# Math Tripos Part IA: Variational Principles

Michael Li

May 28, 2016

Stationary points for functions on $\mathbb{R}^n$. Necessary and sufficient conditions for minima and maxima. Importance of convexity. Variational problems with constraints; method of Lagrange multipliers. The Legendre Transform; need for convexity to ensure invertibility; illustrations from thermodynamics. [4]

The idea of a functional and a functional derivative. First variation for functionals, Euler-Lagrange equations, for both ordinary and partial differential equations. Use of Lagrange multipliers and multiplier functions. [3]

Fermat's principle; geodesics; least action principles, Lagrange's and Hamilton's equations for particles and fields. Noether theorems and first integrals, including two forms of Noether's theorem for ordinary differential equations (energy and momentum, for example). Interpretation in terms of conservation laws. [3]

Second variation for functionals; associated eigenvalue problem. [2]

# Contents

# 0 Multivariate calculus

## 0.1 Stationary points

The quest of minimization starts with finding stationary points.

**Definition.** *Stationary points* are points in $\mathbb{R}^n$ for which $\nabla f = 0$.

But knowing $\nabla f = 0$ isn't enough, as we see from the taylor expansion:

$$f(\mathbf{x}) = f(\mathbf{0}) + \mathbf{x}\cdot\nabla f + \frac{1}{2}\sum_{i,j} x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j} + O(x^3).$$

$$= f(\mathbf{0}) + \frac{1}{2}\sum_{i,j} x_i x_j \frac{\partial^2 f}{\partial x_i \partial x_j} + O(x^3) = f(\mathbf{0}) + \frac{1}{2} x_i H_{ij} x_j + O(\mathbf{x}^3).$$

The last step involves defining the important second term:

**Definition** (Hessian matrix)**.** The *Hessian matrix* is

$$H_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

Since $H$ is real symmetric, after diagonalizing, $H$ becomes diagonal with $\lambda_i$ (real) ,the eigenvalues of $H$, as the diagonal terms. In our new coordinate system, we have

$$f(\mathbf{x}) - f(\mathbf{0}) = \frac{1}{2}\sum_{i=0}^{n} \lambda_i (x_i')^2$$

Using this, if all eigenvalues $\lambda_i$ are positive/negative, then it is a minimum/maximum. If there are mixed signs, then we have a saddle point, as the value can both increase/decrease. If some $\lambda = 0$, then we have a *degenerate stationary point.* We need to look at higher derivatives to find the nature of this stationary point.

When $n = 2$, we know that $\det H$ is the product of the eigenvalues, and $\operatorname{tr} H$ is the sum. So if $\det H < 0$, then it is a saddle. If $detH > 0$, we look at the trace to determine maximum/minimum. (cf. vector calculus).

## 0.2 Convex functions

*Convex functions* is a class of functions with awesome properties. You'll see.

**Definition.** A set $S \subseteq \mathbb{R}^n$ is *convex* if for any distinct $\mathbf{x}, \mathbf{y} \in S, t \in (0, 1)$, we have $(1-t)\mathbf{x} + t\mathbf{y} \in S$ meaning any line joining two points in $S$ lies completely within $S$.

**Definition.** A function $f : S \to \mathbb{R}$ is *convex* if $S$ is convex and for $x, y \in S$ and $\delta \in [0, 1]$, $\delta f(x) + (1-\delta)f(y) \geq f(\delta x + (1-\delta)y)$ is always satisfied.

A function is *strictly convex* if the inequality is strict. ($\geq$ replaced with $>$) A function $f$ is *(strictly) concave* iff $-f$ is (strictly) convex.

**Example.**

(i) $f(x) = x^2$ is strictly convex.

(ii) $f(x) = |x|$ is convex, but not strictly.

### 0.2.1 First-order convexity condition

Suppose $f$ is convex. For fixed $\mathbf{x}, \mathbf{y}$, we define:

$$h(t) = (1-t)f(\mathbf{x}) + tf(\mathbf{y}) - f((1-t)\mathbf{x} + f(\mathbf{y})).$$

By the definition of convexity of $f$, $h(t) \geq 0$. Also, trivially $h(0) = 0$. So

$$\frac{h(t) - h(0)}{t} \geq 0, \quad \text{and thus} \quad h'(0) \geq 0.$$

We can also differentiate $h$, evaluate at 0, and combine it with above:

$$h'(0) = f(\mathbf{y}) - f(\mathbf{x}) - (\mathbf{y} - \mathbf{x})\nabla f(\mathbf{x}) \quad \text{so} \quad f(\mathbf{y}) \geq f(\mathbf{x}) + (\mathbf{y} - \mathbf{x}) \cdot \nabla f(\mathbf{x}) \;\; (\dagger)$$

This condition also implies convexity. $f(\mathbf{x}) + (\mathbf{y} - \mathbf{x})\nabla f(\mathbf{x})$ defines the tangent plane of $f$ at $\mathbf{x}$. Thus a convex differentiable function lies above all its tangent planes. So:

**Corollary.** A stationary point of a convex function is a global minimum. If the function is strictly convex, there can be at most one global minimum.

We can rewrite $(\dagger)$ into the form

$$(\mathbf{y} - \mathbf{x}) \cdot [\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})] \geq f(\mathbf{x}) - f(\mathbf{y}) - (\mathbf{x} - \mathbf{y}) \cdot \nabla f(\mathbf{y}) \geq 0.$$

The right hand side is $\geq 0$ by $(\dagger)$. So $(\mathbf{y} - \mathbf{x}) \cdot [\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})] \geq 0$ is a first order condition, equivalent to other conditions, that says $\nabla f(\mathbf{x})$ is non-decreasing.

### 0.2.2 Second-order convexity condition

If the function is twice differentiable, our condition becomes:

$$(\mathbf{y} - \mathbf{x}) \cdot [\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})] \geq 0, \text{ becomes } \mathbf{h} \cdot (\nabla f(\mathbf{x} + \mathbf{h}) - \nabla f(\mathbf{x})) \geq 0.$$

Writing $\mathbf{y} = \mathbf{x} + \mathbf{h}$. Expand the left in Taylor series with suffix notation:

$$h_i[h_j \nabla_j \nabla_i f + O(h^2)] = h_i H_{ij} h_j + O(h^3) \geq 0.$$

As $\nabla_j \nabla_i f = H_{ij}$. Hence convexity implies that the Hessian matrix is positive for all $\mathbf{x} \in D(f)$. Strict convexity implies that it is positive definite (all eigenvalues positive).

The converse is also true - if the Hessian is positive, then $f$ is convex.

## 0.3 Legendre transform

The Legendre transform is an important tool in classical physics.

