SIAM REVIEW
Vol. 49, No. 3, pp. 491–507

As Flat As Possible*

Jon Jacobsen†

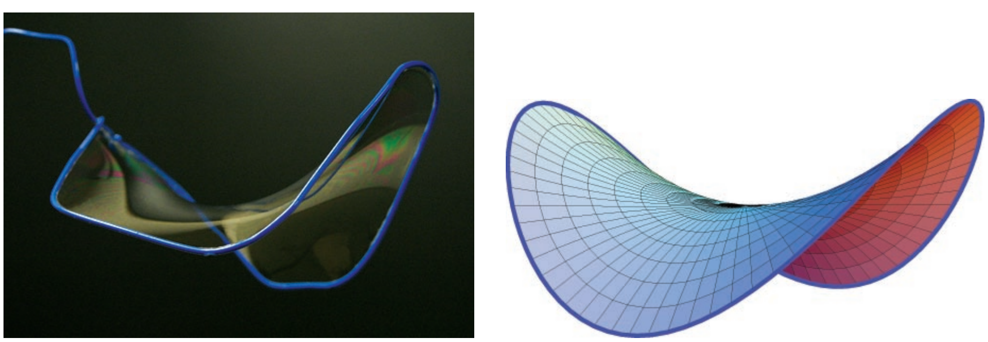
Abstract. How does one determine a surface which is as flat as possible, such as those created by soap film surfaces? What does it mean to be *as flat as possible*? In this paper we address this question from two distinct points of view, one local and one global in nature. Continuing with this theme, we put a temporal twist on the question and ask how to evolve a surface so as to flatten it as efficiently as possible. This elementary discussion provides a platform to introduce a wide range of advanced topics in partial differential equations and helps students build geometric and analytic understanding of solutions of certain elliptic and parabolic partial differential equations.

Key words. pointwise estimates, calculus of variations, gradient flow, harmonic functions, minimal surfaces, Sobolev gradients

AMS subject classifications. 31A05, 35J25, 49-01, 53A05

DOI. 10.1137/060649495

1. Introduction. Here is a fun experiment: bend a piece of wire into a closed curve and dip the resulting shape into a soap film solution to produce a surface. The surfaces created are endlessly fascinating—they must bend and stretch to conform to the boundary data, but they do so in an efficient way. In fact, the surfaces seem to be as flat as possible, given the predefined boundary data. A typical soap film surface is shown in Figure 1.1(a).[1]



(a) (b)

Fig. 1.1 (a) *Soap film surface.* (b) *Mathematical model—is it correct?*

*Received by the editors January 9, 2006; accepted for publication (in revised form) October 26, 2006; published electronically July 30, 2007.
http://www.siam.org/journals/sirev/49-3/64949.html

†Mathematics Department, Harvey Mudd College, 301 Platt Blvd., Claremont, CA 91711 (jacobsen@math.hmc.edu).

[1]For making your own soap film surfaces the following recipe from Andrew Belmonte (The Pennsylvania State University) works well: 2450 mL H_2O; 50 mL Dawn soap; 500 mL glycerol.

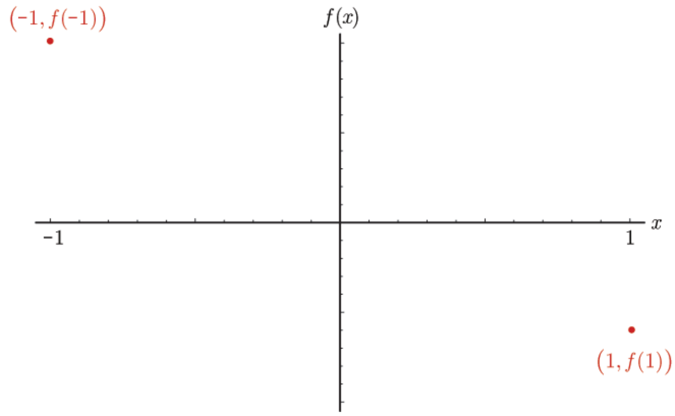


Fig. 2.1 *Given boundary data $f(-1)$ and $f(1)$, what should $f(0)$ be if f is to be as flat as possible?*

A natural question arises as to the precise shape of the surface. For example, given the values of the boundary data (i.e., the blue wire frame curve in Figure 1.1(a)), can we determine the exact surface that appears, say, in terms of the graph of some explicit function $u(x, y)$? How would we find $u(x, y)$? Figure 1.1(b) shows a sample mathematical model. It certainly looks flat, but is it correct? Is it as flat as possible? Is the soap film surface really as flat as possible? What does it mean to be as flat as possible? This is the first fundamental question of this paper:

How do we find a surface, with given boundary data, which is *as flat as possible*?

Such ambiguous language might give the reader pause; yet, this type of scenario is actually quite common in mathematics, and mathematical modeling in particular, where one often has to assign precise and rigorous mathematical meaning to vague notions and fragmentary geometric insights.

In what follows, we investigate two completely different approaches to answering this question. The first approach is local in nature, while the second approach is global. Surprisingly, both approaches lead to the same answer. After exploring these approaches and some of their consequences, we conclude with a temporal twist to the main question and discuss applications to gradient flows in Hilbert spaces.

2. The Local Approach: A Pointwise Constraint. Before tackling the full surface of Figure 1.1, let's consider the following simpler question: Suppose a continuous function $f : [-1, 1] \to \mathbb{R}$ is given, but only the values at $x = -1$ and $x = +1$ are known, as in Figure 2.1. What should the value $f(0)$ be in order for f to be "as flat as possible"?

Even without a precise definition of flat, a moment's thought reveals that a line would surely provide the flattest possible continuous function, in which case the value at the origin would be $f(0) = \frac{f(-1)+f(1)}{2}$, the average of the two endpoint values.

Stepping up a dimension, let's consider the analogous question in the plane. Let $B(0, 1) = \{(x, y) : x^2 + y^2 < 1\}$ denote the open unit ball in the plane and $\partial B(0, 1)$ its boundary. Suppose a continuous function $u : \overline{B(0, 1)} \to \mathbb{R}$ is given, but only the values on $\partial B(0, 1)$ are known, as in Figure 2.2. What should the value $u(0, 0)$ be if the surface defined by u is to be as flat as possible?

