Anthony Gruber

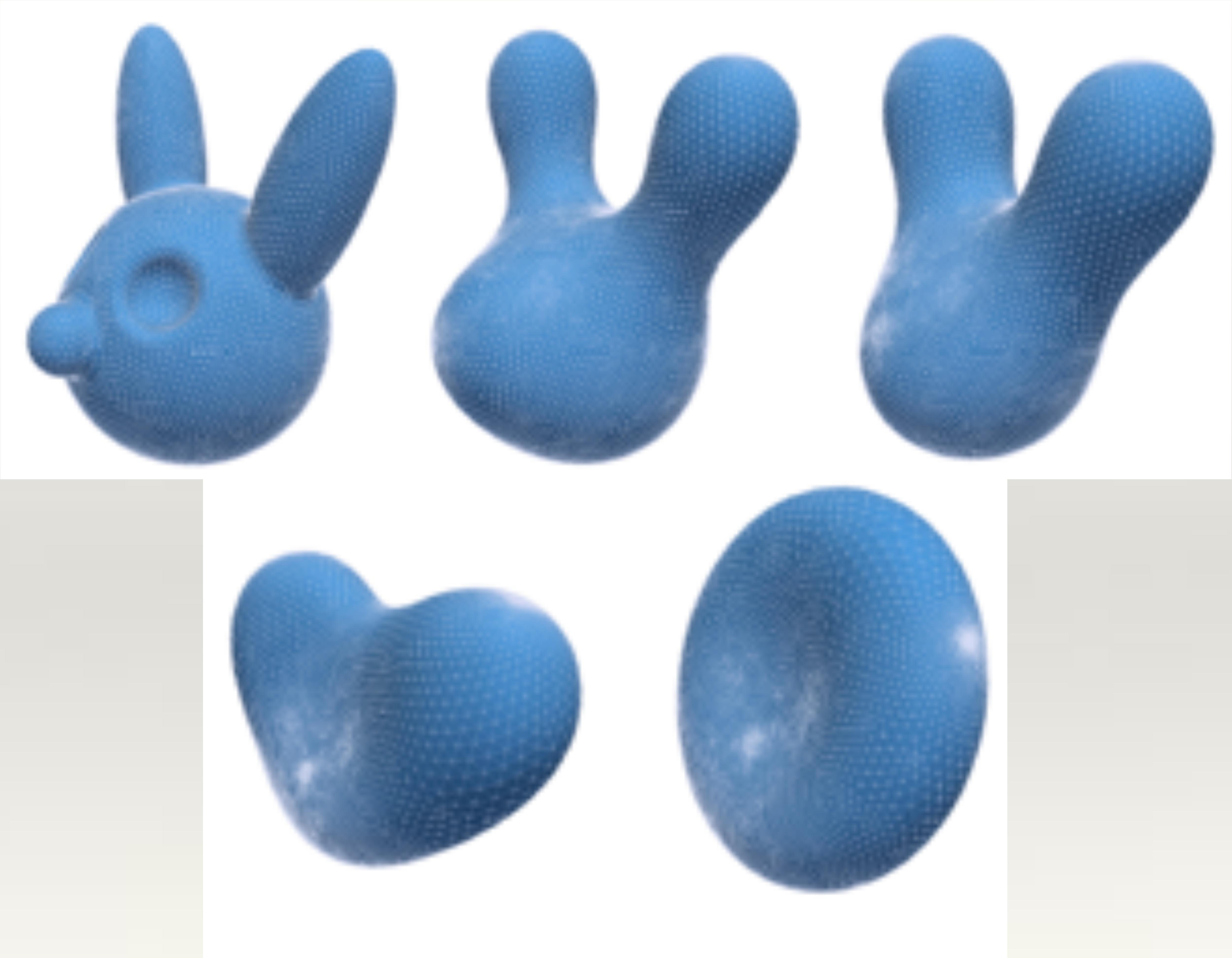
Calculus for Computer Graphics and Data Science

Outline

- Motivation: Why calculus?
- Calculus in Artificial Neural Networks
  - How do they learn?
- Calculus in Curvature Flows
  - How does deformation work?

Motivation

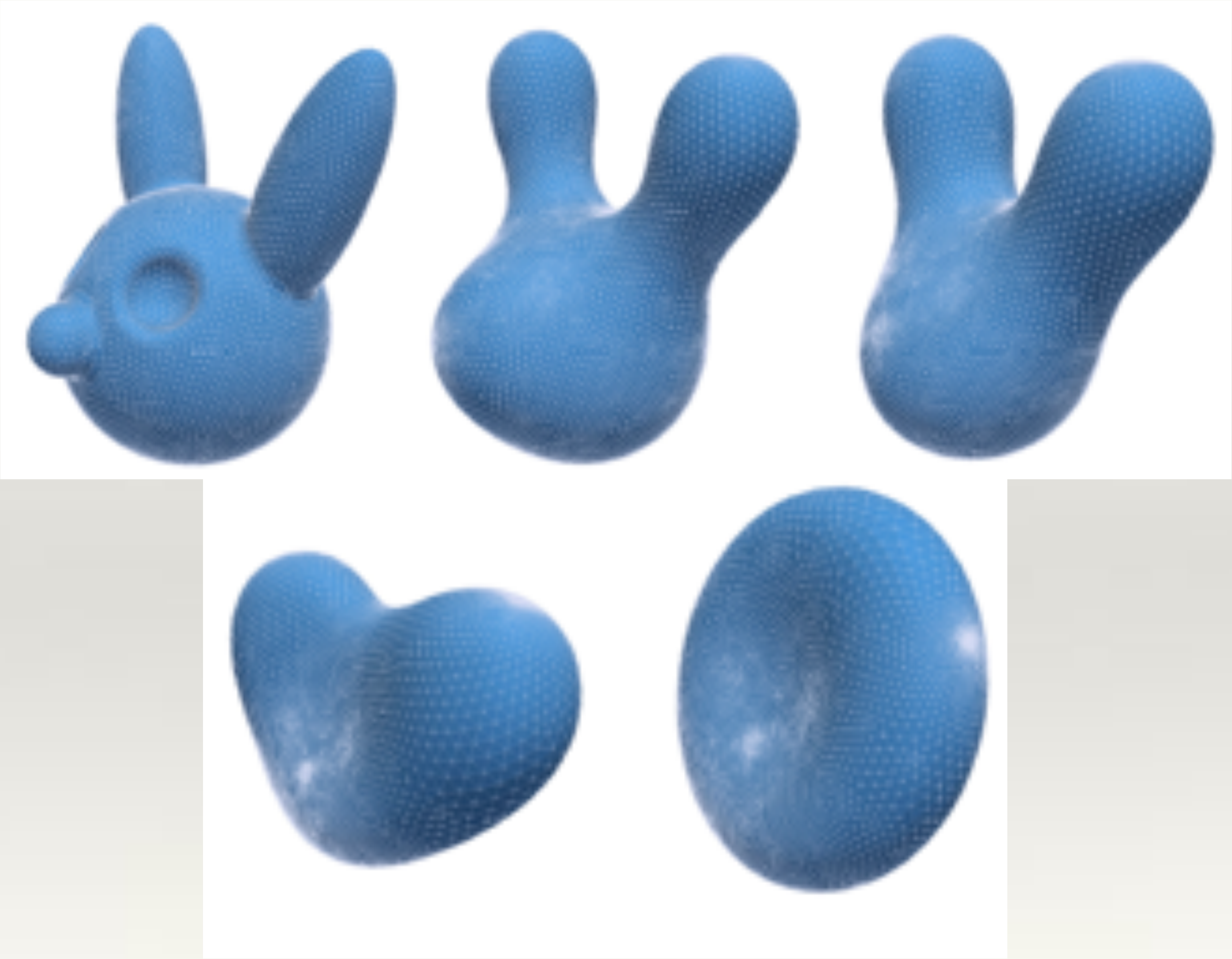
- Animation & Simulation
- Often want *continuous change.*
- Calculus is the study of continuous change!



- Gruber, A. and Aulisa, E. *ACM Trans. Graph.* (2020)

Motivation

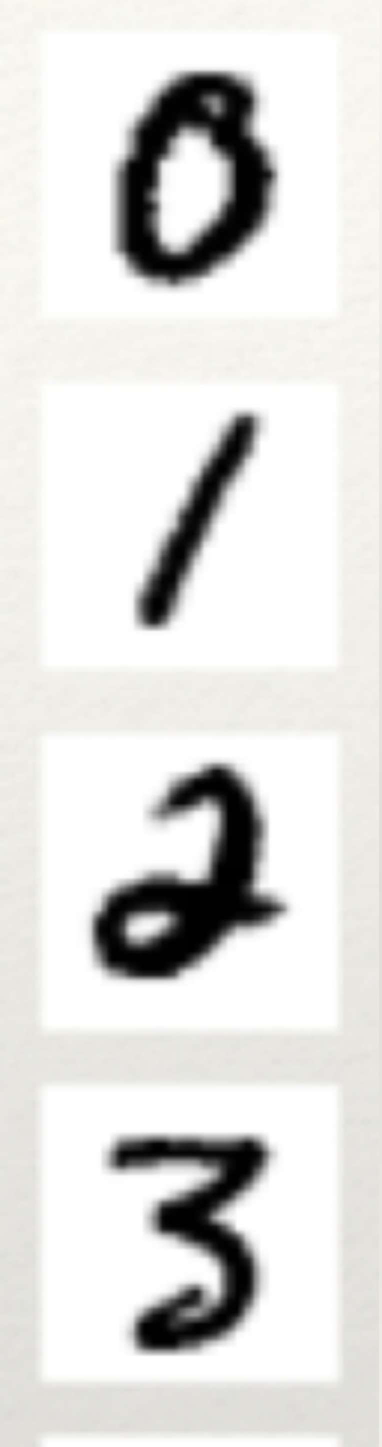
- Quantitative change must be prescribed.
- Pictured: solving a *differential equation.*
- (Certain curvature is minimized over time.)



- Gruber, A. and Aulisa, E. *ACM Trans. Graph.* (2020)

Motivation

- Consider *artificial neural networks* (ANNs).
- ANNs change to learn a predefined task.
- Learn by decreasing error as fast as possible.



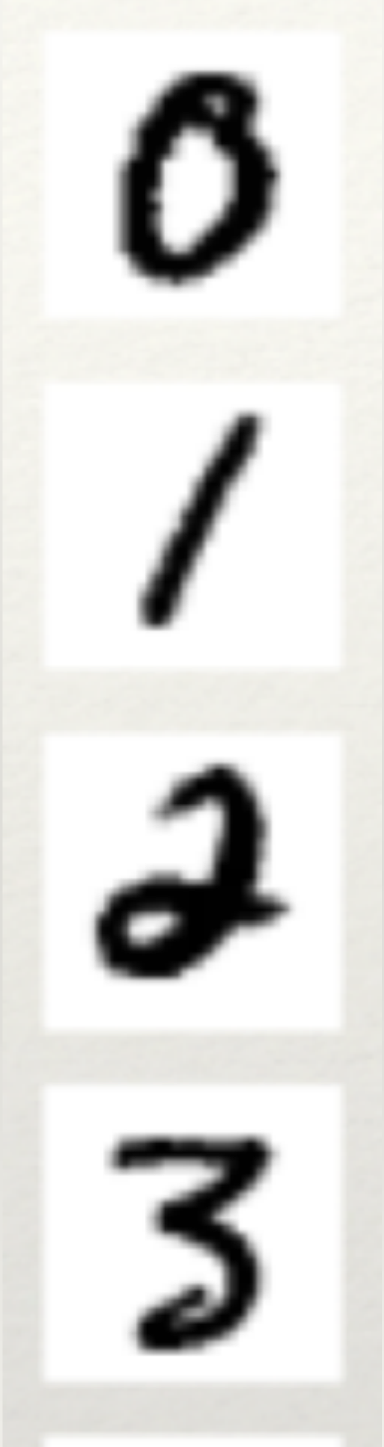
“Zero”

“One”

“Five”

“Eight”

Untrained



“Zero”

“One”

“Two”

“Three”

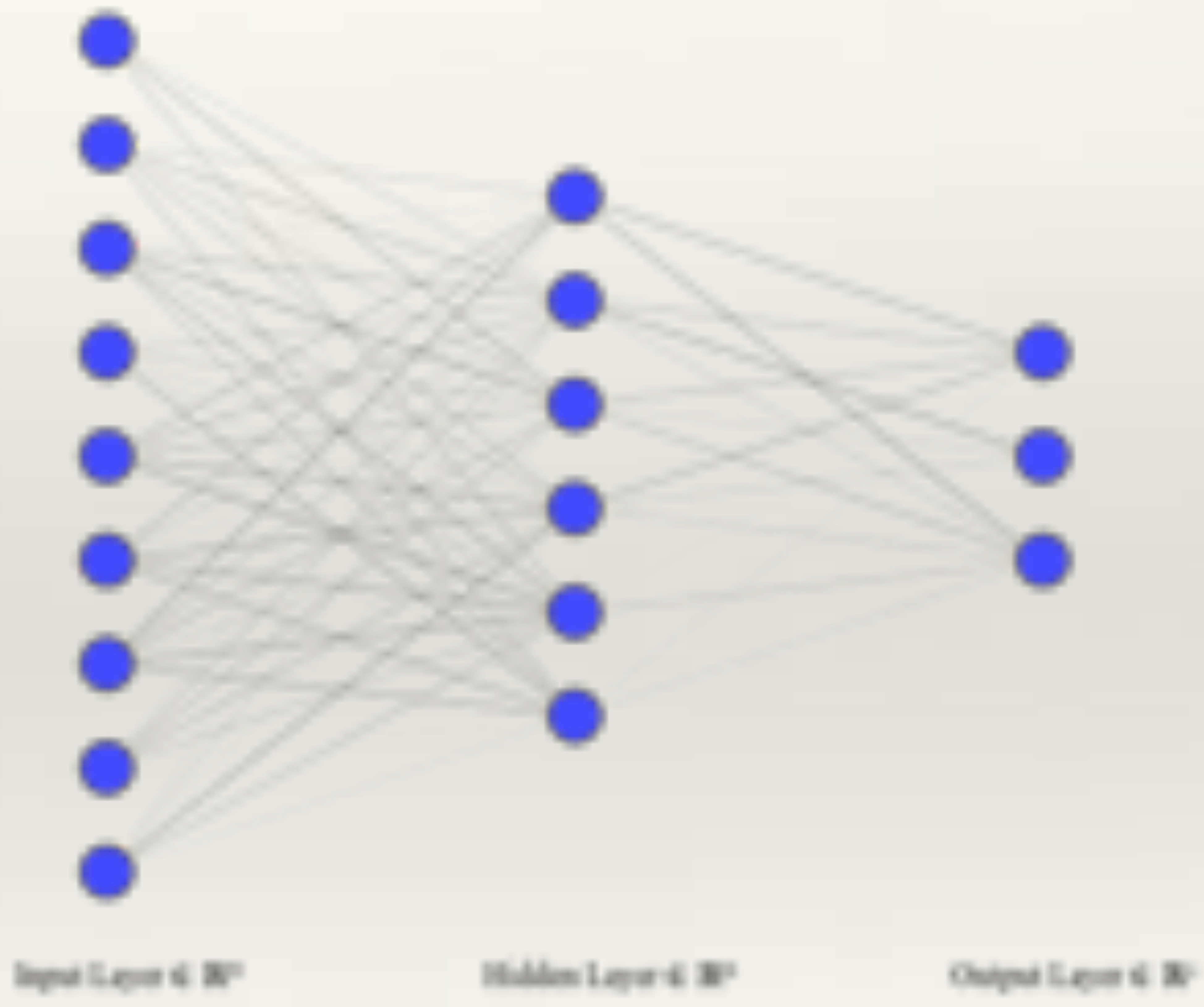
Trained

What is a Neural Network?

- A function approximation machine loosely modeled on biological systems.

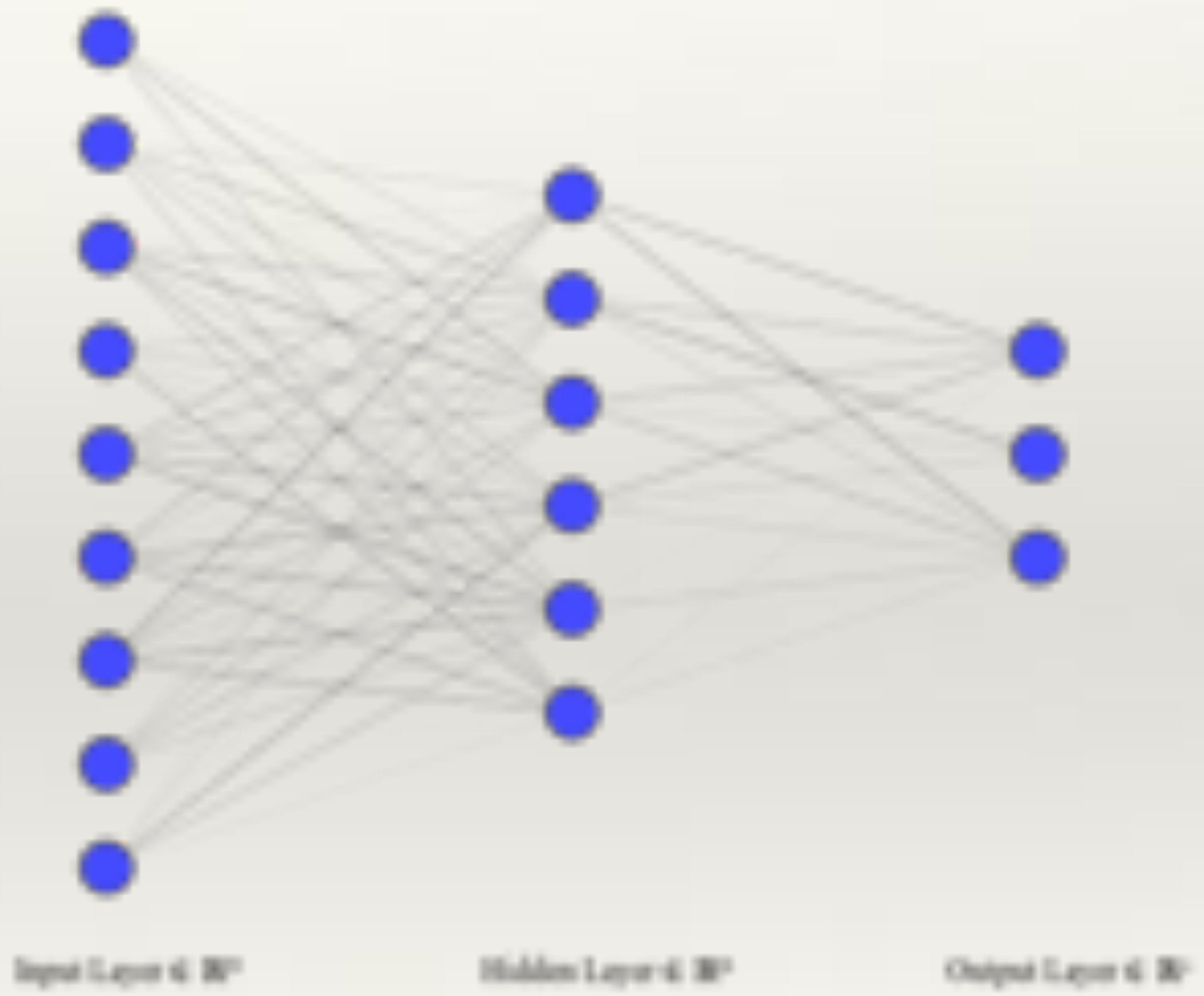
- Each layer contains neurons (nodes) with *learnable parameters* (edge weights).

