Euler Lagrange Equation

Different notations for derivatives
Different notations for derivatives will be used for different occasions.

$$
\begin{aligned}
df(x)/dx &= f_x(x) &&= f^{(1)}(x) &&= f'(x) \\
d^2f(x)/dx^2 &= f_{xx}(x) &&= f^{(2)}(x) &&= f''(x) \\
d^nf(x)/dx^n &= f_{xx...x}(x) &&= f^{(n)}(x) \\
d^2f(x,y)/dxdy &= f_{xy}(x,y)
\end{aligned}
$$

- Notation $f'(x)$ will be used if only 1st and 2nd derivatives are involved.
- Notation $f^{(n)}(x)$ will be used if higher order derivatives are involved.
- Notation $f_{xy}(x,y)$ will be used if there are more than one variables.

Brief review of partial derivative and chain rule
Given a function of x, y and z, **partial derivatives** of the function are defined as the following limits, with all but one variable of interest kept constant.

$$
\begin{aligned}
\frac{\partial f(x,y,z)}{\partial x} &= \lim_{\Delta x \to 0} \frac{f(x+\Delta x,y,z)-f(x,y,z)}{\Delta x} \\
\frac{\partial f(x,y,z)}{\partial y} &= \lim_{\Delta y \to 0} \frac{f(x,y+\Delta y,z)-f(x,y,z)}{\Delta y} \\
\frac{\partial f(x,y,z)}{\partial z} &= \lim_{\Delta z \to 0} \frac{f(x,y,z+\Delta z)-f(x,y,z)}{\Delta z}
\end{aligned}
$$

In case, if y and z are functions of x, we can further define a **total derivative** wrt x as :

$$
\begin{aligned}
\frac{df(x,y,z)}{dx} &= \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y(x+\Delta x), z(x+\Delta x)) - f(x,y(x),z(x))}{\Delta x} \\
\frac{df(x,y,z)}{dx} &\neq \frac{\partial f(x,y,z)}{\partial x}
\end{aligned}
$$

In fact, from chain rule, we have :

$$
\frac{df(x,y,z)}{dx} = \frac{\partial f(x,y,z)}{\partial x} + \frac{\partial f(x,y,z)}{\partial y}\frac{dy(x)}{dx} + \frac{\partial f(x,y,z)}{\partial z}\frac{dz(x)}{dx}
$$

Please note that we use '∂' for partial derivative and 'd' for total derivative in the above equation. If f(y(x), z(x)) contains no x explicitly, it depends on x through y and z, then we have :

$$
\begin{aligned}
\frac{\partial f(x,y,z)}{\partial x} &= 0 \\
\frac{df(x,y,z)}{dx} &= \frac{\partial f(x,y,z)}{\partial y}\frac{dy(x)}{dx} + \frac{\partial f(x,y,z)}{\partial z}\frac{dz(x)}{dx}
\end{aligned}
$$

In other words, $\partial_x f$ denotes the dependence of function f on explicit occurrence of x, while $d_x f$ denotes the dependence of function f on both explicit and implicit occurrence of x.

Functional

Function is a mapping from a value to a value (i.e. f : $\Re^n \rightarrow \Re^1$), where value can be a vector or a complex number. Functional is a mapping from a function to a value (i.e. F : f$\rightarrow \Re^1$), in other words, functional is a function of function. Examples of functional include : mapping from a function to its area, mapping from a function to its arc length etc.

- Area of a curve f : $F : f \rightarrow \int_a^b f(x)dx$ mapping from f to area
- Arc length of a curve f : $F : f \rightarrow \int_a^b \sqrt{1+f_x(x)^2}\,dx$ mapping from f to arc length
- L^p norm of a function f : $F : f \rightarrow \left(\int_a^b |f(x)|^p\,dx\right)^{1/p}$ mapping from f to L^p norm

Here is an explanation for the above 'arc length functional' example.

$$\begin{aligned} dl(x) &= \sqrt{(dx)^2+(df(x))^2} \\ &= \sqrt{(dx)^2+(f_x(x)dx)^2} \\ &= \sqrt{1+f_x(x)^2}\,dx \\ \Rightarrow \int_{x=a}^{x=b} dl(x) &= \int_a^b \sqrt{1+f_x(x)^2}\,dx \end{aligned}$$

Functional usually (but not always) appears as integral form. Here is a typical general form of functional that we usually come across in functional optimization problem, in which the **objective functional** is maximized or minimized with respect to the function, in other words, our target is to solve for an **unknown function** f(x) so as to extremize the objective functional. The general form of objective functional contains x, f(x), and $f_x(x)$, the 1st derivative of f(x) is included in the objective functional because we want to consider the 1st order smoothness of f(x) when we optimize the functional.

$$F : f \rightarrow \int_a^b H(x, f(x), f_x(x))dx$$

or $$F : f \rightarrow \int_a^b H(x, f(x), f_x(x))\mu(dx)$$ (general definition from wiki)

where μ is a function of dx, for example, in finding arc length, we have :

$$F : f \rightarrow \int_a^b \mu(dx)$$ (where $\mu(dx) = \sqrt{(dx)^2+(df(x))^2}$)

Apart from the simplest form in functional 1a (see below), we can generalize the objective functional by adding some higher order derivatives in functional 1b, or adding multiple variables in the unknown function f(x,y) in functional 2, or adding multiple unknown functions with the same variable x in functional 3, or adding multiple unknown functions, each with multiple variables in functional 4. There are 8 combinations in total, some of them are omitted for clarity. The arguments for function are put inside ordinary bracket, like f(x), while the arguments of functional are put inside square bracket, like F[f(x)].

$$F[f(x)] = \int_{x_0}^{x_1} H(x, f(x), f_x(x))dx \qquad f(x) \in C^2_{[x_0,x_1]} \qquad \text{(functional 1a)}$$

$$F[f(x)] = \int_{x_0}^{x_1} H(x, f(x), f_x(x), f_{xx}(x), f_{xxx}(x))dx \qquad f(x) \in C^2_{[x_0,x_1]} \qquad \text{(functional 1b)}$$

$$F[f(x,y)] = \int_{y_0}^{y_1}\int_{x_0}^{x_1} H(x, y, f(x,y), f_x(x,y), f_y(x,y))dxdy \qquad f(x) \in C^2_{\Omega} \qquad \text{(functional 2)}$$

$$F[f(x), g(x)] = \int_{x_0}^{x_1} H(x, f(x), g(x), f_x(x), g_x(x))dx \qquad f(x) \in C^2_{[x_0,x_1]} \qquad \text{(functional 3)}$$

$$F[f(x,y), g(x,y)] = \int_{y_0}^{y_1}\int_{x_0}^{x_1} H(x, y, f(x,y), g(x,y), f_x(x,y), f_y(x,y), g_x(x,y), g_y(x,y))dxdy \qquad f(x) \in C^2_{\Omega} \qquad \text{(functional 4)}$$

In functional optimization, we search for the optimal function f(x) within solution space $f(x) \in C^2_{[x0,x1]}$, where $C^2_{[x0,x1]}$ denotes the set of functions that are twice differentiable with respect to x within range $[x_0, x_1]$, this is necessary because twice differentiability of f(x) is required in deriving the Euler Lagrange equation. Later we will see that the optimal solution of f(x) is given by the Euler Lagrange equation, which is a differential equation of function f(x), hence by solving the differential equation (i.e. Euler Lagrange equation), we will get the optimal function f(x). Furthermore, we do not assume any parameterized model for function f(x), otherwise, the problem will become optimizing the **objective function** with respect to model parameters, which is an ordinary optimization. Functional 1 and 2 are univariate optimization, because there is only one unknown function f(x), while functional 3 and 4 are bivariate optimization, because there are two unknown functions f(x) and g(x). We adopt the following notations :