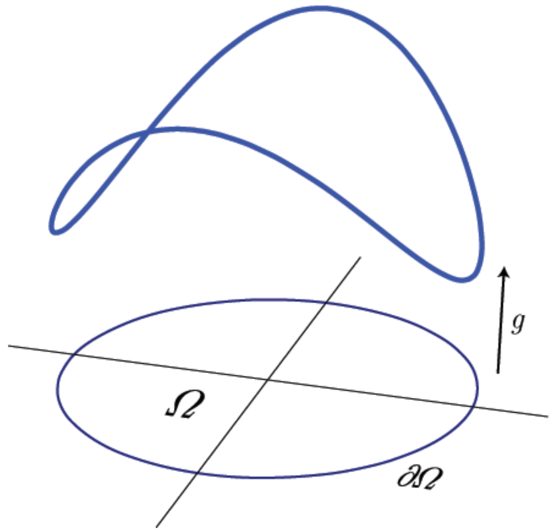


Fig. 2.2 *Given boundary data g, what should $u(0,0)$ be if u is to be as flat as possible?*

In this case the answer is not so immediate. In particular, while a line was a reasonable candidate for the flattest function in one dimension, it is not at all clear what one means by "as flat as possible" in the two-dimensional setting. One reasonable answer to this question is that the value at the center should be the average of the values on the boundary $\partial B(0,1)$. More precisely, if the boundary data is given by a function $g : \partial B \to \mathbb{R}$, then

$$u(0,0) = \frac{1}{2\pi} \int_0^{2\pi} g(\theta)\, d\theta. \tag{2.1}$$

This would allow the surface graph of u to continuously connect to the given boundary data in an efficiently flat manner. Another nice feature of this approach is that it generalizes the previous case of a line. But do soap film surfaces satisfy this property? More generally, does such a continuous surface exist? We'll return to these questions momentarily.

The averaging idea above lends itself to a possible definition of a surface which is *absolutely* as flat as possible. Namely, for every point $\mathbf{x_0} = (x_0, y_0)$ in the domain of u, what if the value $u(\mathbf{x_0})$ was the average of the values of u on any circle centered at $\mathbf{x_0}$? More precisely, what if

$$u(\mathbf{x}) = \frac{1}{2\pi r} \int_{\partial B(\mathbf{x_0}, r)} u(x,y)\, ds \tag{2.2}$$

for any disk $B(\mathbf{x_0}, r)$? Here $ds = r d\theta$ denotes arc length on the circle. Could such a function exist? If it did, then it would be in a sense as flat as possible, since the average or mean-value property defined by (2.2) is a higher-dimensional version of the property that holds for lines in one dimension. The existence of such a function would also resolve the existence question introduced above, where we asked that the property only hold at the origin. Accordingly, we make this our first definition.

Definition 2.1. *A function $u : \Omega \to \mathbb{R}$ with boundary data g is as flat as possible if* (2.2) *holds for all $\mathbf{x_0} \in \Omega$.*

Temporarily postponing the question of existence, suppose we do have a sufficiently smooth function u which satisfies (2.2) at a given point $\mathbf{x_0}$ for all sufficiently close circles centered at $\mathbf{x_0}$. For each such circle the left-hand side of (2.2) is constant (the value of u at $\mathbf{x_0}$), while the right-hand side is a function of the radius r (the integral of u over the circle of radius r). If we differentiate each side with respect to r, then the left-hand side will vanish, but what about the right-hand side? The situation is a bit unusual, in that the region of integration depends on the independent variable r. One way to resolve this is to parametrize the boundary curves in terms of the fixed interval $[0, 2\pi]$. In this way we obtain an integral with limits independent of r (the dependence moved to the integrand):

$$\begin{aligned} u(x_0, y_0) &= \frac{1}{2\pi r} \int_{\partial B(\mathbf{x_0}, r)} u(x, y)\, ds \\ &= \frac{1}{2\pi r} \int_0^{2\pi} u(x_0 + r\cos\theta, y_0 + r\sin\theta)\, r\, d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} u(x_0 + r\cos\theta, y_0 + r\sin\theta)\, d\theta. \end{aligned}$$

Now we can differentiate each side with respect to r to find

$$0 = \frac{1}{2\pi} \int_0^{2\pi} \frac{\partial u}{\partial x} \cos\theta + \frac{\partial u}{\partial y} \sin\theta\, d\theta, \tag{2.3}$$

where the partial derivatives are evaluated at $(x, y) = (x_0 + r\cos\theta, y_0 + r\sin\theta)$. Since the vector $\mathbf{n} = (\cos\theta, \sin\theta)$ represents the unit outer normal vector on the unit circle, the integrand in (2.3) is the dot product of the gradient of u, $\nabla u = (\partial u/\partial x, \partial u/\partial y)$, with the normal vector $\mathbf{n}$ along the boundary. Thus the integral represents the local flux through the boundary. Something interesting happens when we now apply the divergence theorem (dropping the unnecessary factor $1/2\pi$):

$$\begin{aligned} 0 &= \int_0^{2\pi} \frac{\partial u}{\partial x} \cos\theta + \frac{\partial u}{\partial y} \sin\theta\, d\theta \\ &= \frac{1}{r} \int_0^{2\pi} (\nabla u \cdot \mathbf{n})\ r\, d\theta \\ &= \frac{1}{r} \int_{\partial B(\mathbf{x_0}, r)} \nabla u \cdot \mathbf{n}\, ds \\ &= \frac{1}{r} \int_{B(\mathbf{x_0}, r)} \operatorname{div}(\nabla u)\ dx\, dy \\ &= \frac{1}{r} \int_{B(\mathbf{x_0}, r)} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) dx\, dy \\ &= \frac{1}{r} \int_{B(\mathbf{x_0}, r)} \Delta u\, dx\, dy, \end{aligned} \tag{2.4}$$

where $\Delta u = \operatorname{div}(\nabla u) = \nabla \cdot \nabla u = u_{xx} + u_{yy}$ denotes the Laplacian of u. From this calculation we conclude that

$$0 = \int_{B(\mathbf{x_0}, r)} \Delta u\, dx\, dy \tag{2.5}$$

for all sufficiently small disks $B(\mathbf{x_0}, r)$ at $\mathbf{x_0}$. If Δu is a continuous function, then this implies $\Delta u(\mathbf{x_0}) = 0$. If not, suppose $\Delta u(\mathbf{x_0}) > 0$; then by continuity there would exist a small neighborhood of $\mathbf{x_0}$, say $B(\mathbf{x_0}, \varepsilon)$, for which $\Delta u > 0$ remained true. This would imply $\int_{B(\mathbf{x_0},\varepsilon)} \Delta u \, dxdy > 0$, contradicting (2.5). The same argument rules out the possibility that $\Delta u(\mathbf{x_0}) < 0$. If the mean-value property (2.2) holds for all points $\mathbf{x}$ in the domain of u, then we conclude

$$\Delta u = 0 \quad \text{for each } \mathbf{x} \in \Omega. \tag{2.6}$$

Functions which satisfy (2.6) on a domain Ω are called *harmonic functions* on Ω.

We have arrived at our first characterization of the "flat as possible" surfaces. With flatness measured by Definition 2.1 (i.e., measured by the mean-value property (2.2)), it follows that the surface which agrees with g on the boundary and is as flat as possible is defined by the solution of the boundary value problem

$$\begin{cases} \Delta u = 0, & \mathbf{x} \in \Omega, \\ u = g, & \mathbf{x} \in \partial\Omega. \end{cases} \tag{2.7}$$

That is, u is a harmonic function that agrees with g on the boundary. The function u is called the *harmonic extension of* g. In particular, this answers the existence question previously postponed. The existence of a function u satisfying (2.2) throughout Ω is equivalent to the existence of a solution to the partial differential equation (2.7).