- # Nodes = Width,   # Layers = Depth



How do NNs Learn?

- Information flows *forward* from inputs to outputs.
- Outputs produce a *response,* with associated *error.*
- Information flows *backward* through derivative of error.
- Parameters are *updated* accordingly.

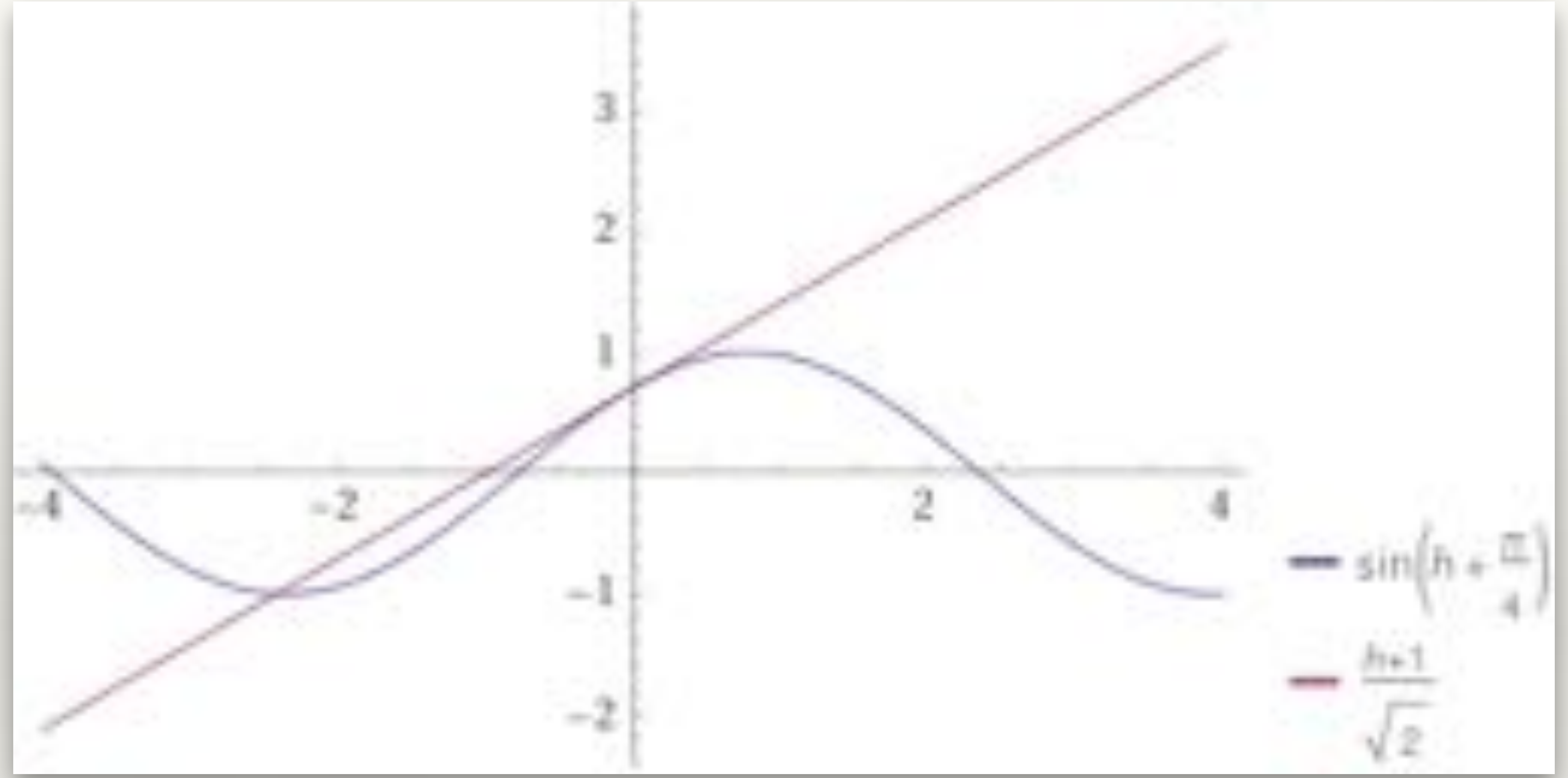


Review: What is a Derivative?

- $f : \mathbb{R} \to \mathbb{R}$ differentiable. Fix $x \in \mathbb{R}$.

- Rate of change:

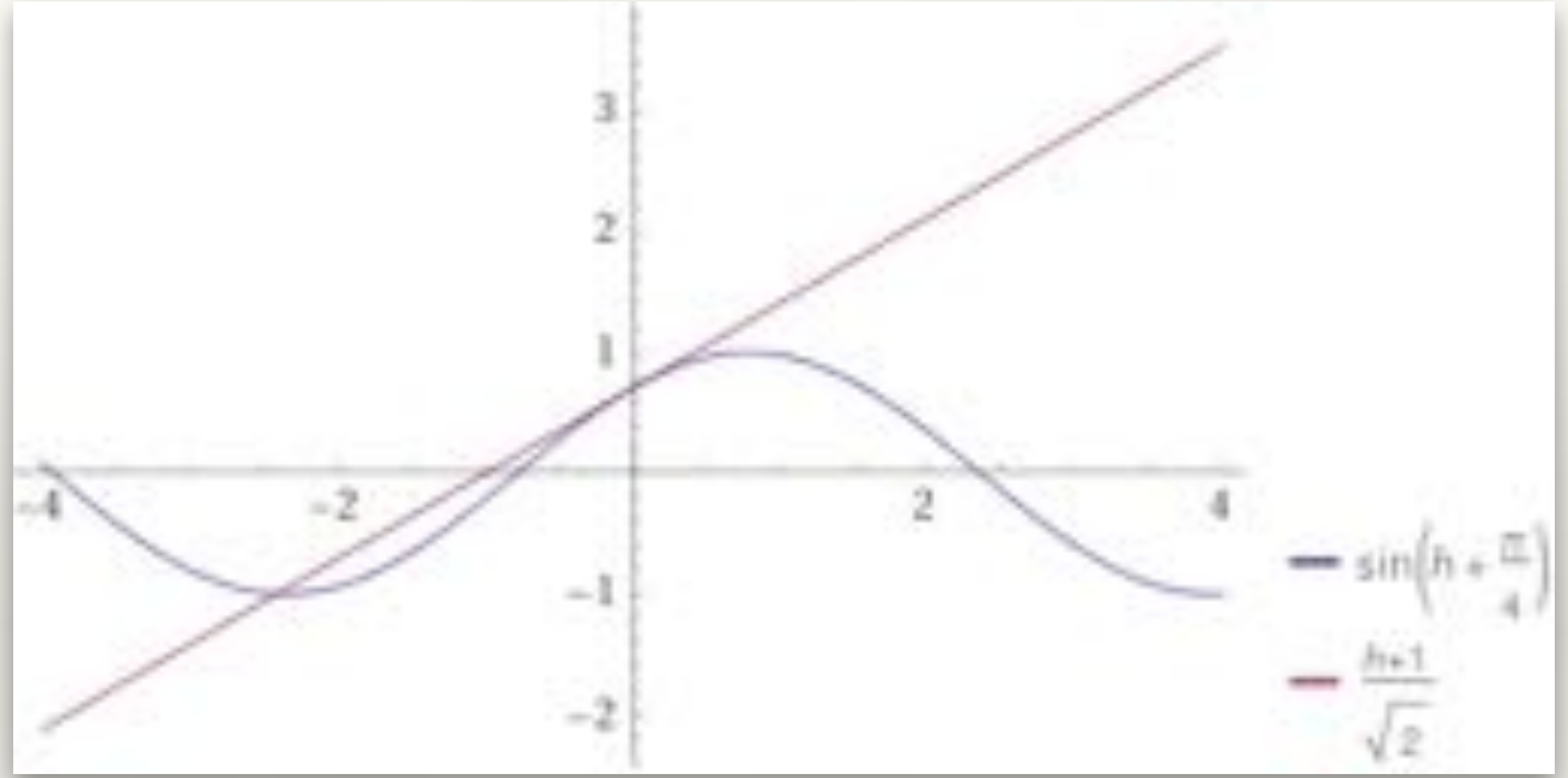
$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$



Ex) $f(x) = \sin x$ around $x = \dfrac{\pi}{4}$

Review: What is a Derivative?

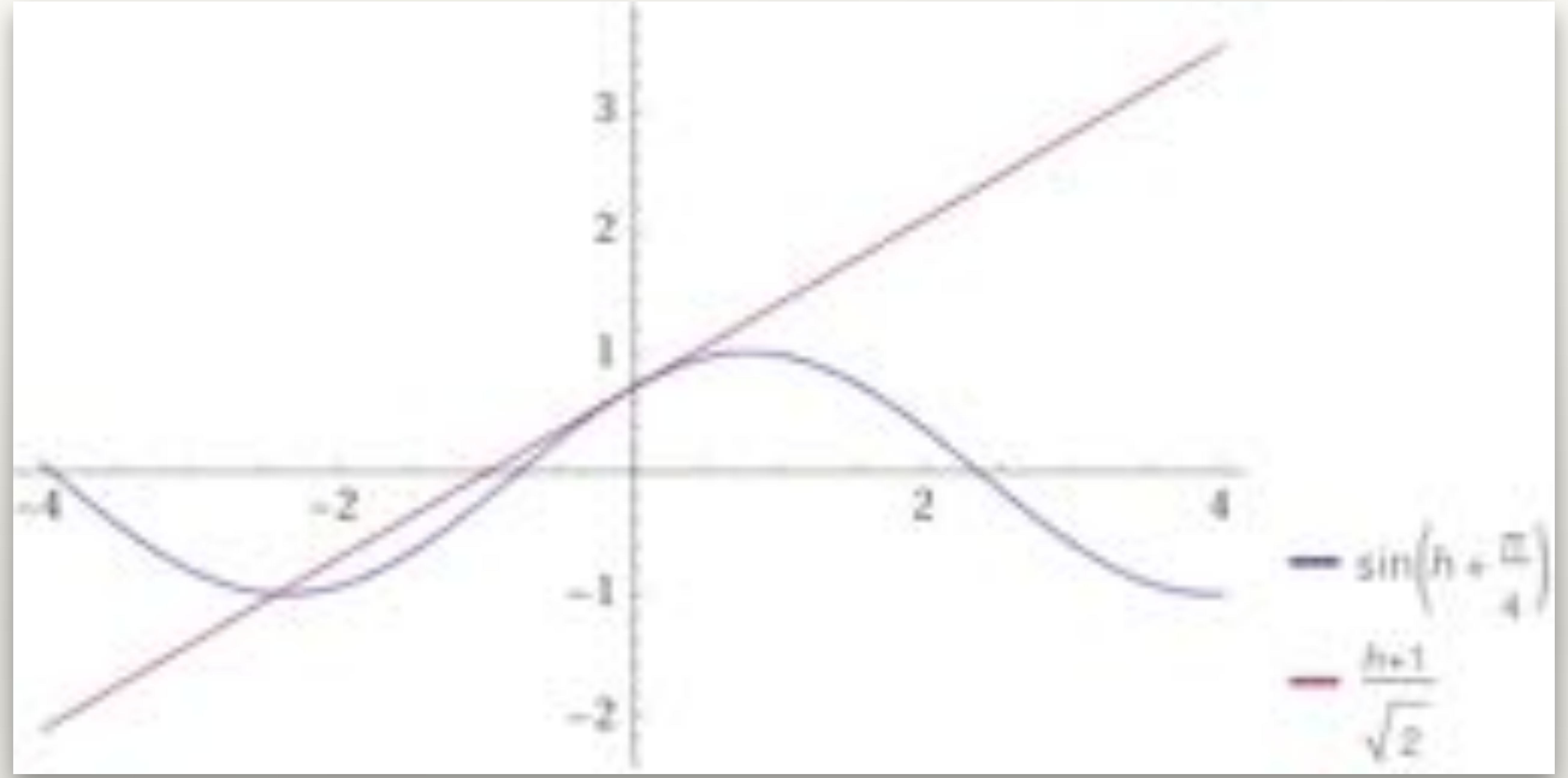
- $f(x+h) = f(x) + f'(x)h + \mathcal{O}(h^2)$
  $= T_x(h) + \mathcal{O}(h^2)$
  - $T_x(h)$ *best linear approximation* to f at x.
  - $f'(x)$ is slope of tangent line to graph.



Ex) $f(x) = \sin x$ around $x = \dfrac{\pi}{4}$

Review: What is a Derivative?

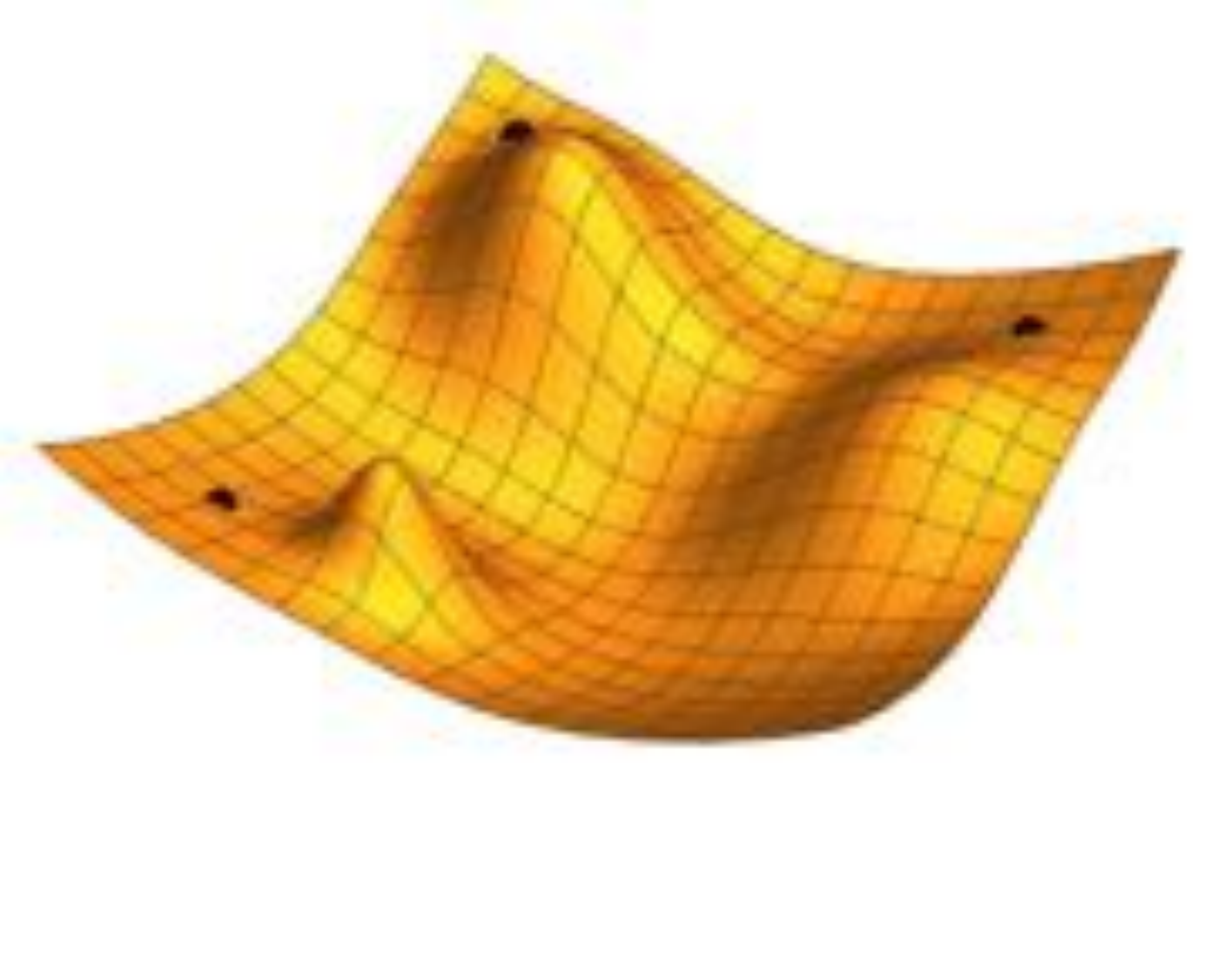
- How can we decrease f?
  - *Roll along the tangents*
- **Fact:** $x_{n+1} = x_n - f'(x_n)h$
  - (for small h)
- Eventually converges to local minimum.



Ex) $f(x) = \sin x$ around $x = \dfrac{\pi}{4}$

Gradient Descent

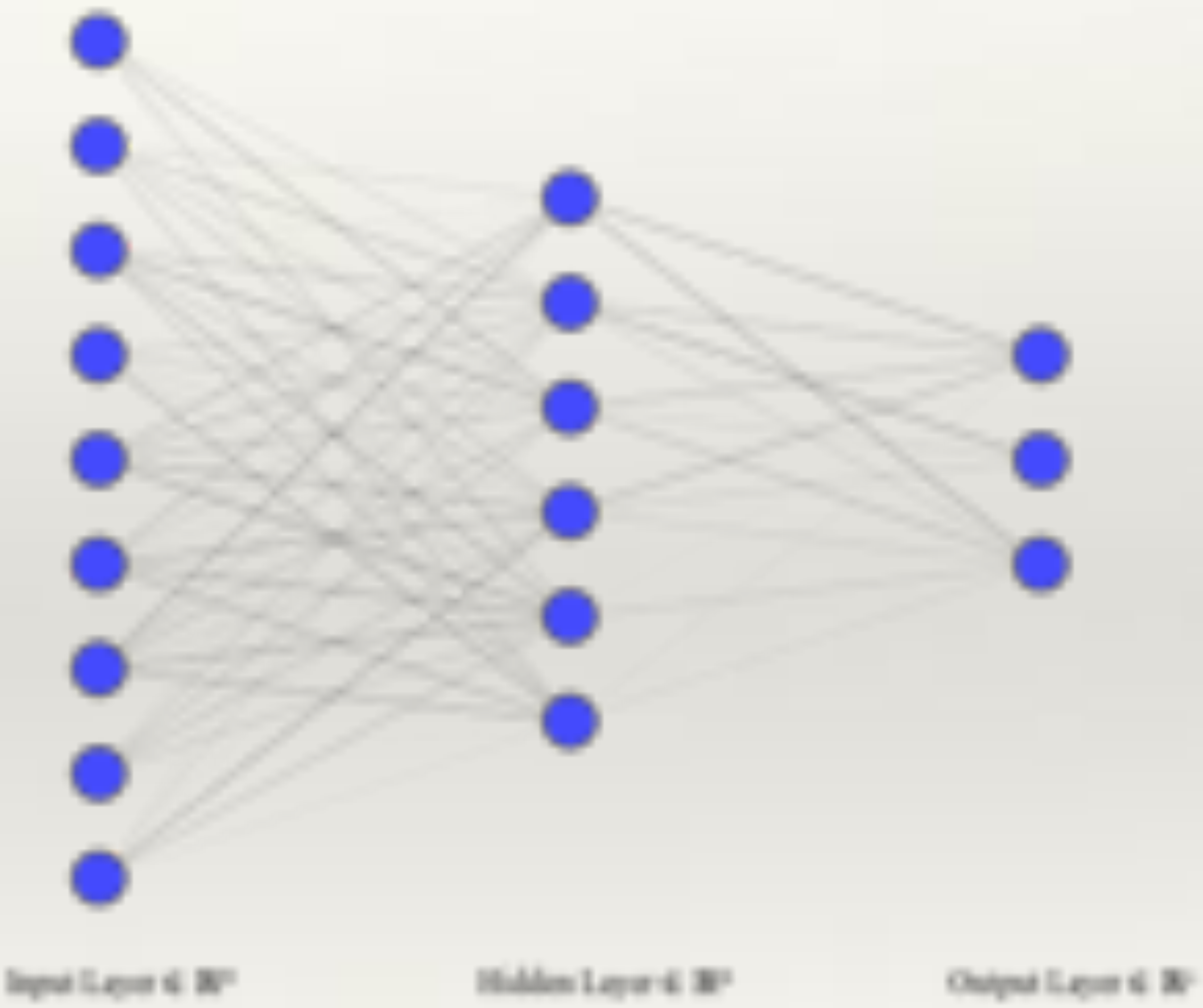
- Consider $f : \mathbb{R}^n \to \mathbb{R}$.
- *gradient field* $\nabla f(\mathbf{x})$ (column vector)
- The maximum rate of change in f occurs in the direction of ∇f.
- Therefore, f decreases maximally when pushed in the direction of $-\nabla f$.



