Lecture 4

Necessary and Sufficient Conditions for Unconstrained Minimization

ME260 Indian Institute of Science

Structural Optimization: Size, Shape, and Topology

G. K. Ananthasuresh

Professor, Mechanical Engineering, Indian Institute of Science, Bengaluru

suresh@iisc.ac.in

Outline of the lecture

Necessary conditions for unconstrained optimization problem

- In one variable
- In two variables
- In multiple variables

What we will learn:

The concept of a local minimum

The premise for writing the necessary condition

The concept of gradient of a function of n variables

And, of course, the necessary conditions of unconstrained optimization problem

Global and local minima: definitions

Simple case: $f(x)$, a function of a single variable, x.

Global minimum

x^* is global minimizer of $f(x)$ if $f(x^*) \le f(x)\ \forall\, x$ in the feasible interval of x.

Local minimum

x^* is a local minimizer of $f(x)$ if $f(x^*) \le f(x)$ in a small neighborhood of x^* in the feasible interval of x.

$$\text{N = small neighborhood} = \left\{ x \,\middle|\, x \in S \text{ with} \left\| x - x^* \right\| < \delta \right\},\ \delta > 0$$

These are just definitions; not conditions.

Definitions of this sort do not let you check if a given value of x is a minimum or not unless you exhaustively check the entire domain of x.

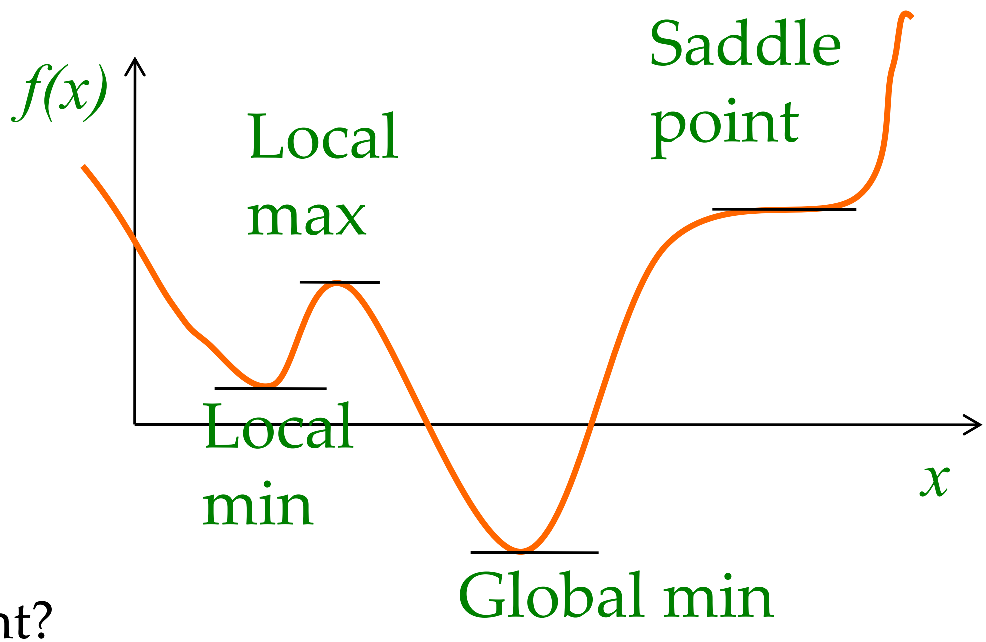
A condition would let you check this easily.

Conditions for a local minimum of $f(x)$

If x^* is a local minimum of $f(x)$, then…

$$\left.\frac{df}{dx}\right|_{x^*}=0$$ Necessary condition

$$\left.\frac{df}{dx}\right|_{x^*}=0 \;\&\; \left.\frac{d^2f}{dx^2}\right|_{x^*}>0$$ Sufficient condition



Why is the necessary condition not sufficient?

Is the sufficient condition also necessary?

Think about the literal meaning of "necessary" and "sufficient".

Why is necessary condition "necessary"?

Consider Taylor series expansion of *f(x)* around x^*.

$$f(x) = f(x^*) + \left.\frac{df}{dx}\right|_{x^*}\left(x - x^*\right) + \frac{1}{2}\left.\frac{d^2 f}{dx^2}\right|_{x^*}\left(x - x^*\right)^2 + O(3)$$

Perturbation around x^*
$\delta x^* = \left(x - x^*\right)$

Zeroth order term | First order term | Second order term | Higher order terms

$$f(x) = f(x^*) + f'(x^*)\,\Delta x^* + \frac{1}{2} f''(x^*)\left(\Delta x^*\right)^2 + O(3)$$

x^* is a local minimizer of $f(x)$ if $f(x^*) \leq f(x)$ in a small neighborhood of x^* in the feasible interval of x .