This is an interesting way to derive a partial differential equation. Starting from the pointwise assumption that u is the average of its neighboring values on nearby circles, we arrived at the conclusion that u must solve the partial differential equation defined by Laplace's equation.[2] Remarkably, for continuous functions the mean-value property (2.2) is *equivalent* to being harmonic and, in particular, such functions are necessarily infinitely differentiable (see, e.g., [6]).

THEOREM 2.2 (mean-value property). *Let $\Omega \subset \mathbb{R}^2$ be open. The continuous function $u : \Omega \to \mathbb{R}$ is harmonic if and only if*

$$u(\mathbf{x}) = \frac{1}{2\pi r} \int_{\partial B(\mathbf{x},r)} u \, ds = \frac{1}{\pi r^2} \int_{B(\mathbf{x},r)} u \, dxdy$$

for each $\mathbf{x} \in \Omega$ and $\overline{B(\mathbf{x}, r)} \subset \Omega$. The statement also holds in $\mathbb{R}^n$, with averages over circles and disks replaced by the appropriate averages over higher-dimensional spheres and balls.

The implications of this theorem are worth dwelling on. Although we have no connection yet between the "flat" soap film surfaces and harmonic functions, we have gained tremendous insight into the properties of harmonic functions. If their graphs bend or change at all, it must always be in a way that preserves the mean-value property. This is even more striking when one considers the many examples of harmonic functions. For instance, it follows from the Cauchy–Riemann equations that the real and imaginary parts of any analytic function are harmonic. This gives us an endless source of interesting harmonic functions—and all of them satisfy the mean-value property! For example, $f(z) = z^2 = (x + iy)^2 = (x^2 - y^2) + i(2xy)$ is everywhere analytic. Thus $u(x, y) = x^2 - y^2$ and $v(x, y) = 2xy$ are harmonic functions on $\mathbb{R}^2$. Another famous analytic function is $e^z = (e^x \cos y) + i\,(e^x \sin y)$, which defines two

[2] For an interesting variation, where u need only have one circle for each $\mathbf{x}$, see [15].

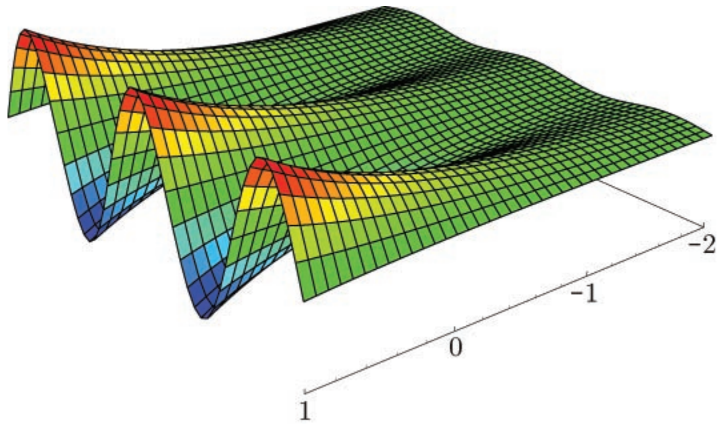


Fig. 2.3 *Harmonic function defined by the real part of e^z—is it a soap film surface?*

harmonic functions $u(x, y) = e^x \cos y$ and $v(x, y) = e^x \sin y$. Figure 2.3 shows a plot of $e^x \cos y$. Notice how the graph appears taut, like a soap film.

For a more exotic example, consider $f(z) = \frac{1}{5}z^5 + \frac{1}{4}z^2 - \frac{i}{5}z^5 \cosh z$. It's not obvious what the real and imaginary parts of f are, but they are harmonic functions. Figure 2.4 shows a graph of the imaginary part of f and reveals an interesting surface which satisfies the mean-value property.

Harmonic functions are fundamental in mathematical physics. For example, Maxwell's equations govern the interaction between electric and magnetic fields [10]. In the static case the equations for the electric field and magnetic field decouple and the electric field $\mathbf{E}$ is governed by the two equations

$$\text{curl } \mathbf{E} = \mathbf{0} \qquad \text{and} \qquad \text{div } \mathbf{E} = 4\pi\rho(x),$$

where $\rho(x)$ is the charge density. Since $\mathbf{E}$ is curl free it follows that $\mathbf{E} = -\nabla\phi$ for some scalar function ϕ, called the *electric potential*. Substituting this into the second equation yields $-\Delta\phi = 4\pi\rho$. Thus in any charge-free region $\rho = 0$ and it follows that $\Delta\phi = 0$. In parlance, the electrostatic potential in a charge-free region is harmonic. This implies that such electrostatic potentials are also as flat as possible, in the sense that (2.2) holds where they are defined.

Harmonic functions also play an essential role in the study of fluid dynamics. In fluid dynamics one is interested in the velocity field $\mathbf{v} = \mathbf{v}(x, y, z, t)$ of a given fluid in motion. If the flow is steady, then the velocity field is independent of time t. If the flow is irrotational (i.e., curl $\mathbf{v} = 0$), then $\mathbf{v} = -\nabla u$ for some scalar function u, called the *velocity potential*. If the flow is incompressible (e.g., constant density), then div $\mathbf{v} = 0$. Therefore, if it is incompressible and irrotational, then div $\mathbf{v} = \text{div}(-\nabla u) = -\Delta u = 0$. Consequently, the velocity potential for an incompressible irrotational fluid is harmonic. This is a very important result in the theory of fluid dynamics, and, once again, the fact that the mean-value property (2.2) holds tells us something interesting about velocity potentials for such flows.

The mean-value property provides a fundamental understanding of the apparent "flatness" of harmonic surfaces. They can bend, but only in such a way as to always preserve the mean-value property. This characterization also implies there can be no interior maxima or minima, for if the surface bends up, then to keep the average

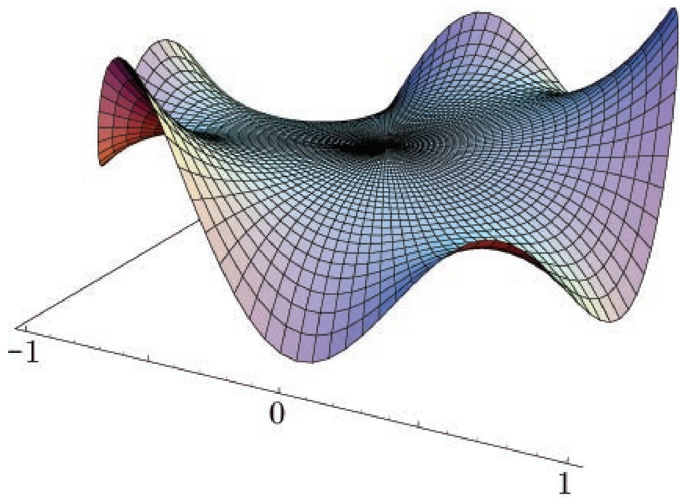


Fig. 2.4 *Exotic harmonic function defined by the imaginary part of* $f(z) = \frac{1}{5}z^5 + \frac{1}{4}z^2 - \frac{i}{5}z^5 \cosh z$.

the same it has to also bend down—a phenomenon evident in the graphs shown in Figures 2.3 and 2.4. That the extremal values of harmonic functions occur on the boundary is known as the *maximum principle*, and we invite the reader to derive a proof based on the mean-value property (2.2).