Video by Jacopo Bertolotti
https://commons.wikimedia.org/wiki/File:Gradient_descent.gif

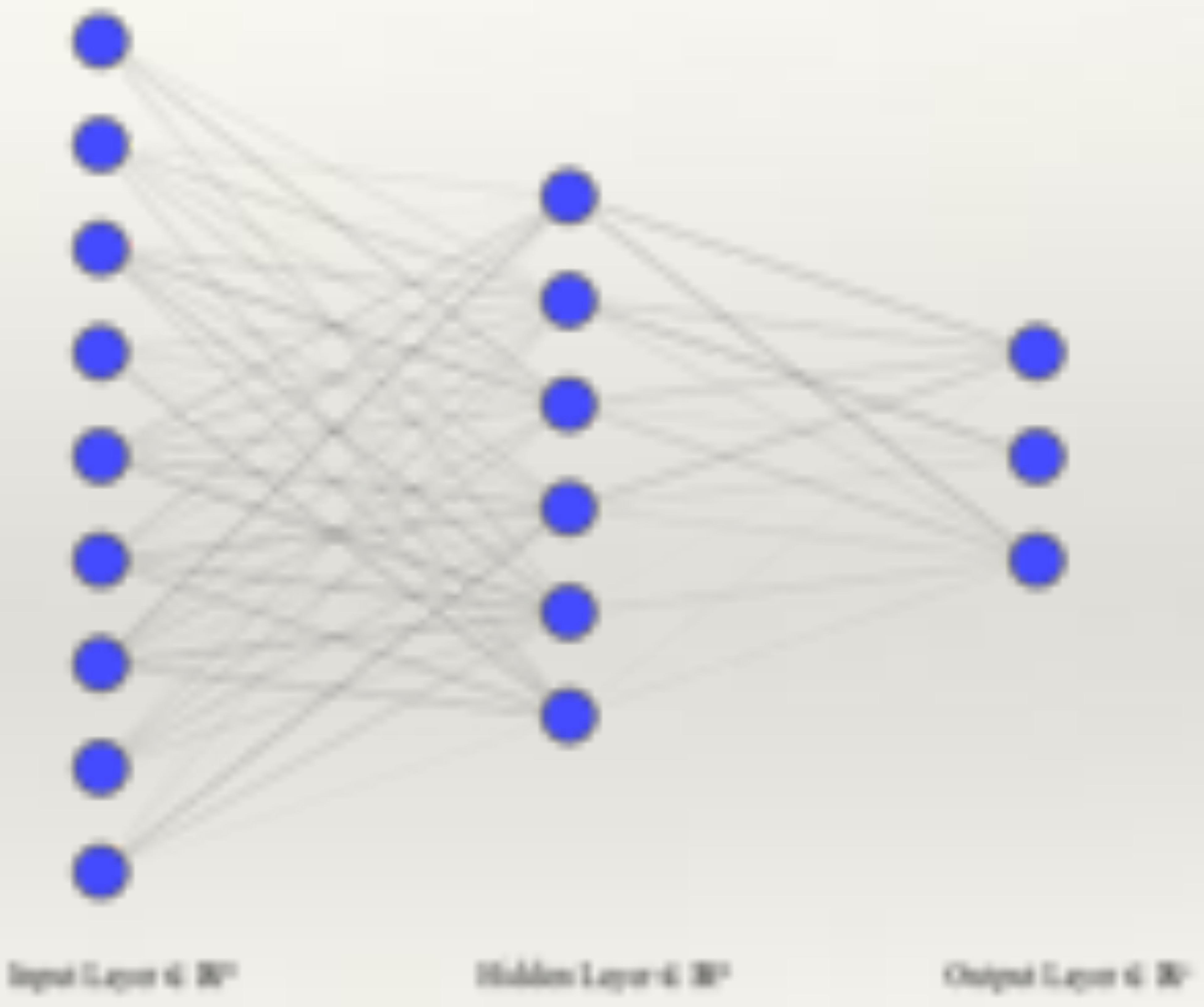
Inside a Neural Network

- Consider a neural network $\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$.
  - Depends on inputs $\mathbf{x}$ and parameters $\boldsymbol{\theta}$.
- Consider a loss (error) function $L(\mathbf{y})$.
- What is $L_{\boldsymbol{\theta}}(\mathbf{y}) = \dfrac{\partial L}{\partial \boldsymbol{\theta}}(\mathbf{y})$?



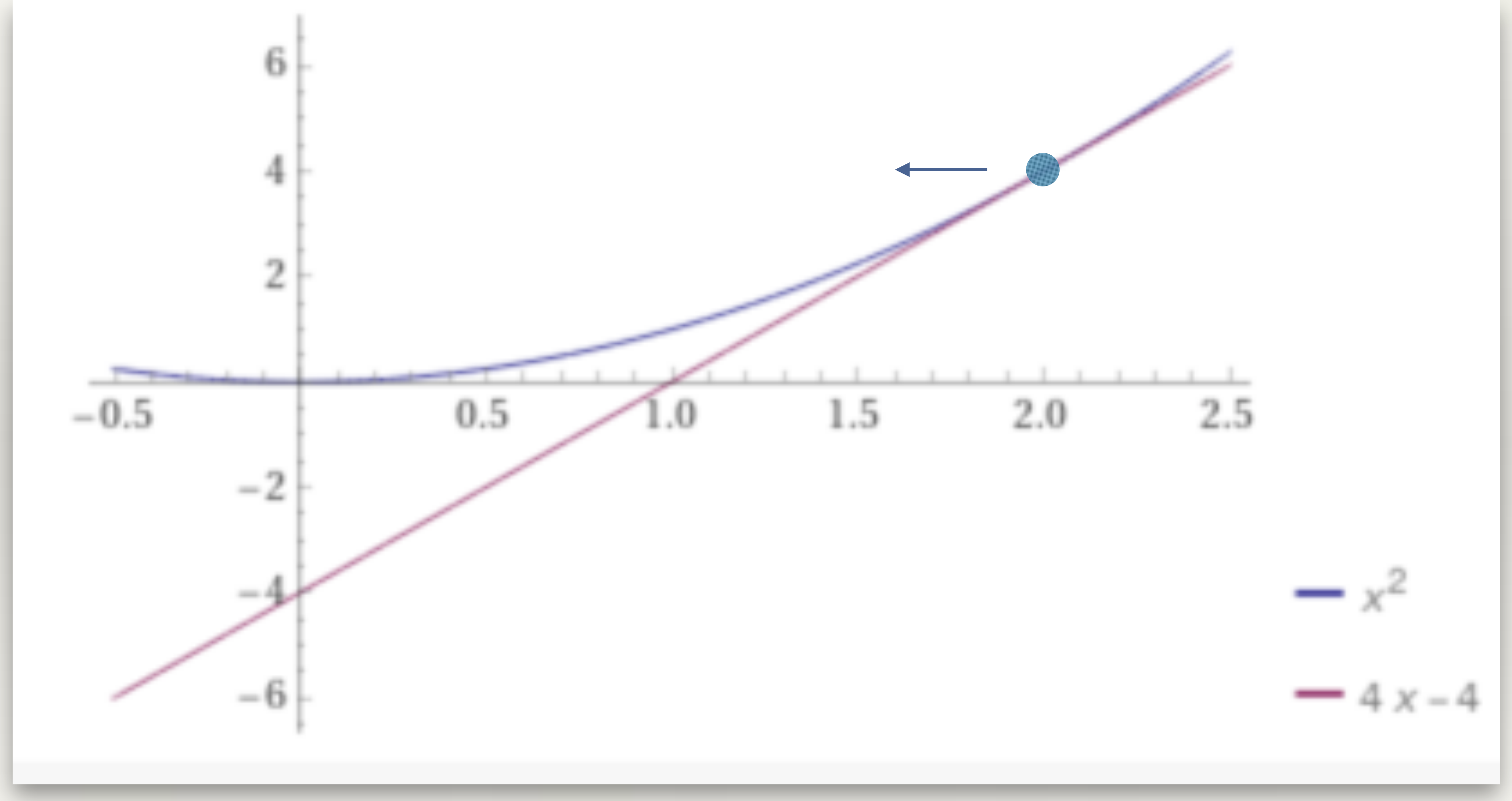
Inside a Neural Network

- Components of $L_{\boldsymbol{\theta}}$: *sensitivities* of loss L to each parameter θ^i.
- Chain rule! $L_{\boldsymbol{\theta}} = L'(\mathbf{y})\mathbf{y}_{\boldsymbol{\theta}}$
  - $L_{\boldsymbol{\theta}} = L_{y^1}\, y^1_{\boldsymbol{\theta}} + L_{y^2}\, y^2_{\boldsymbol{\theta}} + L_{y^3}\, y^3_{\boldsymbol{\theta}}$



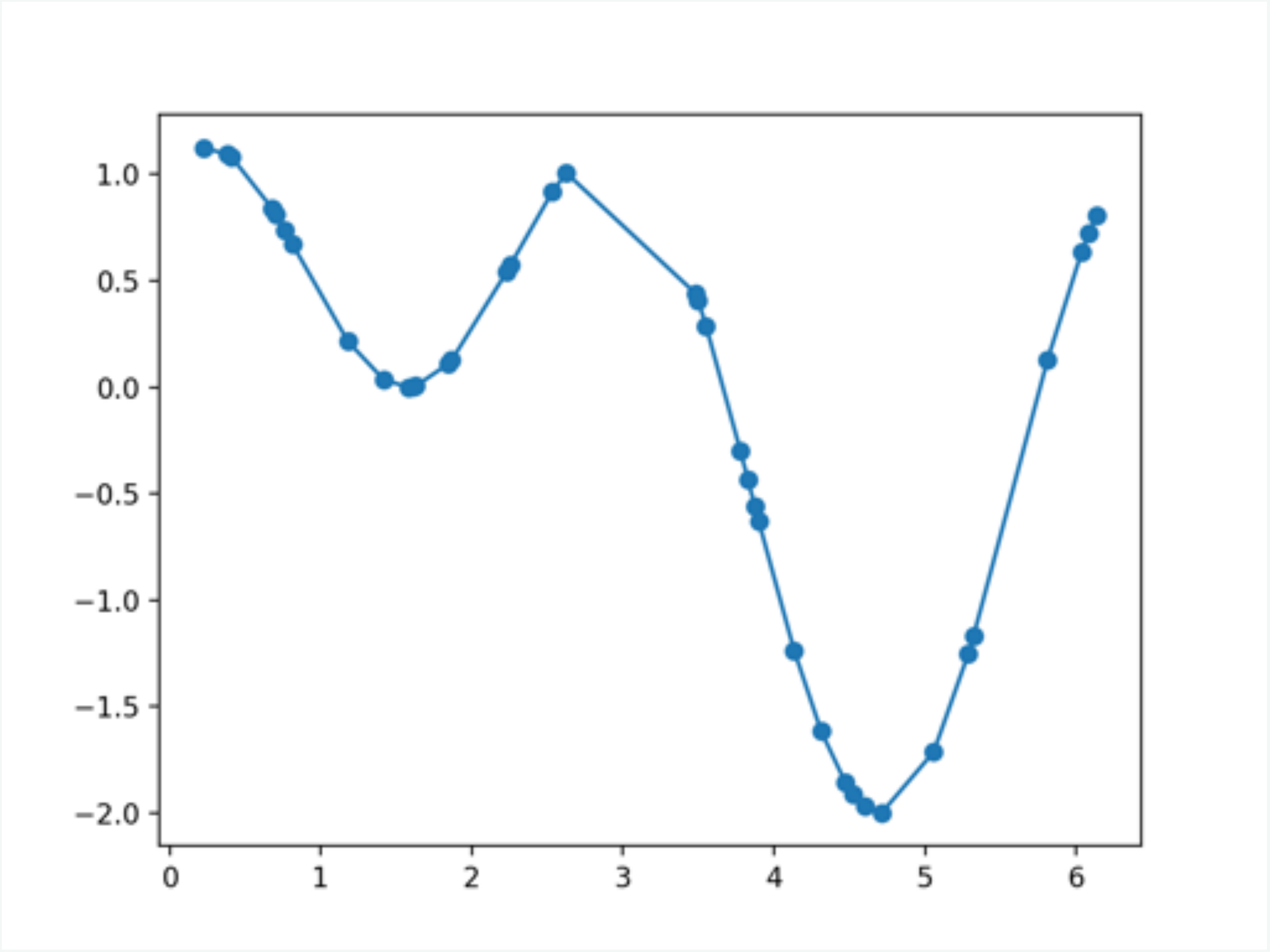
Neural Network Training

- What does the network do with this information?
- Value of L decreases fastest when $\boldsymbol{\theta}$ moves parallel to $-L_{\boldsymbol{\theta}}$.
  - (row version of gradient)
- Can *update* $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - t\,L'(\mathbf{y})\mathbf{y}_{\boldsymbol{\theta}}$ where t is the learning rate.



Example

- Consider $y = \sin x + \cos 2x$
  - Can we learn it?
- Take $x_i \in [0, 2\pi], \quad y_i = \sin x_i + \cos 2x_i$
- Network $f(x, \boldsymbol{\theta})$
- Minimize $L = \sum_i \left| y_i - f(x_i, \boldsymbol{\theta}) \right|^2$

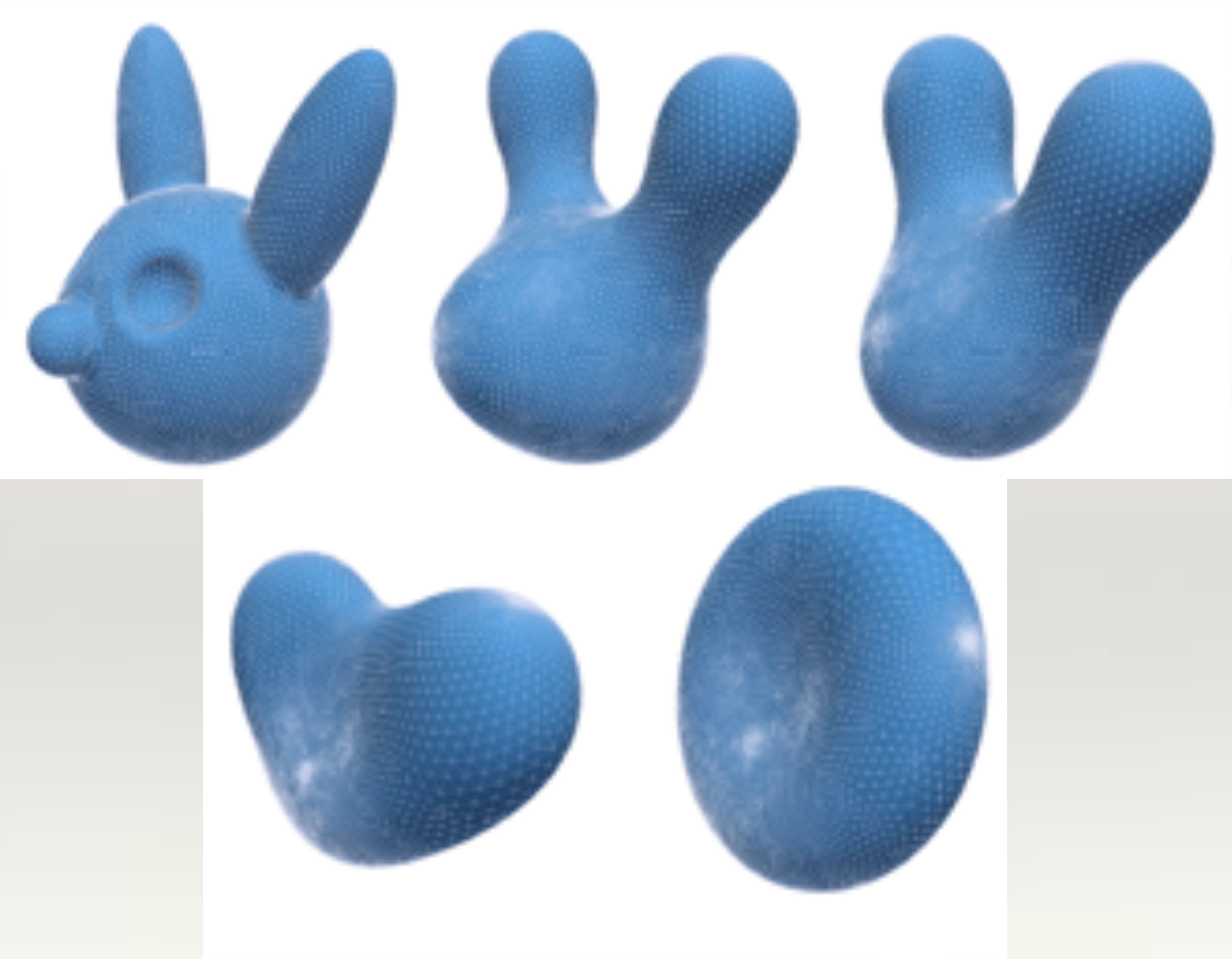


Graphics Applications

- The derivative is inherently linked to *minimization*.
- Willmore energy:

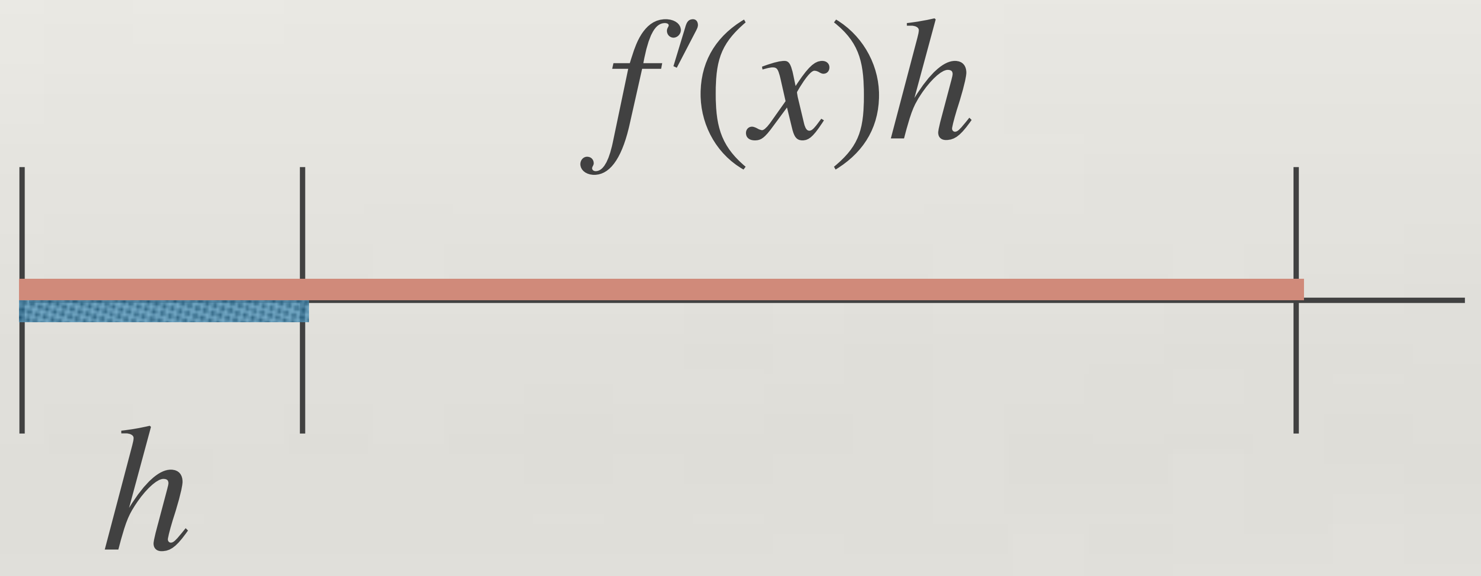
$$\mathscr{W}^2(\mathbf{X}) = \int_M H^2 \, d\mu_g$$

- How can we minimize such functions?



