

# The 19th Problem

## A Leisurely Excursion

Have you ever wondered why soap bubbles are spherical, or why soap films look like they are pulling themselves into the shape of the least possible area, or why the light moving through glass of varying density bends and curves to find the fastest route, not the straightest one.

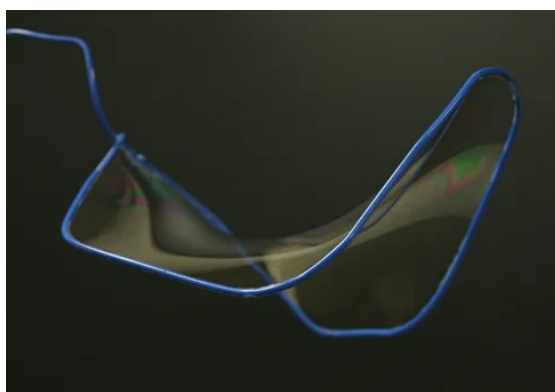


Figure 1: A soap film on a wire frame.

In each case, nature is quietly solving an optimization problem. Something is being minimized, whether it is the area or the time. One thing to note is that even if the wire frame itself (the boundary) has sharp, jagged corners, the soap film suspended inside is perfectly smooth. It is definitely interesting to ponder on why the film develops no gaps, tears, or defects.

*Why should this be true? And more importantly: must it always be true?*

Before answering this, let's rewind to the dawn of the 20<sup>th</sup> century. In the International Congress of Mathematicians held at Paris in August 1900, the German Mathematician David Hilbert proposed 23 problems, most of which would prove to be very influential in the coming century.<sup>1</sup>

While many of his problems arose due to the urge to understand the foundational bedrock of Mathematics like the continuum hypothesis, a few were inspired by physics.

---

<sup>1</sup>He actually had a list of 24 problems but he chose to publish only 23.



Figure 2: David Hilbert (1862–1943)

Belonging to the latter category is the topic of our blog, which is the 19<sup>th</sup> problem. Hilbert was fascinated by the observation that physical principles, usually cast as variational problems (like the principle of least action), seemed to produce solutions of remarkable smoothness.

These are but special cases of a universal variational principle. Hilbert observed that such analytic solutions arise from minimizing functional integrals of the form:

$$J(u) = \int_{B_1} F(\nabla u) dx, \quad (1)$$

where the function  $F$  (often called the Lagrangian) is analytic and *strictly convex*. In the language of linear algebra, convexity here means the Hessian matrix of  $F$  is positive definite (specifically,  $\det(D^2F) > 0$ ).

*The question is this: If the energy function  $F$  is analytic, does it follow that the minimizing configuration  $u$  is also analytic?*

Intuitively, the answer seemed to be yes. To prove it, we translate the minimization problem into an equation. We know from single-variable calculus that if  $f(x)$  has a

minimum at  $x_0$ , then  $f'(x_0) = 0$ . We apply the same logic here.

Suppose  $u(x)$  is the true solution. Imagine we distort this solution by adding a small “perturbation”  $\phi(x)$ , scaled by a tiny number  $\epsilon$ :

$$u_\epsilon(x) = u(x) + \epsilon\phi(x).$$

Since  $u$  is a minimizer, the derivative of the energy with respect to  $\epsilon$  must be zero at  $\epsilon = 0$ :

$$\left. \frac{d}{d\epsilon} J(u + \epsilon\phi) \right|_{\epsilon=0} = 0.$$

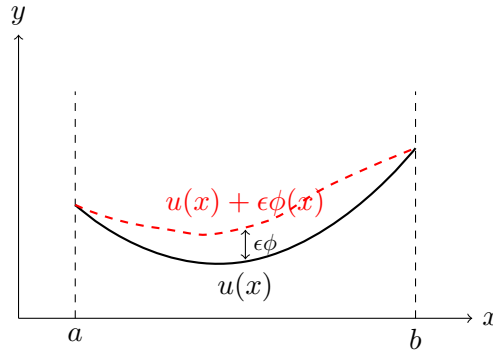


Figure 3

Calculating this (using the Chain Rule inside the integral) and integrating by parts gives the celebrated Euler-Lagrange equation:

$$\text{Div}(\nabla F(\nabla u)) = 0. \quad (2)$$

Now, we expand this. Applying the Chain Rule to the term  $\nabla F(\nabla u)$ , we get:

$$\sum_{i,j=1}^n F_{ij}(\nabla u) u_{ij} = 0, \quad (3)$$

where  $F_{ij}$  are entries of the Hessian of  $F$ , and  $u_{ij}$  are the second partial derivatives of  $u$ .