The calculation leading to (2.2) also provides a physical and geometric interpretation of the Laplacian Δu. For any given function u (not necessarily harmonic), if we let $M_r[u; \mathbf{x_0}] = \frac{1}{2\pi r}\int_{\partial B(\mathbf{x_0},r)} u\, ds$ denote the mean value of u on the circle centered at $\mathbf{x_0}$ of radius r, then the calculation leading to (2.4) shows that

$$\text{(2.8)} \qquad \frac{d}{dr}M_r[u; \mathbf{x_0}] = \frac{1}{2\pi r}\int_{B(\mathbf{x_0},r)} \Delta u\, dx\, dy.$$

If u is continuous at $\mathbf{x_0}$, then $\lim_{r\to 0} M_r[u; \mathbf{x_0}] = u(\mathbf{x_0})$. Therefore, integrating (2.8) from 0 to r yields

$$\text{(2.9)} \qquad M_r[u; \mathbf{x_0}] - u(\mathbf{x_0}) = \int_0^r \left(\frac{1}{2\pi s}\int_{B(\mathbf{x_0},s)} \Delta u\, dx\, dy\right) ds.$$

Since the left-hand side is the difference between $u(\mathbf{x_0})$ and its local average, it follows from the integral on the right-hand side that the Laplacian measures how u deviates from its average on nearby circles. For example, if we consider a sufficiently small circle of radius ε for which $\Delta u(x) \approx \Delta u(\mathbf{x_0})$ for $x \in B(\mathbf{x_0}, \varepsilon)$, this yields

$$M_\epsilon[u; \mathbf{x_0}] - u(\mathbf{x_0}) \approx \frac{\varepsilon^2}{4}\Delta u(\mathbf{x_0}).$$

While we have found one answer to the first fundamental *as flat as possible* question and gained much understanding of harmonic functions and the Laplacian, we still don't know if these are what we observe with soap films. For example, if we make a wire frame in the shape of the boundary curve of Figure 2.4 and dip it in soap film solution, will we see Figure 2.4? Do the real and imaginary parts of analytic functions represent soap film surfaces? To gain further insight we turn to a completely different approach to answering the question.

3. A Global Approach: Integrating Change. Recall the first fundamental question of this paper:

How do we find a surface, with given boundary data, which is *as flat as possible*?

For our second approach, rather than define a pointwise constraint such as (2.2), we consider assigning a positive numerical measure of flatness to a given surface u. One reasonable choice for smooth functions is to integrate the square of the gradient over the surface:

$$E[u] = \frac{1}{2}\int_\Omega |\nabla u|^2\, dy\, dx. \tag{3.1}$$

The value $E[u]$ is a crude measure of a surface's flatness. For example, if $E[u] = 0$, then $|\nabla u(x)| = 0$ for all $x \in \Omega$, in which case u is constant, which *is* as flat as possible. On the other hand, typically g is nonconstant, in which case $E[u] > 0$ for any function u that agrees with g on $\partial\Omega$. This approach of integrating the change over the entire surface lends itself to a second possible definition of a surface which is *absolutely* as flat as possible. Namely, the surface with minimal value of E.

DEFINITION 3.1. *A function $u : \Omega \to \mathbb{R}$ with boundary data g is as flat as possible if it minimizes $E[u]$ over all possible functions with boundary data g.*

Once again postponing the question of existence, suppose that for a given g defined on $\partial\Omega$, the function u minimizes E over the set of all functions that agree with g on $\partial\Omega$. For ϕ smooth, with $\phi = 0$ on $\partial\Omega$, notice that $u + \phi$ also agrees with g on $\partial\Omega$, in which case, since u is the minimizer, it follows that

$$E[u] \le E[u + \phi].$$

In fact, for each such ϕ we can define a nonnegative scalar function i_ϕ by

$$i_\phi(\varepsilon) = E[u + \varepsilon\phi].$$

The fact that u minimizes E implies that the function i_ϕ has a minimum at $\varepsilon = 0$, or $i'_\phi(0) = 0$. But i_ϕ is a single variable function whose derivative we can compute:

$$\begin{aligned}
i'_\phi(0) &= \lim_{h\to 0} \frac{i_\phi(h) - i_\phi(0)}{h} \\
&= \lim_{h\to 0} \frac{E[u + h\phi] - E[u]}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{2}\int_\Omega |\nabla u + h\nabla\phi|^2\, dy\, dx - \frac{1}{2}\int_\Omega |\nabla u|\, dy\, dx}{h} \\
&= \lim_{h\to 0} \frac{1}{2h}\left(\int_\Omega (\nabla u + h\nabla\phi)\cdot(\nabla u + h\nabla\phi)\; dy\, dx - \int_\Omega |\nabla u|^2\, dy\, dx\right) \\
&= \lim_{h\to 0} \frac{1}{2h}\left(\int_\Omega |\nabla u|^2 + 2h\,\nabla u\cdot\nabla\phi + h^2|\nabla\phi|^2\, dy\, dx - \int_\Omega |\nabla u|^2\, dy\, dx\right) \\
&= \lim_{h\to 0} \frac{1}{2h}\left(2h\int_\Omega \nabla u\cdot\nabla\phi\, dy\, dx + h^2\int_\Omega |\nabla\phi|^2\, dy\, dx\right) \\
&= \int_\Omega \nabla u\cdot\nabla\phi\; dy\, dx.
\end{aligned}$$

Since $i'_\phi(0) = 0$, this calculation shows that

$$0 = \int_\Omega \nabla u \cdot \nabla \phi \; dy\, dx.$$

Integrating by parts (recall $\phi = 0$ on $\partial\Omega$) yields

$$0 = \int_\Omega (\Delta u)\, \phi \, dy\, dx. \tag{3.2}$$

Curiously, the Laplacian has made its second appearance in our quest to understand flatness. Since this calculation is independent of the choice of ϕ we conclude that u minimizes E over all functions that agree with g on $\partial\Omega$ only if

$$\int_\Omega (\Delta u)\, \phi \, dy\, dx = 0$$

for all such ϕ. If Δu is continuous, then since ϕ is arbitrary it follows that

$$\Delta u = 0 \qquad \text{for each } \mathbf{x} \in \Omega. \tag{3.3}$$

The proof of this follows in the spirit of the proof of (2.6). For instance, if $\Delta u(\mathbf{x_0}) > 0$ for some $\mathbf{x_0} \in \Omega$, then by continuity there exists a small open neighborhood $U \subset \Omega$ for which $\Delta u(\mathbf{x}) > 0$ for $\mathbf{x} \in U$. Then by choosing a smooth function ϕ such that $\phi > 0$ in U and $\phi = 0$ on the complement of U it would follow that

$$\int_\Omega \Delta u\, \phi \, dy\, dx = \int_U \Delta u\, \phi \, dy\, dx > 0,$$

which contradicts (3.3). The same argument also works if $\Delta u(\mathbf{x}) < 0$.