Suppose we have a differentiable function $f(x)$. We want to transform it into a function of the conjugate variable $p = \frac{\mathrm{d}f}{\mathrm{d}x}$ for some reason, usually because of physical significance. For example, in classical dynamics, if $L$ is the Lagrangian, then $p = \frac{\partial L}{\partial \dot{x}}$ is the conjugate momentum, which is more interesting.

But the obvious option $f^*(p) = f(x(p))$ is not the transform we want, because it is *ugly* and doesn't have nice properties. In particular, we want our $f^*(p)$ to satisfy:

$$\frac{\mathrm{d}f^*}{\mathrm{d}p} = x.$$

This says that if $x$ is the conjugate of $p$. In terms of differentials we want:

$$\mathrm{d}f = \frac{\mathrm{d}f}{\mathrm{d}x}\,\mathrm{d}x = p\,\mathrm{d}x. \quad \text{and} \quad \mathrm{d}f^* = x\,\mathrm{d}p.$$

How can we obtain this? From the product rule, we know that $\mathrm{d}(xp) = x\,\mathrm{d}p + p\,\mathrm{d}x.$. So if we define $f^* = xp - f$, we have the desired relation.

**Definition** (Legendre transform)**.** Given a function $f : \mathbb{R}^n \to \mathbb{R}$, its *Legendre transform* $f^*$ (the "conjugate" function) of the conjugate variable $\mathbf{p}$ is defined by

$$f^*(\mathbf{p}) = \sup_x(\mathbf{p}\cdot\mathbf{x} - f(\mathbf{x})),$$

The domain of $f^*$ is the set of $\mathbf{p} \in \mathbb{R}^n$ such that the supremum is finite.

This is what we said as the sup of $\mathbf{p}\cdot\mathbf{x} - f(\mathbf{x})$ is obtained when its derivative is zero, ie. $\mathbf{p} = \nabla f(\mathbf{x})$. But here we don't need differentiability. We can conclude that:

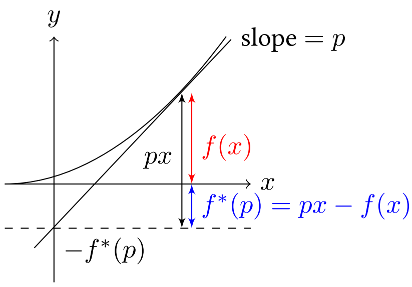
**Lemma.** $f^*$ is always convex.

*Proof.*

$$\begin{aligned} f^*((1-t)\mathbf{p} + t\mathbf{q}) &= \sup_x \big[(1-t)(\mathbf{p}\cdot\mathbf{x} - f(\mathbf{x})) + t(\mathbf{q}\cdot\mathbf{x} - f(\mathbf{x}))\big] \\ &\le (1-t)\sup_x[\mathbf{p}\cdot\mathbf{x} - f(\mathbf{x})] + t\sup_x[\mathbf{q}\cdot\mathbf{x} - f(\mathbf{x})] \\ &= (1-t)f^*(\mathbf{p}) + tf^*(\mathbf{q}) \end{aligned}$$

Note that we cannot immediately say that $f$ is convex, since we have to show that the domain is convex. But by the above bounds, $f^*((1-t)\mathbf{p} + t\mathbf{q})$ is bounded by the sum of two finite terms, which is finite. So $(1-t)\mathbf{p} + t\mathbf{q}$ is also in the domain of $f$. □

It is the intersection of the tangent line of $f$ at point $x$ and the $y$ axis:




**Example.** Let $f = cx$ for $c > 0$. This is convex but not strictly convex. Then $px - f(x) = (p-c)x$. This has no maximum unless $p = c$. So the domain of $f^*$ is simply $\{c\}$. One point. So $f^*(p) = 0$. So a line goes to a point.

**Theorem.** If $f$ is convex, differentiable with Legendre transform $f^*$, then $f^{**} = f$.

*Proof.* We have $f^*(\mathbf{p}) = (\mathbf{p}\cdot\mathbf{x}(\mathbf{p}) - f(\mathbf{x}(\mathbf{p}))$ where $\mathbf{x}(\mathbf{p})$ satisfies $\mathbf{p} = \nabla f(\mathbf{x}(\mathbf{p}))$.

Differentiating with respect to $\mathbf{p}$, we have

$$\nabla_i f^*(\mathbf{p}) = x_i + p_j\nabla_i x_j(\mathbf{p}) - \nabla_i x_j(\mathbf{p})p_j = x_i.$$

This means that the conjugate variable of $\mathbf{p}$ is our original $\mathbf{x}$. So

$$f^{**}(\mathbf{x}) = (\mathbf{x}\cdot\mathbf{p} - f^*(\mathbf{p}))|_{\mathbf{p}=\mathbf{p}(\mathbf{x})} = \mathbf{x}\cdot\mathbf{p} - (\mathbf{p}\cdot\mathbf{x} - f(\mathbf{x})) = f(\mathbf{x}).$$

□

Note convexity is required for Legendre transform in the theorem to work.

### 0.3.1 Application to thermodynamics

Given a system of particles, the energy of it is a function of entropy and volume:

$$E = E(S, V).$$

There are two ways that can affect the energy: we can push the piston and do $-p\,\mathrm{d}V$ amount of work. Or, we can heat it up for an energy change $T\,\mathrm{d}S$. So:

$$\mathrm{d}E = T\,\mathrm{d}S - p\,\mathrm{d}V.$$

But what the hell is entropy? I know temperature, which is $T = \frac{\partial E}{\partial S}$. (chain rule) So we use the (negative) Legendre transform to obtain *Helmholtz free energy*:

$$F(T, V) = \inf_S [E(S, V) - TS] = E(S, V) - S\frac{\partial E}{\partial S},$$

If we take the Legendre transform with respect to $V$, we get the enthalpy instead.

## 0.4 Lagrange multipliers

Cf. optimization, a problem of *constrained maximization* would be as follows: we have a path $p$ defined by $p(x, y) = 0$. What is the highest point along the path $p$?

We need $\mathrm{d}f = \nabla f \cdot \mathrm{d}\ell = 0$, but $\mathrm{d}\ell$ is *not* arbitrary. We only consider the $\mathrm{d}\ell$ parallel to the path. Alternatively, $\nabla f$ has to be entirely perpendicular to the path. Since we know that the normal to the path is $\nabla p$, our condition becomes

$$\nabla f = \lambda \nabla p \quad , \quad p = 0$$

for the three variables $x, y, \lambda$. We can also change this into a single problem of *unconstrained* extremization by asking the stationary points of the function:

$$\phi(x, y, \lambda) = f(x, y) - \lambda p(x, y)$$

When we maximize against the variables $x$ and $y$, we obtain the $\nabla f = \lambda \nabla p$ condition, and maximizing against $\lambda$ gives the condition $p = 0$.