Here comes the trouble. If we differentiate Eq. (3) with respect to  $x_k$ , letting  $v = u_{x_k}$ , we obtain a linear elliptic equation for the partial derivatives of  $u$ :

$$\sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left( a_{ij}(x) \frac{\partial v}{\partial x_j} \right) = 0, \quad \text{where } a_{ij}(x) = F_{ij}(\nabla u(x)). \quad (4)$$

However, the coefficients  $a_{ij}(x) = F_{ij}(\nabla u(x))$  of this new linearized equation Eq. (4) depend intrinsically on the unknown gradient. If we assume only that  $u$  has bounded slope so that the gradient  $\nabla u$  is bounded but not known to be continuous, then these



Figure 4: Charles Morrey Jr. (1907–1984)

coefficients are bounded, but they may be discontinuous. We are thus caught in a vicious circle: to prove the coefficients are smooth, we need the solution’s gradient to be smooth; but to prove the solution is smooth, the classical theory demands the coefficients be smooth. This “gap” between having a bounded slope versus a continuous slope was what Morrey later called the “sad state of affairs” that hindered progress for decades. The solution was finally achieved by bridging this gap. Let us illustrate the strategy in 2D, using geometric intuition.

You are likely familiar with the gradient  $\nabla u$  as a vector field pointing in the direction of steepest increase. But for this problem, we need a shift in perspective. Imagine mapping every point  $x$  in our domain not to its height  $u(x)$ , but to its slope vector  $\nabla u(x)$ . This creates a “gradient map” from the physical domain into a “plane of slopes.”

*The core question of regularity is simply: Is this map continuous? Or can the slope jump abruptly from one value to another, creating a sharp crease or kink in our soap film?*

Imagine the image of the gradient map  $\nabla u$  as a region (a “blob”) in the plane. Fix a direction  $e$  (a unit vector) and consider the strip in the gradient plane between two parallel lines perpendicular to  $e$ , defined by values  $a < b$ .

Let  $v(x) = \nabla u(x) \cdot e$  be the corresponding directional derivative. If the gradient image  $\nabla u(B_r)$  crosses this strip, it means  $v$  takes values both below  $a$  and above  $b$  within the ball  $B_r$ .

Here we use the Maximum Principle. The Maximum Principle is a powerful property of elliptic equations. You’ve seen a special case already: Laplace’s equation  $\nabla^2 u = 0$ . Its solutions (harmonic functions) obey the Mean Value Property and the Maximum Principle: the maximum and minimum occur on the boundary. Our linearized equation Eq. (4) is a generalization of Laplace’s equation with variable coefficients  $a_{ij}(x)$ . The principle still holds: the oscillation of  $v$  inside a ball is controlled by its oscillation on the boundary-like how the temperature at the center of a room is bounded by the

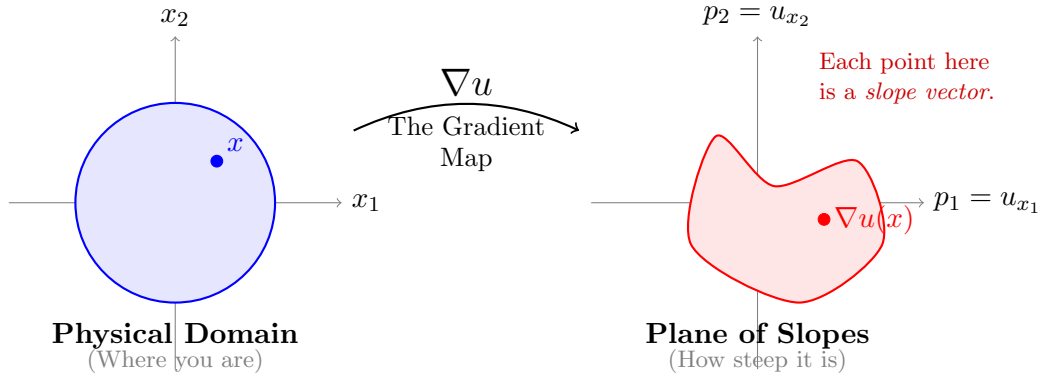


Figure 5: **The Gradient Map.** This map takes a location  $x$  in the physical domain (left) and sends it to its corresponding slope vector  $\nabla u(x)$  in the plane of slopes (right). If the solution is smooth, the red “blob” of slopes will be connected. If the solution has a crease, this blob will be torn apart.