In other words, we have shown that harmonic functions are precisely the minimizers of the integral (3.1). Not only are they as flat as possible in the mean-value property sense, they also minimize the integral of $|\nabla u|^2$ over all possible functions that agree with g on $\partial\Omega$. Surprisingly, the two completely different definitions, one based on a local pointwise estimate and the other based on a global integration, have led to the same partial differential equation.

This second approach is an example of the *calculus of variations* and is a very active area of research (see [6] for a modern introduction). Typically, in the calculus of variations one is interested in the critical points of a given scalar-valued function of functions like $E[u]$, called a *functional*. In the example above the functional $E[u]$ is known as the *Dirichlet energy integral*, from the fact that if u represents a velocity, then $E[u]$ represents the kinetic energy. We have shown that another characterization of harmonic functions is as minimizers of the Dirichlet energy integral.

It may have occurred to the reader that the choice of the Dirichlet energy integral for our second definition of flatness, while reasonable, was somewhat arbitrary. However, given a different numerical measure of flatness, we can still proceed as above and obtain a partial differential equation for its minimizers, which are the flattest possible surfaces for that measure. For example, minimizers of

$$E[u] = \frac{1}{p} \int_\Omega |\nabla u|^p \, dy\, dx \qquad \text{for } p > 1$$

are called p-harmonic functions and solve

$$\Delta_p u := \operatorname{div}(|\nabla u|^{p-2} \nabla u) = 0.$$

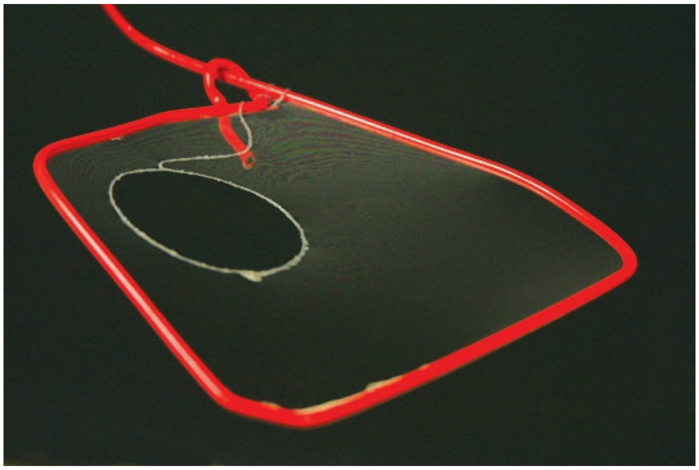


Fig. 3.1 *A loose loop of string was embedded in the soap film surface. When the film is poked inside the loop, the loop rapidly expands into a circle. Since a circle encloses the maximal area for a fixed perimeter, this demonstration shows that surface tension acts to minimize the surface area.*

The operator $\Delta_p u$ is known as the p-Laplacian and is an important operator with a growing body of literature and applications (see, e.g., [8]).

It is this second global approach to measuring flatness that yields the key insight into soap film surfaces. From physical principles we know that surface tension acts to minimize the surface area of the soap film. A beautiful demonstration of this is provided by placing a small loop of string on a flat soap film surface and gently popping the fluid inside the loop. The loop will rapidly expand to form the shape of a circle, maximizing the region of air inside, while simultaneously minimizing the complementary surface area of the soap film. Figure 3.1 shows the result immediately after popping the portion of the soap film inside the loop.

Thus, the soap film surface that spans a given wire frame is the surface that minimizes the surface area integral

$$SA[u] = \int_\Omega \sqrt{1 + |\nabla u|^2}\, dy\, dx \tag{3.4}$$

over all possible functions that agree with g on $\partial\Omega$. Performing the earlier computations for $i'_\phi(0)$ with the surface area integral (3.4) yields the partial differential equation

$$\operatorname{div}\left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}}\right) = 0, \qquad \mathbf{x} \in \Omega, \tag{3.5}$$

which is known as the *minimal surface equation* [6]. For example, in the two-dimensional case this takes the form

$$(1 + u_y^2)u_{xx} - 2u_x u_y u_{xy} + (1 + u_x^2)u_{yy} = 0. \tag{3.6}$$

For a given function $u(x, y)$, the expression on the left-hand side of (3.5) measures twice the mean curvature of the surface at a point $(x, y, u(x, y))$. Thus solutions to

(3.5) have the property that their mean curvature vanishes everywhere. Such surfaces with zero mean curvature are called *minimal surfaces* [5].

With this approach, we see that the function whose graph defines the soap film surface is not harmonic, but rather a minimal surface. Moreover, it does not satisfy the mean-value property, but it can come close. Indeed, from the Taylor approximation $\sqrt{1+x} \approx 1 + \frac{x}{2}$, it follows that

$$E[u] = \int_\Omega \sqrt{1 + |\nabla u|^2}\, dy\, dx \approx \int_\Omega 1 + \frac{1}{2}|\nabla u|^2\, dy\, dx.$$

Thus within the context of this approximation (i.e., for $|\nabla u|^2$ small) minimizing the surface area is equivalent to minimizing the Dirichlet energy integral.

However, there is a much stronger connection between minimal surfaces and harmonic functions, which is provided through parametrization. Rather than viewing the surface as the graph of a given function on the domain Ω, consider the surface as the parametric image of a given region in the plane. For example, the parametrization corresponding to the graph of $u(x, y)$ is $H(r, s) = (r, s, u(r, s))$ for $(r, s) \in \Omega$. We have shown that if the resulting surface is minimal, then u solves (3.6). However, the parametrization of a surface is not unique. A parametrization $X(r, s) = (x(r, s), y(r, s), z(r, s))$ is *isothermal* if $|X_r|^2 = |X_s|^2$ and $X_r \cdot X_s = 0$. An important result from the theory of surfaces is that a surface with isothermal parametrization is minimal if and only if the component functions x, y, and z are harmonic [5]. Thus, the component functions x, y, and z of an isothermal parametrization of a minimal surface *are* as flat as possible. Viewing surfaces through parametrization also has the advantage of opening up the notion of a minimal surface beyond those that are restricted to be the graph of some function, which carries with it a predetermined coordinate system. Indeed, in the words of one expert, parametrization "shows that minimal surfaces are truly harmonic, in a geometric sense, independent of the tyranny of a particular coordinate system!" We refer the reader to [2, 5, 12, 13] for more on minimal surfaces.