**Example.** For $x \in \mathbb{R}^n$, find the minimum of $f(x) = x_i A_{ij} x_j$ on $|\mathbf{x}|^2 = 1$.

(i) The constraint imposes a normalization condition $\mathbf{x}$. So if we define:

$$\Lambda(\mathbf{x}) = \frac{f(\mathbf{x})}{g(\mathbf{x})}, \quad g(\mathbf{x}) = |\mathbf{x}|^2,$$

the problem is equivalent to minimization of $\Lambda(\mathbf{x})$ without constraint. Then

$$\nabla_i \Lambda(\mathbf{x}) = \frac{2}{g}\left[A_{ij}x_j - \frac{f}{g}x_i\right] \quad \text{or} \quad A\mathbf{x} = \Lambda\mathbf{x}$$

So $\Lambda(\mathbf{x})$ takes extremes at eigenvalues of $A$. So $\Lambda_{\mathrm{min}}$ is the lowest eigenvalue.

(ii) Let's do it with Lagrange multipliers. $\phi(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda(|\mathbf{x}|^2 - 1)$, so:

$$0 = \nabla\phi \Rightarrow A_{ij}x_j = \lambda x_i$$

Then differentiating with respect to $\lambda$ gives $|\mathbf{x}|^2 = 1$. So it is the same.

# 1 Euler-Lagrange equation

## 1.1 Functional derivatives

**Definition.** A *functional*, $F[x]$, is a function that takes a real-valued function as an argument, or a meta-function,. We say $F[x]$ is a functional of the function $x(t)$.

This is general, but we actually only care about one class of functionals: Given a function $x(t)$ defined for $\alpha \leq t \leq \beta$, we have, for some function $f$

$$F[x] = \int_\alpha^\beta f(x, \dot{x}, t)\ \mathrm{d}t$$

We want the fixed points of $F[x]$. Varying $x(t)$ by $\delta x(t)$ changes $F[x]$ by $\delta F[x]$:

$$\delta F[x] = F[x + \delta x] - F[x] = \int_\alpha^\beta \big(f(x + \delta x, \dot{x} + \delta \dot{x}, t) - f(x, \dot{x}, t)\big)\ \mathrm{d}t$$

We then Taylor expand, and integrate the second term by parts to obtain:

$$= \int_\alpha^\beta \left(\delta x \frac{\partial f}{\partial x} + \delta \dot{x} \frac{\partial f}{\partial \dot{x}}\right)\ \mathrm{d}t + o(\delta^2) = \int_\alpha^\beta \delta x \left[\frac{\partial f}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right)\right]\ \mathrm{d}t + \left[\delta x \frac{\partial f}{\partial \dot{x}}\right]_\alpha^\beta.$$

We always choose boundary conditions to let the last term, *boundary* term, be 0. Then

$$\delta F[x] = \int_\alpha^\beta \left(\delta x \frac{\delta F[x]}{\delta x(t)}\right)\ \mathrm{d}t$$

**Definition.**

$$\frac{\delta F[x]}{\delta x} = \frac{\partial f}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right)$$

is the *functional derivative* of $F[x]$.

If we want to find a stationary point of $F$, then we need $\frac{\delta F[x]}{\delta x} = 0$. So

**Definition.** The *Euler-Lagrange* equation is, for $\alpha \leq t \leq \beta$, and all $i$:

$$\frac{\partial f}{\partial x_i} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}_i}\right) = 0$$

But everything we just said is so abstract. So lets get down to earth:

**Example.** What is the curve $C$ of minimal length between points $A, B$ in the Euclidean plane? The length is $L = \int_C\ \mathrm{d}\ell$ where $\mathrm{d}\ell = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$. We have:

Let's parameterize $\mathbf{r} = (x(t), y(t))$ for $t \in [0, 1]$ such that $\mathbf{r}(0) = A$, $\mathbf{r}(1) = B$. So $\mathrm{d}\ell = \sqrt{\dot{x}^2 + \dot{y}^2}\ \mathrm{d}t$. Then

$$L[x, y] = \int_0^1 \sqrt{\dot{x}^2 + \dot{y}^2}\ \mathrm{d}t. \quad \text{with} \quad \frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0.$$

Solving Euler-Lagrange:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{y}}\right) = 0. \quad \text{gives} \quad \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = c, \quad \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = s$$

where $c$ and $s$ are constants. But the two constants are not independent. as we must have $c^2 + s^2 = 1$. So let $c = \cos\theta$, $s = \sin\theta$. Then we have conditions equivalent to:

$$(\dot{x}\sin\theta)^2 = (\dot{y}\sin\theta)^2. \quad \text{or} \quad \dot{x}\sin\theta = \pm\dot{y}\cos\theta.$$

We can choose a $\theta$ such that we have a positive sign. So

$$y\cos\theta = x\sin\theta + A$$

for a constant $A$. This is a straight line with slope $\tan\theta$.

## 1.2 First integrals

In above, $f$ did not depend on $x$, so $\frac{\partial f}{\partial x} = 0$. Then Euler-Lagrange gives:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right) = 0. \quad \text{or} \quad \frac{\partial f}{\partial \dot{x}} = \text{constant}.$$

We call this the *first integral*. We like first integrals as they are simple, and in physics we get a conserved quantity, which is always good.

There is also a more complicated first integral when $f$ does not (explicitly) depend on $t$:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial f}{\partial x} + \frac{\mathrm{d}\dot{x}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{x}} = \frac{\partial f}{\partial t} + \dot{x}\frac{\partial f}{\partial x} + \ddot{x}\frac{\partial f}{\partial \dot{x}}.$$

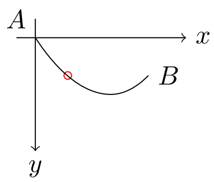
Substituting the Euler Lagrange for $\frac{\partial f}{\partial x}$:

$$= \frac{\partial f}{\partial t} + \dot{x}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right) + \ddot{x}\frac{\partial f}{\partial \dot{x}} = \frac{\partial f}{\partial t} + \frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{x}\frac{\partial f}{\partial \dot{x}}\right).$$

Then from our assumption:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(f - \dot{x}\frac{\partial f}{\partial \dot{x}}\right) = \frac{\partial f}{\partial t} = 0 \quad \text{so} \quad f - \dot{x}\frac{\partial f}{\partial \dot{x}} = \text{constant}.$$

**Example.** The Brachistochrone problem, Greek *brákhistos* ("shortest") and *khrónos* ("time"), is a famous problem in variational principles. What shape would minimize traveling time from $A$ to $B$, starting from rest, with no friction and only gravity?




