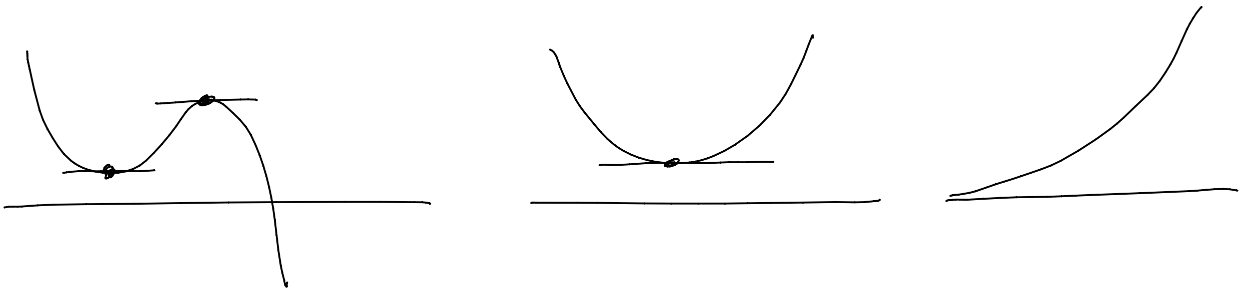
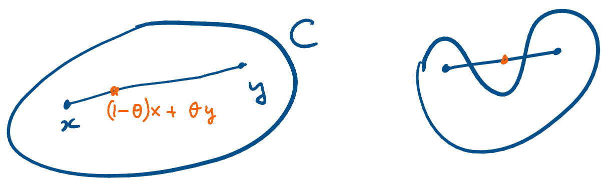
Convexity



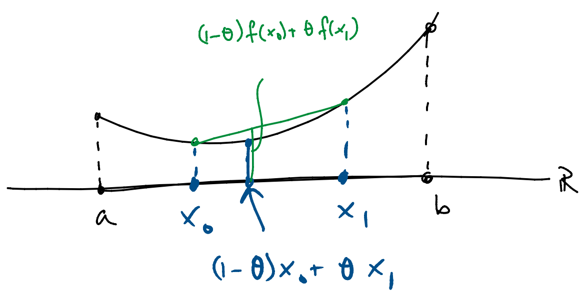
Definitions Let V be a real vector space

A subset $C \subset V$ is convex if for all $x, y \in C$ and all $\theta \in [0,1]$ one has $(1-\theta)x + \theta y \in C$



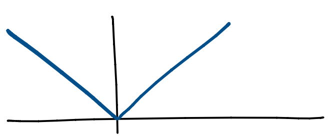
If $C \subset V$ is a convex subset then a function $f: C \to \mathbb{R}$ is called (strictly) convex if for all $x, y \in C$ and $\theta \in (0,1)$

$$f((1-\theta)x + \theta y) \underset{<}{\leq} (1-\theta) f(x) + \theta f(y)$$



$C = [a,b]$

Example. $f: \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$ is convex
not strictly convex.

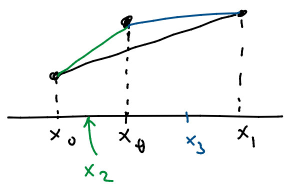


Theorem (from Calculus).

① If $f: [a,b] \to \mathbb{R}$ is C^2 then

f is convex $\Leftrightarrow f''(x) \geq 0$ for all $x \in [a,b]$

Proof idea (421) for $f''(x) \overset{>}{\geq} 0 \Rightarrow f$ convex.



$x_\theta = (1-\theta)x_0 + \theta x_1$

Suppose $f(x_\theta) \overset{\geq}{>} (1-\theta)f(x_0) + \theta f(x_1)$

Picture + MVThm $\Rightarrow \exists x_2 \in (x_0, x_\theta)$

with $f'(x_2) \geq \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}$

and $\exists x_3 \in (x_\theta, x_1)$ with $f'(x_3) \leq \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}$

MVThm again $\Rightarrow$ $\exists x_4 \in (x_2, x_3)$ with

$$f''(x_4) = \frac{f'(x_3) - f'(x_2)}{x_3 - x_2} \leq 0 \qquad \lightning$$

So: if $f''(x) > 0$ for all x then f is strictly convex.

② If $C \subset \mathbb{R}^n$ is convex & open and if $f: C \to \mathbb{R}$ is C^2 then

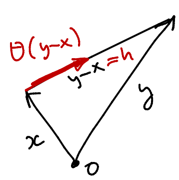
f is convex $\Leftrightarrow$ $\underbrace{D^2 f(x) \geq 0}_{}$ for all $x \in C$
$D^2 f$ is positive semidefinite.