4. Evolution and Gradient Flows. We next consider a temporal twist on our original question. Suppose a nonlinear curve is given on $\Omega = (-1, 1)$ with $u(-1) = u(1) = 0$, as in Figure 4.1. Here is the second fundamental question of this paper:

> What partial differential equation will flatten u as efficiently as possible?

In other words, if we evolve the curve according to a partial differential equation

$$u_t = A[u],$$

what should the operator A be in order to flatten u as fast as possible?

The ideas of the last section offer one possible approach to answering this question. Recall that we had

$$E[u] = \frac{1}{2}\int_\Omega |\nabla u|^2\, dx$$

as a measure of the flatness of u. Since u also now depends on t we should actually consider the time-dependent function $E[u](t)$,

$$E[u](t) = \frac{1}{2}\int_\Omega |\nabla u(x, t)|^2\, dx. \tag{4.1}$$

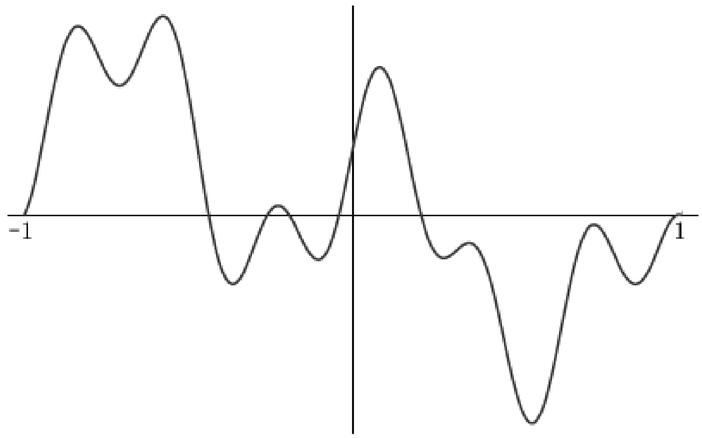


Fig. 4.1 *What evolutionary partial differential equation will flatten u as efficiently as possible?*

Here the gradient of u is taken with respect to the spatial variable x only.[3] Thus the partial differential equation that decreases E as fast as possible will have the effect of flattening the graph of u as fast as possible. Given this perspective, we can recast the second fundamental question as follows:

What partial differential equation will decrease $E[u](t)$ as fast as possible?

In order to understand this question, let's consider an analogous, but finite-dimensional, scenario. Suppose you are hiking and you happen to know the mountains' height above sea level is defined by the scalar function $h(x_1, x_2)$. How should you hike in order to decrease h as rapidly as possible?

If $\mathbf{x}(t) = (x_1(t), x_2(t))$ represents your path on the map as you walk, then the instantaneous change in height as you walk is defined by

$$\frac{d}{dt}h(\mathbf{x}(t)) = \frac{\partial h}{\partial x_1}(\mathbf{x}(t))\, x_1{}'(t) + \frac{\partial h}{\partial x_2}(\mathbf{x}(t))\, x_2{}'(t) = \nabla h(\mathbf{x}(t)) \cdot \mathbf{x}'(t),$$

or, in terms of the inner product $(\cdot\,,\cdot)$ in $\mathbb{R}^2$,

$$\frac{d}{dt}h(\mathbf{x}(t)) = (\nabla h, \mathbf{x}'(t)). \tag{4.2}$$

Thus knowing the gradient ∇h allows us to precisely measure our rate of change of height along any given path $\mathbf{x}(t)$. In particular, it shows that we can decrease $h(\mathbf{x}(t))$ most efficiently by making the right-hand side of (4.2) as negative as possible. Since $\mathbf{x}\cdot\mathbf{y} = |\mathbf{x}|\,|\mathbf{y}|\cos\theta$, where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$, it follows that $\nabla h(\mathbf{x}(t))\cdot\mathbf{x}'(t)$ is most negative when $\theta = \pi$, or $\mathbf{x}'(t)$ points opposite the direction of ∇h. In other words, we should take the path $\mathbf{x}(t)$ defined by

$$\mathbf{x}'(t) = -\nabla h(\mathbf{x}(t)). \tag{4.3}$$

[3]Although we consider here a one-dimensional example, the ideas of this section apply equally to surfaces and their higher-dimensional counterparts, and for this reason we use ∇u instead of $u'(x)$ throughout this section.

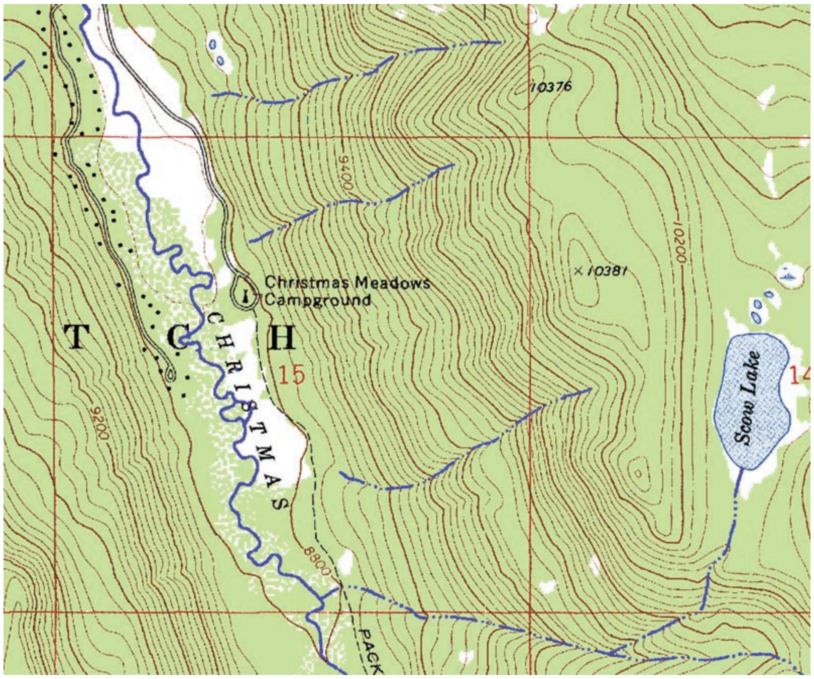


Fig. 4.2 *Topographic map near Christmas Meadows Campground, Utah. Note that the river lines run orthogonal to the level curves, providing a demonstration of the steepest descent flow. (Courtesy of Topozone, www.topozone.com.)*

The evolutionary law defined by (4.3) is known as the *law of steepest descent* and was introduced in 1847 by Cauchy [4] as a numerical tool for finding roots of a function $H(x_1, \ldots, x_n)$. The basic idea is to start at any point $(x_1, \ldots, x_n)$ and use steepest descent to decrease along the surface $y = H(x_1, \ldots, x_n)$. Since the gradient is orthogonal to the level curves of the surface, paths of steepest descent also proceed orthogonal to level curves. This phenomenon can often be seen on topographic trail maps, where the river lines tend to run orthogonal to the contour lines of constant altitude (see Figure 4.2).