Conservation of energy implies that

$$\frac{1}{2}mv^2 = mgy. \quad \text{or} \quad v = \sqrt{2gy}$$

We want to minimize

$$T = \int \frac{\mathrm{d}\ell}{v} = \frac{1}{\sqrt{2g}}\int \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{\sqrt{y}} = \frac{1}{\sqrt{2g}}\int \sqrt{\frac{1 + (y')^2}{y}}\,\mathrm{d}x$$

Since there is no explicit dependence on $x$, we have the first integral

$$f - y'\frac{\partial f}{\partial y'} = \frac{1}{\sqrt{y(1+(y')^2)}} = \text{constant} \quad \text{or} \quad y(1+(y')^2) = c$$

for some positive constant $c$. The solution of this ODE is, in parametric form,

$$x = c(\theta - \sin\theta) \quad \text{and} \quad y = c(1 - \cos\theta).$$

Note that this has $x = y = 0$ at $\theta = 0$. This describes a cycloid.

## 1.3 Constrained variation of functionals

What if $x$ is constrained? For example, we constrain $x$ to be on the surface $g(x(t)) = 0$ and then minimize $F[x]$. Here, we can again use Lagrange multipliers:

**Example.** The *Sturm-Liouville problem* is a very general class of problems: let $\rho(x)$, $\sigma(x)$ and $w(x)$ be real functions of $x$ defined on $a \le x \le b$, where $\rho$ and $w$ are positive on $a < x < b$. Our objective is to find stationary points of the functional

$$F[y] = \int_\alpha^\beta (\rho(x)(y')^2 + \sigma(x)y^2)\,\mathrm{d}x \quad \text{subject to} \quad G[y] = \int_\alpha^\beta w(x)y^2\,\mathrm{d}x = 1.$$

Using the Euler-Lagrange equation, the functional derivatives of $F$ and $G$ are

$$\frac{\delta F[y]}{\delta y} = 2\big(-(\rho y')' + \sigma y\big) \quad \text{and} \quad \frac{\delta G[y]}{\delta y} = 2(wy).$$

So the Euler-Lagrange of $\Phi_\lambda[y] = F[y] - \lambda(G[y] - 1)$ (Lagrange multipliers) is

$$-(\rho y')' + \sigma y - \lambda w y = 0 \quad \Leftrightarrow \quad \mathcal{L}y = \lambda y. \quad \text{where} \quad \mathcal{L} = \frac{1}{w}\left(-\frac{\mathrm{d}}{\mathrm{d}x}\left(\rho\frac{\mathrm{d}}{\mathrm{d}x}\right)\right) + \sigma.$$

$\mathcal{L}y = \lambda y$ is a *Sturm-Liouville eigenvalue* problem, $\mathcal{L}$ the *Sturm-Liouville operator* and $w$ the *weight function.* $\mathcal{L}y = \lambda w y$ is linear in $y$, so $y$ is a solution $\Rightarrow Ay$ is one, but $G[y] = 1 \Rightarrow G[Ay] = A^2$, so $G[y] = 1$ just normalizes. We thus ask the minimum of:

$$\Lambda[y] = \frac{F[y]}{G[y]}$$

unconstrained. We can't apply Euler-Lagrange here but we can try to vary it directly:

$$\delta\Lambda = \frac{1}{G}\delta F - \frac{F}{G^2}\delta G = \frac{1}{G}(\delta F - \Lambda\delta G) = 0 \quad \Leftrightarrow \quad \frac{\delta F}{\delta y} = \Lambda\frac{\delta G}{\delta y} \quad \Leftrightarrow \quad \mathcal{L}y = \Lambda y.$$

So at stationary values of $\Lambda[y]$, $\Lambda$ is the associated Sturm-Liouville eigenvalue.

**Example.** Suppose that we have a surface in $\mathbb{R}^3$ defined by $g(x) = 0$, and we want to find the path of shortest distance between two points, known as *geodesics.*

We can impose the condition $g(x(t)) = 0$ with a Lagrange multiplier. However, since we want the constraint to be satisfied for *all* $t$, we need a Lagrange multiplier *function* $\lambda(t)$. Then our problem would be to find stationary values of

$$\Phi[x, \lambda] = \int_0^1 \big(|\dot{\mathbf{x}}| - \lambda(t)g(x(t))\big)\,\mathrm{d}t$$

# 2 Hamilton's principle

Lagrange and Hamilton reformulated Newtonian dynamics with the first important concept being a *configuration space*, a vector space containing *generalized coordinates* $\xi(t)$ that captures *all* information of the system in one vector. So if we have $N$ free particles, the space has $3N$ dimensions, due to the 3 coordinates from each particle.

However, note that *generalized* coordinates need *not* be the Cartesian coordinates!

## 2.1 The Lagrangian

Hamilton showed the solutions of ODEs for particle paths from Lagrange are extremal points of an *action* $S$, a variational principle minimized in motion:

$$S[\xi] = \int L \, \mathrm{d}t \quad \text{where} \quad L = T - V$$

is the *Lagrangian*, $T$ the kinetic energy and $V$ the potential energy.

**Law** (Hamilton's principle)**.** The path $\xi(t)$ taken by a particle makes $S$ stationary.

**Example.** Suppose we have 1 particle in Euclidean 3-space. The configuration space is simply the coordinates of the particle. We can choose Cartesian coordinates $\mathbf{x}$. Then

$$T = \frac{1}{2}m|\dot{\mathbf{x}}|, \quad V = V(\mathbf{x}, t), \quad S[\mathbf{x}] = \int_{t_A}^{t_B} \left( \frac{1}{2}m|\dot{\mathbf{x}}|^2 - V(\mathbf{x}, t) \right) \mathrm{d}t.$$

We apply the Euler-Lagrange equations to obtain

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\mathbf{x}}}\right) - \frac{\partial L}{\partial \mathbf{x}} = m\ddot{\mathbf{x}} + \nabla V. \quad \text{so} \quad m\ddot{\mathbf{x}} = -\nabla V$$

This is Newton's law. But, Lagrangian mechanics has the advantage that it does not care what coordinates you use. Moreover, Lagrangian mechanics applies even when $V$ is time-dependent (and thus so is $L$). Then we can obtain a first integral from $S$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(L - \dot{\mathbf{x}} \cdot \frac{\partial L}{\partial t}\right) = \frac{\partial L}{\partial t} = 0 \quad \text{or} \quad \dot{\mathbf{x}} \cdot \frac{\partial L}{\partial \dot{\mathbf{x}}} - L = E$$

For some constant $E$. For one particle, this represents total energy (substitute L).