`index N and n` = order of derivative
`index M and m` = dimension of x
`index K and k` = dimension of function
`function e(x)` = function of x (not necessarily an exponential)

Euler Lagrange equation

Euler Lagrange equation states that if f(x) is a function such that $f(x) \in C^2_{[a,b]}$, which minimizes the functional :

$$F[f(x)] = \int_a^b H(x, f, f')dx \qquad \text{where } f \in C^2_{[a,b]}$$

then the following differential equation must be satisfied :

$$\frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f'} = 0$$ known as Euler Lagrange equation, please identify '∂' and 'd'

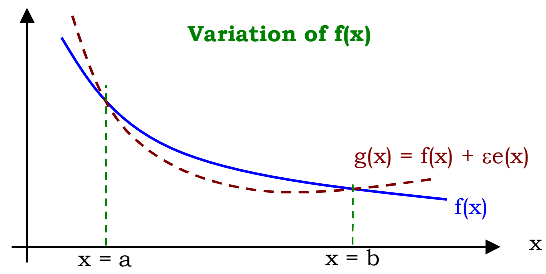
<u>Proof</u>

Lets define variation of f(x) is defined as :

$$g(x) = f(x) + \varepsilon e(x)$$
$$g'(x) = f'(x) + \varepsilon e'(x)$$

where function $e(x) \in C^2_{[a,b]}$, such that :

$$e(a) = e(b) = 0$$
$$\Rightarrow g(a) = f(a)$$
$$\Rightarrow g(b) = f(b)$$



We construct a univariate function G (with parameter ε) to replace the functional F.

$$G(\varepsilon) = F[g(x)] = \int_a^b H(x, g(x), g'(x))dx$$

Note that G(ε) is a function of ε, while F[g(x)] is a functional of g(x). It is given that f(x) optimizes F[g(x)], hence F[g(x)] is optimized when ε is zero, and since G(ε) = F[g(x)], G(ε) is optimized when ε is zero. G(ε) is a function, we can apply ordinary calculus, i.e. dG(ε)/dε = 0 when ε = 0. In the following proof, please notice when we can use total derivative 'd', and when we can use partial derivative '∂'.

$$\frac{dG(\varepsilon)}{d\varepsilon} = \frac{d}{d\varepsilon}\int_a^b H(x, g(x), g'(x))dx$$

$$= \int_a^b \frac{d}{d\varepsilon} H(x, g(x), g'(x))dx$$ (by Leibniz rule : $\frac{d}{dt}\int_a^b f(x,t)dx = \int_a^b \frac{\partial}{\partial t} f(x,t)dx$)

$$= \int_a^b \left(\frac{\partial H}{\partial x}\frac{\partial x}{\partial \varepsilon} + \frac{\partial H}{\partial g}\frac{\partial g}{\partial \varepsilon} + \frac{\partial H}{\partial g'}\frac{\partial g'}{\partial \varepsilon} \right) dx$$ (by chain rule)

$$= \int_a^b \left(\frac{\partial H}{\partial g} e + \frac{\partial H}{\partial g'} e' \right) dx$$ (since $\frac{\partial x}{\partial \varepsilon} = 0$, $\frac{\partial g}{\partial \varepsilon} = e$ and $\frac{\partial g'}{\partial \varepsilon} = e'$)

$$\left.\frac{dG(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = \left.\int_a^b \left(\frac{\partial H}{\partial g} e + \frac{\partial H}{\partial g'} e' \right) dx\right|_{\varepsilon=0}$$

$$= \int_a^b \left(\frac{\partial H}{\partial f} e + \frac{\partial H}{\partial f'} e' \right) dx$$ (see remark 1)

$$= \int_a^b \frac{\partial H}{\partial f} e dx - \int_a^b \left(\frac{d}{dx}\frac{\partial H}{\partial f'} \right) e dx$$ (see remark 2)

$$0 = \int_a^b \left(\frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f'} \right) e dx$$ (since $\left.\frac{dG(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = 0$)

$$\Rightarrow \boxed{0 = \frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f'}}$$ (see remark 3)

Note : Symbol 'e' does not stand for exponential.

Euler Lagrange equation with higher order derivatives

Euler Lagrange equation with higher order derivative states that if f(x) is a function such that $f(x) \in C^{N+1}_{[a,b]}$, which minimizes the functional :

$$F[f(x)] = \int_a^b H(x, f, f^{(1)}, f^{(2)}, ..., f^{(N)}) dx \qquad \text{where } f \in C^{N+1}_{[a,b]}$$

then the following differential equation must be satisfied :

$$\frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f^{(1)}} + \frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(2)}} - ... + (-1)^N \frac{d^N}{dx^N}\frac{\partial H}{\partial f^{(N)}} = 0$$

Proof

Lets define variation of f(x) is defined as :

$$\begin{aligned} g(x) &= f(x) + \varepsilon e(x) \\ g^{(n)}(x) &= f^{(n)}(x) + \varepsilon e^{(n)}(x) \qquad \forall n \in [1, N] \end{aligned}$$

where function $e(x) \in C^{N+1}_{[a,b]}$, such that :

$$\begin{array}{lllllll} e(a) = e(b) = 0 & \Rightarrow & g(x) & = & f(x) & \text{for} & x = a,b \\ e^{(1)}(a) = e^{(1)}(b) = 0 & \Rightarrow & g^{(1)}(x) & = & f^{(1)}(x) & \text{for} & x = a,b \\ e^{(2)}(a) = e^{(2)}(b) = 0 & \Rightarrow & g^{(2)}(x) & = & f^{(2)}(x) & \text{for} & x = a,b \\ \dots & \Rightarrow & \dots & & & & \\ e^{(N-1)}(a) = e^{(N-1)}(b) = 0 & \Rightarrow & g^{(N-1)}(x) & = & f^{(N-1)}(x) & \text{for} & x = a,b \end{array}$$

while there is no limitation for $e^{(N)}(a)$ and $e^{(N)}(b)$. We can then define function $G(\varepsilon)$ as :

$$\begin{aligned} G(\varepsilon) &= F[g(x)] \\ &= \int_a^b H(x, g, g^{(1)}, g^{(2)}, g^{(3)}, ..., g^{(N)}) dx \end{aligned}$$

