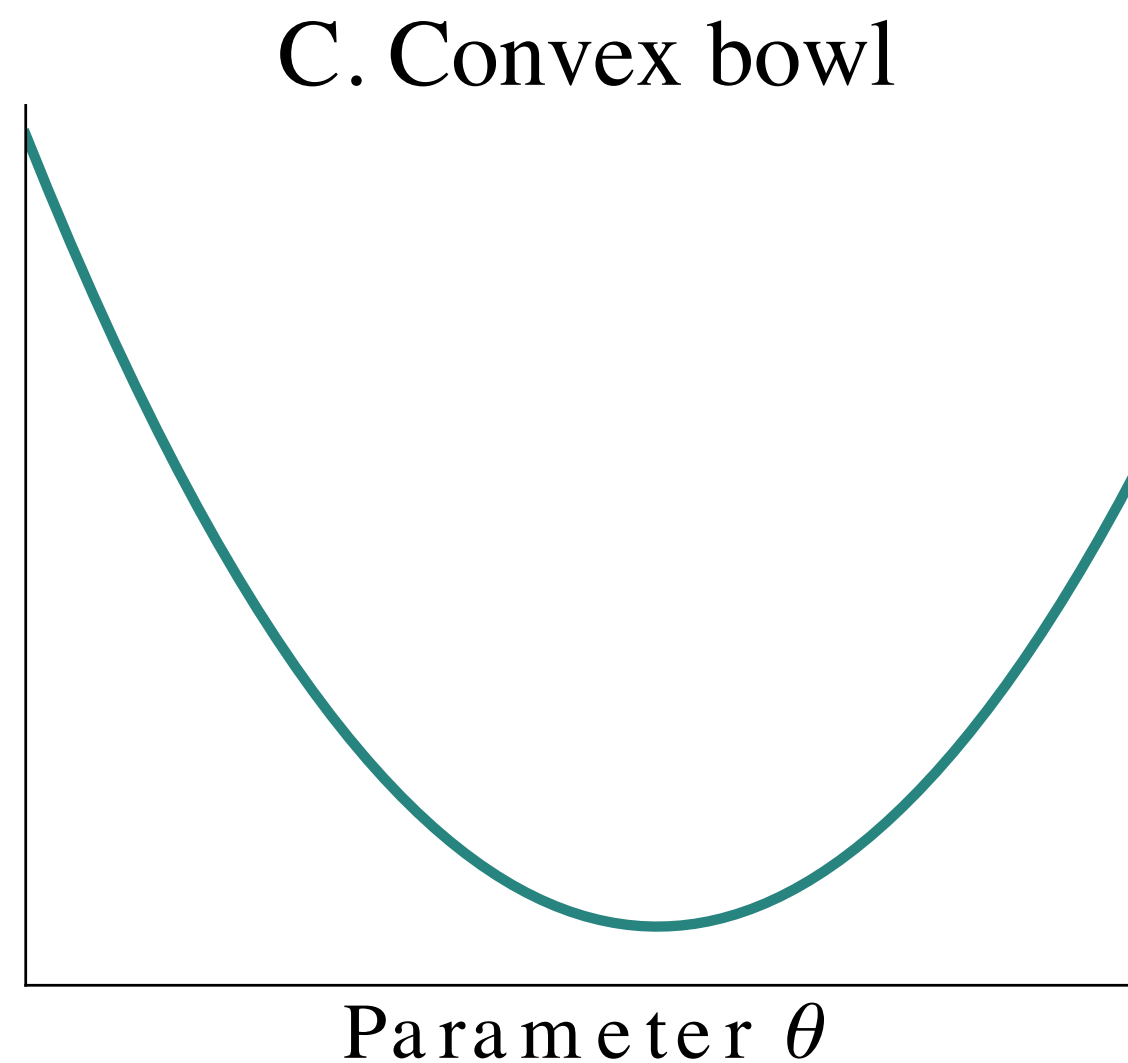
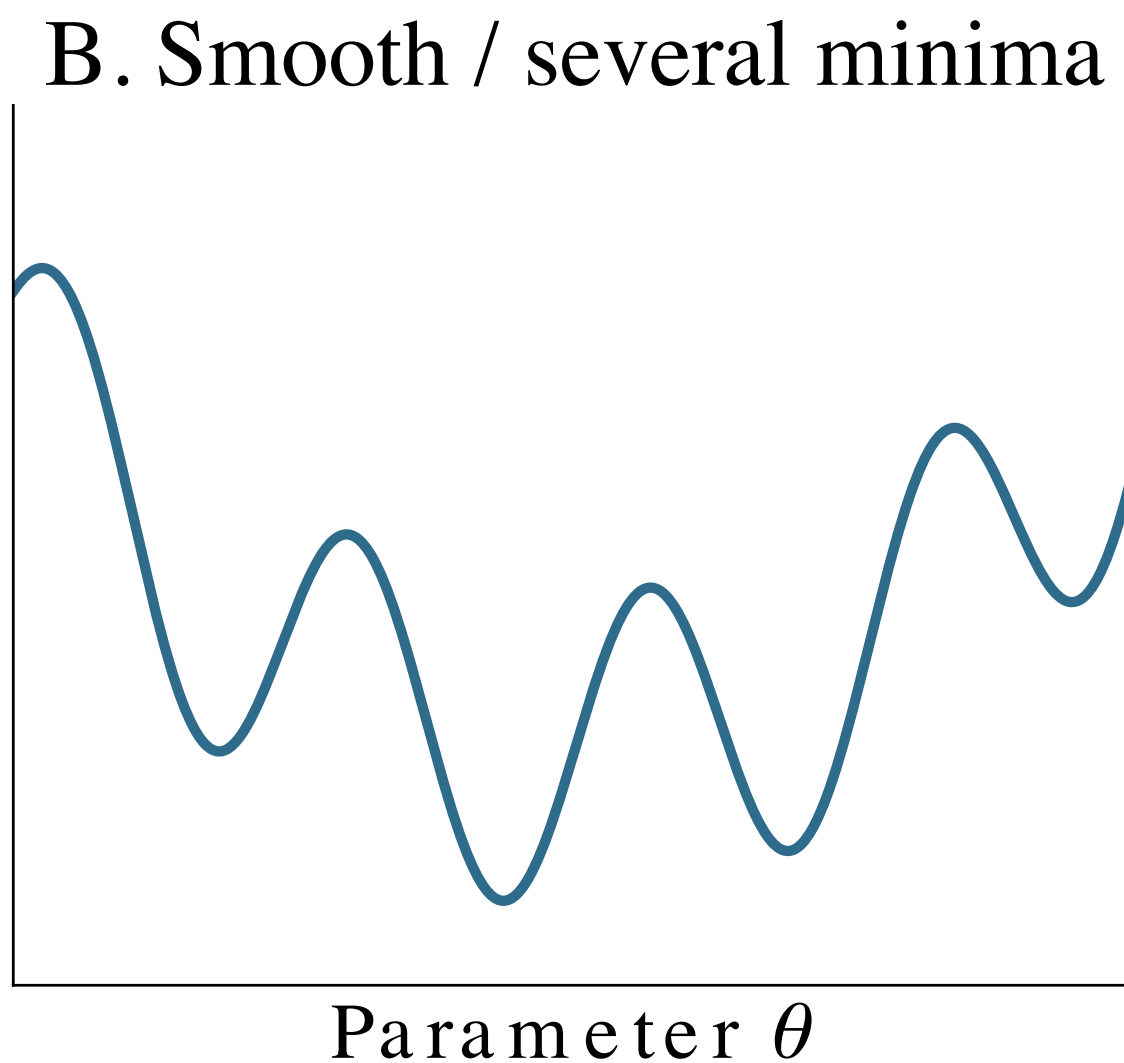