Proof of "$\Leftarrow$" Given $x, y \in C$ consider the function

$$g: [0,1] \to \mathbb{R} \quad \text{with} \quad g(\theta) = f((1-\theta)x + \theta y)$$

We show that $g''(\theta) \geq 0$ for $0 < \theta < 1$

Use $(1-\theta)x + \theta y = x + \theta(y - x) = x + \theta h$ where $h = y - x$



Then $g(\theta) = f(x + \theta h) =$

$= f(x_1 + \theta h_1, x_2 + \theta h_2, \ldots, x_n + \theta h_n)$

Differentiate:

$$g'(\theta) = h_1 \cdot \frac{\partial f}{\partial x_1}(x_1 + \theta h_1, \ldots, x_n + \theta h_n) + \cdots + h_n \frac{\partial f}{\partial x_n}(-)$$

$$g''(\theta) = h_1 h_1 \frac{\partial^2 f}{\partial x_1 \partial x_1}(-) + h_1 h_2 \frac{\partial^2 f}{\partial x_1 \partial x_2}(-) + \cdots + h_1 h_n \frac{\partial^2 f}{\partial x_1 \partial x_n}(-)$$

$$+ h_2 h_1 \frac{\partial^2 f}{\partial x_2 \partial x_1}(--) + h_2 h_2 \frac{\partial^2 f}{\partial x_2 \partial x_2}(-) + \cdots + h_2 h_n \frac{\partial^2 f}{\partial x_2 \partial x_n}(-)$$

$$\vdots$$

$$+ h_n h_1 \frac{\partial^2 f}{\partial x_n \partial x_1}(\ldots) + h_n h_2 \frac{\partial^2 f}{\partial x_n \partial x_2}(\ldots) + \cdots + h_n h_n \frac{\partial^2 f}{\partial x_n \partial x_n}(\ldots)$$

$$= \langle h, D^2 f(\ldots)\, h \rangle$$

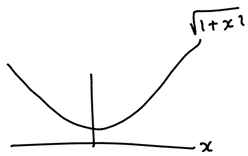
where $h = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$ and $D^2 f = \begin{pmatrix} f_{x_1 x_1} & \cdots & f_{x_1 x_n} \\ \vdots & & \vdots \\ f_{x_n x_1} & \cdots & f_{x_n x_n} \end{pmatrix}$ is

the Hessian matrix of f

$D^2 f$ is a symmetric $n \times n$ matrix.

Definition from linear algebra. A symmetric $n\times n$ matrix A is positive semi definite if $\langle h, Ah\rangle \geq 0$ for all h

A is positive semi definite $\iff$ all eigenvalues λ of A satisfy $\lambda \geq 0$.



Examples

$f(x) = |x|$

$f(x) = \sqrt{1+x^2}$ $\quad f'(x) = \dfrac{x}{\sqrt{1+x^2}}$ $\quad f''(x) = \dfrac{1}{\sqrt{1+x^2}} + x\cdot\left(-\dfrac{1}{\cancel{2}}\right)\dfrac{\cancel{2}x}{(1+x^2)^{3/2}}$

$$\Rightarrow f''(x) = \frac{1+x^2}{(1+x^2)^{3/2}} - \frac{x^2}{(1+x^2)^{3/2}} = \frac{1}{(1+x^2)^{3/2}} > 0$$

So $f(x) = \sqrt{1+x^2}$ is strictly convex.

$f(x_1, \dots, x_n) = \sqrt{1+x_1^2+\dots+x_n^2}$ is strictly convex

$$\frac{d^2}{d\theta^2} f(x_1+\theta h_1, \dots, x_n+\theta h_n) = \frac{d}{d\theta}\left(\frac{d}{d\theta}\sqrt{1+(x_1+\theta h_1)^2+\dots+(x_n+\theta h_n)^2}\right)$$

$$= \frac{d}{d\theta}\left(\frac{(x_1+\theta h_1)h_1+\dots+(x_n+\theta h_n)h_n}{\sqrt{1+\dots}}\right)$$