Here the small neighborhood is Δx^*. When it is small, it is the first order term that matters more than the second order term.
Unless $f'(x^*)$ is zero, we cannot be sure that the definition is satisfied. More in the next slide.

Why is necessary condition *necessary*? (contd.)

For small Δx^* (as small as you can imagine…)

$$f(x) = f(x^*) + f'(x^*)\Delta x^* + \frac{1}{2} f''(x^*)\left(\Delta x^*\right)^2 + O(3)$$

$f'(x^*)\Delta x^*$: May be positive or negative depending on the sign of *f(x*)* as Δx^* can be positive or negative.

$\frac{1}{2} f''(x^*)\left(\Delta x^*\right)^2 + O(3)$: Negligible

So, for $f(x^*) \leq f(x)$ it is necessary to have $f'(x^*) = 0$

Why is necessary condition not sufficient?

Note that we only talk about the function “not being less in the small neighborhood of the minimizing point” to say that it has a local minimum.

So, the condition is that the first order term is zero for any small perturbation. This necessitates the first order derivative to be zero.

This condition is necessary, as noted in the previous slide.

But…

The necessary condition is true for a local minimum and a local maximum. So, it is not sufficient to conclude that a given value of x is a local minimum.

Is sufficient condition also necessary?

Consider…

$$f(x) = x^4$$

$$f'(x) = 4x^3 \quad \Rightarrow \quad x^* = 0$$

$$f''(x) = 12x^2 = 0$$

$$f'''(x) = 24x = 0 \quad \Rightarrow x^* = 0$$

$$f^{iv}(x) = 24 > 0$$

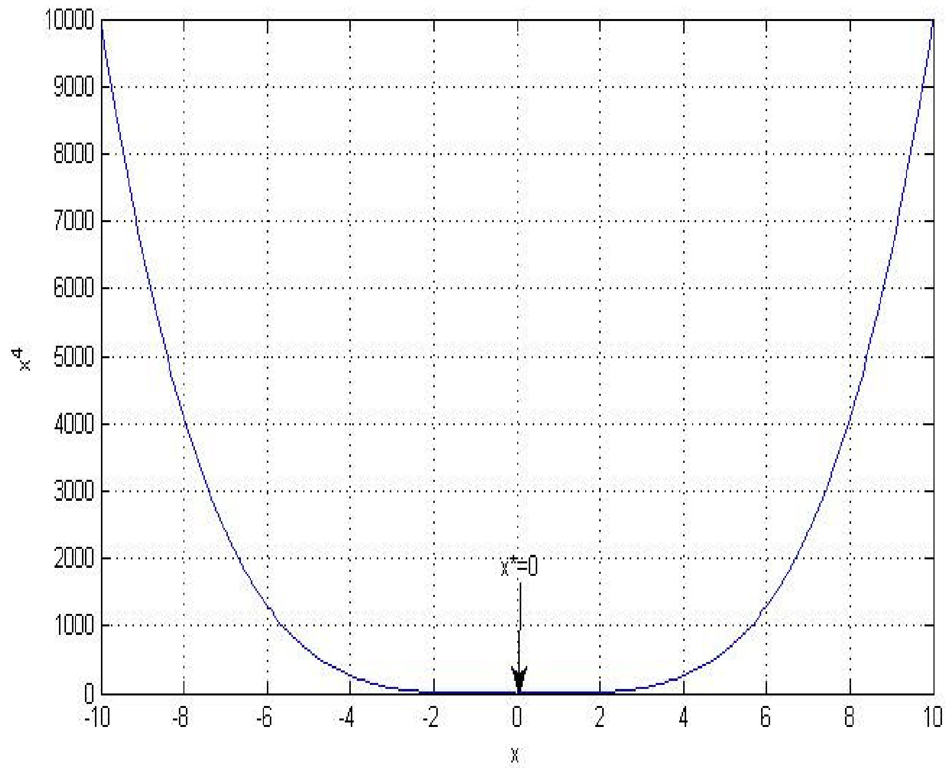
$$x^* = 0$$

$$\left.\frac{df}{dx}\right|_{x^*} = 0$$

Necessary condition is satisfied.

$$\left.\frac{d^2 f}{dx^2}\right|_{x^*} = 0$$

Sufficient condition is **not** satisfied.



But $x^* = 0$, is a minimizer here! So, sufficient condition is not necessary.

Understand “necessary” and “sufficient” well.

The logic of necessary and sufficient conditions should be clearly understood.

They actually mean what they say but it can be confusing and misleading sometimes.

What is necessary may not be sufficient.

What is sufficient may not be necessary.