Review: What is a Derivative?

- What does $f'(x)$ do to the number h?
  - The map $h \mapsto f'(x)h$ is multiplication!
  - The derivative $f'(x) : \mathbb{R} \to \mathbb{R}$ *dilates* h.
- $f'(x) = 0 \iff f'(x)h = 0$ for all $h \in \mathbb{R}$.

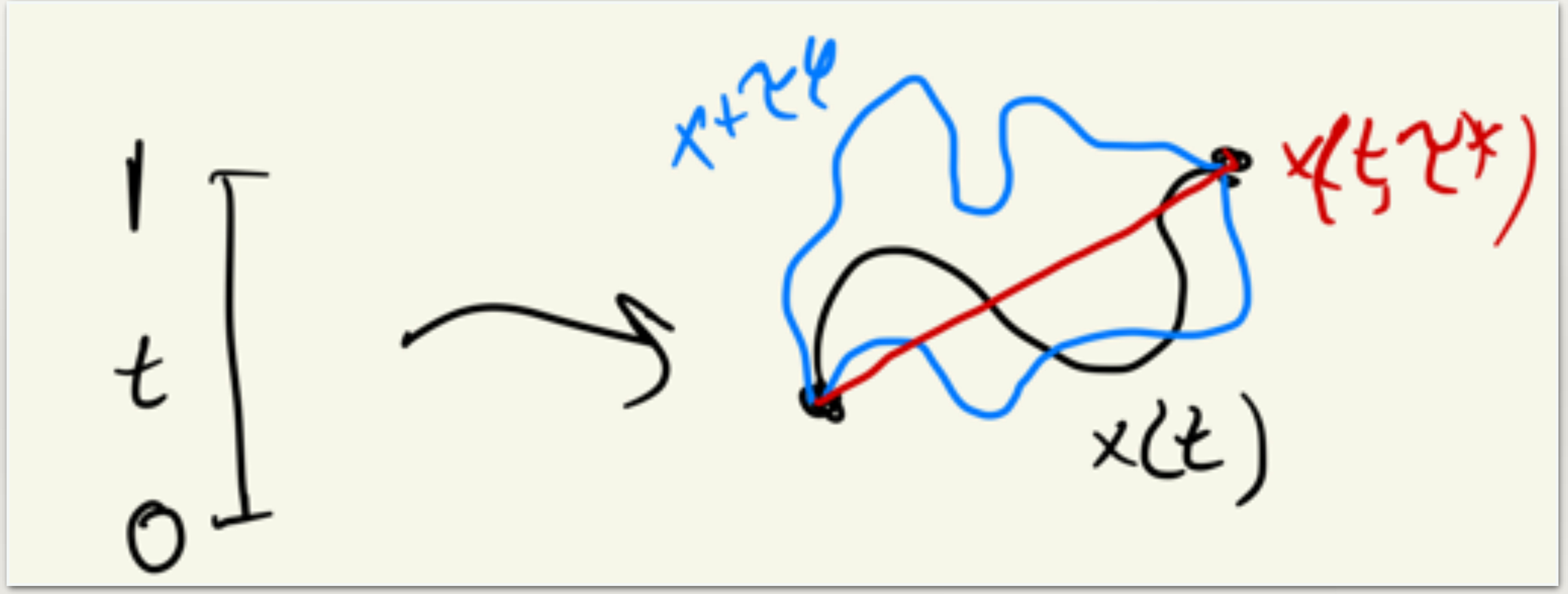


Example: Shortest length

- Curve $\mathbf{x} : [0,1] \to \mathbb{R}^2$.
- Length functional is

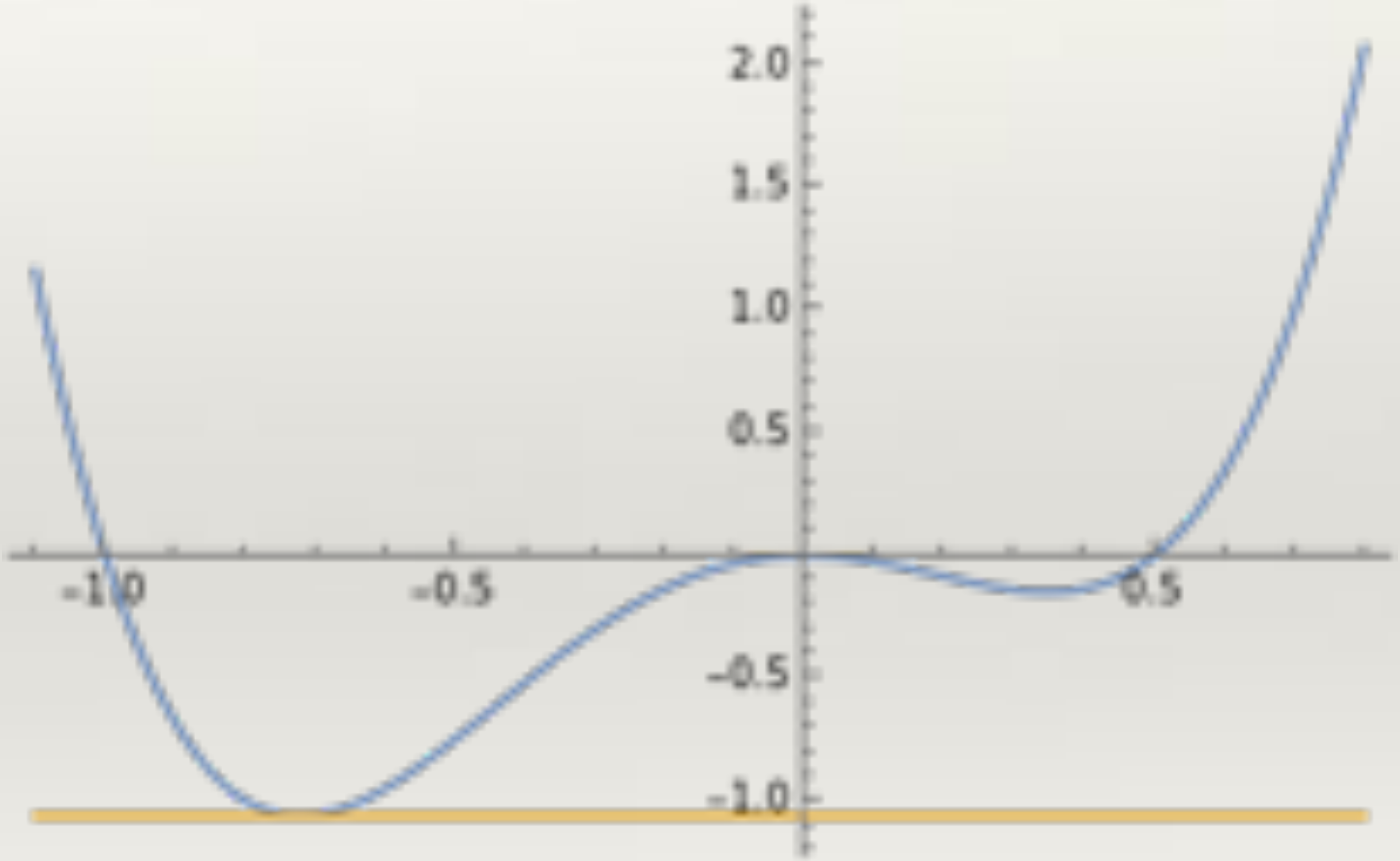
$$\mathscr{L}(\mathbf{x}) = \int_C ds = \int_0^1 |\,\mathbf{x}'(t)\,|\,dt$$

- $\mathbf{x}(t,\tau) = \mathbf{x}(t) + \tau\boldsymbol{\varphi}(t)$
  variation of curve.



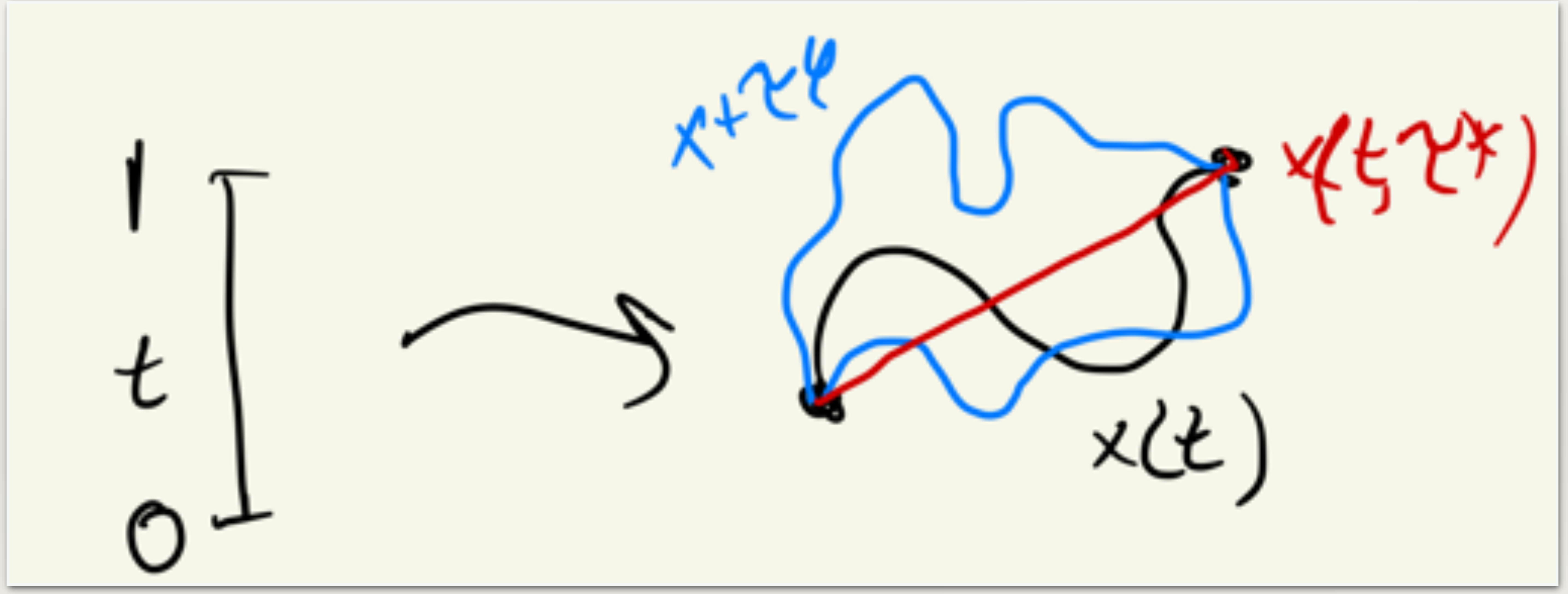
First Variation of Length

- Want to find $\mathbf{x}$ where $\mathscr{L}$ is *stationary*.
  - Means
  $$\delta\mathscr{L}(\mathbf{x})\boldsymbol{\varphi} = \frac{d}{d\tau}\Big|_{\tau=0} \mathscr{L}(\mathbf{x} + \tau\boldsymbol{\varphi}) = 0.$$
  - Must hold for all admissible $\boldsymbol{\varphi}$.
- This is called taking the *first variation* of $\mathscr{L}$.



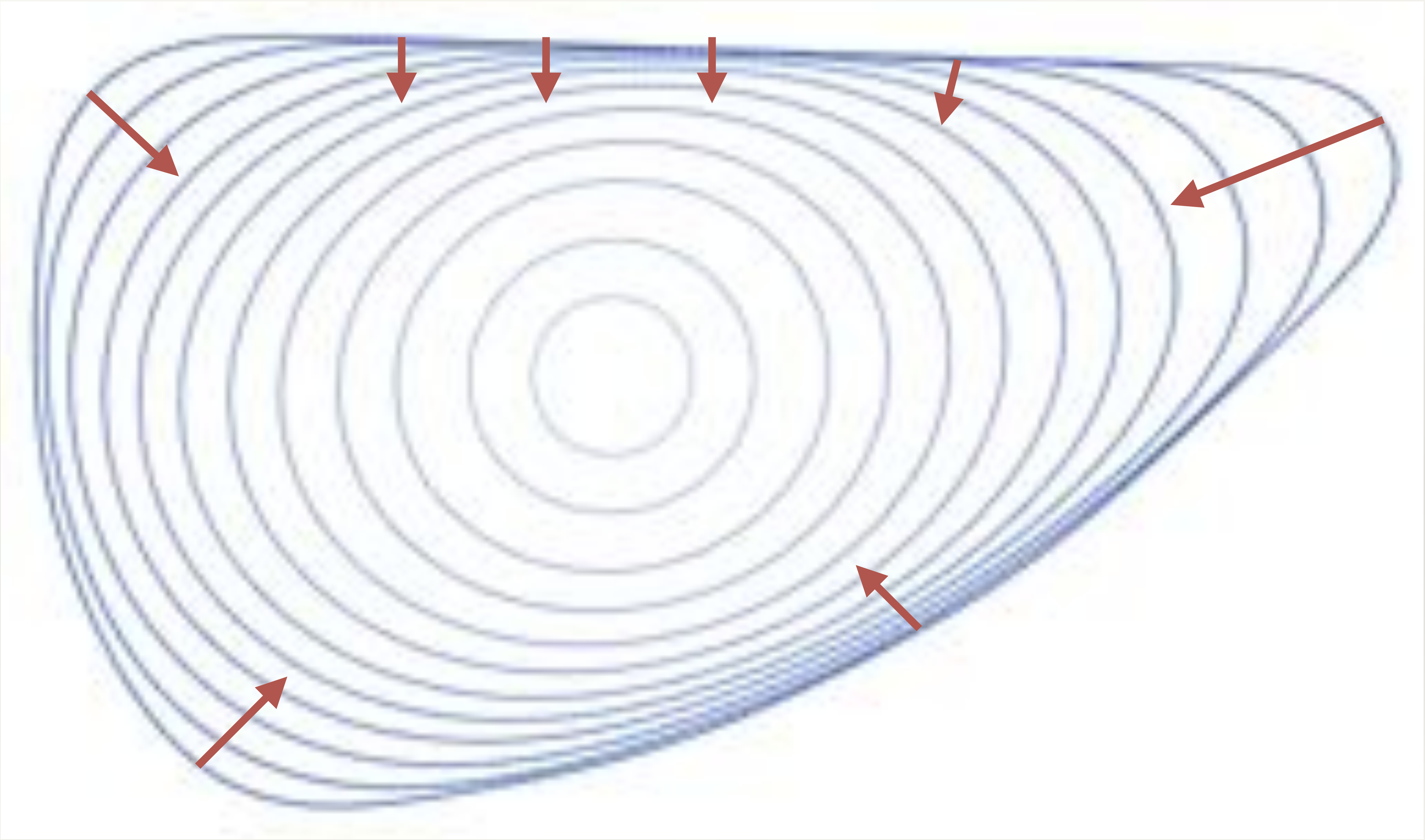
Example: Shortest length

- When is $\mathscr{L}$ stationary?
- $\left.\dfrac{d}{d\tau}\right|_{\tau=0} \mathscr{L}(\mathbf{x} + \tau\boldsymbol{\varphi}) = 0.$
- If $\boldsymbol{\varphi}(0) = \boldsymbol{\varphi}(1) = 0$, this implies
  - $\displaystyle\int_0^1 \kappa\,\mathbf{N}\cdot\boldsymbol{\varphi}\,dt = 0 \;\text{ for all } \boldsymbol{\varphi}.$
- Curvature κ must be 0!