**Example.** Consider a central force field $\mathbf{F} = -\nabla V$, where $V = V(r)$ is independent of time. We use spherical polar coordinates $(r, \theta, \phi)$, so

$$L = T - V = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right) - V(r) = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\dot{\phi}^2 - V(r).$$

as $|\dot{\mathbf{x}}|^2 = \dot{r}^2 + r^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right)$ and motion is planar (aka angular momentum conserved). So wlog $\theta = \frac{\pi}{2}$, or $\sin\theta = 0$. Then the Euler Lagrange equations give

$$m\ddot{r} - mr\dot{\phi}^2 + V'(r) = \frac{\mathrm{d}}{\mathrm{d}t}\left(mr^2\dot{\phi}\right) = 0$$

So $r^2\dot{\phi} = h$ is a constant (angular momentum per unit mass). Then $\dot{\phi} = h/r^2$. So

$$m\ddot{r} - \frac{mh^2}{r^3} + V'(r) = 0 \quad \Leftrightarrow \quad m\ddot{r} = -V'_{\text{eff}}(r) \quad \text{where} \quad V_{\text{eff}} = V(r) + \frac{mh^2}{2r^2}$$

## 2.2 The Hamiltonian

In 1833, Hamilton took Lagrangian mechanics further. (aka not doing much)

**Definition.** The *Hamiltonian* of a system is the legendre transform of the Lagrangian:

$$H(\mathbf{x}, \mathbf{p}) = \mathbf{p} \cdot \dot{\mathbf{x}} - L(\mathbf{x}, \dot{\mathbf{x}}),$$

where $\dot{\mathbf{x}}$ is a function of $\mathbf{p}$ that is the solution to $\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{x}}}$. $\mathbf{p}$ is the *conjugate momentum* of $\mathbf{x}$. The space containing the variables $\mathbf{x}, \mathbf{p}$ is known as the *phase space*.

Since the Legendre transform is its self-inverse, the Lagrangian is the Legendre transform of the Hamiltonian with respect to $\mathbf{p}$. So

$$L = \mathbf{p} \cdot \mathbf{x} - H(\mathbf{x}, \mathbf{p}) \quad \text{with} \quad \dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}}. \qquad S[\mathbf{x}, \mathbf{p}] = \int (\mathbf{p} \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p}))\ \mathrm{d}t.$$

This is the *phase-space form* of the action. The Euler-Lagrange equations for these are

$$\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{x}}$$

For a single particle, the Lagrangian and the conjugate momentum is given by

$$L = \frac{1}{2} m |\dot{\mathbf{x}}|^2 - V(\mathbf{x}, t)., \quad \mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{x}}} = m\dot{\mathbf{x}},$$

Usual momentum. But as it is *generalized coordinates*, $\mathbf{p}$ changes: the conjugate momentum of the angle is angular momentum in polar. Substituting this, we have:

$$H(\mathbf{x}, \mathbf{p}) = \mathbf{p} \cdot \frac{\mathbf{p}}{m} - \frac{1}{2} m \left(\frac{\mathbf{p}}{m}\right)^2 + V(\mathbf{x}, t) = \frac{1}{2m} |\mathbf{p}|^2 + V.$$

So the Hamiltonian is the total energy, but expressed in terms of $\mathbf{x}, \mathbf{p}$, not $\mathbf{x}, \dot{\mathbf{x}}$.

## 2.3 Symmetries and Noether's theorem

For $F[x] = \int_\alpha^\beta f(x, \dot{x}, t)\ \mathrm{d}t$, suppose we change variables by the transformation $t \mapsto t^*(t)$ and $x \mapsto x^*(t^*)$. Then this gives:

$$F[x] \mapsto F^*[x^*] = \int_{\alpha^*}^{\beta^*} f(x^*, \dot{x}^*, t^*)\ \mathrm{d}t^*$$

with $\alpha^* = t^*(\alpha)$ and $\beta^* = t^*(\beta)$. Some interesting transformations:

**Definition.** If $F^*[x^*] = F[x]\ \forall x$, $\alpha$ and $\beta$, then the transformation $*$ is a *symmetry*.

**Example.** Consider the transformation $t \mapsto t$ and $x \mapsto x + \varepsilon$ for some small $\varepsilon$. Then

$$F^*[x^*] = \int_\alpha^\beta f(x + \varepsilon, \dot{x}, t)\ \mathrm{d}x = \int_\alpha^\beta \left(f(x, \dot{x}, t) + \varepsilon \frac{\partial f}{\partial x}\right)\ \mathrm{d}x$$

by the chain rule. Hence this transformation is a symmetry if $\frac{\partial f}{\partial x} = 0$.

However, if $\frac{\partial f}{\partial x} = 0$, then we have first integral

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right) = 0.$$

. So $\frac{\partial f}{\partial \dot{x}}$ is conserved.

We see that for the simple symmetry we have above,we have a conserved quantity. Noether's theorem is a powerful generalization of this.

**Theorem** (Noether's theorem)**.** For every continuous symmetry of $F[x]$, the solutions (ie. the stationary points of $F[x]$) will have a corresponding conserved quantity.

We will consider symmetries that involve only the $x$ variable. Up to first order, we can write the symmetry as $t \mapsto t, x(t) \mapsto x(t) + \varepsilon h(t)$, for some $h(t)$ representing the symmetry transformation (and $\varepsilon$ a small number). The change in $F[x]$ is given by

$$\begin{aligned}\delta F &= \int \big(f(x + \varepsilon h, \dot{x} + \varepsilon \dot{h} + \dot{\varepsilon} h, t) - f(x, \dot{x}, t)\big)\ \mathrm{d}t \\ &= \int \varepsilon \left(\frac{\partial f}{\partial x} h + \frac{\partial f}{\partial \dot{x}} \dot{h}\right)\ \mathrm{d}t + \int \dot{\varepsilon}\left(\frac{\partial f}{\partial \dot{x}} h\right)\ \mathrm{d}t.\end{aligned}$$

First consider the case where $\varepsilon$ is a constant. Then the second integral vanishes. And since this is a symmetry, $\delta F = 0$. So we obtain

$$\varepsilon \int \left(\frac{\partial f}{\partial x} h + \frac{\partial f}{\partial \dot{x}} \dot{h}\right)\ \mathrm{d}t = 0 \text{ so } \frac{\partial f}{\partial x} h + \frac{\partial f}{\partial \dot{x}} \dot{h} = 0. \text{ Thus } \delta F = \int \dot{\varepsilon}\left(\frac{\partial f}{\partial \dot{x}} h\right)\ \mathrm{d}t.$$

Then consider a variable $\varepsilon$ that is non-constant but vanishes at end-points. Then

$$\int \dot{\varepsilon}\left(\frac{\partial f}{\partial \dot{x}} h\right)\ \mathrm{d}t = \int \varepsilon \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}} h\right)\ \mathrm{d}t = 0. \quad \text{so} \quad \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}} h\right) = 0.$$

as it is true for *any* $\varepsilon$ vanishing at end-points. (Left side is integration by parts)

So $\frac{\partial f}{\partial \dot{x}} h$ is a conserved quantity. We can encode any other transformation, such as $t \mapsto t - \varepsilon$. into one with $x$ variable only: $x(t) \mapsto x(t - \varepsilon)$. In general, we find $\delta F$ assuming that $\varepsilon$ depends on time. Then the coefficient of $\dot{\varepsilon}$ is the conserved quantity.