Following the same argument as the 1st order derivative case, we have :

$$\begin{aligned} \frac{dG(\varepsilon)}{d\varepsilon} &= \frac{d}{d\varepsilon}\int_a^b H(x, g, g^{(1)}, g^{(2)}, g^{(3)}, ..., g^{(N)}) dx \\ &= \int_a^b \frac{d}{d\varepsilon} H(x, g, g^{(1)}, g^{(2)}, g^{(3)}, ..., g^{(N)}) dx && \text{(by Leibniz rule)} \\ &= \int_a^b \left(\frac{\partial H}{\partial x}\frac{\partial x}{\partial \varepsilon} + \frac{\partial H}{\partial g}\frac{\partial g}{\partial \varepsilon} + \frac{\partial H}{\partial g^{(1)}}\frac{\partial g^{(1)}}{\partial \varepsilon} + \frac{\partial H}{\partial g^{(2)}}\frac{\partial g^{(2)}}{\partial \varepsilon} + ... + \frac{\partial H}{\partial g^{(N)}}\frac{\partial g^{(N)}}{\partial \varepsilon} \right) dx && \text{(by chain rule)} \\ &= \int_a^b \left(\frac{\partial H}{\partial g} e + \frac{\partial H}{\partial g^{(1)}} e^{(1)} + \frac{\partial H}{\partial g^{(2)}} e^{(2)} + ... + \frac{\partial H}{\partial g^{(N)}} e^{(N)} \right) dx && \left(\text{since } \frac{\partial x}{\partial \varepsilon} = 0 \text{ and } \frac{\partial g^{(n)}}{\partial \varepsilon} = e^{(n)}\right) \\ \left.\frac{dG(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} &= \left.\int_a^b \left(\frac{\partial H}{\partial g} e + \frac{\partial H}{\partial g^{(1)}} e^{(1)} + \frac{\partial H}{\partial g^{(2)}} e^{(2)} + ... + \frac{\partial H}{\partial g^{(N)}} e^{(N)} \right) dx \right|_{\varepsilon=0} \\ &= \int_a^b \left(\frac{\partial H}{\partial f} e + \frac{\partial H}{\partial f^{(1)}} e^{(1)} + \frac{\partial H}{\partial f^{(2)}} e^{(2)} + ... + \frac{\partial H}{\partial f^{(N)}} e^{(N)} \right) dx && \text{(see remark 1)} \\ &= \int_a^b \frac{\partial H}{\partial f} e dx - \int_a^b \left(\frac{d}{dx}\frac{\partial H}{\partial f^{(1)}} \right) e dx + \int_a^b \left(\frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(2)}} \right) e dx - ... + (-1)^N \int_a^b \left(\frac{d^N}{dx^N}\frac{\partial H}{\partial f^{(N)}} \right) e dx && \text{(see remark 2)} \\ 0 &= \int_a^b \left(\frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f^{(1)}} + \frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(2)}} - ... + (-1)^N \frac{d^N}{dx^N}\frac{\partial H}{\partial f^{(N)}} \right) e dx && \left(\text{since } \left.\frac{dG(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = 0\right) \\ \Rightarrow \quad 0 &= \frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f^{(1)}} + \frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(2)}} - ... + (-1)^N \frac{d^N}{dx^N}\frac{\partial H}{\partial f^{(N)}} && \text{(see remark 3)} \end{aligned}$$

Euler Lagrange equation for multi-functions

Euler Lagrange equation for multi-functions states that if $f_k(x)$ $\forall k\in[1,K]$ are functions such that $f(x)\in C^{N+1}_{[a,b]}$, which minimizes the functional :

$$F[f_1,f_2,...,f_K]=\int_a^b H(x,f_1,f_2,...,f_K,f_1^{(1)},f_2^{(1)},...,f_K^{(1)},f_1^{(2)},f_2^{(2)},...,f_K^{(2)},...,f_1^{(N)},f_2^{(N)},...,f_K^{(N)})dx \qquad \text{where } f_k\in C^{N+1}_{[a,b]},\ \forall k\in[1,K]$$

then the following differential equation must be satisfied :

$$\frac{\partial H}{\partial f_k}-\frac{d}{dx}\frac{\partial H}{\partial f_k^{(1)}}+\frac{d^2}{dx^2}\frac{\partial H}{\partial f_k^{(2)}}-...+(-1)^N\frac{d^N}{dx^N}\frac{\partial H}{\partial f_k^{(N)}}=0 \qquad \forall k\in[1,K]$$

Proof

Lets define variation of f(x) is defined as :

$$\begin{aligned}
g_k(x) &= f_k(x)+\varepsilon_k e(x) && \forall k\in[1,K]\\
g_k^{(n)}(x) &= f_k^{(n)}(x)+\varepsilon_k e^{(n)}(x) && \forall k\in[1,K] \text{ and } \forall n\in[1,N]
\end{aligned}$$

where function $e(x)\in C^{N+1}_{[a,b]}$, such that :

$$\begin{aligned}
e(a) &= e(b) = 0 &&\Rightarrow & g_k(x) &= f_k(x) && \forall k\in[1,K] \text{ and for } x=a,b\\
e^{(1)}(a) &= e^{(1)}(b) = 0 &&\Rightarrow & g_k^{(1)}(x) &= f_k^{(1)}(x) && \forall k\in[1,K] \text{ and for } x=a,b\\
e^{(2)}(a) &= e^{(2)}(b) = 0 &&\Rightarrow & g_k^{(2)}(x) &= f_k^{(2)}(x) && \forall k\in[1,K] \text{ and for } x=a,b\\
... & &&\Rightarrow & ... & &&\\
e^{(N-1)}(a) &= e^{(N-1)}(b) = 0 &&\Rightarrow & g_k^{(N-1)}(x) &= f_k^{(N-1)}(x) && \forall k\in[1,K] \text{ and for } x=a,b
\end{aligned}$$

while there is no limitation for $e^{(N)}(a)$ and $e^{(N)}(b)$. We can then define function $G(\varepsilon)$ as :

$$\begin{aligned}
G(\varepsilon_1,...,\varepsilon_K) &= F[g_1(x),g_2(x),...,g_K(x)]\\
&= \int_a^b H(x,g_1,g_2,...,g_K,g_1^{(1)},g_2^{(1)},...,g_K^{(1)},g_1^{(2)},g_2^{(2)},...,g_K^{(2)},...,g_1^{(N)},g_2^{(N)},...,g_K^{(N)})dx
\end{aligned}$$

Following the same argument as the 1st order derivative case, we have :