Sometimes, a condition can be necessary and sufficient.

Note all of this we are saying only in the context of a local minimum.

For a global minimum, there is no “operationally useful” definition or condition.

Let us move to a function of more than one variable next…

A function of two variables…

Taylor series expansion of *f(x,y)* around *(x*,y*)*:

$$f(x,y) = f(x^*,y^*) + f_x(x^*,y^*)\Delta x^* + f_y(x^*,y^*)\Delta y^* +$$

$$+\frac{1}{2}\left\{ f_{xx}(x^*,y^*)(\Delta x^*)^2 + 2 f_{xy}(x^*,y^*)\Delta x^* \Delta y^* + f_{yy}(x^*,y^*)(\Delta y^*)^2 \right\} + O(3)$$

where $f_x = \frac{\partial f}{\partial x};\ f_y = \frac{\partial f}{\partial y};\ f_{xx} = \frac{\partial^2 f}{\partial x^2};\ f_{xy} = \frac{\partial^2 f}{\partial x \partial y};\ f_{yy} = \frac{\partial^2 f}{\partial y^2}$

In matrix form

$$f(x,y) = f(x^*,y^*) + \left\{ f_x(x^*,y^*) \quad f_y(x^*,y^*) \right\} \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} +$$

$$+\frac{1}{2}\left\{ \Delta x^* \quad \Delta y^* \right\} \begin{bmatrix} f_{xx}(x^*,y^*) & f_{xy}(x^*,y^*) \\ f_{xy}(x^*,y^*) & f_{yy}(x^*,y^*) \end{bmatrix} \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} + O(3)$$

Two-variable function (contd.)

$$f(x,y) = f(x^*,y^*) + \left\{ f_x(x^*,y^*) \quad f_y(x^*,y^*) \right\} \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} +$$

$$+ \frac{1}{2} \left\{ \Delta x^* \quad \Delta y^* \right\} \begin{bmatrix} f_{xx}(x^*,y^*) & f_{xy}(x^*,y^*) \\ f_{xy}(x^*,y^*) & f_{yy}(x^*,y^*) \end{bmatrix} \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} + O(3)$$

$$= f(x^*,y^*) + \nabla f(x^*,y^*)^T \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} + \frac{1}{2} \left\{ \Delta x^* \quad \Delta y^* \right\} \mathbf{H}(x^*,y^*) \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} + O(3)$$

where $\nabla f(x^*,y^*) = \begin{Bmatrix} f_x(x^*,y^*) \\ f_y(x^*,y^*) \end{Bmatrix}$ and $\mathbf{H}(x^*,y^*) = \begin{bmatrix} f_{xx}(x^*,y^*) & f_{xy}(x^*,y^*) \\ f_{xy}(x^*,y^*) & f_{yy}(x^*,y^*) \end{bmatrix}$

Gradient

Hessian

Necessary conditions for the minimum of *f(x,y)*

Second order term is negligible and the first order term should be zero for small $\left(\delta x^*, \delta y^*\right)$

So,

$$\nabla f(x^*, y^*) = \begin{Bmatrix} f_x(x^*, y^*) \\ f_y(x^*, y^*) \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

This has two scalar equations in it.

Two variables (x^*,y^*) to be found using two scalar equations!

$$f_x(x^*, y^*) = 0$$
$$f_y(x^*, y^*) = 0$$

Sufficient condition for the minimum of *f(x,y)*

$$\frac{1}{2}\left\{\Delta x^* \quad \Delta y^*\right\}\begin{bmatrix} f_{xx}(x^*,y^*) & f_{xy}(x^*,y^*) \\ f_{xy}(x^*,y^*) & f_{yy}(x^*,y^*) \end{bmatrix}\begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} > 0 \text{ for any } \left(\Delta x^*, \Delta y^*\right)$$

$$\frac{1}{2}\left\{\Delta x^* \quad \Delta y^*\right\}\mathbf{H}(x^*,y^*)\begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} > 0 \text{ for any } \left(\Delta x^*, \Delta y^*\right)$$

A matrix that has this property is said to be **positive definite**.

Positive definite, and other definite things…

For any $(\Delta x^*, \Delta y^*)$

$$\frac{1}{2}\begin{Bmatrix} \Delta x^* & \Delta y^* \end{Bmatrix} \mathbf{H}(x^*, y^*) \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} > 0$$

Positive definite **H**; minimum.

$$\frac{1}{2}\begin{Bmatrix} \Delta x^* & \Delta y^* \end{Bmatrix} \mathbf{H}(x^*, y^*) \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} \geq 0$$

Positive semi-definite **H**; minimum or flat.

$$\frac{1}{2}\begin{Bmatrix} \Delta x^* & \Delta y^* \end{Bmatrix} \mathbf{H}(x^*, y^*) \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} < 0$$

Negative definite **H**; maximum.

$$\frac{1}{2}\begin{Bmatrix} \Delta x^* & \Delta y^* \end{Bmatrix} \mathbf{H}(x^*, y^*) \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} \leq 0$$

Negative semi-definite **H**; maximum or flat.

