Proceedings of the AMS Centennial Symposium
August 8–12, 1988

Regularity of Solutions and Level Surfaces of Elliptic Equations

LUIS A. CAFFARELLI

In many instances, the regularity theory of solutions to second-order equations may be thought of as a stability question; that is, as how a perturbation propagates along a solution surface.

For instance, it is well known that if one slightly perturbs a solution of the wave equation, for instance by changing the data in part of the boundary, the perturbation propagates only on certain directions or regions and therefore one may not expect local regularization effects. That is, regularity, in the few instances in which it can be proven, has to come from somewhere else, i.e., from a regular data.

On the other hand, for uniformly elliptic linear equations, small perturbations propagate all over the surface, in fact in a quantitative fashion, and that implies regularity and stability of such surfaces.

I would like to discuss today a series of nonlinear problems where degeneracies and discontinuities make the question of regularity of solutions and level surfaces (and this associated idea of propagation of perturbations) a very challenging one.

Elliptic equations and interior regularity. We start by discussing the notion of ellipticity, and the circle of ideas surrounding Harnack type inequalities.

In the nonvariational context one may loosely say that a continuous function u or surface is an elliptic equation if one may control the smallest (more negative) eigenvalue of its Hessian $D_y u$ buy its largest one. For instance, in the sense that

$$|\lambda_{\min}| \le F(\lambda_{\max}, x).$$

Of course, a continuous function has not necessarily a Hessian but one may avoid that by using the "viscosity method, i.e., requiring that such control exists for any C^2 function φ, whose graph manages to touch" the graph of u by above at x_0.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 35B30; Secondary 35B50, 35B05, 35B45, 35B65.

Of course, u is a solution if both u and $-u$ are subsolutions.

A remarkable theorem of Krylov-Safonov (the Harnack inequality) states that if the equation is "uniformly elliptic with a right-hand side in L^n," i.e.,

$$f(x) = \sup_{\lambda > 0} \frac{F(\lambda, x)}{\lambda + 1}$$

belongs to L^n, then, for any ball B_{2R} in the domain of definition of u, $\sup_{B_R} u \leq C \inf_{B_R} u + R^2 (f_{B_{2R}} |f|^n)^{1/n}$.

This is a very powerful theorem that implies, for instance, the Holder continuity of u at a point given the appropriate controlled growth of $\int f^n$.

In fact, this theorem is in turn very much inspired in its character and proof by DeGiorgi's work on the regularity of solutions of variational problems which is one of the great papers on partial differential equations. There, the ellipticity condition is given in "energy" terms, i.e., instead of considering functions u for which the eigenvalues of the Hessian are somewhat comparable, one looks at functions u whose energy is locally under control for the function and its truncations, i.e., for all λ,

$$\int (\nabla (u - \lambda)^+)^2 \leq C R^{-2} \int [(u - \lambda)^+]^2 B_R + R^{\alpha - 2}.$$

One may wonder at this point what is the relation between the first and second family of functions.

In the first case one is trying to say that at each point x

$$|\lambda_{\min}| \leq C(\lambda_{\max} + f(x)) \quad \text{and} \quad \lambda_{\max} \leq C(\lambda_{\min} + f(x)),$$

or fixing coordinates

$$|a_{ij}(x) D_{ij} u| \leq C f(x)$$

for $a_{ij}(x)$ a positive definite matrix changing discontinuously from point to point.

In the second case, one is saying that

$$|D_i(a_{ij} D_j u)| \leq C f(x)$$

with f of controlled growth.

