{x})\, dth + O(h^2) = \int_a^b F_x(t, x, \dot{x}) - \frac{d}{dt} F_{\dot{x}}(t, x, \dot{x})\, dth + O(h^2)$. The first variation is zero if $F_x(t, x, p) = \frac{d}{dt} F_p(t, x, p)$ for all t. These are the **Euler-Lagrange equations**.

INHOMOGENOUS PLANE. If $\gamma : t \mapsto (t, x(t))$ is a curve in the plane, we can look at $\int_a^b F(t, x, \dot{x})\, dt = \int_a^b \sqrt{1 + \dot{x}(t)^2}\, dt$. The Euler equations show that $\dot{x}/\sqrt{1 + \dot{x}(t)^2}$ is time independent. Therefore $\ddot{x}$ is constant and consequently, the optimal curve is a straight line. In the inhomogeneous case, the Euler-Lagrange equations for $F(t, x, \dot{x}) = g(t)\sqrt{1 + \dot{x}^2}$ are $0 = \frac{d}{dt}\left(\frac{\dot{x}(t) g(t)}{\sqrt{1 + \dot{x}(t)^2}}\right)$. This proves

> SNELLS THEOREM. $g(t)\dot{x}/\sqrt{1 + \dot{x}^2} = g(t) \sin(\alpha(x))$ is constant, where $\alpha(x)$ is the angle the curve makes with the x axes.

SNELLS LAW. A limiting situation is when the medium has two densities like air and water. In this situation, the Euler-Lagrange equations do not help. But the Archimedes approach still works. If $g = u$ on the left hand side and $g = v$ on the right hand side, then $\sin(\alpha_i) = \sin(\alpha_{i+1})$ as before in the left or the right region and $u(y_i - y_{i-1})/l_i - v(y_{i+1} - y_i)/l_i = u \sin(\alpha_i) - v \sin(\alpha_{i+1}) = 0$ at the boundary. Therefore, the shortest path is a line with angle α on the left hand side and angle β on the right hand side and

$$u \sin(\alpha) = v \sin(\beta).$$



This is called **Snells law** named after **Willebrord Snel**, who had discovered this refraction law. Descartes and Fermats thought about this too. Their dispute about this is described in Nahins book "When least is best". For a more general density distribution Archimedes proof also gives that $g(t) \sin(\alpha(x))$ is constant. **Archimedes proof is more powerful: it leads to a result for nonsmooth** $g(t)$**.**

AN INITIAL VALUE PROBLEM. With the assumption that a particle moves without an influence of an external force and minimize the action, we are lead to a dynamical system. Start at a point P and a direction v. The extremization requirement leads to a **Newton law**, which is a differential equation of the form $\ddot{x} = f(x, \dot{x}, t)$. One can actually derive all of Newtons law from a minimization principle. Extremization of action is one of the most important principles in physics: Newton equations, Maxwell equations, Einstein equations can be derived like this.