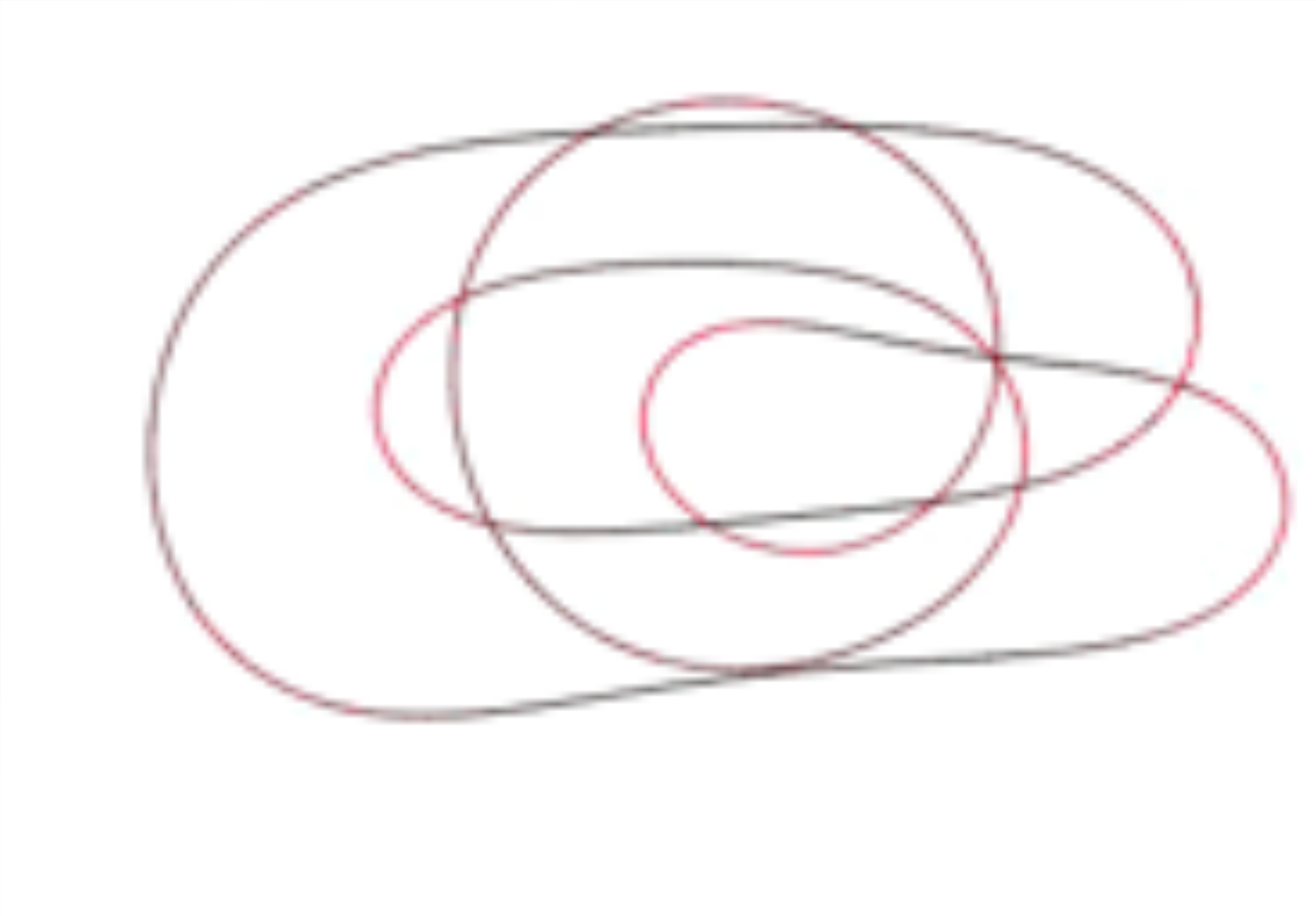


Derivative of Length Functional

- This shows $\kappa\mathbf{N}$ is the *gradient* of the length.
- What if we solve $\dot{\mathbf{x}} = -\kappa\mathbf{N}$
  - Curve-shortening flow!
- *Fastest way* to decrease length.



Curve-Shortening Flow



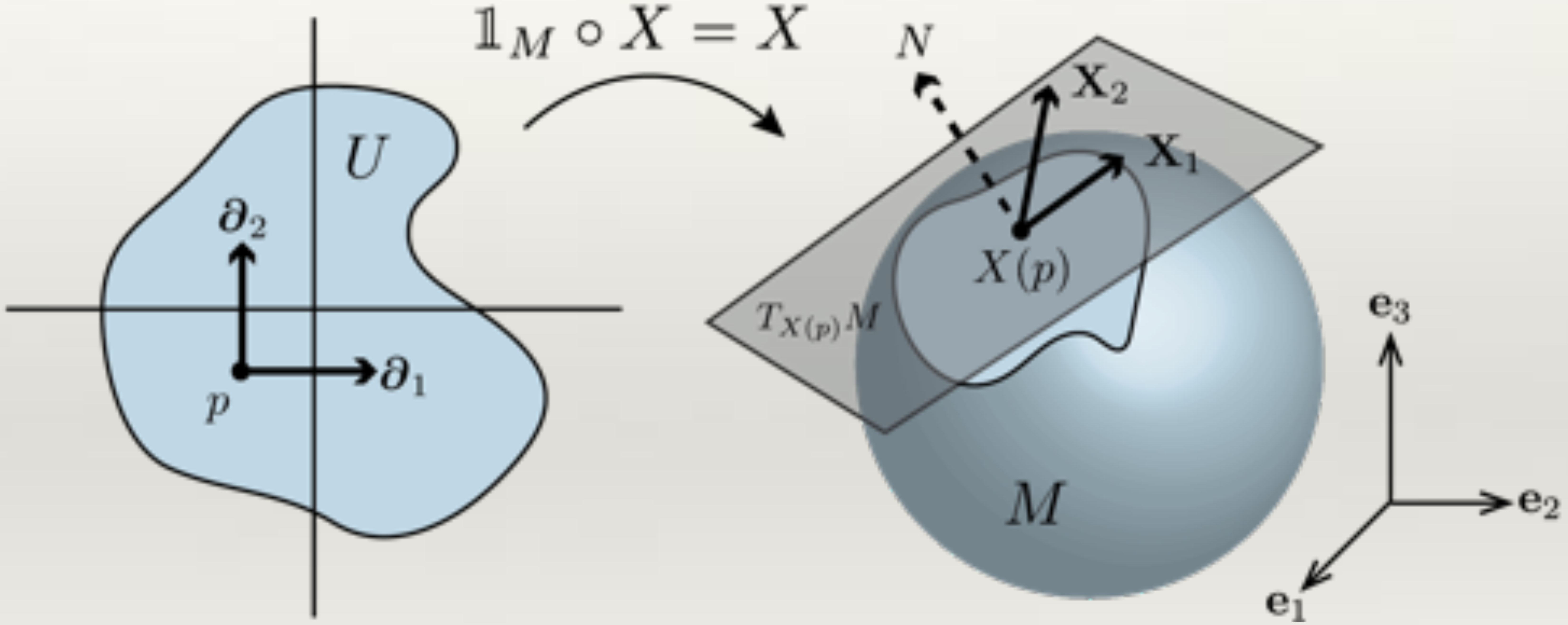
Code by Anthony Carapetis: https://github.com/acarapetis/curve-shortening-demo

What About in 2-D?

- Can you do the same in higher dimensions?
  - Yes! *Mean curvature flow* $\dot{\mathbf{X}} = \Delta_g \mathbf{X} = -2H\mathbf{N}.$
- Fastest way to decrease surface area.

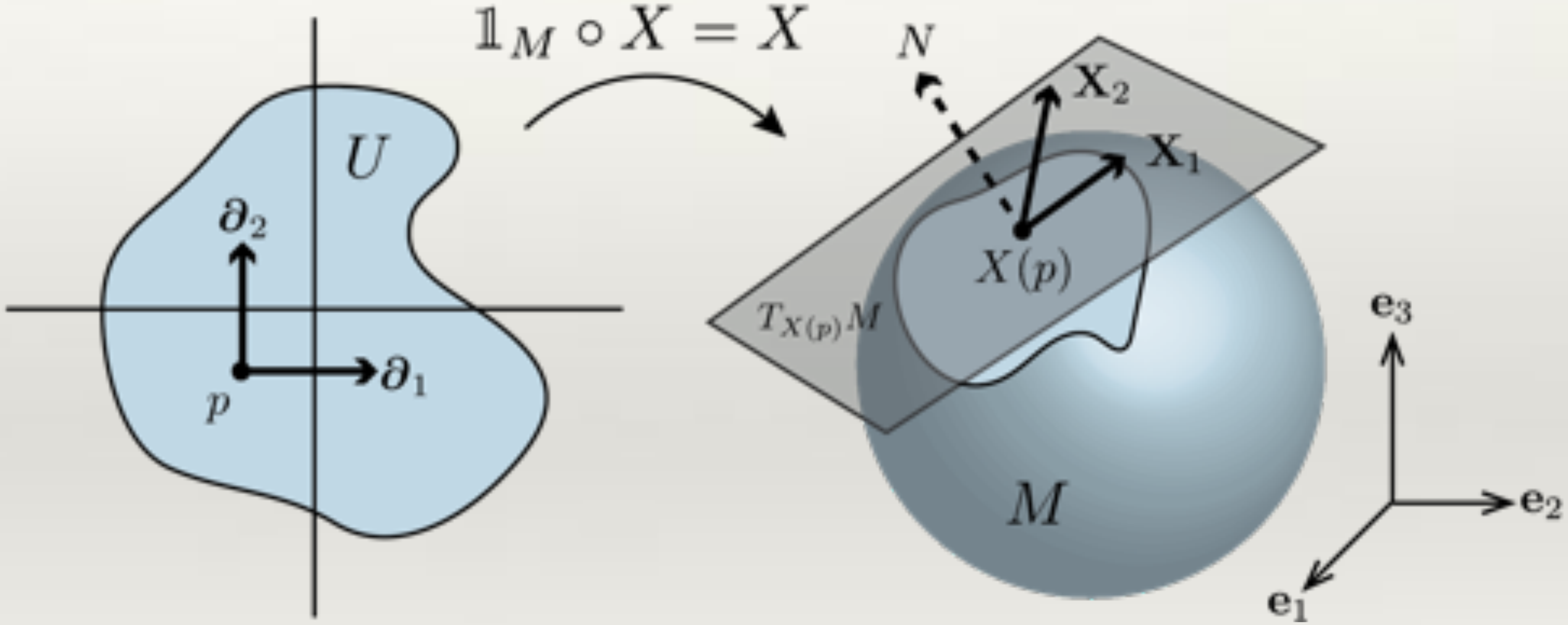
Geometry on a Surface

- How do we measure distances on M?
- We need a *Riemannian metric* g.
- g measures the angle between tangent vectors!



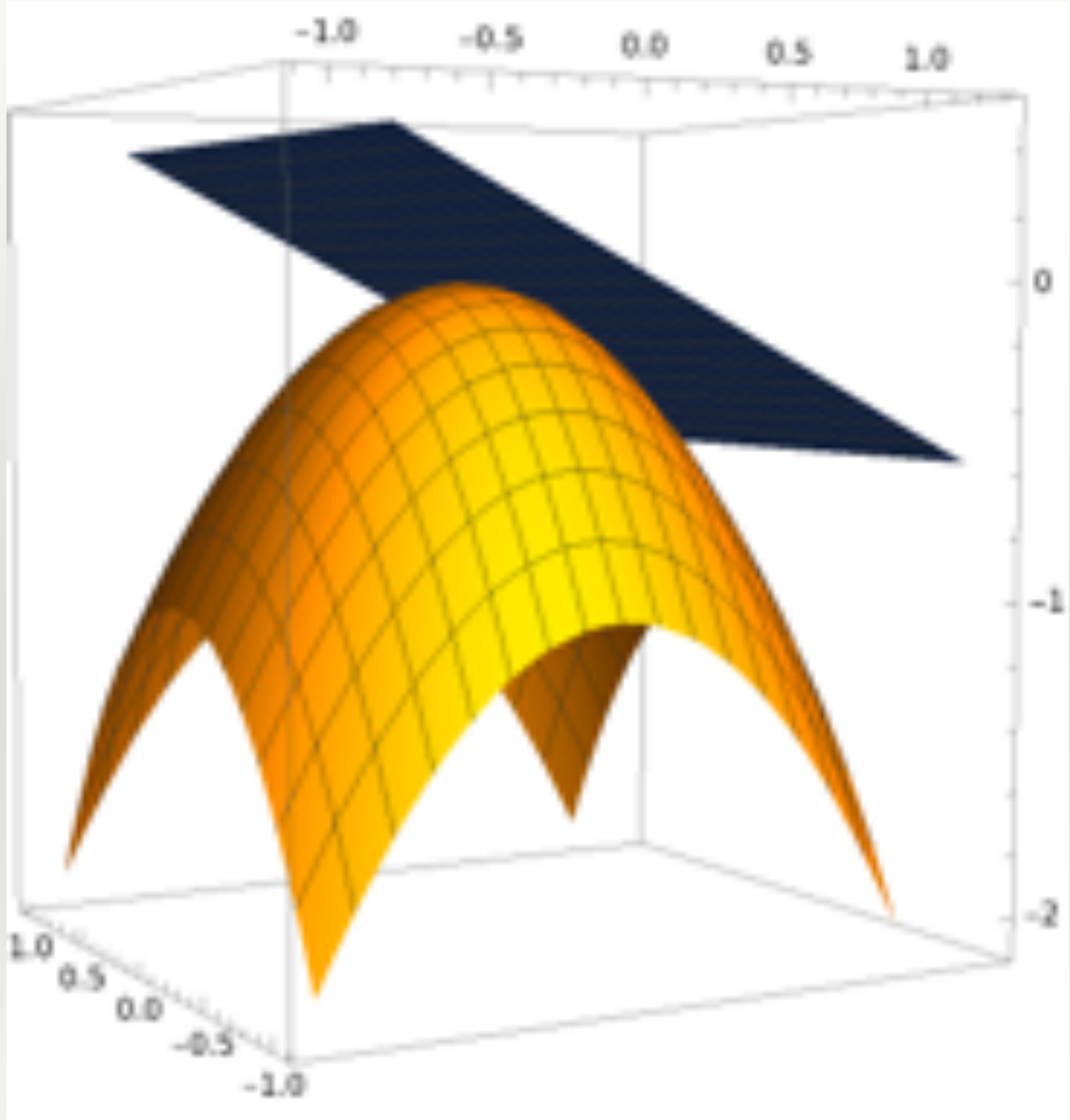
Geometry on a Surface

- $g_x(\mathbf{u}, \mathbf{v}) = \mathbf{X}'(\mathbf{x})\mathbf{u} \cdot \mathbf{X}'(\mathbf{x})\mathbf{v}$ gives a shape for M.
- Using linearity, $g_{ij} = \mathbf{X}'\partial_i \cdot \mathbf{X}'\partial_j = \mathbf{X}_i \cdot \mathbf{X}_j$
- Then, $g_x(\mathbf{u}, \mathbf{v}) = \sum_{ij} g_{ij} u^i v^j = \mathbf{u}^\top \mathbf{G}\mathbf{v}$



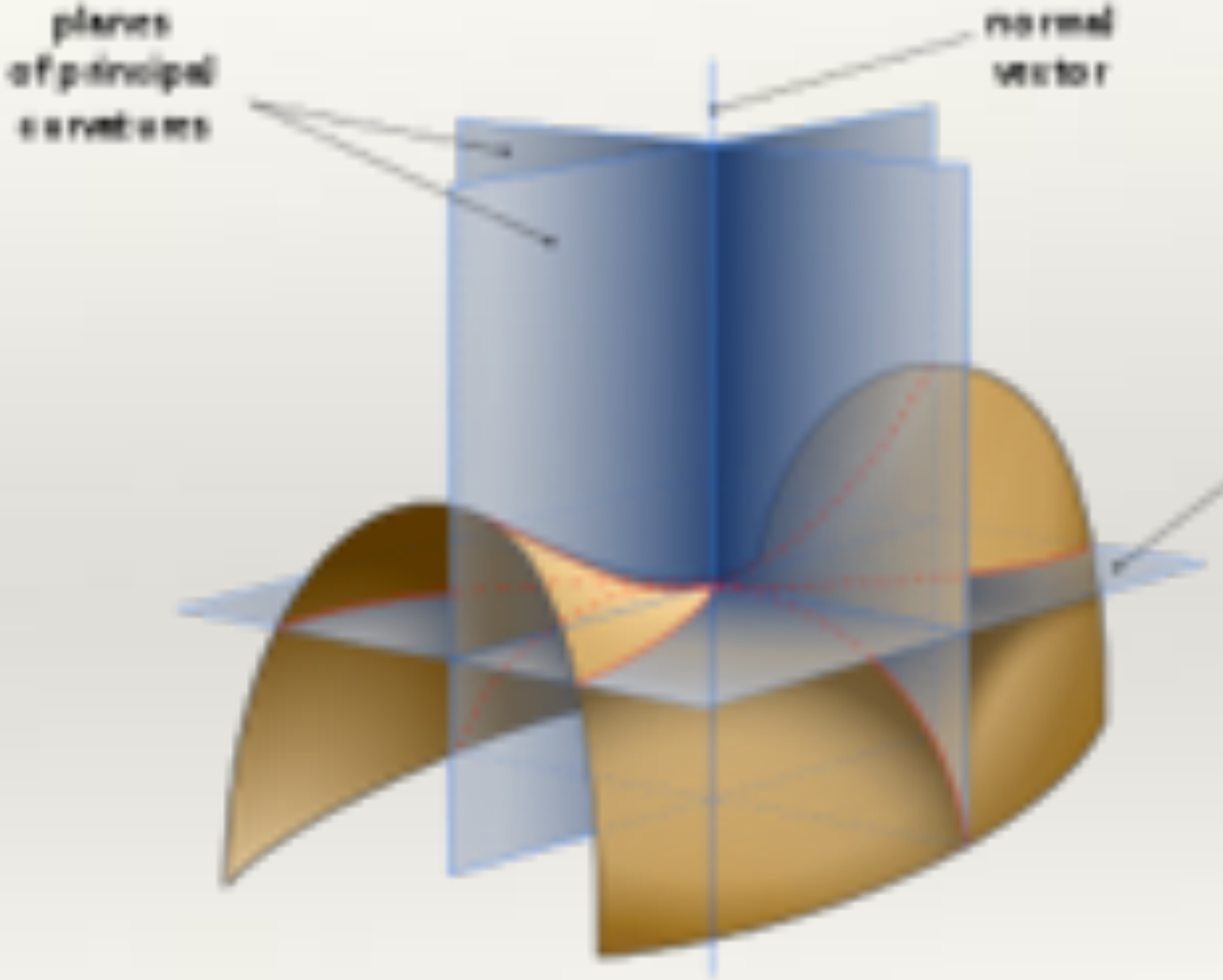
Example

- $\mathbf{X}(x_1, x_2) = \begin{pmatrix} x_1 & x_2 & -x_1^2 - x_2^2 \end{pmatrix}^\top.$
- $$\mathbf{X}'(x_1, x_2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -2x_1 & -2x_2 \end{pmatrix}$$
- $$\mathbf{G} = \begin{pmatrix} 1 + 4x_1^2 & 4x_1x_2 \\ 4x_1x_2 & 1 + 4x_2^2 \end{pmatrix}$$



What is the Mean Curvature?

- H is an extrinsic measure of how M bends in $\mathbb{R}^3$.
  - Depends on how $\mathbf{N}$ changes, i.e. $\mathbf{N}' : TM \to T\mathbb{S}^2$.
- Eigenvalues of $-\mathbf{N}'$ are the *principal curvatures* κ_1, κ_2.
  - $H = (1/2)(\kappa_1 + \kappa_2)$.



What is Mean Curvature Flow?

- Extend $\mathbf{X} = \mathbf{X}_0$ to a variation $\mathbf{X} : M \times \mathbb{R} \to \mathbb{R}^3$, $\mathbf{X} = \mathbf{X}_0 + t\boldsymbol{\varphi}$.
  - $\boldsymbol{\varphi} : M \to \mathbb{R}^3$ is the *velocity field* of the variation $\mathbf{X}$.
- Consider the area functional:

$$\mathscr{A}(\mathbf{X}) = \int_M 1 \, d\mu_g \, .$$

- How does $\mathscr{A}$ change as we change t ?