The power of these regularity results can be understood when one applies them to the study of nonlinear equations of respectively nondivergence or divergence type, i.e., equations of the form

$$F(D^2 u) = 0 \quad \text{or} \quad D_i(F_i(\nabla u)) = 0.$$

In the first case, we say that the relation is elliptic if $F(M)$ is monotone in the space of symmetric matrices, and strictly elliptic if we have the further quantitative estimate (for N positive definite)

$$F(M) + C_1 \|N\| \leq F(M + N) \leq F(M) + C_2 \|N\|.$$

In the second case, if the vector field is coercive, then

$$\langle \vec{F}(p) - \vec{F}(q), p - q \rangle \geq 0,$$

or if it is strictly coercive, then

$$C_1 \|p - q\|^2 \geq \langle \vec{F}(p) - \vec{F}(q), p - q \rangle \geq C_2 \|p - q\|^2.$$

If one is allowed to derivate u, both definitions correspond to classical nonlinear equations $F^{ij} D_{ij}(u) = 0$, where, in the first case $F^{ij} = F^{ij}(D^2 u)$ and in the second case $F^{ij} = F^{ij}(Du)$ are strictly positive definite bounded matrices.

Here, the DeGiorgi and Krylov theorems become interesting when applied to first derivatives of the functions under consideration.

Indeed, the definitions of both divergence and nondivergence nonlinear equations embody a comparison principle, i.e., two solutions, u_1 and u_2, of the equation $F(D^2 u) = 0$ cannot "touch," i.e., if $u_1 \leq u_2$ for $X \neq X_0$ and $u_1(X_0) = u_2(X_0)$, then at such a point $D_{ij} u_1 < D_{ij} u_2$, contradicting the strict monotony of F.

Of course, this is not entirely correct, but it is so if for instance

$$F(D^2 u_1) = 0, \qquad F(D^2 u_2) \leq -\varepsilon$$

for some positive ε (so one may perform the old trick of looking at $u_2 - \varepsilon |X|^2$).

Now, it has been noted many times that a comparison principle for solutions of a translation invariant operator is related to a maximum principle for the first derivatives.

In our example of $F(D^2 u)$,

$$u_{\lambda, h} = u(X + he) + \lambda h$$

is again a solution of $F(D^2 u) = 0$ and therefore if $u_{\lambda_0, h} \geq u$ along the boundary of B_1, a ball in the domain of definition of both functions, then $u_{\lambda_{0,h}} \geq u$ in the interior of B_1.

Indeed, this is true for λ very large, and the comparison principle tells us that there is no "first λ" ($\lambda > \lambda_0$) for which $u_{\lambda, h}$ may touch u.

It follows that the supremum of the incremental quotient

$$\frac{-u(X + he) + u(X)}{h} = \Delta_h u$$

is attained at the boundary of B_1. One may think of this as the fact that

$$F(D^2 u_{\lambda, h}) = F(D^2 u) = F^{ij}(D^2 u(\xi)) D_{ij}(\Delta_h u).$$

That is, $\Delta_h u$ satisfies an elliptic equation with discontinuous coefficients.

But then Krylov's theorem says much more. It not only says that $\Delta_{\lambda, h} u$ is positive but that it is comparable at any two points of any smaller ball $B_{1/2} \ll B_1$.

That is, Harnack's inequality is a very strong, *quantitative form* of the maximum principle.

It tells us not only that the solution surfaces u and $u_{\lambda,h}$ separate, but that they do so *uniformly* and hence one can further translate u (to $u_{\lambda,h+\cdots}$) and still this translation will remain above the graph of u.

An iteration of this argument implies the Holder continuity of the first derivatives of u, i.e., bounded weak solutions of $F(D^2u) = 0$ are locally $C^{1,\alpha}$.

Let me stress the perturbation view of this result. We have "modified" the boundary data of u (to those of $u_{\lambda,h}$, slightly larger), and this perturbation propagated all over the domain in some uniform way. Since the operator under consideration is translation invariant, this implies $C^{1,\alpha}$ regularity of solutions.

If one wants to push this idea further, to second derivatives, a structural condition (concavity of F) is necessary to ensure that pure second-order incremental quotients are formally subsolutions of the "linearized" differential equations. Then (Evans, Krylov) one combines the fact that D^2u lies in a Lipschitz, elliptic hypersurface with this fact to control its oscillation.