$\dfrac{d}{d\theta}\dfrac{1}{z} = \dfrac{-1}{z^2}\dfrac{dz}{d\theta}$ $\quad z = \sqrt{\dots}$

$$= \frac{h_1^2+h_2^2+\dots+h_n^2}{\sqrt{}} + \left[(x_1+\theta h_1)h_1+\dots+(x_n+\theta h_n)h_n\right]\cdot\frac{d}{d\theta}\frac{1}{\sqrt{}}$$

$$= \frac{h_1^2+\dots+h_n^2}{\sqrt{}} - \frac{(x_1+\theta h_1)h_1+\dots+(x_n+\theta h_n)h_n}{1+(x_1+\theta h_1)^2+\dots+(x_n+\theta h_n)^2}\cdot\frac{(x_1+\theta h_1)h_1+\dots+(x_n+\theta h_n)h_n}{\sqrt{1+\dots}}$$

$$= \frac{\|h\|^2}{\sqrt{}} - \frac{\langle x+\theta h, h\rangle^2}{(1+\|x+\theta h\|^2)\sqrt{}}$$

$$= \frac{(1+\|x+\theta h\|^2)\|h\|^2 - \langle x+\theta h, h\rangle^2}{(1+\|x+\theta h\|^2)^{3/2}}$$

$$\geq \frac{\|h\|^2 + \|x+\theta h\|^2\cdot\|h\|^2 - \|x+\theta h\|^2\cdot\|h\|^2}{(1+\|\dots\|^2)^{3/2}}$$

$$= \frac{\|h\|^2}{(1+\|\dots\|^2)^{3/2}} > 0 \quad \text{if } h \neq 0.$$

Cauchy-Schwarz

$\langle a,b\rangle \leq \|a\|\cdot\|b\| \leq \frac{1}{2}(\|a\|^2+\|b\|^2)$

$\langle x+\theta h, h\rangle \leq \|x+\theta h\|\cdot\|h\|$

Therefore $f(x_1,\dots,x_n) = \sqrt{1+x_1^2+\dots+x_n^2}$ is strictly convex.

Monday.

Let $L: \Omega \times \mathbb{R} \times \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^1 function and define

$$I[u] = \int \cdots \int_{\Omega} L\left(x, u(x), \frac{\partial u}{\partial x_1}(x), \ldots, \frac{\partial u}{\partial x_n}(x)\right) dx_1 \ldots dx_n .$$

Let $g: \partial\Omega \to \mathbb{R}$ be a continuous function.

Consider the problem of finding a minimizer of $I[u]$ among all C^1 functions $u: \Omega \longrightarrow \mathbb{R}$ with $u|_{\partial\Omega} = g$.

Suppose that L satisfies the following condition:

for each $x \in \Omega$ the function

$$(u, v) \in \mathbb{R} \times \mathbb{R}^n \longmapsto L(x, u, v_1, \ldots, v_n)$$

is strictly convex

Examples ① $n=1$ $L(x,u,v) = \frac{1}{2}v^2$ $\qquad I[u] = \int_a^b \frac{1}{2} u'(x)^2\, dx$

$$\frac{\partial^2 L}{\partial v^2}(x,u,v) = 1 > 0 \Rightarrow \quad v \to L(x,u,v) \text{ is strictly convex.}$$

② $L(x,u,v_1,\ldots,v_n) = \sqrt{1+v_1^2+\cdots+v_n^2}$ strictly convex.

③ $L(x,t,u,v_x,v_t) = \frac{1}{2}(v_t^2 - v_x^2)$

E-L equations: $u_{tt} - u_{xx} = 0$ Wave Equation!

$$\begin{bmatrix} L_{v_x v_x} & L_{v_x v_t} \\ L_{v_t v_x} & L_{v_t v_t} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & +1 \end{bmatrix}$$ not positive semidefinite has a <u>negative</u> eigenvalue

Assume $L(x,u,v_1,\ldots,v_n)$ does not depend on u.

<u>Theorem 1</u> There is at most one minimizer of $I[u]$ with $u = g$ on $\partial\Omega$.

<u>Theorem 2</u> If L is C^2, if $u: \Omega \to \mathbb{R}$ is C^2 and satisfies the Euler–Lagrange equation, and if $u = g$ on $\partial\Omega$ then u minimizes $I[u]$ among all C^2 functions $u: \Omega \longrightarrow \mathbb{R}$ with $u = g$ on $\partial\Omega$

<u>Proof of Theorem 1</u> Let $u, \bar{u}: \Omega \to \mathbb{R}$ both be minimizers of $I[u]$ with $u = \bar{u} = g$ on $\partial\Omega$.

This means that for every $\tilde{u}: \Omega \to \mathbb{R}$ with $\tilde{u} = g$ on $\partial\Omega$ we have

$$I[\tilde{u}] \geq I[u] \quad \text{and} \quad I[\tilde{u}] \geq I[\bar{u}]$$

In particular $I[u] \geq I[\bar{u}]$ and $I[\bar{u}] \geq I[u]$

i.e. $I[u] = I[\bar{u}]$.

For any $\theta \in (0,1)$ consider $w = (1-\theta)u + \theta \bar{u}$.

i.e. $w(x) = (1-\theta)u(x) + \theta \bar{u}(x)$ for $x \in R$.