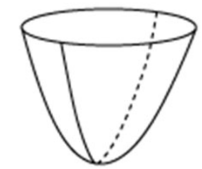
$$\frac{1}{2}\begin{Bmatrix} \Delta x^* & \Delta y^* \end{Bmatrix} \mathbf{H}(x^*, y^*) \begin{Bmatrix} \Delta x^* \\ \Delta y^* \end{Bmatrix} = 0$$

Null-definite **H**; just flat; neither minimum nor maximum.

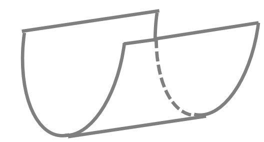
Peaks, valleys, folds, ridges, and flat planes…

The surface represented by $f(x,y)$ locally looks like this at (x^*,y^*).

Positive definite **H**; minimum; Bottom of a valley

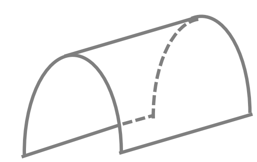


Positive semi-definite **H**; minimum or flat; A valley fold

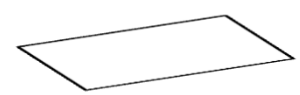


Negative definite **H**; maximum; Peak of a hill

Negative semi-definite **H**; maximum or flat; A ridge



Null-definite **H**; neither minimum nor maximum; Just flat



Indefiniteness of a matrix

$$\frac{1}{2}\left\{\Delta x^* \quad \Delta y^*\right\}\mathbf{H}(x^*,y^*)\begin{Bmatrix}\Delta x^* \\ \Delta y^*\end{Bmatrix} \overset{>}{\underset{<}{?}}\ 0 \text{ for any } \left(\Delta x^*,\Delta y^*\right)$$

It is positive sometimes and negative sometimes… it is **indefinite**. Then, (x^*,y^*) is a **saddle point**.

http://explore.org/photos/2238/horse-saddle

Minimum one way and maximum the other way!

http://www.pringles.com/products

Function of n variables

Taylor's series expansion…

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \sum_{i=1}^{n} \frac{\partial f(\mathbf{x}^*)}{\partial x_i}(x_i - x_i^*) +$$

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f(\mathbf{x}^*)}{\partial x_i \partial x_j}(x_i - x_i^*)(x_j - x_j^*) + O(3)$$

$$= f(\mathbf{x}^*) + \nabla f^T(\mathbf{x}^*)\Delta\mathbf{x}^* + \frac{1}{2}\Delta\mathbf{x}^{*T}\mathbf{H}(\mathbf{x}^*)\Delta\mathbf{x}^* + O(3)$$

Gradient of an *n*-variable function

$$\nabla f(\mathbf{x}^*) = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{Bmatrix}$$

Necessary condition:
n variables;
n equations.

Hessian of an *n*-variable function

$$\mathbf{H}(\mathbf{x}) = \begin{Bmatrix} \dfrac{d^2 f}{dx_1^{\ 2}} & \dfrac{\partial^2 f}{\partial x_1\, \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1\, \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2\, \partial x_1} & \dfrac{d^2 f}{dx_2^{\ 2}} & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n\, \partial x_1} & \cdots & \cdots & \dfrac{\partial^2 f}{\partial x_n\, \partial x_n} \end{Bmatrix}$$

Sufficient condition: **H** should be positive definite for a minimum of *f(**x**)*