One may view this approach geometrically the following way: If F is concave and (possibly degenerate) elliptic, and one envelopes the solution surface u by above by paraboloids of fixed quadratic part (or spheres of fixed radius), then the new surface $\overline{u}$ is a subsolution of the same equation.

Hence, if $D_{ij}u$ are bounded above at the boundary, then $\overline{u} \equiv u$ for a narrow enough choice of paraboloids, i.e., D_{ij} are bounded above in the interior. This is the maximum principle part of the argument.

The Harnack inequality may then be thought of as taking envelopes of the variable quadratic part, so as to improve control of the second derivatives. We will come back to this point later.

Free boundary problems and harmonic analysis in Lipschitz domains. Let us now look at problems where the solution completely degenerates past a certain value of u. For instance, the simplest example is that of minimizers of

$$J(u) = \int (\nabla u)^2 + X_{u>0}\, dx.$$

Such a minimizer is harmonic when positive, or negative, i.e., $F(D^2u) = 0$ with $F = \text{Trace}$, and therefore perturbations propagate "elliptically" in regions where u keeps a "strict sign."

But in view of the previous discussion, the interesting phenomena to study is how a perturbation crosses the surface of discontinuity (for ∇u) $\{u = 0\}$, i.e., how would a perturbation of order ε displace this surface? Would the new surface $\{u_\varepsilon = 0\}$ separate uniformly from $\{u = 0\}$ in the interior of the domain of definition? If so, does this "ellipticity" property of free boundaries imply its regularity?

In thinking about such a problem, we may naturally divide it into two parts.

The first part asks: How does this perturbation reach the boundary? At first the sets $\{u > 0\}$ and $\{u < 0\}$ are completely amorphous. Even for a C^{∞} function u there is not much you can say about how narrow or cuspidal a level set may become—the most elementary geometric obstructions one may find for our perturbation to effectively reach the free boundary.

At this point a beautiful link between the basic geometric properties of minimizers to these variational problems and the theory of harmonic measure in Lipschitz domains occurs.

In terms of perturbations of solutions, this theory says that if you have locally a domain that is, say, the intersection of a Lipschitz surface S (or more generally, a surface with a Harnack chain property (Jerison and Kenig)), then if you have a function u, harmonic and nonnegative, vanishing on S, and you perturb it, this perturbation arrives to the boundary in full. That is, if we have two harmonic functions and $u \leq u_{\varepsilon}$, $u_{\varepsilon}(X_0)/u(X_0) \geq 1 + \varepsilon$, then $u_{\varepsilon}/u \geq 1 + C\varepsilon$ uniformly along (any compact subset of) S. That is, $(u_{\varepsilon})_{\nu} \geq (1 + (\varepsilon))u_{\nu}$.

From the free boundary context, the nondegeneracy properties of u^+ and a curious monotonicity formula (Alt, Friedman, and myself) allow you to assert that the above domain satisfies exactly the Harnack chain condition.

Since the variational term $\chi_{u>0}$ translates into a jump relation between u_{ν}^+ and u_{ν}^-, this makes u_{ε} a strict subsolution of the free boundary problem.

The second part of the problem answers the question: How is this perturbation, whose influence is felt fully along the free boundary, forcing it to drift away.

That is, we are thinking of the perturbation as occurring in two steps. First we lift u^+ somewhere, but force S to stay fixed, and then we let S drift to S_{ε}, so that the energy attains equilibrium.

Since the free boundary still has almost no shape, it appears very difficult to construct such a perturbation. (As a parallel, a general strict comparison theorem for generalized minimal surfaces, due to L. Simon, is recent and delicate.)

Here we return to the question of variable parallel surface perturbation, to which we hinted at the end of the previous section.

Given a solution u to a general translation invariant equation

$$F(D^2 u, Du, u) = 0,$$

then

$$u^h(X) = \sup_{B_h(X)} u$$

is a subsolution to the same equation, and it is (heuristically) correct that u^h is also a subsolution to an "elliptic" free boundary jump condition since it increases u_{ν}^+ and decreases u_{ν}^-.