$w_{x_1} = (1-\theta)u_{x_1} + \cdots + \theta \bar{u}_{x_1}$

Then

$$I[u] = I[\bar{u}] \leq I[w] = \int \cdots \int_R L\left(x, \frac{\partial w}{\partial x_1}, \ldots, \frac{\partial w}{\partial x_n}\right) dx_1 \ldots dx_n$$

$$= \int \cdots \int_R L\left(x, (1-\theta)u_{x_1}(x) + \theta \bar{u}_{x_1}(x), \ldots, (1-\theta)u_{x_n}(x) + \theta \bar{u}_{x_n}(x)\right) dx_1 \ldots dx_n$$

$\downarrow$ $L(x, v_1, \ldots, v_n)$ convex function of $(v_1, \ldots, v_n)$

$$\leq \int \cdots \int_R \left\{(1-\theta) L(x, u_{x_1}, \ldots, u_{x_n}) + \theta\, L(x, \bar{u}_{x_1}, \ldots, \bar{u}_{x_n})\right\} dx_1 \ldots dx_n$$

$$= (1-\theta) \int \cdots \int_R L(x, u_{x_1}, \ldots, u_{x_n})\, dx_1 \ldots dx_n + \theta \int \cdots \int_R L(x, \bar{u}_{x_1}, \ldots, \bar{u}_{x_n})\, dx_1 \ldots dx_n$$

$$= (1-\theta)\, I[u] + \theta\, I[\bar{u}]$$

$$= I[u] \quad \text{because } I[u] = I[\bar{u}]$$

Therefore all inequalities above are equalities.

For all $x \in R$ we have

$$\text{LHS} = L\left(x, (1-\theta)u_{x_1}(x) + \theta \bar{u}_{x_1}(x), \ldots, (1-\theta)u_{x_n}(x) + \theta u_{x_n}(x)\right)$$

$$\leq (1-\theta)\, L(x, u_{x_1}(x), \ldots, u_{x_n}(x)) + \theta\, L(x, \bar{u}_{x_1}(x), \ldots, \bar{u}_{x_n}(x)) = \text{RHS}$$

with equality only if $u_{x_1}(x) = \bar{u}_{x_1}(x), \ldots, u_{x_n}(x) = \bar{u}_{x_n}(x)$.

We know LHS and RHS are continuous functions of x and

$$\int \cdots \int_R (\text{RHS} - \text{LHS})\, dx \overset{*}{=} 0 \quad \text{with } \text{RHS} - \text{LHS} \geq 0.$$

Therefore LHS = RHS.

Strict convexity of L implies $\frac{\partial u}{\partial x_1} = \frac{\partial \bar{u}}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_n} = \frac{\partial \bar{u}}{\partial x_n}$

Hence $\frac{\partial}{\partial x_k}(u - \bar{u}) = 0$ on R.

Since R is connected it follows that $u(x) - \bar{u}(x)$ is constant.

Finally, $u = \bar{u} = g$ on ∂R so $u(x) = \bar{u}(x)$ on R. ////

PROOF of Theorem 2.

Given $u : R \to \mathbb{R}$ is C^2 and u is a solution to the Euler Lagrange equation.

Let $\bar{u} : R \to R$ be C^2 with $\bar{u} = u = g$ on ∂R.

Consider again $w_\theta = (1-\theta)u + \theta \bar{u}$ for $0 \leq \theta \leq 1$.

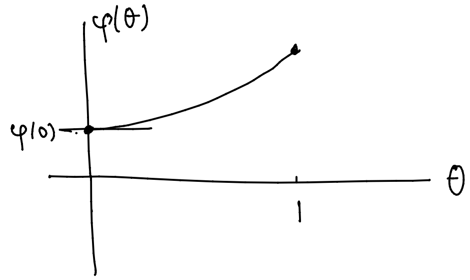
Then:

$$\varphi(\theta) = I[w_\theta] = I[u + \theta \cdot (\bar{u} - u)] = I[u + \theta h]$$

is a convex function of $\theta \in [0,1]$ (proved this in Theorem 1)

u satisfies the Euler Lagrange equation $\Leftrightarrow$

$$\varphi'(0) = \left(\frac{d}{d\theta} I[u + \theta h] \right)_{\theta = 0} = 0 \quad \text{for all } h \text{ with } h = 0 \text{ on } \partial R$$



$\varphi : [0,1] \to \mathbb{R}$ is convex and $\varphi'(0) = 0$

$\Longrightarrow \quad \varphi(1) \geq \varphi(0)$

$\Rightarrow \quad I[\bar{u}] \geq I[u]$.