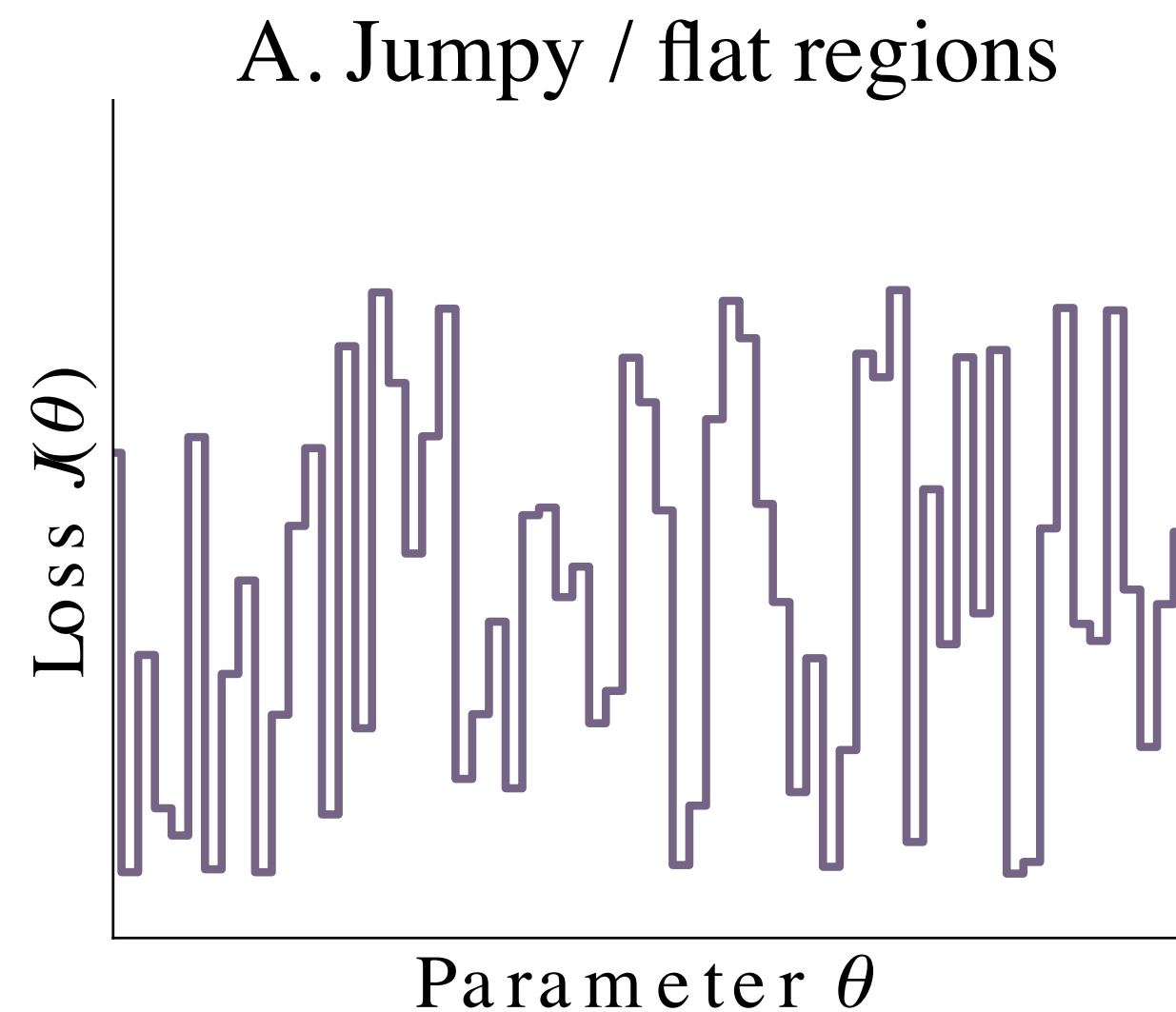