$$\begin{aligned}
\frac{dG(\varepsilon_1,...)}{d\varepsilon_k} &= \frac{d}{d\varepsilon}\int_a^b H(x,g_1,g_2,...,g_K,g_1^{(1)},g_2^{(1)},...,g_K^{(1)},g_1^{(2)},g_2^{(2)},...,g_K^{(2)},...,g_1^{(N)},g_2^{(N)},...,g_K^{(N)})dx && (\forall k\in[1,K])\\
&= \int_a^b \frac{d}{d\varepsilon}H(x,g_1,g_2,...,g_K,g_1^{(1)},g_2^{(1)},...,g_K^{(1)},g_1^{(2)},g_2^{(2)},...,g_K^{(2)},...,g_1^{(N)},g_2^{(N)},...,g_K^{(N)})dx && \text{(by Leibniz rule)}\\
&= \int_a^b\left(\frac{\partial H}{\partial x}\frac{\partial x}{\partial\varepsilon}+\sum_{i=1}^K\left(\frac{\partial H}{\partial g_i}\frac{\partial g_i}{\partial\varepsilon_k}+\frac{\partial H}{\partial g_i^{(1)}}\frac{\partial g_i^{(1)}}{\partial\varepsilon_k}+\frac{\partial H}{\partial g_i^{(2)}}\frac{\partial g_i^{(2)}}{\partial\varepsilon_k}+...+\frac{\partial H}{\partial g_i^{(N)}}\frac{\partial g_i^{(N)}}{\partial\varepsilon_k}\right)\right)dx && \text{(by chain rule)}\\
&= \int_a^b\left(\frac{\partial H}{\partial g_k}e+\frac{\partial H}{\partial g_k^{(1)}}e^{(1)}+\frac{\partial H}{\partial g_k^{(2)}}e^{(2)}+...+\frac{\partial H}{\partial g_k^{(N)}}e^{(N)}\right)dx && \left(\text{since } \frac{\partial x}{\partial\varepsilon_k}=0,\ \frac{\partial g_i^{(n)}}{\partial\varepsilon_k}=\delta_{ik}e^{(n)}\right)\\
\left.\frac{dG(\varepsilon_1,...)}{d\varepsilon_k}\right|_{\varepsilon_k=0} &= \left.\int_a^b\left(\frac{\partial H}{\partial g_k}e+\frac{\partial H}{\partial g_k^{(1)}}e^{(1)}+\frac{\partial H}{\partial g_k^{(2)}}e^{(2)}+...+\frac{\partial H}{\partial g_k^{(N)}}e^{(N)}\right)dx\right|_{\varepsilon_k=0} && (\forall k\in[1,K])\\
&= \int_a^b\left(\frac{\partial H}{\partial f_k}e+\frac{\partial H}{\partial f_k^{(1)}}e^{(1)}+\frac{\partial H}{\partial f_k^{(2)}}e^{(2)}+...+\frac{\partial H}{\partial f_k^{(N)}}e^{(N)}\right)dx && \text{(see remark 1)}\\
&= \int_a^b\frac{\partial H}{\partial f_k}edx-\int_a^b\left(\frac{d}{dx}\frac{\partial H}{\partial f_k^{(1)}}\right)edx+\int_a^b\left(\frac{d^2}{dx^2}\frac{\partial H}{\partial f_k^{(2)}}\right)edx-...+(-1)^N\int_a^b\left(\frac{d^N}{dx^N}\frac{\partial H}{\partial f_k^{(N)}}\right)edx && \text{(see remark 2)}\\
0 &= \int_a^b\left(\frac{\partial H}{\partial f_k}-\frac{d}{dx}\frac{\partial H}{\partial f_k^{(1)}}+\frac{d^2}{dx^2}\frac{\partial H}{\partial f_k^{(2)}}-...+(-1)^N\frac{d^N}{dx^N}\frac{\partial H}{\partial f_k^{(N)}}\right)edx && \left(\text{since } \left.\frac{dG(\varepsilon_1,...)}{d\varepsilon_k}\right|_{\varepsilon_k=0}=0\right)\\
\Rightarrow\quad 0 &= \frac{\partial H}{\partial f_k}-\frac{d}{dx}\frac{\partial H}{\partial f_k^{(1)}}+\frac{d^2}{dx^2}\frac{\partial H}{\partial f_k^{(2)}}-...+(-1)^N\frac{d^N}{dx^N}\frac{\partial H}{\partial f_k^{(N)}} && \text{(see remark 3)}
\end{aligned}$$

Remark 1

$$\left.\frac{\partial H}{\partial g}\right|_{\varepsilon=0} = \left.\lim_{\Delta\to 0}\frac{H(g+\Delta)-H(g)}{\Delta}\right|_{\varepsilon=0} = \lim_{\Delta\to 0}\frac{H(f+\Delta)-H(f)}{\Delta} = \frac{\partial H}{\partial f}$$

$$\left.\frac{\partial H}{\partial g^{(n)}}\right|_{\varepsilon=0} = \left.\lim_{\Delta\to 0}\frac{H(g^{(n)}+\Delta)-H(g^{(n)})}{\Delta}\right|_{\varepsilon=0} = \lim_{\Delta\to 0}\frac{H(f^{(n)}+\Delta)-H(f^{(n)})}{\Delta} = \frac{\partial H}{\partial f^{(n)}}$$

Remark 2

Integration by parts is generalized to cases with n-th order derivative, it is done by applying integration by parts repeatedly. Please note the introduction of total derivative whenever appropriate. Recall : symbol 'e' does not denote exponential.

$$\int_a^b \frac{\partial H}{\partial f^{(n)}} e^{(n)}dx = \int_a^b \left(\frac{\partial H}{\partial f^{(n)}}\right) de^{(n-1)} \qquad \text{(since } e^{(n)}dx = de^{(n-1)}\text{)}$$

$$= \left[\left(\frac{\partial H}{\partial f^{(n)}}\right)e^{(n-1)}\right]_a^b - \int_a^b \left(\frac{d}{dx}\frac{\partial H}{\partial f^{(n)}}\right)e^{(n-1)}dx \qquad \text{(now apply : } e^{(n-1)}(a) = e^{(n-1)}(b) = 0\text{)}$$

$$= -\int_a^b \left(\frac{d}{dx}\frac{\partial H}{\partial f^{(n)}}\right)de^{(n-2)} \qquad \text{(since } e^{(n-1)}dx = de^{(n-2)}\text{)}$$

$$= -\left[\left(\frac{d}{dx}\frac{\partial H}{\partial f^{(n)}}\right)e^{(n-2)}\right]_a^b + \int_a^b \left(\frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(n)}}\right)e^{(n-2)}dx \qquad \text{(now apply : } e^{(n-2)}(a) = e^{(n-2)}(b) = 0\text{)}$$

$$= \int_a^b \left(\frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(n)}}\right)de^{(n-3)} \qquad \text{(since } e^{(n-2)}dx = de^{(n-3)}\text{)}$$

$$= \left[\left(\frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(n)}}\right)e^{(n-3)}\right]_a^b - \int_a^b \left(\frac{d^3}{dx^3}\frac{\partial H}{\partial f^{(n)}}\right)e^{(n-3)}dx \qquad \text{(now apply : } e^{(n-3)}(a) = e^{(n-3)}(b) = 0\text{)}$$

$$= -\int_a^b \left(\frac{d^3}{dx^3}\frac{\partial H}{\partial f^{(n)}}\right)de^{(n-4)} \qquad \text{(since } e^{(n-3)}dx = de^{(n-4)}\text{)}$$

$$= \ldots$$

$$= (-1)^{n-1}\left[\left(\frac{d^{n-1}}{dx^{n-1}}\frac{\partial H}{\partial f^{(n)}}\right)e^{(0)}\right]_a^b + (-1)^n\int_a^b \left(\frac{d^n}{dx^n}\frac{\partial H}{\partial f^{(n)}}\right)e^{(0)}dx \qquad \text{(now apply : } e(a) = e(b) = 0\text{)}$$

$$= (-1)^n\int_a^b \left(\frac{d^n}{dx^n}\frac{\partial H}{\partial f^{(n)}}\right)edx$$

Remark 3

Let f(x) be a continuous function on the interval [a,b]. Suppose for any continuous function e(x) with e(a) = e(b) = 0, such that the inner product is always zero :

$$\int_a^b f(x)e(x)dx = 0$$

then we have f(x) = 0 on the interval [a,b]. Here is the proof. Since e(x) can be any continuous function with e(a) = e(b) = 0, we can choose e(x) to be :

$$e(x) = -f(x)(x-a)(x-b)$$

since f(x) is continuous, g(x) is therefore continuous. Furthermore :

$$(x-a)(x-b) \le 0 \quad \forall x \in [a,b]$$

$$\Rightarrow \quad -f(x)^2(x-a)(x-b) \ge 0 \quad \forall x \in [a,b]$$

$$\Rightarrow \quad f(x)e(x) \ge 0 \quad \forall x \in [a,b] \qquad \text{(since } e(x) = -f(x)(x-a)(x-b)\text{)}$$

The integration of a non negative function f(x)g(x) is zero only if the non negative function is in fact zero.

$$\Rightarrow \quad f(x)e(x) = 0 \quad \forall x \in [a,b]$$

$$\Rightarrow \quad -f(x)^2(x-a)(x-b) = 0 \quad \forall x \in [a,b]$$

$$\Rightarrow \quad f(x) = 0 \quad \forall x \in [a,b] \qquad \text{(since } (x-a)(x-b) < 0 \;\; \forall x \in (a,b) \text{ and } f(x) \text{ is continuous)}$$

Summary of Euler Lagrange equation
Euler Lagrange equation version (1), (2) and (4) have been proved in the previous sections.