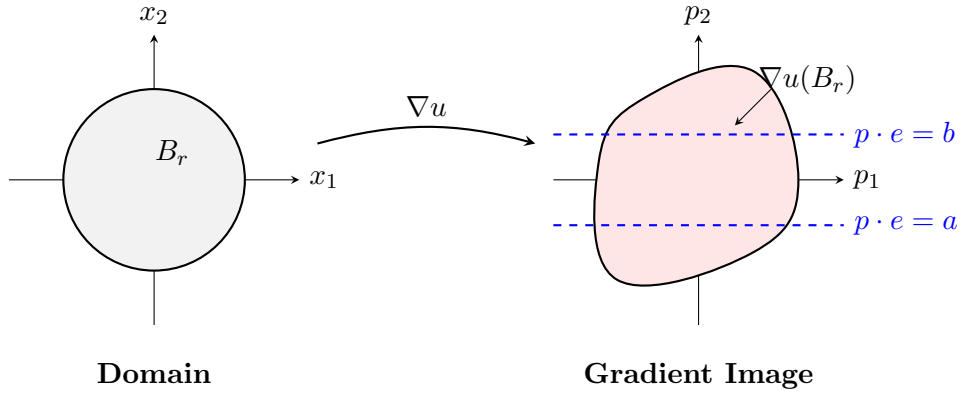


Figure 6: Chopping at the gradient image with lines

temperatures on the walls.

In simple terms, for elliptic equations, the values inside a domain are controlled by the values on the boundary (much like temperature in a room is controlled by the walls). By this principle, if  $v$  oscillates inside  $B_r$ , it must oscillate at least as much on the boundary circle  $\partial B_r$ . This allows us to restrict our attention to the boundary.

To see if this oscillation can persist as  $r \rightarrow 0$ , we quantify its “energy cost” using a geometric estimate:

$$(\text{osc}_{\partial B_r} v)^2 \leq \frac{\pi}{\ln(1/2r)} \int_{B_{1/2}} |\nabla v|^2 dx. \quad (5)$$

2

---

$^2\omega(r) := \text{osc}_{\partial B_r} v$

Suppose, for contradiction, that the oscillation stays large. Specifically, assume that on all circles  $\partial B_r$ , the oscillation is bounded below by a constant  $\delta > 0$ .

Summing these oscillations, due to the term  $\ln(1/r)$ , the required energy must diverge (recall the harmonic series):

$$\text{Energy} \sim \int \frac{1}{r} dr \rightarrow \infty.$$

However, we know there exists a finite upper bound. The differential equation itself provides a “natural” energy bound (Caccioppoli inequality):

$$\int_{B_{1/2}} |\nabla v|^2 dx \leq C \int_{B_1} v^2 dx < \infty. \quad (6)$$

This is a contradiction: Eq. (5) requires infinite energy, but Eq. (6) bounds it finitely.

Therefore, for sufficiently small  $r$ , the gradient image  $\nabla u(B_r)$  cannot cross the strip. It must lie entirely on one side. By “chopping” with strips in various directions, we force the gradient image to localize to a single point as  $r \rightarrow 0$ . Thus,  $\nabla u$  is continuous.

In 3D, we analogously attempt to trap the gradient using parallel planes. If the gradient persists in crossing these planes, it again implies a singularity with characteristic scale  $|\nabla v| \sim \frac{\delta}{r}$ . However, the geometric measure of the boundary (surface area) now scales as  $r^2$ .

The energy integral in 3D becomes:

$$\text{Energy}_{3D} = \int_0^1 \left( \int_{\partial B_r} |\nabla v|^2 d\sigma \right) dr.$$

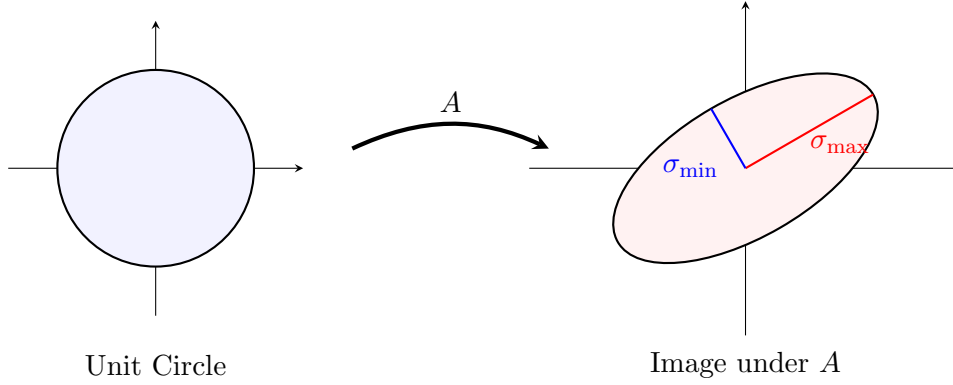
Substituting the 3D scaling factor ( $d\sigma \sim 4\pi r^2$ ) gives us:

$$\text{Energy}_{3D} \sim \int_0^1 \left( \frac{\delta}{r} \right)^2 \cdot (4\pi r^2) dr \sim 4\pi \delta^2 \int_0^1 1 dr = 4\pi \delta^2 < \infty.$$