Inspired by this finite-dimensional calculation, let us return to the partial differential equation problem of decreasing $E[u](t)$ as fast as possible. To decrease the height h as fast as possible one moves in the direction opposite to the gradient ∇h. If we could find a suitable notion for ∇E, then we could make sense of the equation $u_t = -\nabla E[u]$. The key insight to understanding ∇E is (4.2). If u lies in an inner product space H, then if

$$\frac{d}{dt}E[u](t) = (W, u_t)_H \tag{4.4}$$

for some $W \in H$, where $(\cdot, \cdot)_H$ is the given inner product on H, then we can define $\nabla E[u] := W$, just as we define $\nabla h[\mathbf{x}] := (\frac{\partial h}{\partial x_1}(\mathbf{x}), \frac{\partial h}{\partial x_2}(\mathbf{x})) \in \mathbb{R}^2$. In particular, the choice of the gradient $\nabla E[u] = W$ depends on the space H and the element u, just as $\nabla h[\mathbf{x}]$ depends on the space $\mathbb{R}^2$ and the point $\mathbf{x}$. Also note that $\nabla E[u]$ is an element of H, just as $\nabla h[\mathbf{x}]$ is an element of $\mathbb{R}^2$.[4] In this case, we can define the law of steepest descent in H to be

$$u_t = -\nabla E[u]. \tag{4.5}$$

What actual law (4.5) represents depends on the choice of E and the choice of inner product, of which there may be many. Thus, how efficient the method of steepest descent is depends on the mathematical setting, but with these tools we are able to precisely quantify any given setting. For example, a natural first choice is to let $H = L^2(\Omega)$, where

$$L^2(\Omega) = \left\{ w : \Omega \to \mathbb{R} \text{ such that } \int_\Omega w^2\, dx < \infty \right\}$$

is the space of square integrable functions on Ω (where dx is Lebesgue measure). The inner product on L^2 is

$$(u, v)_2 = \int_\Omega uv\, dx,$$

and the associated norm is

$$\|u\|_2 = \sqrt{(u,u)_2} = \sqrt{\int_\Omega u^2\, dx}.$$

With this inner product and norm, $L^2(\Omega)$ is a complete normed inner product space, or a *Hilbert space*. Using this setup we can find $\nabla E[u]$ for $E[u]$ defined by (4.1). We compute

$$\frac{d}{dt}E[u](t) = \frac{1}{2}\int_\Omega \frac{d}{dt}\left(\nabla u \cdot \nabla u\right)\, dx \tag{4.6}$$

$$= \int_\Omega \nabla(u_t) \cdot \nabla u\, dx \tag{4.7}$$

$$= \int_{\partial\Omega} u_t \nabla u \cdot \nu\, dS - \int_\Omega u_t \Delta u\, dx, \tag{4.8}$$

where we have used Green's identity (or integration by parts) in the last step. Since u is fixed on $\partial\Omega$ it follows that $u_t = 0$ and therefore

$$\frac{d}{dt}E[u](t) = \int_\Omega (-\Delta u) u_t\, dx = (-\Delta u, u_t)_2.$$

Comparing with (4.4), we see that we have found the gradient of E at u in $L^2(\Omega)$, namely, $\nabla E[u] = -\Delta u$, the negative of the Laplacian of u! We conclude that to decrease u as efficiently as possible in L^2 one should evolve u according to the rule $u_t = -(-\Delta u)$, or

$$u_t = \Delta u. \tag{4.9}$$

[4]Alternatively, the derivative of the map $h : \mathbb{R}^2 \to \mathbb{R}$ is defined as the best linear approximation $L : \mathbb{R}^2 \to \mathbb{R}$ of h at $\mathbf{x}$. By the Riesz representation theorem it is uniquely represented by an element in the domain $\mathbb{R}^2$, denoted $\nabla h(\mathbf{x})$. In particular, $L(\mathbf{v}) = (\nabla h(\mathbf{x}), \mathbf{v}) = \nabla h(\mathbf{x}) \cdot \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^2$. Similarly, the derivative of $E : H \to \mathbb{R}$ is defined as the best linear approximation of E at u and is also uniquely represented by an element W of H, with $\nabla E[u](v) = (W, v)_H$ for all $v \in H$.

Equation (4.9) is known as the *diffusion* or *heat equation*. In other words,

> The heat equation is the law of steepest descent for the Dirichlet energy integral in the Hilbert space $L^2(\Omega)$.

That is, we can understand heat flowing (or any diffusive process) as evolving the temperature (or density) in L^2 so as to efficiently decrease E, just as we think of rivers flowing down mountains so as to efficiently decrease the height h. As the river flows it traces out a path along the map in $\mathbb{R}^2$. As the temperature function flows, it also traces out a path, now in the function space $L^2(\Omega)$. In one dimension the heat equation takes the form $u_t = u_{xx}$. Thus, the infinitesimal time rate of change of u (i.e., where it flows next in the function space L^2) is governed by the concavity of u. The temperature will decrease at points which are concave down and increase at points which are concave up. The reader is invited to reexamine Figure 4.1 in this light and imagine how the curve would evolve.

What if we made a different choice for H? For example, consider $H = W_0^{1,2}(\Omega)$, the Sobolev space of $L^2(\Omega)$ functions with weak first derivatives in $L^2(\Omega)$ with the inner product

$$(u, v)_{1,2} = \int_\Omega \nabla u \cdot \nabla v \, dx, \tag{4.10}$$

where the derivatives are taken in the weak or generalized sense. In this case, using the same E, we calculate exactly as in (4.6)–(4.8) up to (4.7). At this point observe that (4.7) is precisely the $W_0^{1,2}(\Omega)$ inner product (4.10) of u with u_t. In other words, for this choice of Hilbert space we find

$$\frac{d}{dt}E[u](t) = \int_\Omega \nabla(u_t) \cdot \nabla u \, dx \tag{4.11}$$

$$= (u, u_t)_{1,2}. \tag{4.12}$$

Therefore $\nabla E[u] = u$ in $W_0^{1,2}(\Omega)$ and the law of steepest descent reads

$$u_t = -u.$$

In this case, the temperature decreases wherever u is positive and increases wherever u is negative, and this represents the law of steepest descent in the Sobolev space $W_0^{1,2}(\Omega)$. Thus steepest descent in $L^2(\Omega)$ is governed by the concavity of u, while steepest descent in $W_0^{1,2}(\Omega)$ is governed by the sign of u. Both approaches drive the solution to the equilibrium solution $u = 0$, but each one is efficient in its own space.