(1) single function, single variable, 1st derivative

$$F[f(x)] = \int_a^b H(x, f, f^{(1)})dx \qquad \frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f^{(1)}} = 0$$

(2) single function, single variable, higher order derivatives

$$F[f(x)] = \int_a^b H(x, f, f^{(1)}, f^{(2)}, ..., f^{(N)})dx \qquad \frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f^{(1)}} + \frac{d^2}{dx^2}\frac{\partial H}{\partial f^{(2)}} - ... + (-1)^N \frac{d^N}{dx^N}\frac{\partial H}{\partial f^{(N)}} = 0$$

(3) single function, multi variables, 1st derivative

$$F[f(X)] = \int_\Omega H(x_1, x_2, ..., x_M, f, f_{x_1}, f_{x_2}, ..., f_{x_M})dX \qquad \frac{\partial H}{\partial f} - \frac{d}{dx_1}\frac{\partial H}{\partial f_{x_1}} - \frac{d}{dx_2}\frac{\partial H}{\partial f_{x_2}} - ... - \frac{d}{dx_M}\frac{\partial H}{\partial f_{x_M}} = 0$$

(4) multi function, single variables, 1st derivative

$$F[f_1, f_2, ..., f_K] = \int_a^b H(x, f_1, f_2, ..., f_K, f_1', f_2', ..., f_K')dx \qquad \frac{\partial H}{\partial f_k} - \frac{d}{dx}\frac{\partial H}{\partial f_k'} = 0 \quad \forall k \in [1, K]$$

(K differential equations for solving K unknown functions)

(5) multi function, multi variables, 1st derivative

$$F[f_1, f_2, ..., f_K] = \int_\Omega H\begin{pmatrix} x_1, x_2, ..., x_M, f_1, f_2, ..., f_K, \\ f_{1,x_1}, f_{1,x_2}, ..., f_{1,x_M} \\ f_{2,x_1}, f_{2,x_2}, ..., f_{2,x_M} \\ f_{3,x_1}, f_{3,x_2}, ..., f_{3,x_M} \\ ...... \\ f_{K,x_1}, f_{K,x_2}, ..., f_{K,x_M} \end{pmatrix}dX \qquad \frac{\partial H}{\partial f_k} - \frac{d}{dx_1}\frac{\partial H}{\partial f_{k,x_1}} - \frac{d}{dx_2}\frac{\partial H}{\partial f_{k,x_2}} - ... - \frac{d}{dx_M}\frac{\partial H}{\partial f_{k,x_M}} = 0 \quad \forall k \in [1, K]$$

(6) single function, two variables, higher order derivatives

$$F[f(x,y)] = \int_\Omega H(x, y, f, f_x, f_y, f_{xx}, f_{xy}, f_{yy} ..., f_{\underbrace{xx...x}_N}, ..., f_{\underbrace{yy...y}_N})dX$$

$$\begin{pmatrix} \frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f_x} - \frac{d}{dy}\frac{\partial H}{\partial f_y} + \frac{d^2}{dx^2}\frac{\partial H}{\partial f_{xx}} + \frac{d^2}{dxdy}\frac{\partial H}{\partial f_{xy}} + \frac{d^2}{dy^2}\frac{\partial H}{\partial f_{yy}} \\ -\frac{d^3}{dx^3}\frac{\partial H}{\partial f_{xxx}} + \frac{d^3}{dx^2dy}\frac{\partial H}{\partial f_{xxy}} + \frac{d^3}{dxdy^2}\frac{\partial H}{\partial f_{xyy}} + \frac{d^3}{dy^3}\frac{\partial H}{\partial f_{yyy}} + ... \\ + (-1)^N \frac{d^N}{dx^N}\frac{\partial H}{\partial f_{\underbrace{xx...x}_N}} + ... + (-1)^N \frac{d^N}{dx^N}\frac{\partial H}{\partial f_{\underbrace{yy...y}_N}} \end{pmatrix} = 0$$

Beltrami identity (one function in one variable with 1st order derivative)
Beltrami identity is a special case of Euler Lagrange equation, when H does not depend on x explicitly, which is very common in functional optimization. Lets derive the Beltrami identity starting from chain rule.

$$\begin{aligned} \frac{dH}{dx} &= \frac{\partial H}{\partial x} + \frac{\partial H}{\partial f}f' + \frac{\partial H}{\partial f'}f'' && \text{(since H does not contain x, } \therefore \frac{\partial H}{\partial x} = 0\text{)} \\ &= \frac{\partial H}{\partial f}f' + \frac{\partial H}{\partial f'}f'' \\ &= \boxed{\left(\frac{d}{dx}\frac{\partial H}{\partial f'}\right)f'} + \frac{\partial H}{\partial f'}f'' && \text{(since } \frac{\partial H}{\partial f} = \frac{d}{dx}\frac{\partial H}{\partial f'}\text{, by Euler Lagrange equation)} \\ &= \boxed{\frac{d}{dx}\left(\frac{\partial H}{\partial f'}f'\right) - \frac{\partial H}{\partial f'}f''} + \frac{\partial H}{\partial f'}f'' && \text{(since } \frac{d}{dx}\left(\frac{\partial H}{\partial f'}f'\right) = \boxed{\left(\frac{d}{dx}\frac{\partial H}{\partial f'}\right)}f' + \frac{\partial H}{\partial f'}f''\text{)} \\ &= \frac{d}{dx}\left(\frac{\partial H}{\partial f'}f'\right) \\ \frac{d}{dx}\left(H - \frac{\partial H}{\partial f'}f'\right) &= 0 \end{aligned}$$

finally : $H - \frac{\partial H}{\partial f'}f' = C$ this is known as the Beltrami identity,

and $\frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f'} = 0$ this is Euler Lagrange equation, it is shown here for comparison.

Shortest path between two points on a plane

Shortest path between two points on a plane is a straight line, lets derive it using functional optimization. Suppose the two points are A and B, which have x coordinates x_A and x_B respectively.

$$
\begin{aligned}
F[f(x)] &= \int_A^B dl \\
&= \int_A^B \sqrt{(dx)^2 + (df)^2} \\
&= \int_A^B \sqrt{(dx)^2 + f'^2 (dx)^2} \\
&= \int_{x_A}^{x_B} \sqrt{1+f'^2}\, dx
\end{aligned}
$$