First Variation of Area

- We write $\delta\mathscr{A}(\mathbf{X}_0)\boldsymbol{\varphi} = \frac{d}{dt}\Big|_{t=0} \mathscr{A}(\mathbf{X}_0 + t\boldsymbol{\varphi}).$
- Since $d\mu_g = \sqrt{\det \mathbf{G}}\, dA,$ one can show

$$\delta\mathscr{A}(\mathbf{X})\boldsymbol{\varphi} = \int_M 2H\,\mathbf{N} \cdot \boldsymbol{\varphi}\, d\mu_g$$

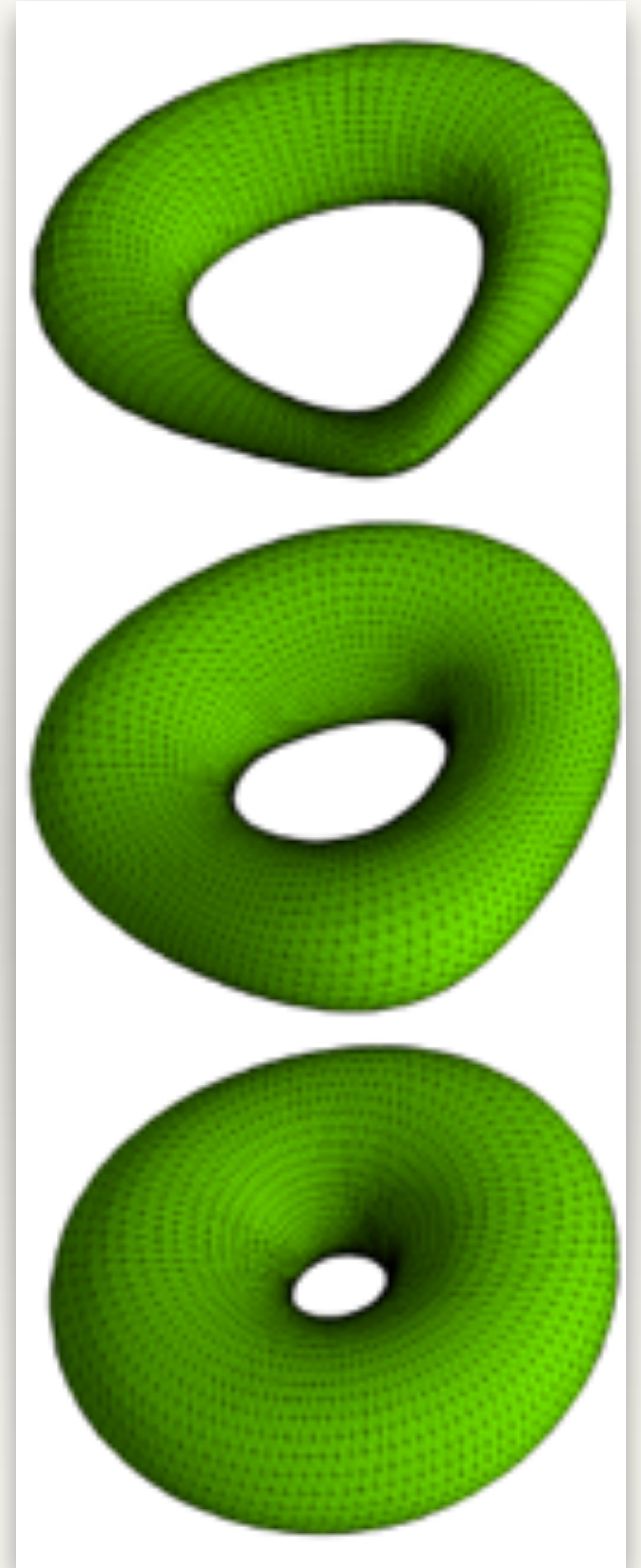
- Change in area is proportional to mean curvature!

Willmore Energy

- Another popular functional for graphics applications.
  - Willmore energy: $\mathscr{W}^2(\mathbf{X}) = \int_M H^2 \, d\mu_g$.
- Conformal invariant (hard for analysis)
- Qualitatively: $\mathscr{W}^2$ measures roundness.

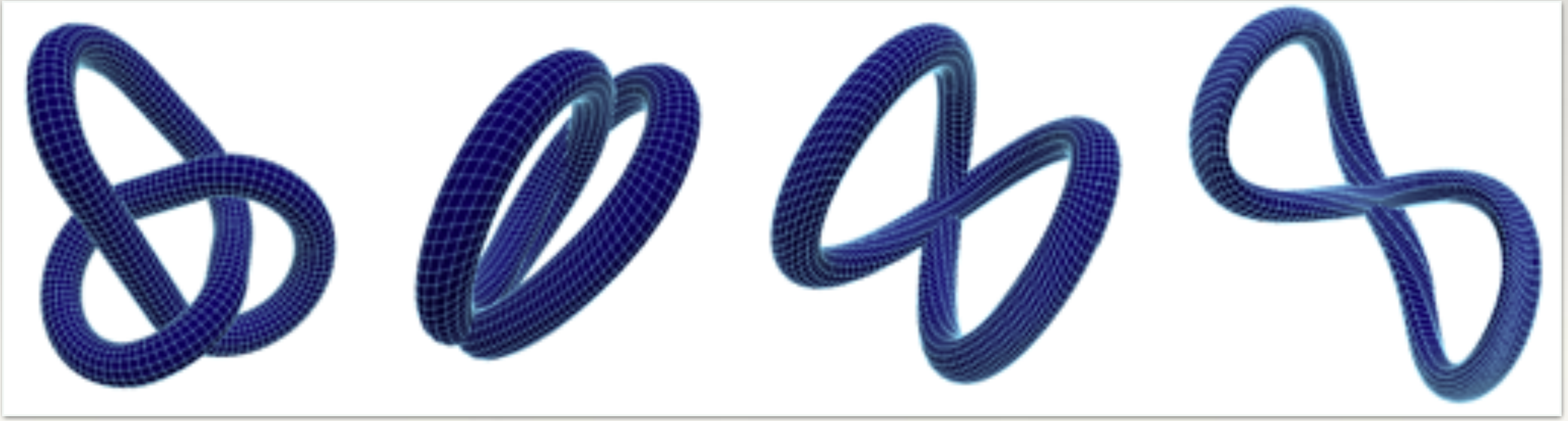
Willmore Flow

- Need to solve $\dot{\mathbf{X}} \cdot \mathbf{N} = \Delta_g H + 2H(H^2 - K)$
- Suppose M is closed. New variable $\mathbf{Y} := \Delta_g \mathbf{X}$ (G. Dziuk, 2012).
  - Weak definition $\int_M \mathbf{Y} \cdot \psi + \langle \mathbf{X}' \cdot \psi' \rangle_g = 0.$
- Willmore flow becomes coupled pair of 2^{nd}-order PDEs for X.
- (G., Aulisa) Extended ideas to p-Willmore energy.



Problems with Moving Domains

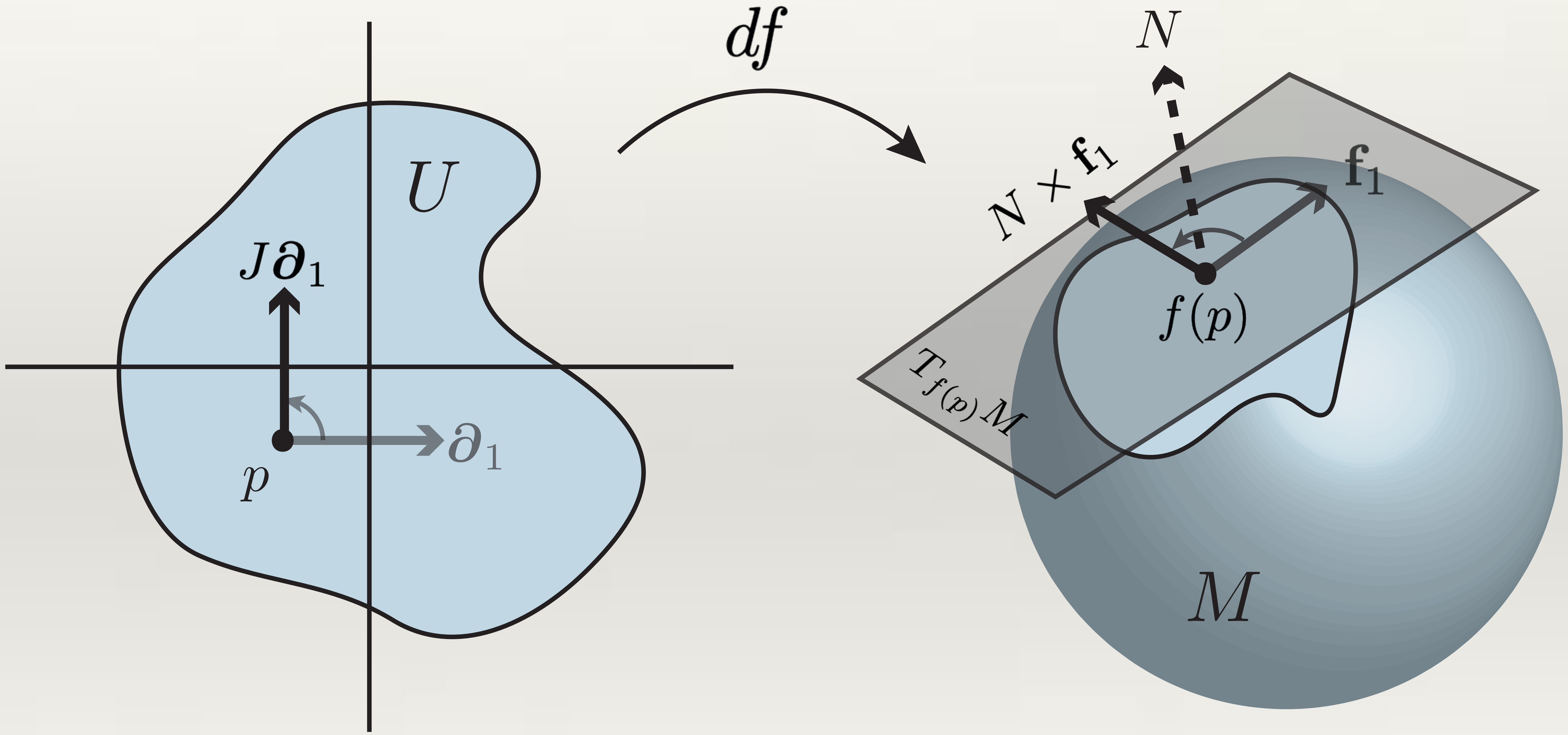
- Mesh can degenerate!
- Need some way to stop this…



What about the Mesh?

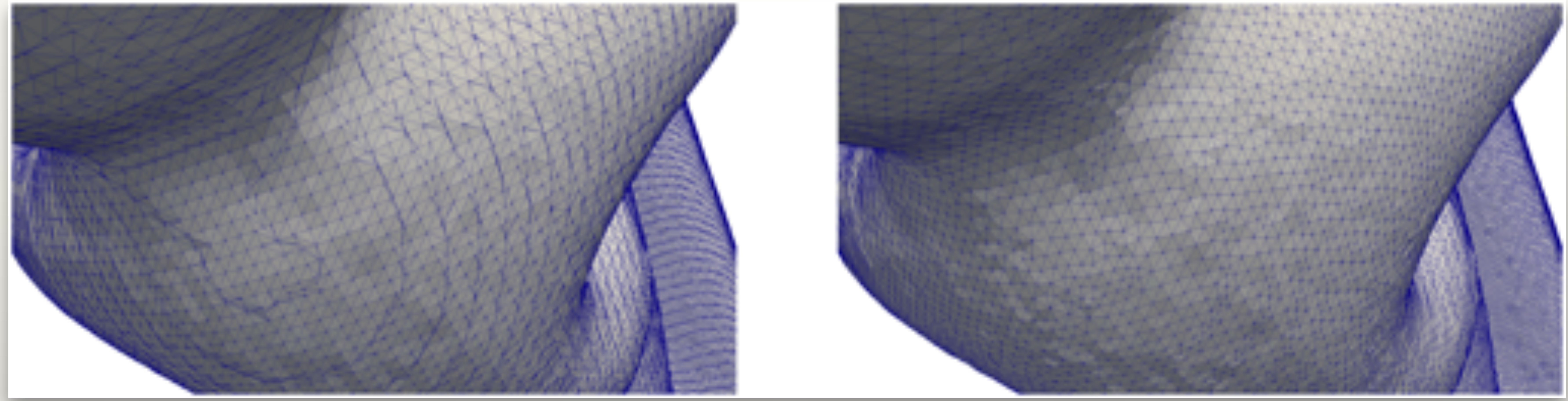
- Can consider least-squares conformal mapping.
- Map $f : (M, g) \to \mathbb{R}^3$ is *conformal* if it preserves angles.
- When $f : M \to \mathbb{R}^3$, equiv. to $\exists\, N$ s.t. $\star\, df = N \times df$

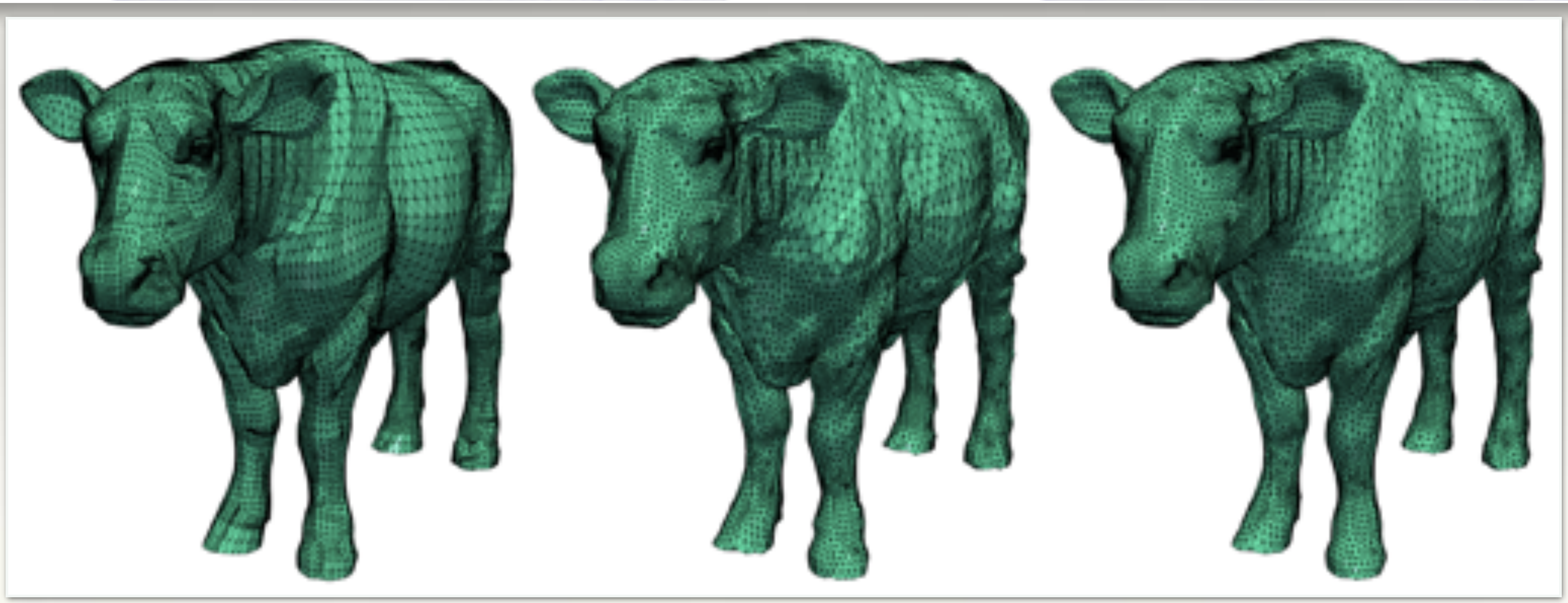
(Kamberov, Pedit, Pinkall 1996).



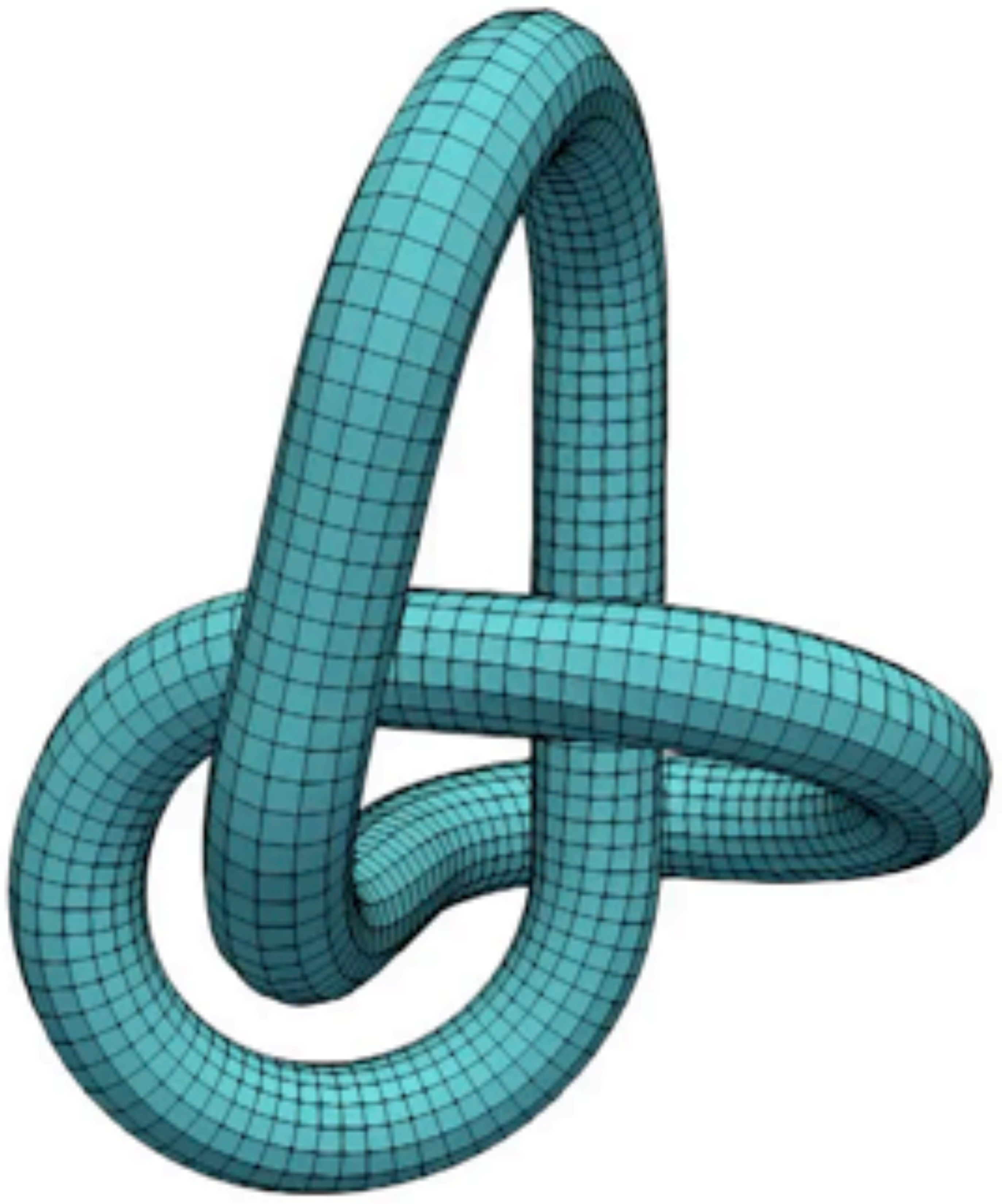
Results

- Can minimize integral of $| \star dX - N \times dX |^2$ with constraint.
- Yields least-squares conformal maps
- Makes triangulations much nicer.



