ICM 2002 · Vol. I · 179–187

Non Linear Elliptic Theory and the Monge-Ampere Equation

Luis A. Caffarelli*

Abstract

The Monge-Ampere equation, plays a central role in the theory of fully non linear equations. In fact we will like to show how the Monge-Ampere equation, links in some way the ideas comming from the calculus of variations and those of the theory of fully non linear equations.

2000 Mathematics Subject Classification: 35J15, 35J20, 35J70.

When learning complex analysis, it was a remarkable fact that the real part u of an analytic function, just because it satisfies the equation:

$$u_{xx} + u_{yy} = \Delta\, u = 0$$

(Laplace's equation) is real analytic, and furthermore, the oscillation of u in any given domain U, controls *all* the derivatives of u, of *any* order, in any subset $\bar{U}$, compactly contained in U.

One can give three, essentially different explanations of this phenomena.

a) Integral representations (Cauchy integral, for instance)**.** This gives rise to many of the modern aspects of real and harmonic analysis: fundamental solutions, singular integrals, pseudo-differential operators, etc. For our discussion, an important consequence of this theory are the Schauder and Calderon-Zygmund estimates.

Heuristically, they say that if we have a solution of an equation

$$A_{ij}(x) D_{ij} u = 0$$

and $A_{ij}(x)$ is, in a given functional space, a small perturbation of the Laplacian then $D_{ij}u$ is actually in the same functional space as A_{ij}. For instance, if $[A_{ij}]$ is Hölder continuous $(C^{\alpha}(\bar{U}))$ and positive definite, we can transform it to the identity (the Laplacian) at any given point x_0 by an affine transformation, and will remain close to it in a neighborhood. Thus $D_{ij}u$ will also be $C^{\alpha}(\bar{U})$.

*Department of Mathematics, University of Texas at Austin, Austin, TX 78712, USA. E-mail: caffarel@math.utexas.edu

b) Energy considerations. Harmonic functions, u, are also local minimizers of the Dirichlet integral

$$E(v) = \int (\nabla v)^2 \, dx \ .$$

That is, if we change u to w, in $\bar{U} \subset\subset U$

$$E(w)|_{\bar{U}} \geq E(u)|_{\bar{U}} \ .$$

This gives rise to the theory of calculus of variations (minimal surface, harmonic maps, elasticity, fluid dynamics).

One is mainly concerned, there, with equations (or systems) of the form

$$D_i F_i(\nabla u, X) = 0 \ . \tag{1}$$

For instance, in the case in which u is a local minimizer of

$$E(u) = \int \mathcal{F}(\nabla u, X) \, dx$$

(1) is simply the Euler-Lagrange equation associated to E:

$$F_i = \nabla_p \mathcal{F} \ .$$

If we attempt to write (1) in second derivatives form, we get

$$F_{i,j}(\nabla u, X) D_{ij} u + \cdots = 0 \ .$$

This strongly suggests that in order for the variational problem to be "elliptic", like the Laplacian, $F_{i,j}$ should be positive definite, that is $\mathcal{F}$ should be strictly convex.

It also leads to the natural strategy of showing that ∇u, that in principle is only in L^2 (finite energy), is in fact Hölder continuous. Reaching this regularity allows us to apply the (linear) Schauder theory.

That implies $D_{ij}u$ is $C^\alpha(\bar{U})$, thus ∇u is $C^{1,\alpha}(\bar{U})$, and so on (the bootstrapping method).

The difficulty with this approach is that solutions, u, are invariant under R^{n+1}-dialations of their graphs.

This fact keeps the class of Lipschitz functions (bounded gradients) invariant. There is no reason, thus, to expect that this equation will "improve" under dialations. The fact that ∇u is indeed Hölder continuous is the celebrated De Giorgi's theorem, that solved the nineteenth Hilbert's problem:

De Giorgi looked at the equation that first derivatives, u_α satisfy

$$D_i F_{ij}(\nabla u) D_j u_\alpha = 0.$$

He thought of $F_{ij}(\nabla u)$ as elliptic coefficients $A_{ij}(x)$ that had no regularity whatsoever, and he proved that any solution w of

$$D_i A_{ij}(x) D_j w = 0$$

was Hölder continuous

$$\|w\|_{C^\alpha(\bar U)} \le C\|w\|_{L^2(U)} \ .$$

De Giorgi's theorem is in fact a linear one, but for a new invariant class of equations. No matter how the solution (and the equation) is renormalized, it stays far from the constant coefficient theory, and a radically new idea surfaces: if we have a class of functions for which at every scale, in some average sense, the function controls its derivatives (the energy inequality), further regularity follows.