Rules of checking positive definiteness

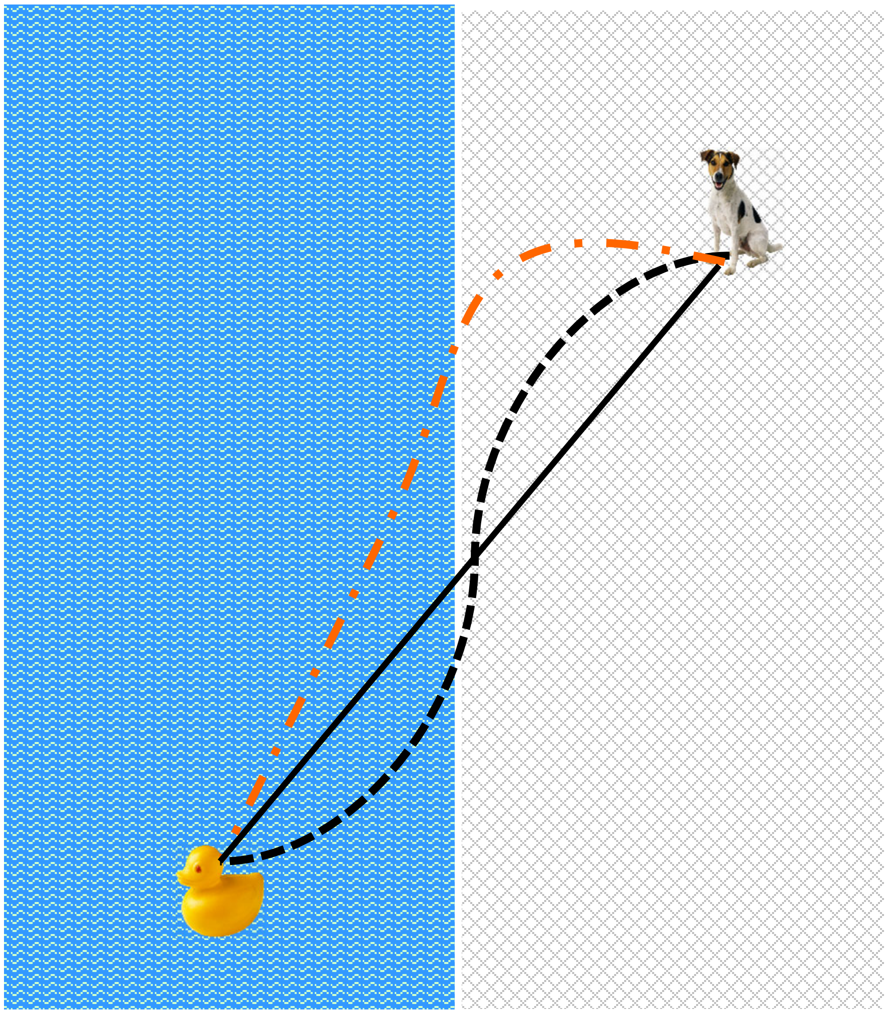
H is positive definite

- All Eigen-values are positive
- All principal minors are positive
- Pivots are positive in the row-echelon form