Why the Shape of the Objective Matters

David I. Inouye

| Which loss would be easiest to optimize?



Which shape gives us the clearest route to a global minimum?

A. The jumpy loss.

B. The smooth loss with several minima.

C. The convex bowl.

C. A local minimum is already global; there are no worse local minima to escape.

| A useful objective lets us **reason about the solution**

- **Jumpy:** a small change may reveal little—or abruptly change the loss.
- **Smooth, several minima:** downhill progress can end at a worse local minimum.
- **Convex:** a local minimum is global, making optimality easier to analyze.

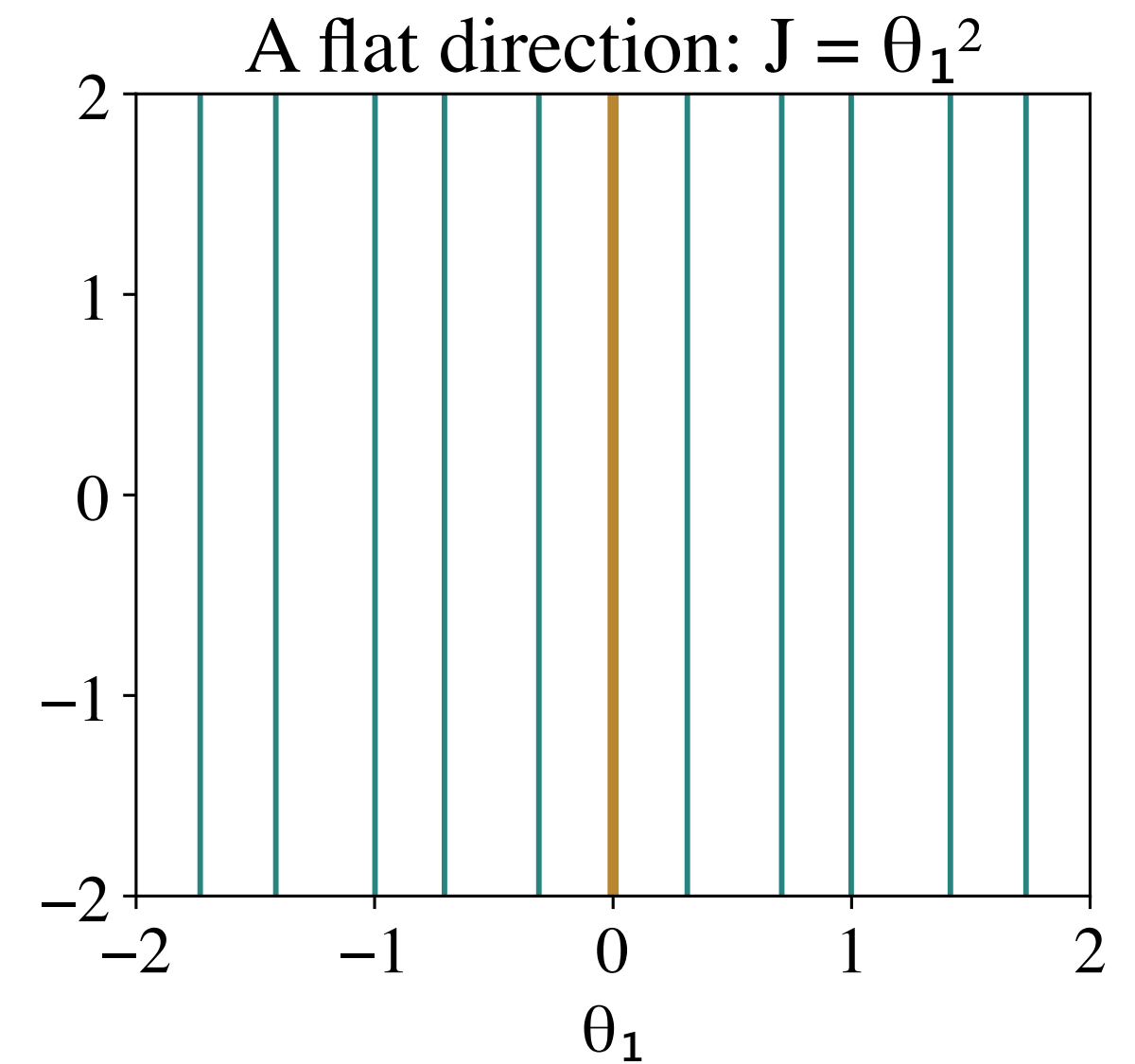
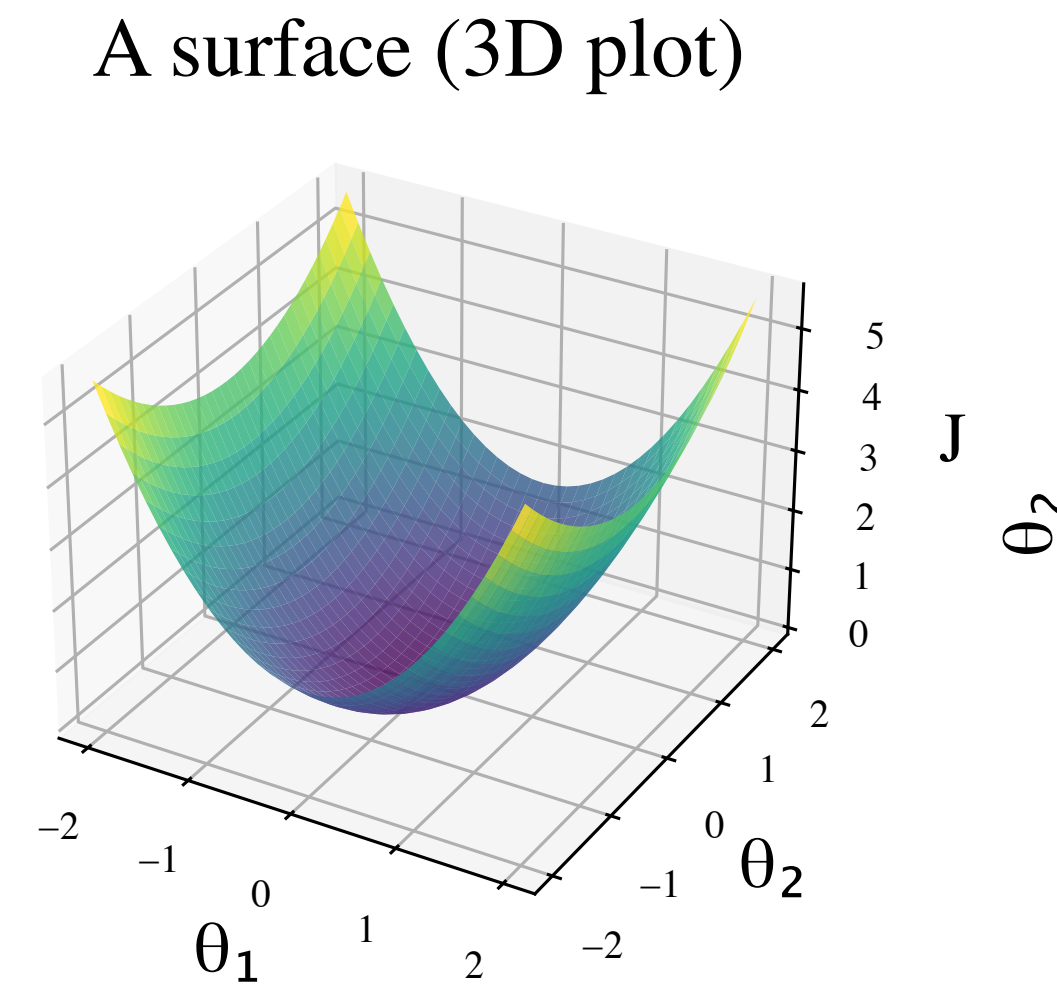
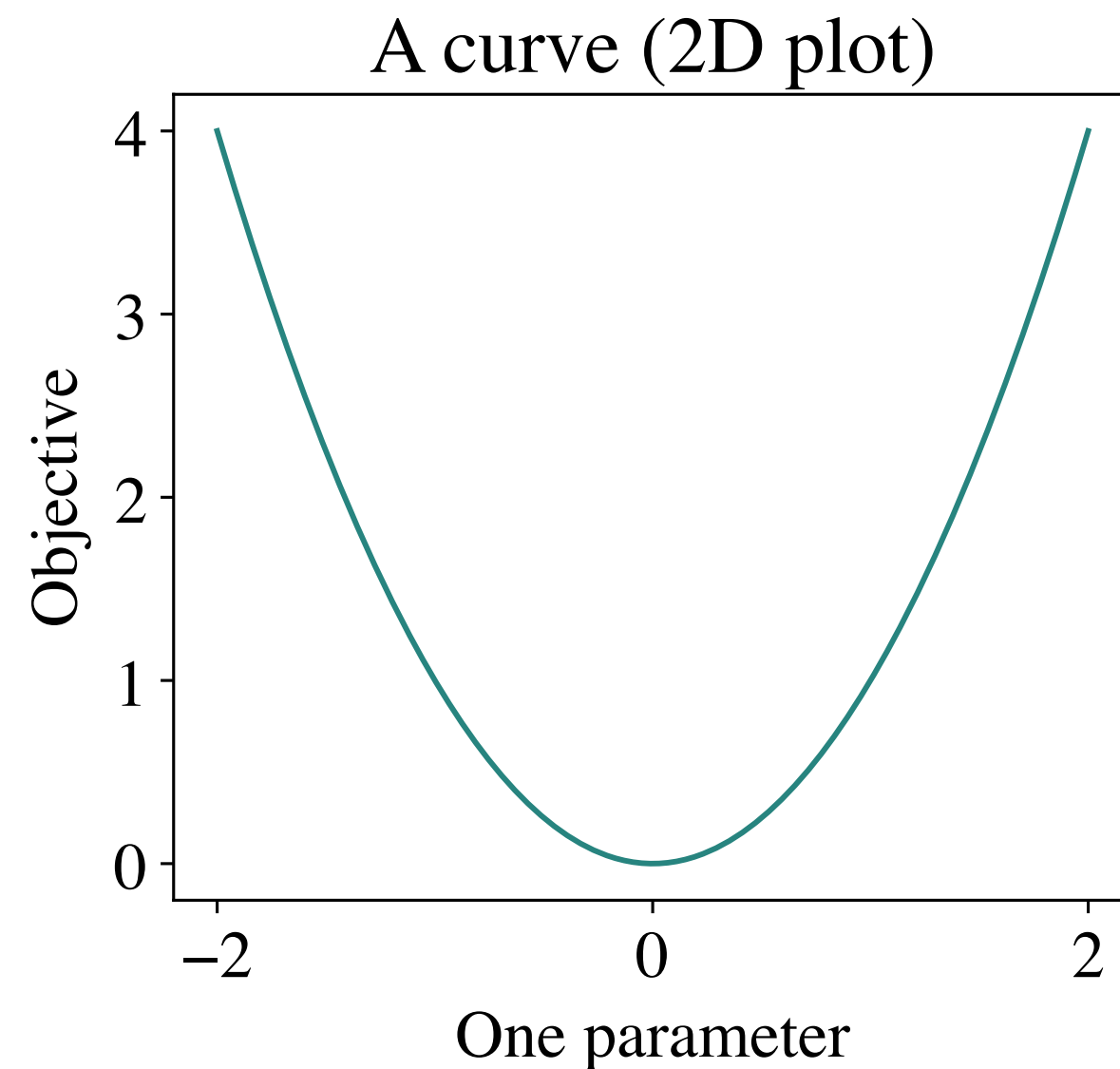
Convexity is a useful class of well-behaved objectives.