Finally, what if we made a different choice for $E[u]$, our measure of flatness? For example, what happens when we use the surface area function $E[u] = \int_\Omega \sqrt{1 + |\nabla u|^2}\, dx$ on the space $L^2(\Omega)$? In this case,

$$\frac{d}{dt}E[u](t) = \int_\Omega \frac{d}{dt}\left(\sqrt{1+|\nabla u|^2}\right) dx \tag{4.13}$$

$$= \int_\Omega \frac{1}{2}\left(1+|\nabla u|^2\right)^{-1/2} \frac{d}{dt}\left[\nabla u \cdot \nabla u\right] dx \tag{4.14}$$

$$= \int_\Omega \frac{\nabla u}{\sqrt{1+|\nabla u|^2}} \cdot \nabla(u_t)\, dx \tag{4.15}$$

Table 4.1

Measure of flatness	u-space	Gradient $\nabla E[u]$	Steepest descent
$E[\mathbf{u}] = h(\mathbf{u})$	$\mathbb{R}^n$	$\left(\frac{\partial h}{\partial u_1}(\mathbf{u}), \ldots, \frac{\partial h}{\partial u_n}(\mathbf{u})\right)$	$\mathbf{u}'(t) = -\nabla h(\mathbf{u}(t))$
$E[u] = \int_\Omega \lvert\nabla u\rvert^2 \, dx$	$L^2(\Omega)$	$-\Delta u$	$u_t = \Delta u$
$E[u] = \int_\Omega \lvert\nabla u\rvert^2 \, dx$	$W_0^{1,2}(\Omega)$	u	$u_t = -u$
$E[u] = \int_\Omega \sqrt{1 + \lvert\nabla u\rvert^2} \, dx$	$L^2(\Omega)$	$-\operatorname{div}\left(\frac{\nabla u}{\sqrt{1+\lvert\nabla u\rvert^2}}\right)$	$u_t = \operatorname{div}\left(\frac{\nabla u}{\sqrt{1+\lvert\nabla u\rvert^2}}\right)$

$$= \int_{\partial\Omega} u_t \frac{\nabla u}{\sqrt{1+|\nabla u|^2}} \cdot \nu \, dS - \int_\Omega \operatorname{div}\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right) u_t \, dx \tag{4.16}$$

$$= \left(-\operatorname{div}\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right), u_t\right)_{L^2(\Omega)}, \tag{4.17}$$

assuming, for instance, $u_t = 0$ on $\partial\Omega$. Thus, the law of steepest descent in $L^2(\Omega)$ with flatness measured by the surface area function is

$$u_t = \operatorname{div}\left(\frac{\nabla u}{\sqrt{1+|\nabla u|^2}}\right). \tag{4.18}$$

The right-hand side of (4.18) represents the mean curvature of the surface u, and this law is known as *flow by mean curvature*. That is, mean-curvature flow is the most efficient way of decreasing the surface area integral in $L^2(\Omega)$. Flow by mean curvature is another active area of research, and we refer the reader to [9, 7, 3] and the references therein for more on this fascinating subject.

The methods indicated here are also of interest for numerical applications. For instance, in numerical methods, a subject of current interest is the choice of Hilbert space, and hence the gradient, and its impact on numerical algorithms for steepest descent methods. Of particular importance for differential equations is the use of Sobolev spaces, their associated Sobolev gradients, and rates of convergence for steepest descent methods [11].

Table 4.1 summarizes our study of the second fundamental question of this paper.

5. Classroom Notes. Before discussing either of the fundamental questions in a particular class, it may be useful to leave students with the question at the end of the previous lecture to allow them sufficient time to develop possible approaches. Although the calculations in section 2 were presented in $\mathbb{R}^2$, depending on the class, the same calculation could be worked out in $\mathbb{R}^n$, the only difference being the surface measure on the boundary on the sphere. One might also want to prove Liouville's theorem for $\mathbb{R}^n$ (a bounded harmonic function $u : \mathbb{R}^n \to \mathbb{R}$ is constant) as an application of the mean-value property of harmonic functions.

Finally, in order to make the material accessible to a broad range of undergraduates, we have been slightly carefree with the introduction and use of function spaces. To properly handle such notions would require a discussion of weak derivatives or generalized functions (see, e.g., [1]). For example, the calculations for the heat equation and mean-curvature equation are rigorous in the function space $W^2(\Omega) \cap W_0^{1,2}(\Omega)$.

These issues lead to interesting properties of the various gradient flows. For instance, the heat equation will instantaneously flow $L^2(\Omega)$ initial data into $W_0^{1,2}(\Omega)$, while the gradient flow $u_t = -u$ requires $W_0^{1,2}(\Omega)$ initial data to begin with, and no further regularity is introduced. Mean-curvature flow can also lead to singularities in finite time [14]. These topics may be nice for class projects, or the instructor may wish to emphasize that the calculations only really hold for sufficiently smooth functions in the given spaces. The hope is that the ideas contained herein will motivate students to seek answers to these deeper questions.

Acknowledgments. The author thanks Paul Fife for introducing him to the notion of thinking *as flat as possible*, and for providing the inspiration for this article. The author also thanks Kevin Mapp for the photography, Darryl Yong for help with the figures, and Michael Orrison and the anonymous referees for their expert advice and insights that improved the paper.

REFERENCES

[1] R. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.
[2] C. V. Boys, *Soap Bubbles, Their Colours and the Forces Which Mold Them*, Dover, New York, 1959.
[3] K. A. Brakke, *The Motion of a Surface by Its Mean Curvature*, Math. Notes 20, Princeton University Press, Princeton, NJ, 1978.
[4] P. L. Cauchy, *Méthod générale pour la resolution des systemes d'équations simultanées*, C. R. Acad. Sci. Paris, 25 (1847), pp. 536–538.
[5] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, NJ, 1976. Translated from the Portuguese.
[6] L. C. Evans, *Partial Differential Equations*, AMS, Providence, RI, 1998.
[7] L. C. Evans and J. Spruck, *Motion of level sets by mean curvature*. I, J. Differential Geom., 33 (1991), pp. 635–681.
[8] J. Heinonen, T. Kilpeläinen, and O. Martio, *Nonlinear Potential Theory of Degenerate Elliptic Equations*, The Clarendon Press, Oxford University Press, New York, 1993.
[9] G. Huisken, *Flow by mean curvature of convex surfaces into spheres*, J. Differential Geom., 20 (1984), pp. 237–266.
[10] J. Marsden and A. Tromba, *Vector Calculus*, 5th ed., W. H. Freeman, New York, 2003.
[11] J. W. Neuberger, *Sobolev Gradients and Differential Equations*, Lecture Notes in Math. 1670, Springer-Verlag, Berlin, 1997.
[12] J. Oprea, *The Mathematics of Soap Films: Explorations with Maple*, Stud. Math. Libr. 10, AMS, Providence, RI, 2000.
[13] R. Osserman, *A Survey of Minimal Surfaces*, 2nd ed., Dover, New York, 1986.
[14] M. Struwe, *Evolution problems in geometry and mathematical physics*, in Prospects in Mathematics (Princeton, NJ, 1996), H. Rossi, ed., AMS, Providence, RI, 1999, pp. 83–101.
[15] W. A. Veech, *A converse to Gauss' theorem*, Bull. Amer. Math. Soc., 78 (1972), pp. 444–446.