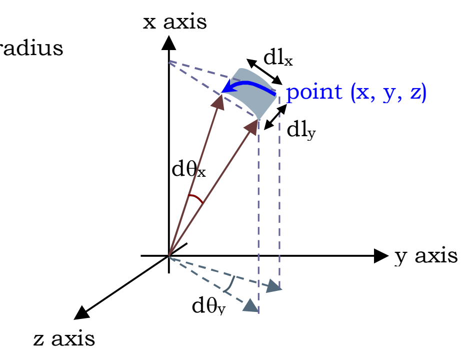
We then apply Euler Lagrange equation :

$$
\begin{aligned}
0 &= \frac{\partial H}{\partial f} - \frac{d}{dx}\frac{\partial H}{\partial f'} \\
&= 0 - \frac{d}{dx}\frac{\partial}{\partial f'}\sqrt{1+f'^2} \\
&= \frac{d}{dx}\frac{1}{2}\frac{2f'}{\sqrt{1+f'^2}} \\
C &= \frac{f'}{\sqrt{1+f'^2}} \\
f'^2 &= C^2 + C^2 f'^2 \\
f' &= \frac{\pm C}{\sqrt{1-C^2}} = C_1 \\
f(x) &= C_1 x + C_2 \qquad \text{which is a straight line}
\end{aligned}
$$

Shortest path between two points on a sphere

Shortest path between two points on a sphere is called a geodesics, lets derive it using functional optimization. The geodesics must be a segment of a great circle (note : great circle of a sphere is defined as the intersection of the sphere with a plane that passes through the sphere's centre). Please read the ***N Dimensional Sphere document***. Suppose the radius of the sphere is r, while θ_x (and θ_y) is the angle between a 3D point and x axis (and y axis respectively), the two points are A and B, which have θ_x coordinates θ_A and θ_B respectively.

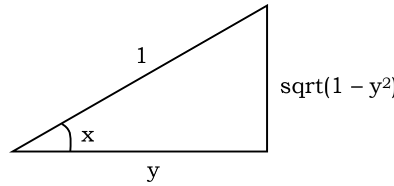
$$
\begin{aligned}
dr &= 0 && \text{lying on a sphere with fixed radius} \\
dl_x &= r d\theta_x && \text{arc length spanned by } \theta_x \\
dl_y &= r \sin\theta_x d\theta_y && \text{arc length spanned by } \theta_y \\
F[f(x)] &= \int_A^B dl \\
&= \int_A^B \sqrt{(dr)^2 + (dl_x)^2 + (dl_y)^2} \\
&= r \times \int_A^B \sqrt{(d\theta_x)^2 + \sin^2\theta_x (d\theta_y)^2} \\
&= r \times \int_{\theta_A}^{\theta_B} \sqrt{1+\sin^2\theta_x f'^2}\, d\theta_x
\end{aligned}
$$



Let $\theta_y = f(\theta_x)$ be the curve of the sphere, which specifies the relationship between the two angles. We then apply Euler Lagrange equation :

$$
\begin{aligned}
0 &= \frac{\partial H}{\partial f} - \frac{d}{d\theta_x}\frac{\partial H}{\partial f'} && \text{note : H depends explicitly on } \theta_x \text{ and f', but not explicitly on f.} \\
&= 0 - \frac{d}{d\theta_x}\frac{1}{2}\frac{2\sin^2\theta_x f'}{\sqrt{1+\sin^2\theta_x f'^2}} \\
C &= \frac{\sin^2\theta_x f'}{\sqrt{1+\sin^2\theta_x f'^2}} && \Rightarrow (1+\sin^2\theta_x f'^2)C^2 = \sin^4\theta_x f'^2 \\
f'^2 &= \frac{C^2}{\sin^4\theta_x - C^2\sin^2\theta_x} \\
f' &= \frac{C}{\sin\theta_x\sqrt{\sin^2\theta_x - C^2}} && \Rightarrow \theta_y = f(\theta_x) = \cos^{-1}\left(\frac{C}{\tan\theta_x\sqrt{1-C^2}}\right) + k
\end{aligned}
$$

Lets verify the solution of the differential equation. First of all, we need to know the derivative of (1) arccos and (2) tan, according to the ***Chain Rule document***, we have :



(1) $$g^{-1}{}'(y) = 1/g'(x) \qquad \text{where } y=g(x)\text{ , i.e. } x=g^{-1}(y)$$

$$\begin{aligned}
\Rightarrow \quad \frac{d}{dy}\cos^{-1} y &= 1/\cos'(x) \qquad \text{where } y=\cos(x)\text{ , i.e. } x=\cos^{-1} y \\
&= -1/\sin(x) \\
&= -1/\sin(\cos^{-1}(y)) \\
&= -1/\sqrt{1-y^2}
\end{aligned}$$

(2) $$\begin{aligned}
\frac{d}{dx}\tan x &= \frac{d}{dx}\frac{\sin x}{\cos x} \\
&= \frac{\cos x}{\cos x} - \frac{\sin x}{\cos^2 x}\times(-\sin x) \\
&= 1+\tan^2 x \\
&= 1/\cos^2 x
\end{aligned}$$

Now, lets verify the solution.

$$\begin{aligned}
f' &= \frac{d}{d\theta_x}(\cos^{-1}(\frac{C}{\tan\theta_x\sqrt{1-C^2}})+k) \\
&= \frac{-1}{\sqrt{1-(\frac{C}{\tan\theta_x\sqrt{1-C^2}})^2}}\frac{-C}{\tan^2\theta_x\sqrt{1-C^2}}\frac{1}{\cos^2\theta_x} \\
&= \frac{1}{\sqrt{1-\frac{C^2}{1-C^2}\frac{1}{\tan^2\theta_x}}}\frac{C}{\sin^2\theta_x\sqrt{1-C^2}} \\
&= \frac{1}{\sqrt{1-C^2-\frac{C^2}{\tan^2\theta_x}}}\frac{C}{\sin^2\theta_x} \\
&= \frac{C}{\sin\theta_x\sqrt{\sin^2\theta_x(1-C^2)-C^2\cos^2\theta_x}} \\
&= \frac{C}{\sin\theta_x\sqrt{\sin^2\theta_x - C^2}}
\end{aligned}$$

Thus, solution to Euler Lagrange equation is verified.

Why does this function describe a great circle?

Lets verify that $\theta_y = \cos^{-1}(\frac{C}{\tan\theta_x\sqrt{1-C^2}})+k$ represents a great circle.

$$\begin{aligned}
\theta_y &= \cos^{-1}(\frac{C}{\tan\theta_x\sqrt{1-C^2}})+k \\
(C_0\tan\theta_x)^{-1} &= \cos(\theta_y - k) \qquad \text{where } C_0=\sqrt{1-C^2}/C \\
(C_0\tan\theta_x)^{-1} &= \cos\theta_y\cos k+\sin\theta_y\sin k \\
r\cos\theta_x &= rC_0\sin\theta_x\cos\theta_y\cos k + rC_0\sin\theta_x\sin\theta_y\sin k \\
r\cos\theta_x &= C_0\cos k(r\sin\theta_x\cos\theta_y) - C_0\sin k(r\sin\theta_x\sin\theta_y) \\
\Rightarrow \quad x &= (C_0\cos k)y-(C_0\sin k)z \\
&= w_y y+w_z z
\end{aligned}$$

which is a 3D plane passes through the origin

where
$$\begin{aligned}
x &= r\cos\theta_x \\
y &= r\sin\theta_x\cos\theta_y \\
z &= r\sin\theta_x\sin\theta_y
\end{aligned}$$

Hence (x, y, z) is a point, lying on a sphere centred at origin with radius r, the solution to Euler Lagrange equation is the intersection of the sphere and a plane that passes through the sphere's centre, and therefore it is a great circle (recall the definition of a great circle : intersection of a sphere with a plane that passes through its centre).