There is no contradiction. The integral converges, meaning the energy cost of a persistent singularity is finite. The gradient can effectively “afford” to keep oscillating across the planes forever without violating the global energy bounds established by the Caccioppoli inequality.

While the “chopping” argument above relies on energy estimates, there is a second, completely different reason why 2D solutions are smooth. This second method relies on a “happy accident of geometry”, a special feature that disappears in higher dimensions. The linearized equation implies that the gradient map  $\nabla u : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  behaves like a linear transformation with bounded distortion. Recall from linear algebra that a matrix  $A$  maps the unit circle to an ellipse. The “distortion” is determined by the condition number (the ratio of eigenvalues):

$$K = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}. \quad (7)$$



$$\text{Distortion } K = \frac{\sigma_{\max}}{\sigma_{\min}}$$

Figure 7: Action of a linear map on the unit circle and the induced distortion.

You might wonder: *Why is this distortion  $K$  bounded?*

In 2D, the linearized equation  $\sum a_{ij}v_{ij} = 0$  acts as a tight “lock” on the eigenvalues  $\mu_1, \mu_2$  of the Hessian matrix  $D^2u$ . Since the coefficient matrix  $a_{ij}$  is positive definite, for the weighted sum of eigenvalues to be zero,  $\mu_1$  and  $\mu_2$  must have opposite signs. More importantly, the uniform ellipticity constants  $[\lambda, \lambda^{-1}]$  force their magnitudes to be comparable:

$$|\mu_1| \approx |\mu_2|.$$

Because they are “locked” together, one eigenvalue cannot grow large without the other growing continuously with it. This ensures that their ratio  $K = \frac{|\mu_{\max}|}{|\mu_{\min}|}$  stays finite. Geometrically, this means the gradient map is *quasiconformal*: it transforms infinitesimal circles into ellipses of bounded eccentricity, a property known to strictly enforce continuity.

In higher dimensions, this “safety net” vanishes. Let’s take 3D as an example, we still have a single scalar equation  $\sum a_{ij}v_{ij} = 0$ , but now it must constrain three eigenvalues  $\mu_1, \mu_2, \mu_3$ . This single constraint is too weak to lock them together. Imagine a scenario where one eigenvalue is massive, and the other two are small and negative. The equation can still be satisfied, but the distortion ratio may blow up. This allows the map to stretch space infinitely in one direction while remaining a valid solution to the equation. Consequently, the gradient map in 3D loses the quasiconformal property, and the geometric guarantee of regularity is lost.

How does preventing a circle from stretching into a thin ellipse guarantee that the solution is smooth? The link is provided by Mori’s Theorem. The theorem states that any map with bounded distortion  $K$  (a  $K$ -quasiconformal map) is automatically

Hölder continuous with an exponent  $\alpha$  determined by the dimension and the distortion. Specifically, for dimension  $n \geq 2$ :

$$\alpha = K^{\frac{1}{1-n}}.$$

In our specific 2D, this exponent simplifies elegantly:

$$\alpha = K^{\frac{1}{1-2}} = K^{-1} = \frac{1}{K}.$$

This provides a direct translation from geometry to analysis: if the distortion  $K$  is finite, the Hölder exponent  $\alpha$  is positive. Intuitively, this means that because the map cannot stretch space infinitely, it cannot “tear” values apart. Points that are close in the domain are forced to map to values that are close in the range.

Mathematically, both the “chopping” argument and the “bounded distortion” argument enforce the same decay estimate. If we shrink the radius by a factor  $\delta$ , the oscillation drops by a fixed fraction:

$$\omega(\delta r) \leq \frac{1}{2}\omega(r) \implies \omega(\delta^k) \leq 2^{-k}\omega(1). \quad (8)$$

This implies that the oscillation decays as a power of the radius:

$$\omega(r) \leq Cr^\alpha. \quad (9)$$

Equation (9) guarantees that the gradient  $\nabla u$  is continuous with a controlled rate of convergence (Hölder continuity). Thus, in the planar case, the specific geometry of 2D forces regularity. The wild oscillations are controlled by the very cost of their existence.