It gives a global guarantee; speed still depends on the problem and algorithm.

| Objective shape answers **three linked questions**

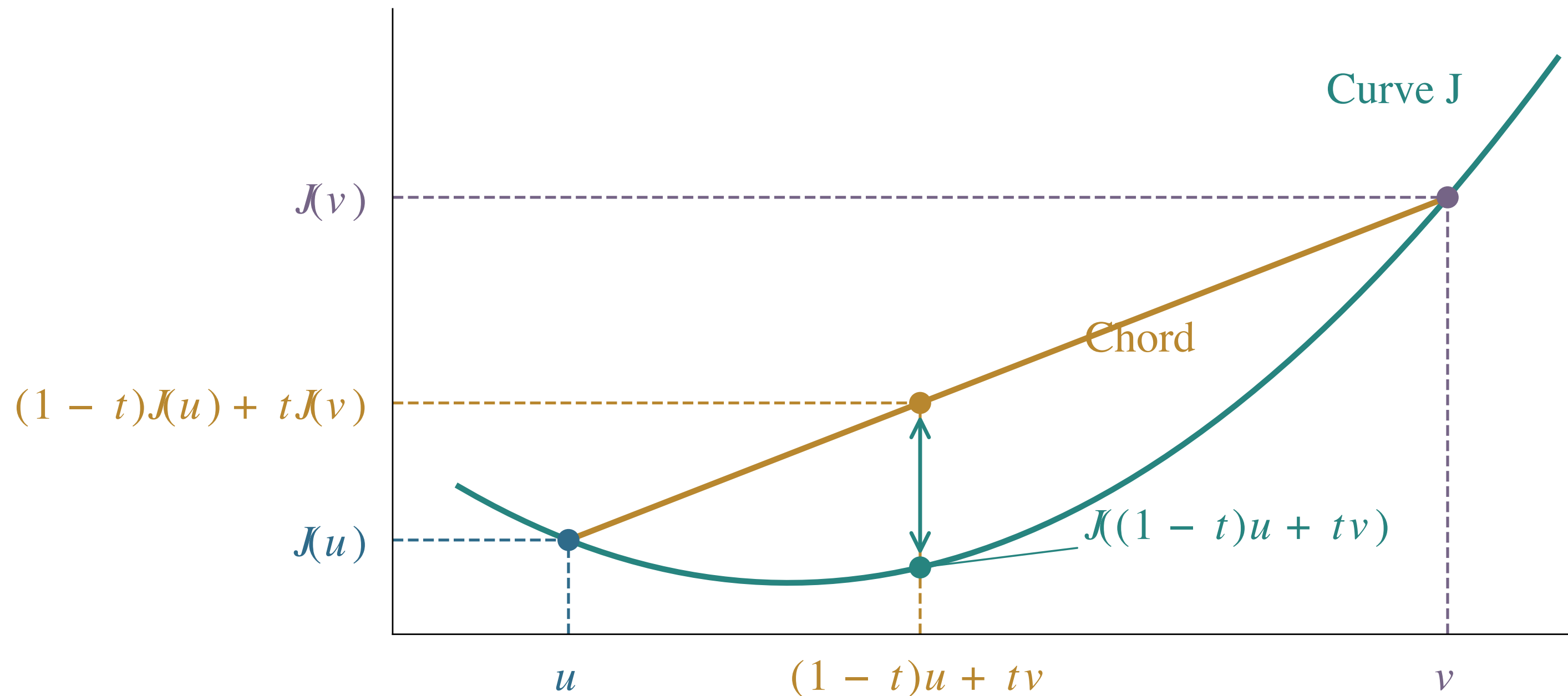
1. **Define:** what does convexity mean geometrically?
2. **Check:** are our OLS and logistic objectives convex?
3. **Interpret:** what does that guarantee—and what else do we need?