Brachistochrone problem (or quickest descent problem)
Suppose two points A and B, with A lying higher than B, are connected by a trail. Brachistochrone problem requires us to solve for the function of the trail, so that it minimizes the time needed for a frictionless ball falling from A to B, drived only by gravity along the trail.

$$\begin{aligned}
F[f(x)] &= \int_A^B dt \\
&= \int_A^B \frac{1}{v(x)} ds \\
&= \int_A^B \frac{\sqrt{(dx)^2+(dy)^2}}{v(x)} \\
&= \int_{x_A}^{x_B} \frac{\sqrt{1+f'^2(x)}}{v(x)} dx && \text{where } f(x) \text{ is the y coordinate, depending x, i.e. shape of trail}
\end{aligned}$$

Since sum of kinetic energy and potential energy conserve, we have the relation between velocity v(x) and trail f(x).

$$\begin{aligned}
\frac{mv^2(x)}{2} &= mg(f(x_A)-f(x)) && \text{where } f(x_A) > f(x_B) \text{ as point A is higher than point B} \\
v(x) &= \sqrt{2g(f(x_A)-f(x))} \\
F[f(x)] &= \int_{x_A}^{x_B} \sqrt{\frac{1+f'^2(x)}{2g(f(x_A)-f(x))}} dx \\
&= \int_{x_A}^{x_B} \sqrt{\frac{1+f'^2(x)}{-2gf(x)}} dx && \text{assume } f(x_A)=0 \text{ for convenience}
\end{aligned}$$

We then apply the Beltrami identity :

$$\begin{aligned}
C &= H - \frac{\partial H}{\partial f'} f' \\
&= \sqrt{\frac{1+f'^2}{-2gf}} - f'\frac{\partial}{\partial f'}\sqrt{\frac{1+f'^2}{-2gf}} \\
&= \sqrt{\frac{1+f'^2}{-2gf}} - f'\frac{1}{2}\frac{2f'}{\sqrt{-2gf(1+f'^2)}} \\
&= \frac{1+f'^2-f'^2}{\sqrt{-2gf(1+f'^2)}} \\
C^2 &= \frac{1}{-2gf(1+f'^2)} && \text{hence } f(1+f'^2) = -\frac{1}{2gC^2} \\
f(1+f'^2) &= -2r && \text{assume } r = \frac{1}{g(2C)^2}, \text{ i.e. } 2r = \frac{1}{2gC^2} \\
h(1+h'^2) &= 2r && \text{assume } h(x) = -f(x) \text{ and } h'(x) = -f'(x), \text{ i.e. flipping the y axis}
\end{aligned}$$

Function $f(x)$ represents fall, while function $h(x)$ represents height. Solution to the ODE in parametric form is :

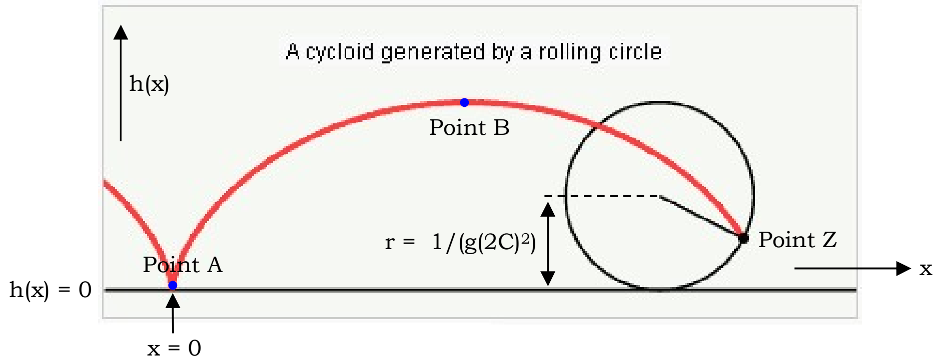
$$x(\theta) = r(\theta - \sin\theta)$$

and $$h(\theta) = r(1-\cos\theta)$$

Lets verify the above solution by substituting into the ODE.

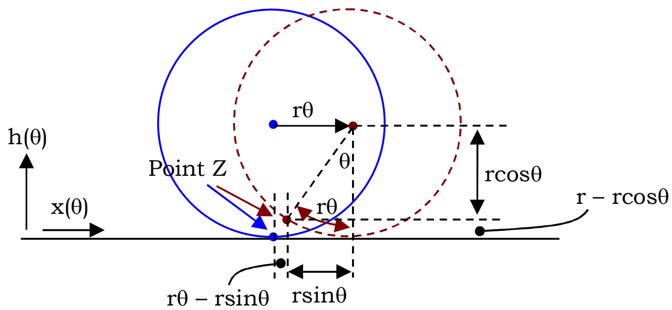
$$\begin{aligned}
h' &= \frac{dh(\theta)}{d\theta}\frac{dx(\theta)}{d\theta}^{-1} \\
&= \frac{r\sin\theta}{r(1-\cos\theta)} \\
h(1+h'^2) &= r(1-\cos\theta)\left(1+\frac{\sin^2\theta}{(1-\cos\theta)^2}\right) \\
&= r(1-\cos\theta)\frac{1-2\cos\theta+\cos^2\theta-\sin^2\theta}{(1-\cos\theta)^2} \\
&= r\frac{2\cos^2\theta-2\cos\theta}{1-\cos\theta} \\
&= 2r && \text{Thus, solution to Beltrami identity is verified.}
\end{aligned}$$

The solution is known as a cycloid. Cycloid is a curve traced by a point on the rim of a circular wheel as the wheel rolls along a straight line. Here is an example of a cycloid. Please note that the y axis represents function h(x), which is identical to vertically flipped functon f(x).



<u>Why does the parametric form solution describe a cycloid?</u>
Lets look at the selected point Z on the rim of the wheel. As the wheel rotates an angle θ, the wheel centre moves to RHS with displacement rθ, while the displacement of point Z (relative to the wheel centre) along x direction is rsinθ towards LHS, thus the displacement of point Z (relative to the ground) along x direction is rθ - rsinθ towards RHS. At the same time, point Z rises by r – rcosθ along y direction. In the following figure, blue and red color denote the wheel before and after the movement. Hence we have :

$$x(\theta) = r(\theta - \sin\theta)$$
and
$$h(\theta) = r(1 - \cos\theta)$$



Optical flow – Horn Schunck's method
Horn Schunck's method can find optical flow between two given images using Euler Lagrange equation, based on the brightness constancy assumption, which means, if a point object (i.e. a pixel) moves from location (x_1, y_1) in image 1 taken at time t_1 to location (x_2, y_2) in image 2 taken at time t_2, its brightness doesn't change. Optical flow is a vector field represented as : $f(x,y)\hat{x} + g(x,y)\hat{y}$.

$$I(x+f(x,y), y+g(x,y), t+\Delta t) = I(x,y,t) \quad \text{brightness constancy}$$
$$I(x+f(x,y), y+g(x,y), t+\Delta t) = I(x,y,t) + I_x f(x,y) + I_y g(x,y) + I_t \Delta t + \ldots \quad \text{Taylor series}$$
$$\Rightarrow \quad I_x f(x,y) + I_y g(x,y) + I_t \Delta t = 0$$
$$\Rightarrow \quad I_x f(x,y) + I_y g(x,y) = I(x,y,t_2) - I(x,y,t_1)$$
$$\Rightarrow \quad \nabla I \cdot [f(x,y), g(x,y)] = I(x,y,t_2) - I(x,y,t_1)$$