Therefore if u, v are solutions of a free boundary problem

$$F(D^2u, Du, u) = 0 \quad \text{for } |u| > 0$$

and $u_\nu^+ = G(u_\nu^-, \nu)$ with F elliptic (monotone in D^2u), G elliptic (monotone in u_ν^-), and $u^h \leq v$ on ∂B_1, then $u^h \leq v$ all over B_1.

But this is only a maximum principle, with no quantitative separation among u^h and v.

Suppose further that $(1+\varepsilon)u^h \leq v$ somewhere away from the free boundary. Can we now assert that the surfaces $\{u^h = 0\}$ and $\{v = 0\}$ are ε-away?

The answer is (in very loose terms) yes, provided that F has a Harnack inequality "up to boundary," and G is strictly monotone. This is done by what we could call variable level surface perturbations; that is, defining $u^\varphi = \sup_{\varphi(X)} u$ and asking when it is true that u^φ is again a subsolution of $F(D^2u, Du, u)$.

If F is uniformly elliptic with a Harnack inequality, one can see that it is enough for φ to satisfy an inequality of the type (L a Pucci extremal operator)

$$\varphi L\varphi > C|\nabla\varphi|^2$$

for u^φ to be a solution on the region $u > 0$.

This allows us to choose a $\varphi = h$ near $\partial\Omega$ (where we have only our original information) and a $\varphi > h$ (where we know that u_h is strictly less than $(1-\varepsilon)v$), and solve the inequality

$$\varphi L\varphi \geq C|\nabla|^2$$

in between, allowing the perturbation to travel *across* the free boundary. (φ variable distorts the free boundary relation, and u^φ has to be corrected using the "up to the boundary" Harnack inequality.)

The common setting. What is, then, the common setting for these problems? It is easy to approximate free boundary problems and problems of generalized surfaces of prescribed curvature relations by one-parameter families of solutions of operators,

$$F_\lambda(D^2u, Du, u) = 0,$$

that degenerate along a level surface $u = 0$. (For instance, solutions of $\Delta u = \beta_\varepsilon(u)$ with β_ε properly chosen, converge for ε going to zero to solutions of the obstacle problem, i.e., minimizers of $\int(\nabla u)^2 + u^+$, or the cavitational flow problem, i.e., minimizers of

$$\int(\nabla u)^2 + \chi_{u>0},$$

or of sets of minimal perimeter, i.e., characteristic functions χ_Ω that locally minimize "$\int|\nabla\chi_\Omega|\,dx$."

Further, the operation $u^h = \sup_{B_h} u$ constructs a new function u^h, whose level surfaces are parallel surfaces to those of u, i.e., the level surface $u_h = t$ is the surface of those points in $\{u < t\}$ whose distance to $\{u = t\}$ is exactly h, i.e., is the h-level surface of the distance function to $\{u = t\}$, and it is well known that the tangential Hessian increases along level surfaces of the distance function. In fact, it does so dramatically (recall the formula $x_i^h = x_i/(1 - x_i h)$ for the curvatures of the level surfaces of the distance function) if the curvatures of the original surface are large. Can we then, by looking at variable supremums u^φ or, what is related, looking at variable normal perturbations $d(X, S) = g(X)$, study how a smoothing effect propagates uniformly along level surfaces of u_λ (solution of F_λ) independently of λ?

Is it possible to infer regularity for level surfaces of F_λ independently of λ?

How elliptic (or hyperbolic) is a problem in λ; which type of perturbations travel in which direction?

How does a transient problem behave: Do perturbations travel fully in finite time to a free boundary. And many other questions related to Liouville type problems of elliptic or parabolic equations that would answer, after appropriate scaling, the fine structure of free boundaries, surfaces of prescribed curvature relations, and conservation laws.

To close, many efforts are under way that seem to indicate that there is indeed some substance behind these general comments.

School of Mathematics, Institute for Advanced Study, Princeton, New Jersey 08540