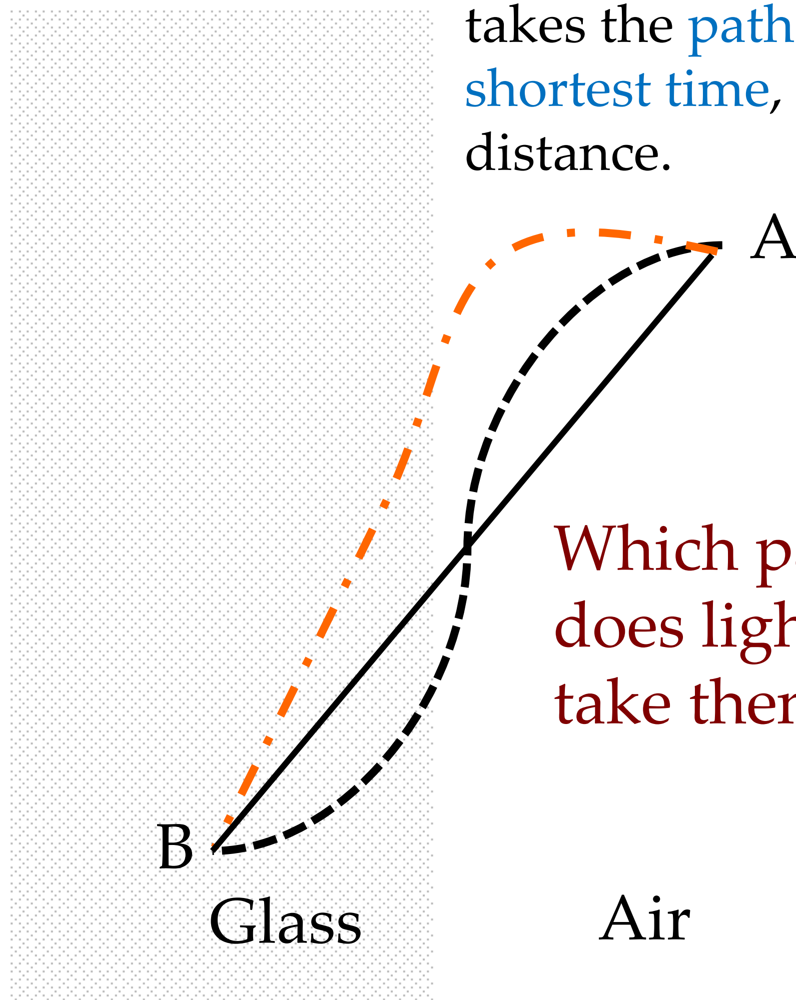
For other "definitenesses"…

Quadratic form $\delta \mathbf{x}^{*T} \mathbf{H}(\mathbf{x}^*) \delta \mathbf{x}^*$	$\mathbf{H}$	Eigen values of $\mathbf{H}$	Nature of $\mathbf{x}^*$
Positive	Positive definite	All are positive	Local min
Negative	Negative definite	All are negative	Local max
Non-negative	Positive semi-definite	Some zero, others positive	A valley fold
Non-positive	Negative semi-definite	Some zero, others negative	A ridge
Any sign	Indefinite	Mixed signs	Saddle point

Fermat's problem

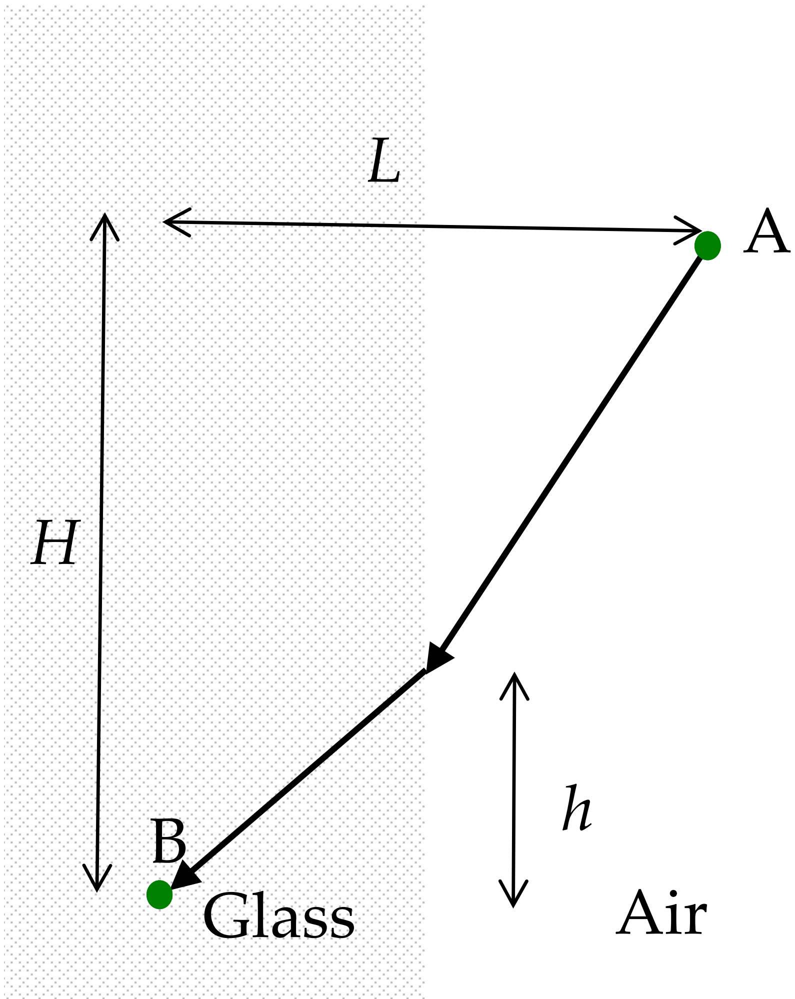


Light travels faster in air than in water. So, to go from A to B, light takes the path of shortest time, not distance.



Which path does light ray take then?