| Convexity is a statement about the entire objective shape



A convex objective may have a single bottom—or a flat set of equally good solutions.

| Definition: a convex function lies below every chord



J is **convex** on a convex domain if, for **every** u, v in the domain and $0 \leq t \leq 1$,

$$J\left((1-t)u + tv\right) \leq (1-t)J(u) + tJ(v).$$

Loss at the averaged input $\leq$ average of the endpoint losses.

| Convexity makes **every local minimum global**

Suppose u is a local minimum, but some v has lower loss.

$$J((1 - t)u + tv) \leq (1 - t)J(u) + tJ(v).$$

Since $J(v) < J(u)$, the right side is **less than** $J(u)$ for every $t > 0$.

Choosing an arbitrarily small t gives a nearby point with lower loss—a contradiction.

Therefore, a local minimum must be global.

For a twice differentiable function, **convexity means nonnegative curvature**

Let J be twice continuously differentiable on an **open convex domain**.

Parameters	Equivalent criterion for convexity
One scalar θ	$J''(\theta) \geq 0$ everywhere
A vector θ	$H(\theta) = \nabla^2 J(\theta) \succeq 0$ everywhere

The **Hessian** is the matrix of second partial derivatives: $H_{jk} = \frac{\partial^2 J}{\partial \theta_j \partial \theta_k}$.

Positive semidefinite (PSD) means $v^\top H v \geq 0$ for every direction v . Equivalently, **all eigenvalues of H are ≥ 0** : no eigenvector direction curves downward; zero allows a flat direction at that point.

Derivation hints: in one dimension, $J'' \geq 0$ makes slopes nondecreasing, giving the chord inequality. In several dimensions, restrict to a line:

$$g(t) = J(\theta + tv), \quad g''(t) = v^\top H(\theta + tv)v.$$

| OLS and logistic log loss have **nonnegative directional curvature**

Append the intercept to $\widetilde{X}$ and the coefficients to θ .

Objective	Hessian $H = \nabla^2 J$
OLS	$\frac{1}{n} \widetilde{X}^\top \widetilde{X}$
Logistic log loss	$\frac{1}{n} \widetilde{X}^\top W \widetilde{X}$, with $W_{ii} = a_i(1 - a_i) \geq 0$

For logistic loss, $\ell(z, y) = \log(1 + e^z) - yz$ gives $\ell' = \sigma(z) - y$ and $\ell'' = a(1 - a)$.

$$\boldsymbol{v}^\top H \boldsymbol{v} = \frac{1}{n} \sum_i a_i(1 - a_i) (\widetilde{\boldsymbol{x}}_i^\top \boldsymbol{v})^2 \geq 0.$$

For OLS, replace the weights by 1. **Both objectives are convex in the coefficients.**

| Convexity makes a local optimum **globally meaningful**

- **Local minimum** → **global minimum**. Zero gradient also certifies a global minimum for a differentiable convex objective.
- **Uniqueness needs more**: strict convexity excludes multiple minimizers, if a minimizer exists.
- **Speed needs more**: smoothness, learning rate, and conditioning govern convergence.

For an L -smooth convex objective with a minimizer, gradient descent with step $1/L$ has an $O(1/t)$ objective-gap bound. Strong convexity can give geometric convergence.

Next: a useful shape does not itself tell us how to compute the solution.