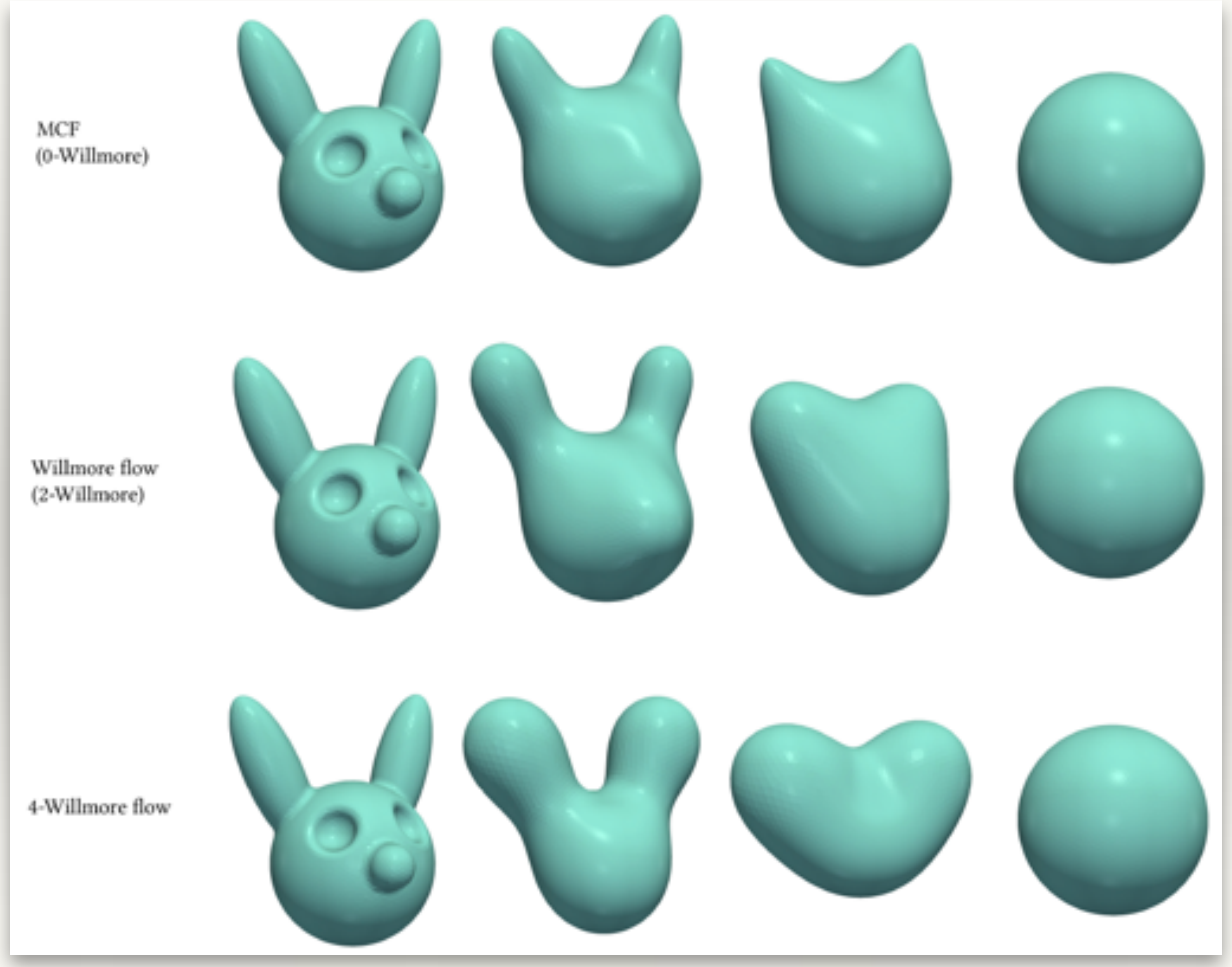


Trefoil Knot Unwinding



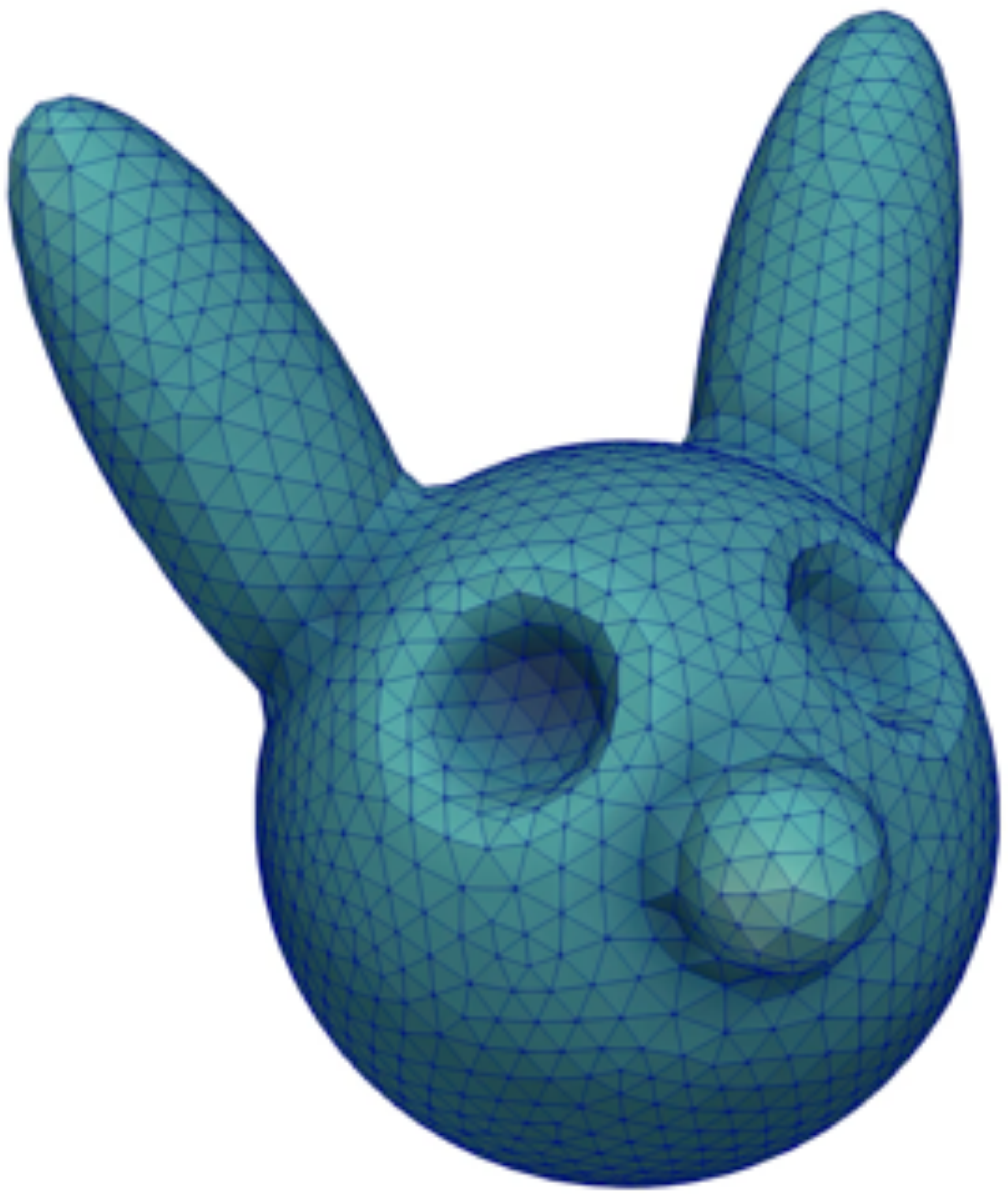
Comparison: p-Willmore

- Volume-constrained flows.
- Higher p — more rounding behavior.
- Can show that p > 2 Willmore minimizers resemble minimal surfaces in some aspects. *

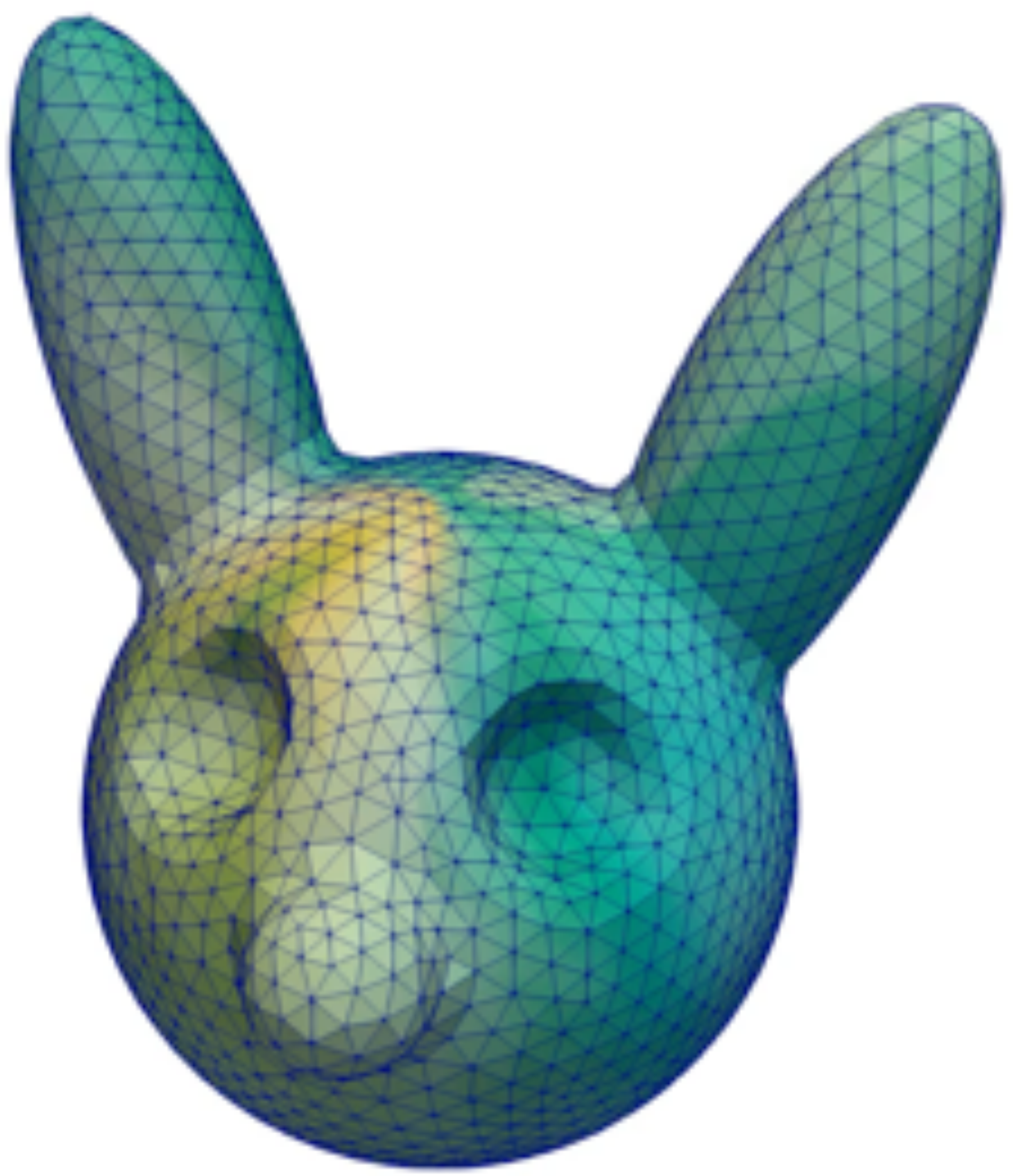


* Gruber, A., Toda M., Tran, H. *Ann. Glob. Anal. Geom.* (2019).
* Gruber, A., Pampano, A., Toda, M. *Ann. Mat. Pura. Appl.* (2021).

Influence of the Constraint



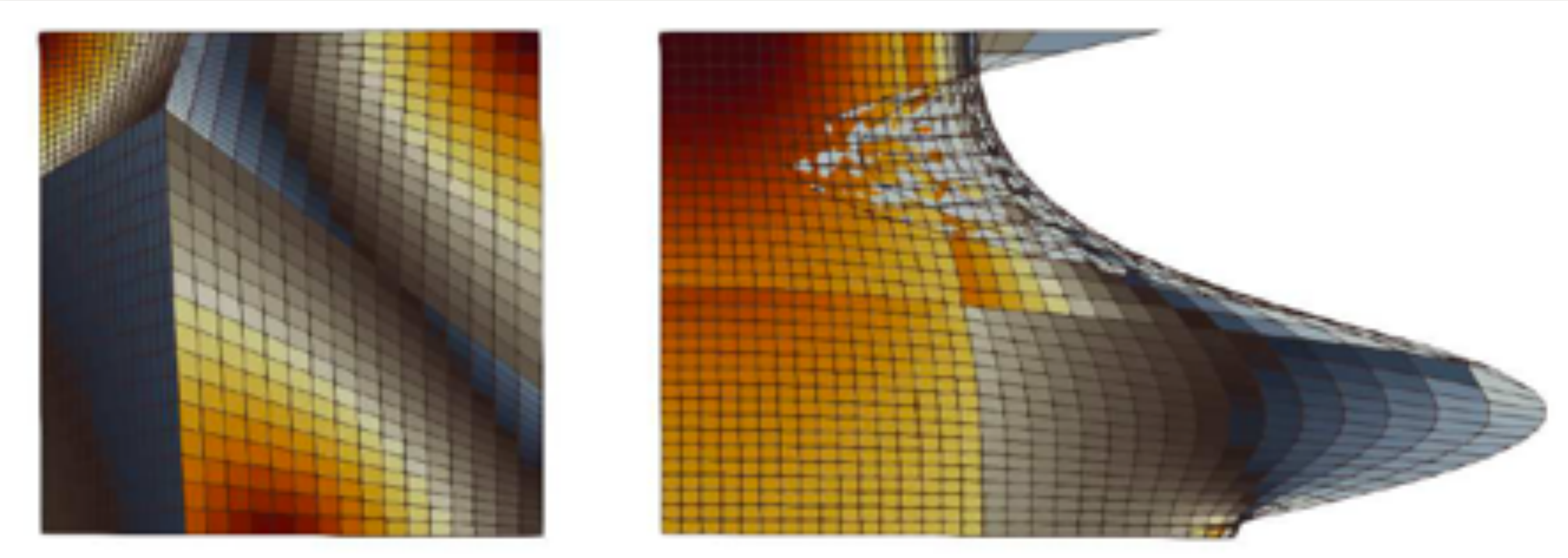
Volume preserving 2-Willmore flow



Volume and area preserving 2-Willmore flow

A Problem with LSCM?

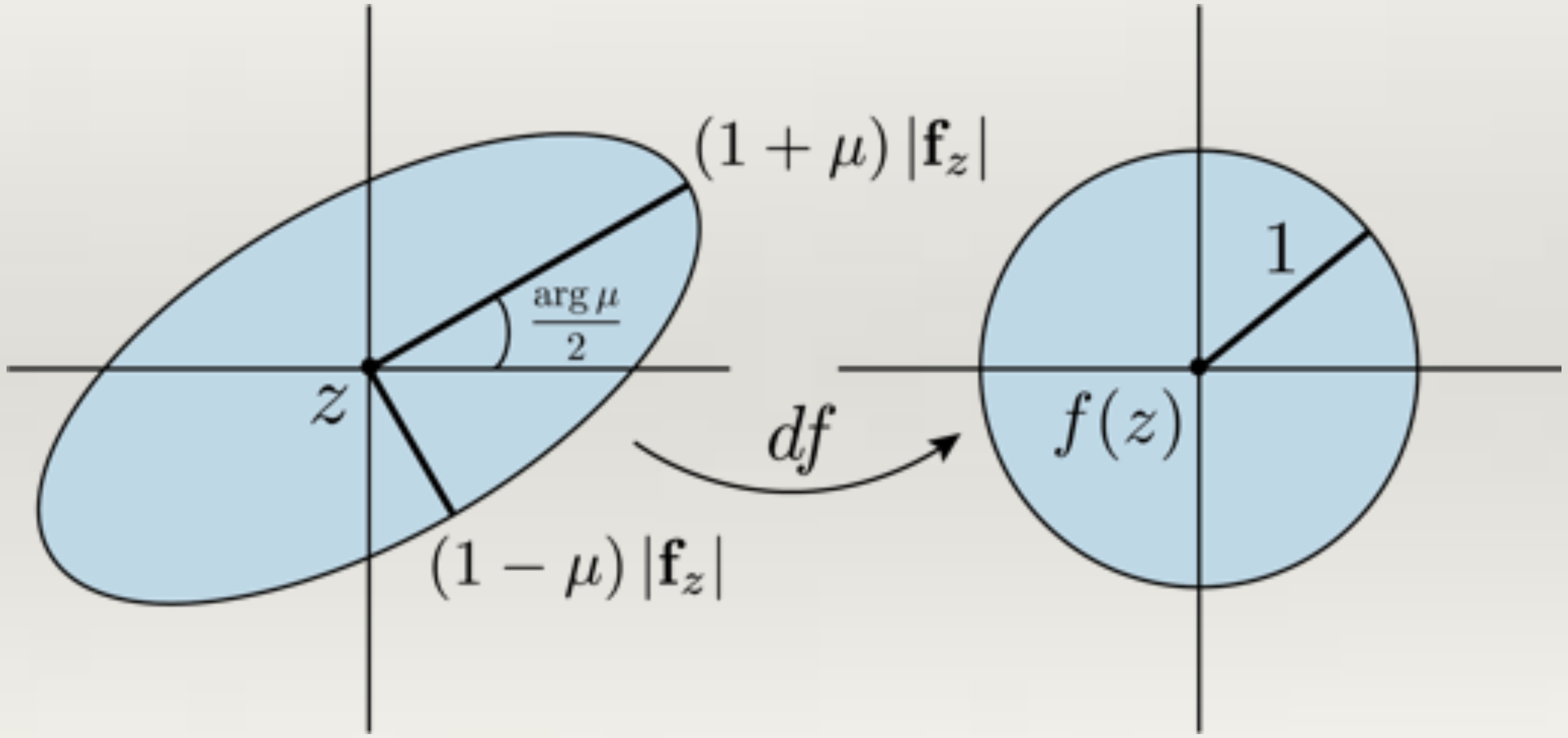
- Not so good for mappings with boundary correspondence.
- Conformal mappings are too restrictive for this.