I_x and I_y can be found by taking gradient of image 1, I_t can be found by taking difference between image 2 and image 1. However, We cannot make the above equality to hold true for all sample points (x, y) ∊ Ω, instead we try to minimize the sum of error square, and set up an objective functional like the following.

$$F[f,g] = \iint_\Omega (I_x f + I_y g - I_t)^2 dxdy$$

This is not enough, we need to consider the tradeoff between minimizing sum of error square and minimizing irregularity of the optical flow. Therefore of we add gradient magnitude of the optical flow into the objective functional, while λ is the parameter that controls the tradeoff.

$$F[f,g] = \iint_\Omega ((I_x f + I_y g - I_t)^2 + \lambda(|\nabla f|^2 + |\nabla g|^2))dxdy$$
$$= \iint_\Omega ((I_x f + I_y g - I_t)^2 + \lambda(f_x^2 + f_y^2 + g_x^2 + g_y^2))dxdy$$

This is a bifunction and bivariate problem. Applying Euler Lagrange equation, we have two PDEs :

(1) $$\frac{\partial H}{\partial f}-\frac{d}{dx}\frac{\partial H}{\partial f_x}-\frac{d}{dy}\frac{\partial H}{\partial f_y} = 0$$

(2) $$\frac{\partial H}{\partial f}-\frac{d}{dx}\frac{\partial H}{\partial g_x}-\frac{d}{dy}\frac{\partial H}{\partial g_y} = 0$$

where $H = (I_x f+I_y g-I_t)^2+\lambda(f_x^2+f_y^2+g_x^2+g_y^2)$

$$\frac{\partial H}{\partial f} = 2(I_x f+I_y g-I_t)I_x \quad \text{and} \quad \frac{\partial H}{\partial f_x} = 2\lambda f_x \quad \frac{\partial H}{\partial f_y} = 2\lambda f_y$$

$$\frac{\partial H}{\partial g} = 2(I_x f+I_y g-I_t)I_y \quad \text{and} \quad \frac{\partial H}{\partial g_x} = 2\lambda g_x \quad \frac{\partial H}{\partial g_y} = 2\lambda g_y$$

$$\Rightarrow \quad 2(I_x f+I_y g-I_t)I_x-2\lambda f_{xx}-2\lambda f_{yy} = 0 \quad \Rightarrow \quad (I_x f+I_y g-I_t)I_x = \lambda\nabla^2 f$$

$$\Rightarrow \quad 2(I_x f+I_y g-I_t)I_y-2\lambda g_{xx}-2\lambda g_{yy} = 0 \quad \Rightarrow \quad (I_x f+I_y g-I_t)I_y = \lambda\nabla^2 g$$

which can be solved by numerical method.

Thin plate spline

Thin plate spline is a regularized regression on a set of given points (x_n, y_n) $\forall n\in[1,N]$ by an unknown function f(x), which minimizes sum of error square, while keeping f(x) as smooth as possible to avoid overfitting. Hence, we have the following objective functional, which contains two parts : (1) sum of error square and (2) smoothness of f(x), the non smoothness (or irregularity) of f(x) is measured by the integral of Laplacian square, while the tradeoff is controlled by parameter λ.

$$F[f,g] = \frac{1}{2}\sum_{n=1}^N (f(x_n)-y_n)^2+\frac{\lambda}{2}\int f_{xx}^2 dx$$

$$= \frac{1}{2}\int\sum_{n=1}^N (f(x)-y_n)^2\delta(x-x_n)dx+\frac{\lambda}{2}\int f_{xx}^2 dx \qquad \text{(technique : convert data sum to function integral)}$$

We then apply Euler Lagrange equation :

$$0 = \frac{\partial H}{\partial f}-\frac{d}{dx}\frac{\partial H}{\partial f_x}+\frac{d^2}{dx^2}\frac{\partial H}{\partial f_{xx}}$$

where $H = \frac{1}{2}\sum_{n=1}^N (f(x)-y_n)^2\delta(x-x_n)+\frac{\lambda}{2}f_{xx}^2$

$$\frac{\partial H}{\partial f} = \sum_{n=1}^N (f(x)-y_n)\delta(x-x_n)$$

$$\frac{\partial H}{\partial f_x} = 0$$

$$\frac{\partial H}{\partial f_{xx}} = \lambda f_{xx}$$

$$\Rightarrow \quad 0 = \sum_{n=1}^N (f(x)-y_n)\delta(x-x_n)-\frac{d}{dx}0+\frac{d^2}{dx^2}\lambda f_{xx}$$

$$\Rightarrow \quad f_{xxxx} = \frac{1}{\lambda}\sum_{n=1}^N (y_n-f(x))\delta(x-x_n) \qquad \text{(equation 1)}$$

To solve this ODE, we need the Green's function for linear differential operator d^4/dx^4 .

$$G^{(4)}(x,x_0) = \delta(x-x_0)$$

$$G(x,x_0) = |x-x_0|^3+o(x^2) \qquad \text{(equation 2)}$$

Intuitively, the RHS of equation 2 is in order 3 of x, hence its 4th derivative is zero for all x, except when x = x_0, where a discontinuity occurs. The solution to ODE is thus the convolution of Green's function with the RHS of equation 1.

$$f(x) = \int\frac{1}{\lambda}\sum_{n=1}^N (y_n-f(s))\delta(s-x_n)G(x,s)ds$$

$$= \int\frac{1}{\lambda}\sum_{n=1}^N (y_n-f(s))\delta(s-x_n)|x-s|^3 ds$$

$$= \frac{1}{\lambda}\sum_{n=1}^N (y_n-f(x_n))|x-x_n|^3$$

$$= \sum_{n=1}^N w_n G(x,x_n) \qquad \text{where } w_n=(y_n-f(x_n))/\lambda \text{ and } G(x,x_n)=|x-x_n|^3$$

Now, we want to solve for f(x), and the above intermediate solution tells us that it is a linear combination of Green's function locating at sampled x, while the weights depends on f(x) recursively. Therefore, we proceed by replacing all f(x) with the weights, and by moving all the weights to one side of the equation, we can finally solve the problem by solving for the values of the weights.

$$
\begin{aligned}
w_n &= (y_n - f(x_n))/\lambda && \forall n \in [1, N] \\
y_n &= \lambda w_n + f(x_n) && \forall n \in [1, N] \\
&= \lambda w_n + \textstyle\sum_{m=1}^{N} w_m G(x_n, x_m) && \forall n \in [1, N] \\
\Rightarrow \quad y &= (G + \lambda I) w \\
\Rightarrow \quad w &= (G + \lambda I)^{-1} y
\end{aligned}
$$

where

$$
\begin{aligned}
y &= [y_1, y_2, y_3, \ldots, y_N]^T \\
w &= [w_1, w_2, w_3, \ldots, w_N]^T \\
G &= \begin{bmatrix} G(x_1, x_1) & G(x_1, x_2) & G(x_1, x_3) & \ldots & G(x_1, x_N) \\ G(x_2, x_1) & G(x_2, x_2) & G(x_2, x_3) & \ldots & G(x_2, x_N) \\ G(x_3, x_1) & G(x_3, x_2) & G(x_3, x_3) & \ldots & G(x_3, x_N) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ G(x_N, x_1) & G(x_N, x_2) & G(x_N, x_3) & \ldots & G(x_N, x_N) \end{bmatrix}
\end{aligned}
$$

Reference

[1] Calculus of variation 2, Numerical geometry of images, Guy Rosman.
[2] Regularization theory and neural network architecture, Girosi, Jones and Poggio.