**Example.** We can apply this to Hamiltonian mechanics. The motion of the particle is the stationary point of

$$S[\mathbf{x}, \mathbf{p}] = \int (\mathbf{p} \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{p}))\ \mathrm{d}t, \quad \text{where} \quad H = \frac{1}{2m}|\mathbf{p}|^2 + V(\mathbf{x}).$$

Suppose that we have a potential $V(|\mathbf{x}|)$ that only depends on radius. Then this has a rotational symmetry. Choose any favorite axis of rotational symmetry $\boldsymbol{\omega}$:

$$\mathbf{x} \mapsto \mathbf{x} + \varepsilon \boldsymbol{\omega} \times \mathbf{x} \qquad \mathbf{p} \mapsto \mathbf{p} + \varepsilon \boldsymbol{\omega} \times \mathbf{p},$$

Our rotation does not affect $|\mathbf{x}|$ and$|\mathbf{p}|$. So $H(\mathbf{x}, \mathbf{p})$ is unaffected. Then

$$\begin{aligned}\delta S &= \int \left(\mathbf{p} \cdot \frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{x} + \varepsilon \boldsymbol{\omega} \times \mathbf{x}) - \mathbf{p} \cdot \dot{\mathbf{x}}\right)\ \mathrm{d}t \\ &= \int \left(\mathbf{p} \cdot \frac{\mathrm{d}}{\mathrm{d}t}(\varepsilon \boldsymbol{\omega} \times \mathbf{x})\right)\ \mathrm{d}t \\ &= \int (\mathbf{p} \cdot [\boldsymbol{\omega} \times (\dot{\varepsilon} \mathbf{x} + \varepsilon \dot{\mathbf{x}})])\ \mathrm{d}t = \int (\dot{\varepsilon} \mathbf{p} \cdot (\boldsymbol{\omega} \times \mathbf{x}) + \varepsilon \mathbf{p} \cdot (\boldsymbol{\omega} \times \dot{\mathbf{x}}))\ \mathrm{d}t\end{aligned}$$

Since $\mathbf{p}$ is parallel to $\dot{\mathbf{x}}$, we are left with

$$\int (\dot{\varepsilon} \mathbf{p} \cdot (\boldsymbol{\omega} \times \mathbf{x}))\ \mathrm{d}t = \int \dot{\varepsilon} \boldsymbol{\omega} \cdot (\mathbf{x} \times \mathbf{p})\ \mathrm{d}x.$$

So $\boldsymbol{\omega} \cdot (\mathbf{x} \times \mathbf{p})$ is a constant of motion. Since this is true for all $\boldsymbol{\omega}$, $\mathbf{L} = \mathbf{x} \times \mathbf{p}$ must be a constant of motion, and this is the angular momentum.

# 3 Multivariate calculus of variations

What if we have multiple variables? We will consider the general $\mathbf{y}(x_1, \cdots, x_m) \in \mathbb{R}^n$ that maps $\mathbb{R}^m \to \mathbb{R}^n$ The functional will be a multiple integral of the form

$$F[\mathbf{y}] = \int \cdots \int f(\mathbf{y}, \nabla \mathbf{y}, x_1, \cdots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_m,$$

where $\nabla \mathbf{f}$ is the second-rank tensor defined as

$$\nabla \mathbf{f} = \left( \frac{\partial \mathbf{y}}{\partial x_1}, \cdots, \frac{\partial \mathbf{y}}{\partial x_m} \right).$$

Now there are two ways of thinking about this. The first way is Euler Lagrange. We can extend the Euler Lagrange into $n$ dimensions by considering the following equation:

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \frac{\partial f}{\partial \mathbf{y}_{x_i}} - \frac{\partial f}{\partial \mathbf{y}} = 0$$

Where the subscript indicates partial derivatives.

However, for most cases we should consider variations $\delta \mathbf{y}$ of $\mathbf{y}$ rather than a complicated Euler-Lagrange.

**Example** (Minimal Surface)**.** Suppose $(x, y)$ takes values in $D \subseteq \mathbb{R}^2$. We want to minimize the surface area on $S$ defined $z = h(x, y)$, where $h$ is the *height function*. Denoting partial differentiation by suffices: $h_x = \frac{\partial h}{\partial x}$, the area is:

$$A[h] = \int_D \sqrt{1 + h_x^2 + h_y^2} \, \mathrm{d}A.$$

(cf. Vector Calculus) Consider $h(x, y)$: $h \mapsto h + \delta h(x, y)$. Then by Taylor:

$$A[h + \delta h] - A[h] = \delta A = \int_D \left( \frac{h_x (\delta h)_x + h_y (\delta h)_y}{\sqrt{1 + h_x^2 + h_y^2}} + O(\delta h^2) \right) \mathrm{d}A$$

We integrate by parts to obtain

$$\delta A = -\int_D \delta h \left( \frac{\partial}{\partial x} \left( \frac{h_x}{\sqrt{1 + h_x^2 + h_y^2}} \right) + \frac{\partial}{\partial y} \left( \frac{h_y}{\sqrt{1 + h_x^2 + h_y^2}} \right) \right) \mathrm{d}A + O(\delta h^2)$$

plus some boundary terms. So our minimal surface will satisfy

$$\frac{\partial}{\partial x} \left( \frac{h_x}{\sqrt{1 + h_x^2 + h_y^2}} \right) + \frac{\partial}{\partial y} \left( \frac{h_y}{\sqrt{1 + h_x^2 + h_y^2}} \right) = 0$$

Simplifying, we have

$$(1 + h_y^2) h_{xx} + (1 + h_x^2) h_{yy} - 2 h_x h_y h_{xy} = 0.$$

Lets look at some solutions to this non-linear PDE:

- A plane. Boring.
- If $|\nabla h|^2 \ll 1$, then $h_x^2$ and $h_y^2$ are small. So we have $h_{yy} + h_{yy} = 0$, or $\nabla^2 h = 0$. So harmonic functions are (approximately) minimal-area.
- We might want a cylindrically-symmetric solution, ie. $h(x, y) = z(r)$, where $r = \sqrt{x^2 + y^2}$. Then we are left with:

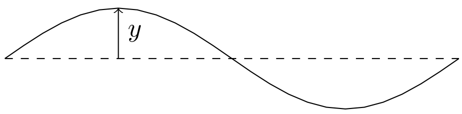
  $$rz'' + z' + z'^3 = 0.$$

  The general solution is

  $$z = A^{-1} \cosh(Ar) + B,$$

  a *catenoid*.