Fermat's conjecture



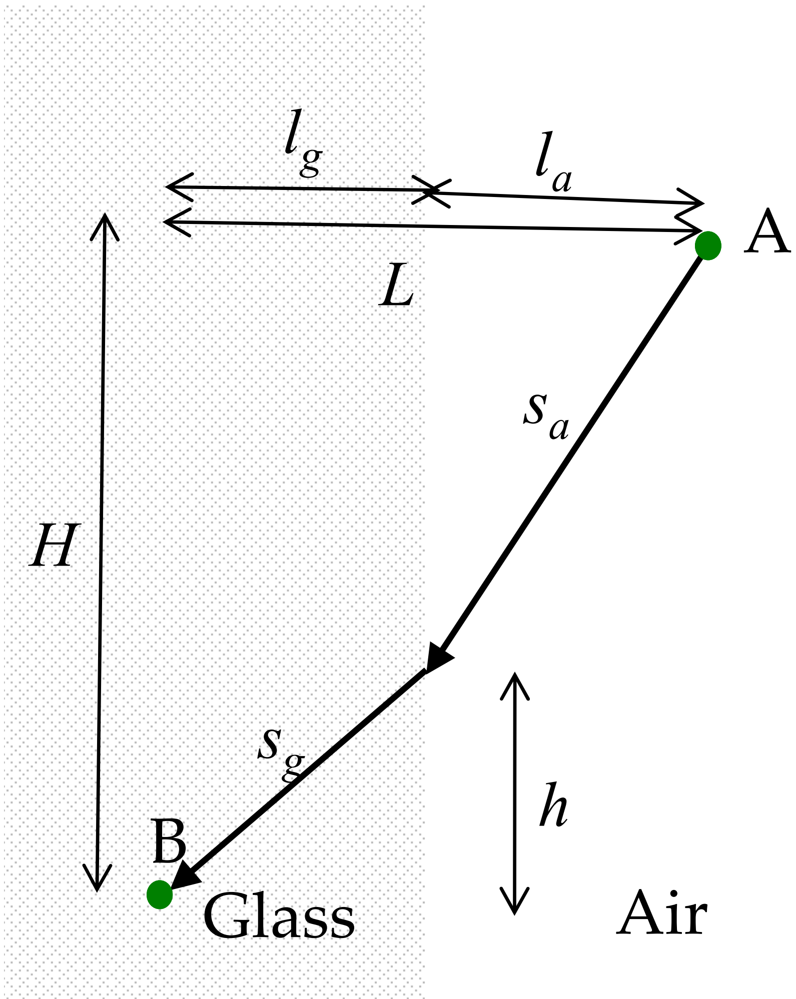
17^{th} century amateur mathematician, Fermat (1601-1665), had conjectured that light rays take the shortest-time paths and **not** shortest-distance paths.

This is clear from the fact that light rays bend when they pass from one medium to another but will travel straight in one medium.

Here, light will take a slightly longer path in air than in water because it can travel faster in air.

How do we find *h*?

Bending of light ray (refraction)



Let v_a be the speed of light in air and v_g be that in glass.

Then, time of going from A to B =

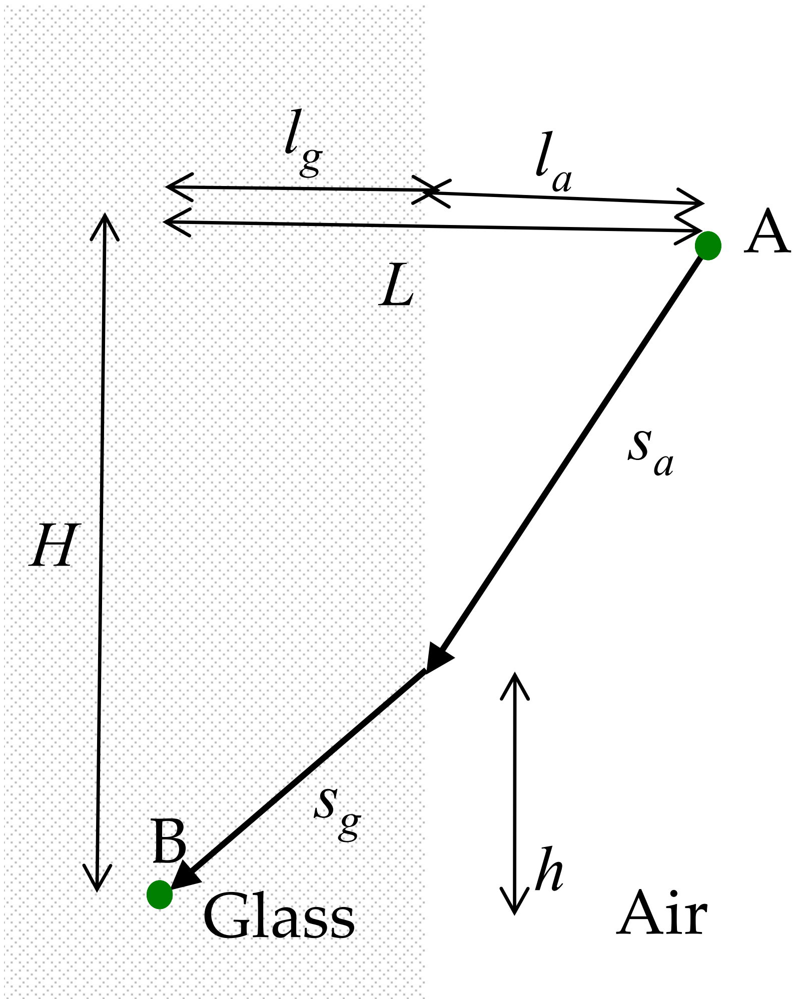
$$T = \frac{s_a}{v_a} + \frac{s_g}{v_g}$$

$$s_a = \sqrt{l_a^2 + (H-h)^2} \qquad s_g = \sqrt{l_g^2 + h^2}$$

$$\underset{h}{\text{Min}} \ T = \frac{\sqrt{l_a^2 + (H-h)^2}}{v_a} + \frac{\sqrt{l_g^2 + h^2}}{v_g}$$