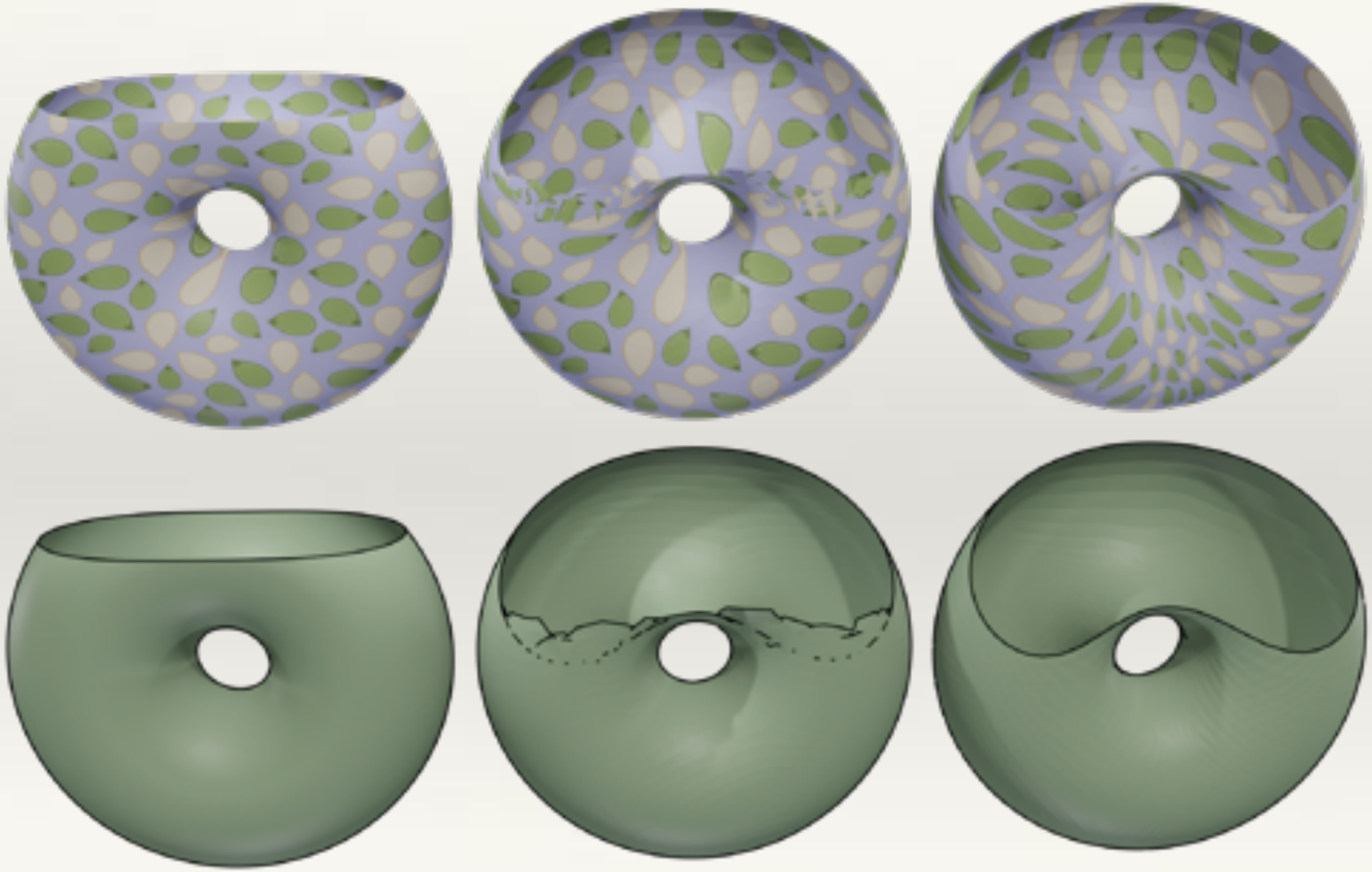
Conformal vs. Quasiconformal

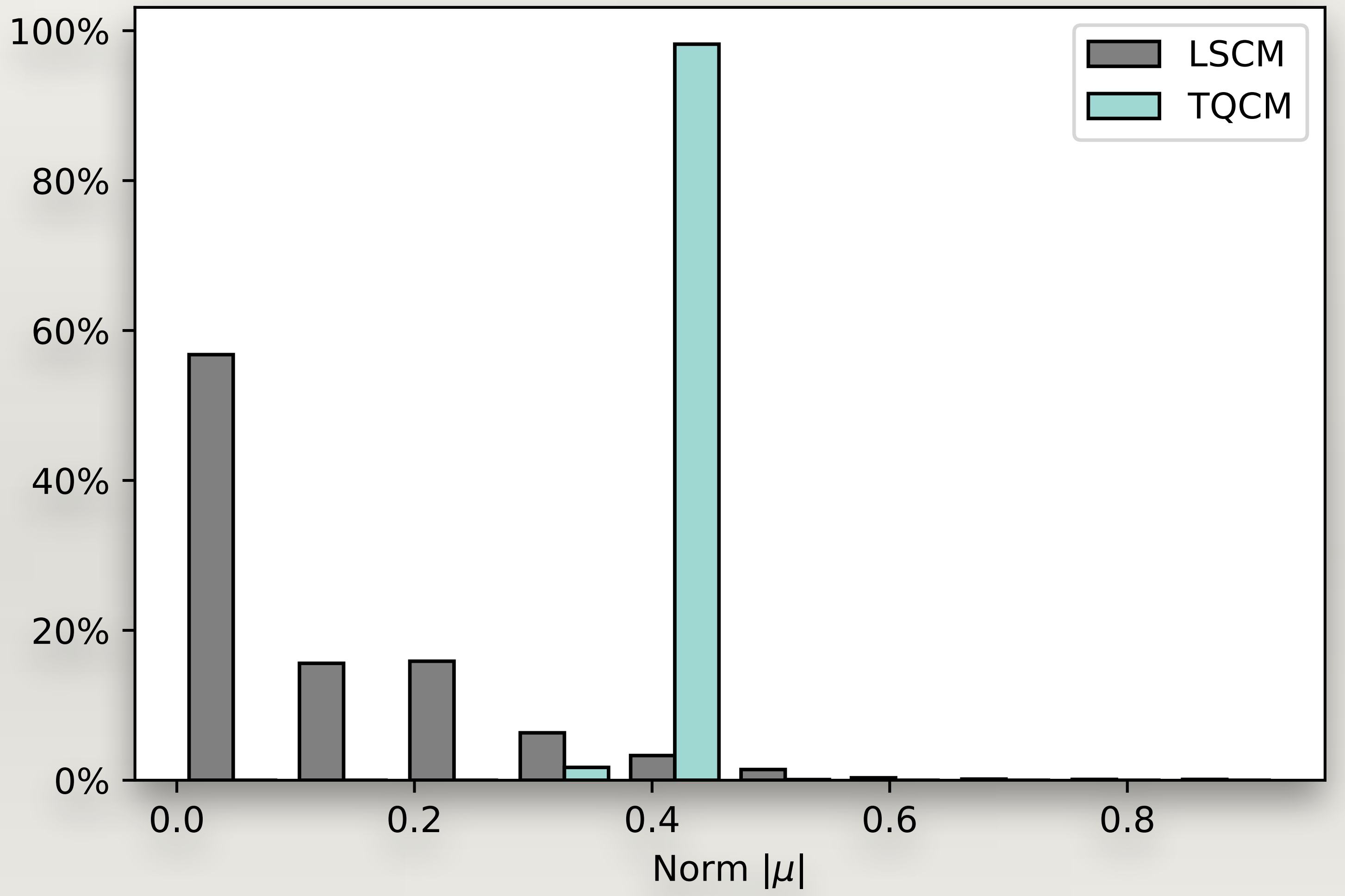
- Must allow bounded shearing distortion.
- In *quaternionic* setting, this means:
  - $df^- = \mu\, df^+$
    - (anticonformal/conformal parts).
  - Conformal iff Beltrami coefficient $\mu = 0$.
  - μ is conjugate-dual to the *Hopf differential* $Q = df^+\overline{df^-}$.

$$df^{\pm} = \frac{1}{2}\left(df \mp N \star df\right)$$



Comparison: TM vs. LSCM

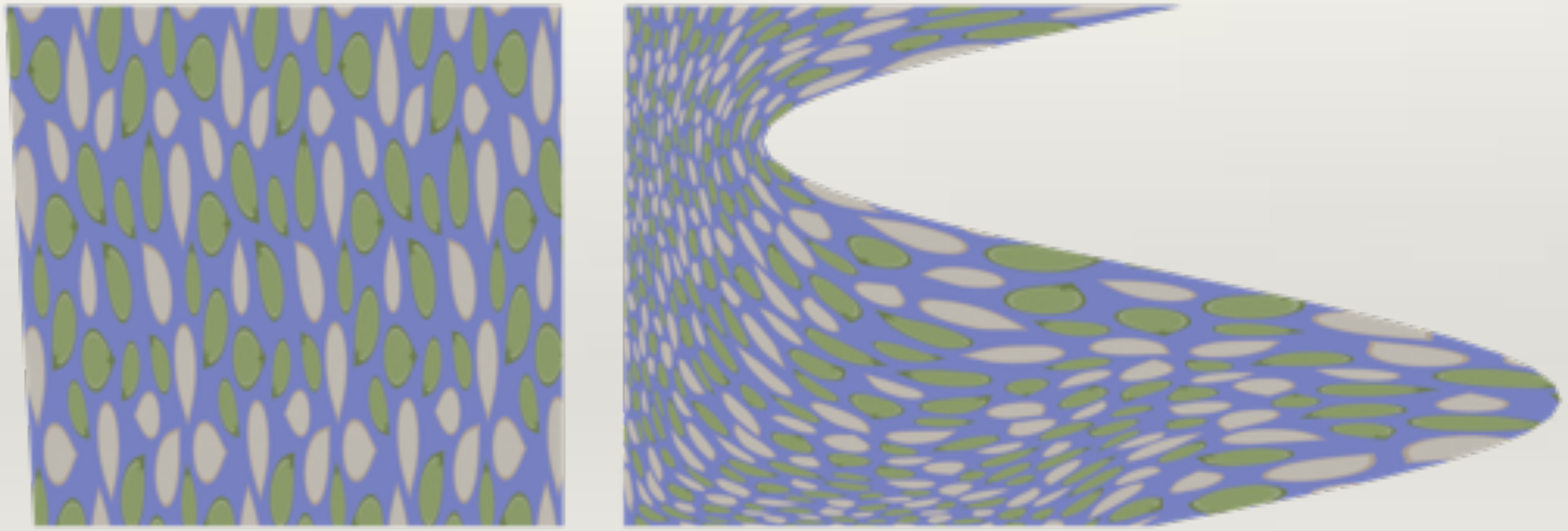




- Gruber, A. and Aulisa, E. 2021 (under review)

Computing TM mappings

- Minimize
  $$\mathscr{QC}(f) = \int_M |df^- - \mu\, df^+|^2 \, d\mu_g$$
  alternatively over f, μ.
  - 1) Minimize for f given μ.
  - 2) Compute $\mu = df^- \left(df^+\right)^{-1}$.
  - 3) Locally adjust μ, moving it toward TM form (next slide)
  - Repeat steps 1-3 until convergence.



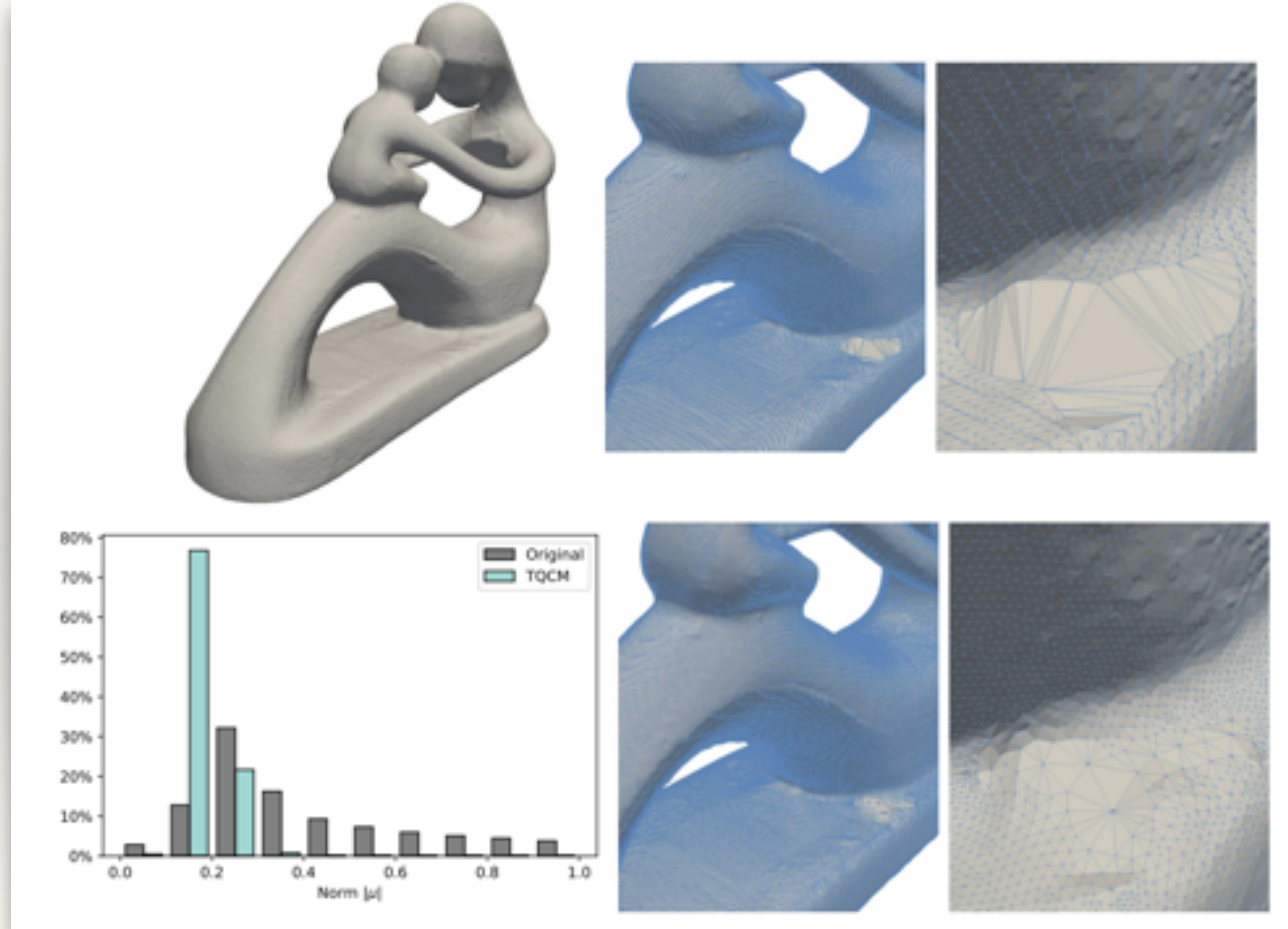
Why does it work?

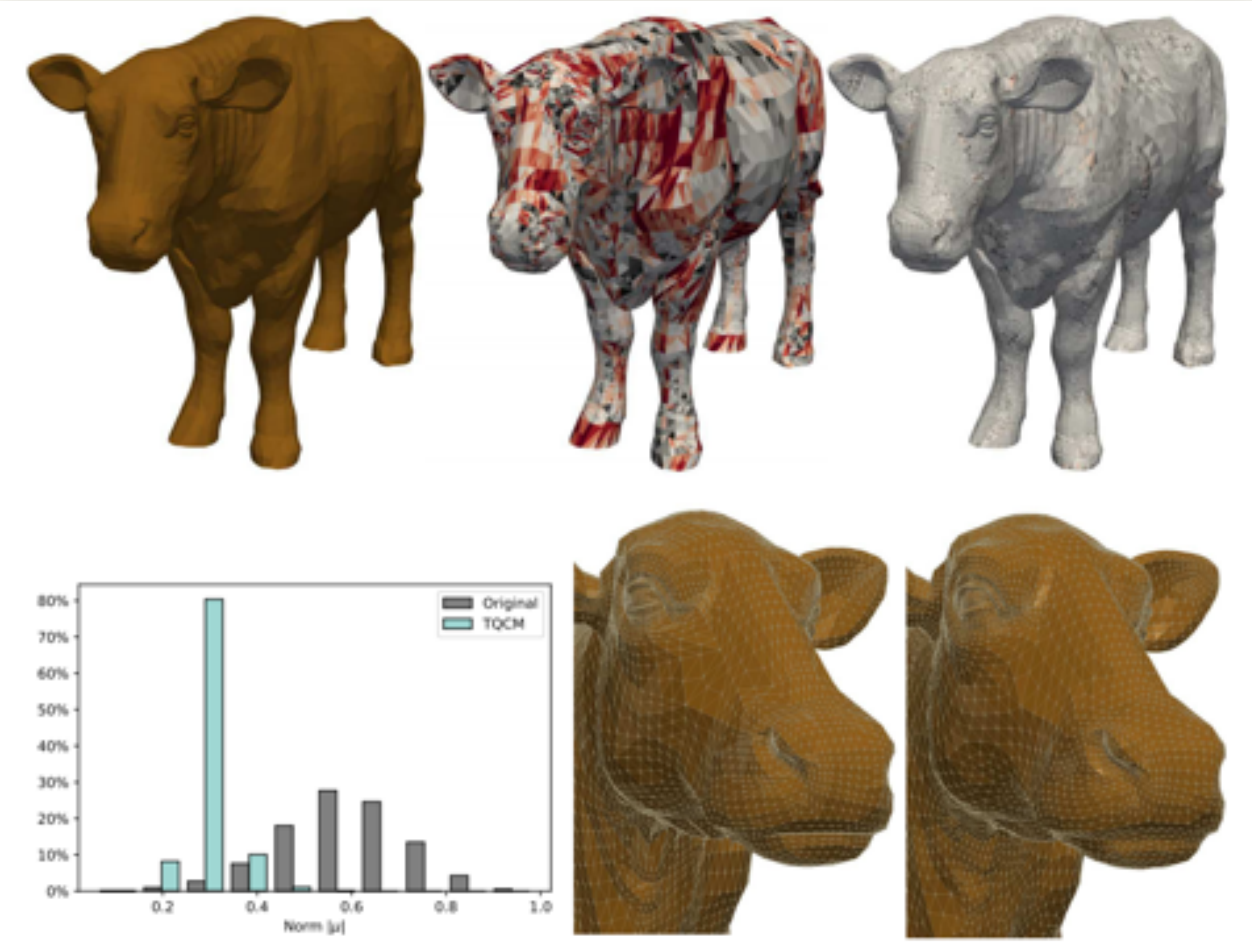
- Dirichlet energy w.r.t. any metric decomposes into *area functional* and *conformal distortion*.

$$\mathscr{D}_g(f) = \int_M |df|^2 = \int_M |df^+|^2 - |df^-|^2 + 2\int_M |df^-|^2 = \mathscr{A}(f) + 2\mathscr{C}\mathscr{D}(f)$$

- Can show that $\mathscr{Q}\mathscr{C}(f)$ is the conformal part of $\mathscr{D}_{g(\mu)}(f)$.
- (Metric $g(\mu)$ induced by QC mapping with BC μ.)

More Examples





Conclusions

- Many popular technologies rely on continuous change.
- Quantitatively, change must be prescribed.
- Often done through minimization of an appropriate function.

- *If you want to be a scientist… Pay attention in calculus class!*

Thank You!