**Example.** Suppose we have a string with uniform constant mass density $\rho$ with uniform tension $T$.




Suppose we pull the line between $x = 0$ and $x = a$ with some tension $T$. Then we set it into motion such that the amplitude is given by $y(x; t)$. Then the kinetic energy is

$$T = \frac{1}{2} \int_0^a \rho v^2 \, \mathrm{d}x = \frac{\rho}{2} \int_0^a \dot{y}^2 \, \mathrm{d}x.$$

The potential energy is the tension times the length. So (ignoring higher terms):

$$V = T \int \mathrm{d}\ell = T \int_0^a \sqrt{1 + (y')^2} \, \mathrm{d}x = (Ta) + \int_0^a \frac{1}{2} T(y'^2) \, \mathrm{d}x.$$

Note that $y'$ is wrt $x$ while $\dot{y}$ is wrt time. The $Ta$ term can be seen as the *ground-state energy*.We ignore this as it doesn't affect fixed points. So we have:

$$S[y] = \int \int_0^a \left( \frac{1}{2} \rho \dot{y}^2 - \frac{1}{2} T (y')^2 \right) \, \mathrm{d}x \, \mathrm{d}t$$

So through integration by parts (boundary term assume vanishes):

$$\delta S[y] = \int \int_0^a \left( \rho \dot{y} \frac{\partial}{\partial t} \delta y - T y' \frac{\partial}{\partial x} \delta y \right) \, \mathrm{d}x \, \mathrm{d}t = \int \int_0^a \delta y (\rho \ddot{y} - T y'') \, \mathrm{d}x \, \mathrm{d}t = 0$$

So we have with $v^2 = T/\rho$:

$$\ddot{y} - v^2 y'' = 0,$$

This is the wave equation in two dimensions, simplified. The general solution is:

$$y(x, t) = f_+(x - vt) + f_-(x + vt),$$

A superposition of a wave travelling rightwards and a wave travelling leftwards.

# 4 The second variation

## 4.1 The second variation

We looked at the "first derivatives" of functionals for fixed points. But for the nature of these points, We need the *second variations.* Suppose $x(t) = x_0(t)$ is a stationary point of $F[x]$. Let $\delta x(t) = \varepsilon\xi(t)$ with constant $\varepsilon \ll 1$. We only consider functionals of the form with fixed boundaries $\xi(\alpha) = \xi(\beta) = 0$:

$$F[x] = \int_\alpha^\beta f(x, \dot{x}, t)\, \mathrm{d}t$$

We consider a variation $x \mapsto x + \delta x$ and expand the integrand to second order:

$$\begin{aligned}
&f(t + \varepsilon\xi, \dot{x} + \varepsilon\dot{\xi}, t) - f(x, \dot{x}, t)\\
&= \varepsilon\left(\xi\frac{\partial f}{\partial \dot{x}} + \dot{\xi}\frac{\partial f}{\partial \dot{x}}\right) + \frac{\varepsilon^2}{2}\left(\xi^2\frac{\partial^2 f}{\partial x^2} + 2\xi\dot{\xi}\frac{\partial^2 f}{\partial x\partial \dot{x}} + \dot{\xi}^2\frac{\partial^2 f}{\partial \dot{x}^2}\right) + O(\varepsilon^3)
\end{aligned}$$

Noting that $2\xi\dot{\xi} = (\xi^2)'$ and integrating by parts, we obtain

$$= \varepsilon\xi\left[\frac{\partial f}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] + \frac{\varepsilon^2}{2}\left\{\xi^2\left[\frac{\partial^2 f}{\partial x^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 f}{\partial x\partial \dot{x}}\right)\right] + \dot{\xi}^2\frac{\partial f}{\partial \dot{x}^2}\right\}.$$

plus some boundary terms which vanish. So

$$F[x + \varepsilon\xi] - F[x] = \int_\alpha^\beta \left\{\epsilon\xi\left[\frac{\partial f}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \dot{x}}\right)\right] + \frac{\varepsilon^2}{2}\delta^2 F[x, \xi] + O(\varepsilon^3)\right\}\, \mathrm{d}t,$$

where

$$\delta^2 F[x, \xi] = \int_\alpha^\beta \left\{\xi^2\left[\frac{\partial^2 f}{\partial x^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 f}{\partial x\partial \dot{x}}\right)\right] + \dot{\xi}^2\frac{\partial^2 f}{\partial \dot{x}^2}\right\}\, \mathrm{d}t$$

is a functional of both $x(t)$ and $\xi(t)$. This is analogous to $\delta\mathbf{x}^T H(\mathbf{x})\delta\mathbf{x}$ appearing in the expansion of a normal $f(\mathbf{x})$. Similar to normal functions, if $\delta^2 F[x, \xi] > 0$ for all non-zero $\xi$ and all allowed $x$, then a solution $x_0(t)$ of $\frac{\delta F}{\delta x} = 0$ is an absolute minimum.

**Example.** Recall we showed a straight line is a stationary point for the curve-length functional, but what if it is a maximum and the shortest distance is to the sun and back? (oh no!) Remember $f = \sqrt{1 + (y')^2}$:

$$\frac{\partial f}{\partial y} = 0, \quad \frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + (y')^2}}, \quad \frac{\partial^2 f}{\partial y'^2} = \frac{1}{\sqrt{1 + (y')^2}^3},$$

with the other second derivatives zero. So we have

$$\delta^2 F[y, \xi] = \int_\alpha^\beta \frac{\dot{\xi}^2}{(1 + (y')^2)^{3/2}}\, \mathrm{d}x > 0$$

As the straight line is a stationary function of the boundary, it's the minimum (Aw..).

But what if the condition doesn't hold? For these, we need to consider

$$\delta^2 F[x_0, \xi] = \int_\alpha^\beta (\rho(t)\dot{\xi}^2 + \sigma(t)\xi^2)\, \mathrm{d}t,$$

where

$$\rho(t) = \left.\frac{\partial^2 f}{\partial \dot{x}^2}\right|_{x=x_0}, \quad \sigma(t) = \left[\frac{\partial^2 f}{\partial x^2} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial^2 f}{\partial x \partial \dot{x}}\right)\right]_{x=x_0}.$$

This is the same form as the Sturm-Liouville problem. For $x_0$ to minimize $F[x]$ locally, we need $\delta^2 F[x_0, \xi] > 0$. A necessary condition, called *Legendre condition*, for this is

$$\rho(t) \geq 0,$$

The intuition (proof tedious, omitted) is: let $\rho(t)$ be negative in some interval $I \subseteq [\alpha, \beta]$. Then we can find a $\xi(t)$ that makes $\delta^2 F[x_0, \xi]$ negative. Make $\xi$ zero outside $I$, and small but oscillating inside $I$. Then $\dot{x}^2$ will be large while $\xi^2$ tiny. So we can make $\delta^2 F[y, \xi]$ arbitrarily negative. But this condition is *not* sufficient. Of course, $\rho(t) > 0, \sigma(t) \geq 0$ is sufficient, but it isn't interesting.