$Data: \ l_a, l_g, H, v_a, v_g$

Finding h



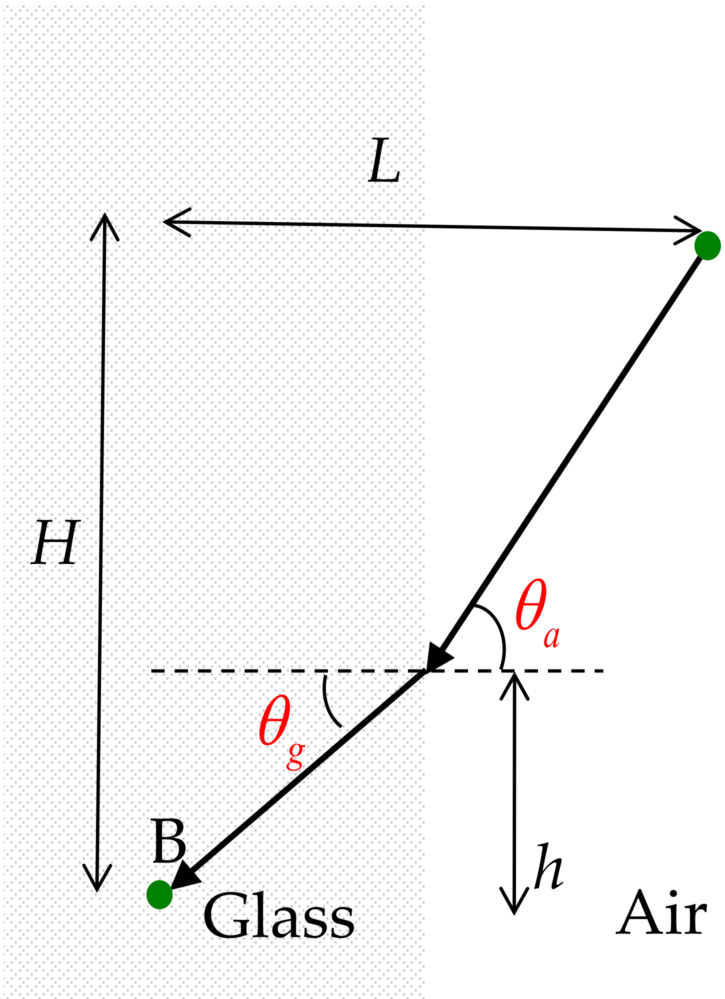
$$\underset{h}{\text{Min}}\ T = \frac{\sqrt{l_a^2 + (H-h)^2}}{v_a} + \frac{\sqrt{l_g^2 + h^2}}{v_g}$$

$$\frac{dT}{dh} = 0$$ (first derivative is zero for a minimum)

$$\Rightarrow \frac{-2(H-h)}{v_a\sqrt{l_a^2 + (H-h)^2}} + \frac{2h}{v_g\sqrt{l_g^2 + h^2}} = 0$$

$$\Rightarrow \frac{(H-h)}{v_a\sqrt{l_a^2 + (H-h)^2}} = \frac{h}{v_g\sqrt{l_g^2 + h^2}}$$

Light ray follow's Snell's rule; the rule follows from optimization.



$$\Rightarrow \frac{(H-h)}{v_a\sqrt{l_a^2+(H-h)^2}} = \frac{h}{v_g\sqrt{l_g^2+h^2}}$$

$$\Rightarrow \frac{\sin\theta_a}{v_a} = \frac{\sin\theta_g}{v_g}$$

Snell's law of refraction.

So, light rays optimize the time taken for them to go from a point to another. Reflection too follows the same optimal path. **Try it.**

We solved a calculus of variations problem as a finite variable optimization problem because we assumed that light follows straight paths in air and glass. **It is a non-smooth path.** We will re-visit it later from the viewpoint of calculus of variations.

The end note

Unconstrained finite-variable optimization

- Only local minimum can have conditions that can easily checked.
  Global minimum does not have "operationally useful" definition or conditons.
- Necessary condition: first order derivative is zero.
  Sufficient condition: second order derivative is positive (or positive definite)
- Gradient
  Hessian
- In two variables, we have peaks, valleys, fold, ridges, and flat planes…
- Rules for checking positive definiteness of a matrix.

And we derived Snell's rule of refraction using optimization.

Thanks