### 4.2 Jacobi condition for local minima of $F[x]$

Legendre tried to prove that $\rho > 0$ is a sufficient condition for $\delta^2 F > 0$, but he failed. However, turned out he was close:Assuming his condition, the strong Legendre condition, $\rho(t) > 0$ for $\alpha < t < \beta$ and boundary conditions $\xi(\alpha) = \xi(\beta) = 0$, when is this sufficient for $\delta^2 F > 0$?

First of all, notice that for any smooth function $w(x)$, we have

$$\int_\alpha^\beta (w\xi^2)' \;\mathrm{d}t = \int_\alpha^\beta (2w\xi\dot{\xi} + \dot{w}\xi^2)\;\mathrm{d}t = 0$$

since this is a total derivative and evaluates to $w\xi(\alpha) - w\xi(\beta) = 0$. This allows us to rewrite $\delta^2 F$ as

$$\delta^2 F = \int_\alpha^\beta \left(\rho\dot{\xi}^2 + 2w\xi\dot{\xi} + (\sigma + \dot{w})\xi^2\right)\;\mathrm{d}t.$$

Now complete the square in $\xi$ and $\dot{\xi}$. So

$$\delta^2 F = \int_\alpha^\beta \left[\rho\left(\dot{\xi} + \frac{w}{\rho}\xi\right)^2 + \left(\sigma + \dot{w} - \frac{w^2}{\rho}\right)\xi^2\right]\;\mathrm{d}t$$

This is non-negative if

$$w^2 = \rho(\sigma + \dot{w}). \tag{$*$}$$

So as long as we have a solution to this, we know $\delta^2 F$ is non-negative. Could it be that $\delta^2 F = 0$? Turns out not. If yes, then $\dot{\xi} = -\frac{w}{\rho}\xi$. Solving this:

$$\xi(x) = C\exp\left(-\int_\alpha^x \frac{w(s)}{\rho(s)}\;\mathrm{d}s\right).$$

We know $\xi(\alpha) = 0$. But $\xi(\alpha) = Ce^0$. So $C = 0$, thus $\xi = 0$.

So all we need to do is to find a solution to $(*)$, and we are sure that $\delta^2 F > 0$. We can convert this into a linear equation by defining $w$ in terms of a new function $u$ by $w = -\rho\dot{u}/u$. Then $(*)$ becomes

$$\rho\left(\frac{\dot{u}}{u}\right)^2 = \sigma - \left(\frac{\rho\dot{u}}{u}\right) = \sigma - \frac{(\rho\dot{u})'}{u} + \rho\left(\frac{\dot{u}}{u}\right)^2. \quad \text{or} \quad -(\rho\dot{u})' + \sigma u = 0.$$

This is the *Jacobi accessory equation*, a second-order linear ODE.

But not every solution $u$ will do. Recall $u$ is used to produce $w$ via $w = -\rho\dot{u}/u$. So within $[\alpha, \beta]$, we cannot have $u = 0$. If we can find a non-zero $u(x)$ satisfying the equation, then $\delta^2 F > 0$ for $\xi \neq 0$, and hence $y_0$ is a local minimum of $F$.

A suitable solution will always exists for sufficiently small $\beta - \alpha$, but may not exist if $\beta - \alpha$ is too large, as from the following intuition: If $\beta - \alpha$ is large, we can make $\xi$ very negative without a lot of slope, so $\delta^2 F < 0$ given $\sigma < 0$.

**Example** (Geodesics on unit sphere)**.** For any curve $C$ on the sphere, we have

$$L = \int_C \sqrt{\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2}.$$

If $\theta$ or $\phi$ is a good parameter of the curve, then

$$L[\phi] = \int_{\theta_1}^{\theta_2} \sqrt{1 + \sin^2\theta(\phi')^2}\,\mathrm{d}\theta. \quad \text{or} \quad L[\theta] = \int_{\phi_1}^{\phi_2} \sqrt{(\theta')^2 + \sin^2\theta}\,\mathrm{d}\phi.$$

We will look at the second case. We have

$$f(\theta, \theta') = \sqrt{(\theta')^2 + \sin^2\theta}.$$

So

$$\frac{\partial f}{\partial \theta} = \frac{\sin\theta\cos\theta}{\sqrt{(\theta')^2 + \sin^2\theta}}, \quad \frac{\partial f}{\partial \theta'} = \frac{\theta'}{\sqrt{(\theta')^2 + \sin^2\theta}}.$$

Since $\frac{\partial f}{\partial \phi} = 0$, we have the first integral

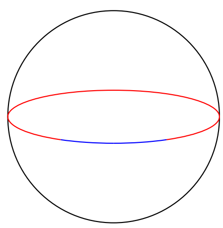
$$\text{const} = f - \theta'\frac{\partial f}{\partial \theta'} = \frac{\sin^2\theta}{\sqrt{(\theta')^2 + \sin^2\theta}}$$

So a solution is

$$c\sin^2\theta = \sqrt{(\theta')^2 + \sin^2\theta}.$$

Here we need $c \geq 1$ for the equation to make sense.

We will consider the case where $c = 1$ (in fact, we can show that we can always orient our axes such that $c = 1$). This occurs when $\theta$ is a constant. Then our first integral gives $\sin^2\theta = \sin\theta$. So $\sin\theta = 1$ and $\theta = \pi/2$a curve on the equator. (we ignore the case $\sin\theta = 0$) There are two equatorial solutions to the Euler-Lagrange equations. Which, if any, minimizes $L[\theta]$?



We have

$$\left.\frac{\partial^2 f}{\partial(\theta')^2}\right|_{\theta=\pi/2} = 1, \quad \frac{\partial^2 f}{\partial\theta\partial\theta'} = -1, \quad \frac{\partial^2}{\partial\theta\partial\theta'} = 0.$$

So $\rho(x) = 1$ and $\sigma(x) = -1$. So

$$\delta^2 F = \int_{\phi_1}^{\phi_2} ((\xi')^2 - \xi^2) \, \mathrm{d}\phi.$$

The Jacobi accessory equation is $u'' + u = 0$. So the general solution is $u \propto \sin\phi - \gamma\cos\phi$. This is equal to zero if $\tan\phi = \gamma$.

Looking at the graph of $\tan\phi$, we see that $\tan$ has a zero every $\pi$ radians. Hence if the domain $\phi_2 - \phi_1$ is greater than $\pi$ (ie. we go the long way from the first point to the second), it will always contain some values for which $\tan\phi$ is zero. So we cannot conclude that the longer path is a local minimum (obviously not a global minimum). But we also cannot conclude that it is *not* a local minimum, since condition is sufficient and not necessary. On the other hand, if $\phi_2 - \phi_1$ is less than $\pi$, then we will be able to pick a $\gamma$ such that $u$ is non-zero in the domain.