Finally, the third approach is

c) Comparison principle. Two solutions u_1, u_2 of $\Delta u = 0$ cannot "touch without crossing". That is, if $u_1 - u_2$ is positive it cannot become zero in some interior point, X_0, of U.

Again, heuristically, this is because the function

$$F(D^2u) = \Delta u = \mathrm{Trace}[D^2u]$$

is a monotone function of the Hessian matrix $[D_{ij}u]$ and, thus, in some sense, we must have $F(D^2u_1)$ "$>$" $F(D^2u_2)$ at X_0 (or nearby).

The natural family of equations to consider in this context, is then

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

for F a strictly monotone function of D^2u.

Such type of equations appear in differential geometry. For instance, the coefficients of the characteristic polynomial of the Hessian

$$P(\lambda) = \det(D^2u - \lambda I)$$

are such equations if we restrict D^2u to stay in the appropriate set of $R^{n\times n}$. If λ_i denote the eigenvalues of D^2u

$$\begin{aligned} C_1 &= \Delta u = \sum \lambda_i \qquad \text{(Laplace)} \\ C_2 &= \sum_{i\neq j} \lambda_k \lambda_j \ldots \\ C_n &= \prod \lambda_i = \det D^2u \qquad \text{(Monge-Ampere)} \ . \end{aligned}$$

In the case of $C_n = \det D^2u = \prod \lambda_i$ is a monotone function of the Hessian provided that all λ_i's are positive. That is, provided that the function, u, under consideration is convex.

If $F(D^2u, X)$ is *uniformly elliptic*, that is, if F is strictly monotone as a function of the Hessian, or in differential form,

$$F_{ij}(M) = D_{m_{ij}}F$$

is uniformly positive definite, then solutions of $F(D^2u)$ are $C^{1,\alpha}(\bar U)$. As in the divergence case, this is because first derivatives u_α satisfy an elliptic operator,

$$F_{ij}(D^2u)D_{ij}u_\alpha = 0$$

now in non divergence form. As long as we do not have further information on D^2u, we must think again of F_{ij} as bounded measurable coefficients.

The De Giorgi type theorem for $a_{ij}(x)D_{ij}u_\alpha = 0$ is due to Krylov and Safanov, and states again that solutions of such an equation are Hölder continuous.

We point out that, again this result has "jumped" invariance classes. Rescaling of $a_{ij}(x)$ does not improve them. Unfortunately, this is not enough to "bootstrap", as in the divergence case: The coefficients, $A_{ij}(x) = F_{ij}(D^2u)$, depend on second derivatives. If we will manage to prove that D^2u is Hölder continuous, then, from equation (1), $D_\alpha u$ would be $C^{2,\alpha}(\bar{U})$, i.e., u would be $C^{3,\alpha}(\bar{U})$ and we could improve and improve.

To prove this, once more convexity reappears. If $F(D^2u)$ is concave (or convex) then all pure second derivatives are sub (or super) solutions of the linearized operator. This, together with the fact that D^2u lies in the surface $F(D^2u)$, implies the Hölder continuity of D^2u, and, by the bootstrapping argument u is as smooth as F allows.

The Monge-Ampere equation and optimal transportation

We would like now to turn our attention to the Monge-Ampere equation

$$\det D^2u = \prod \lambda_i = f(x, u, \nabla u)\ .$$

As pointed out before, the equation fits in the context of elliptic equations provided that we consider convex solutions. That is, provided that f is positive. Further $\log\det D^2u = \sum \log\lambda_i$ is concave as function of the λ_i and thus is a concave function of D^2u. Unfortunately $\det D^2u$ *is not* uniformly strictly convex.

For instance if we prescribe

$$\det D^u = \prod \lambda_i = 1$$

ellipticity deteriorates as one of the λ's goes to infinity and some other is forced to go to zero. This difficulty is compensated by two fundamental facts.

1) The rich family of invariances that the Monge-Ampere equation enjoys.
2) Its "hidden" divergence structure.

The divergence structure is due to the fact that $\det D^2u$ can be thought of as the Jacobian of the gradient map: $X \to \nabla u$. Thus for any domain $\bar{U}$

$$\int_{\bar{U}} \det D^2u\, dx = \mathrm{Vol}(\nabla u(\bar{U})).$$

But if $\bar{U} \subset\subset U$, u being convex implies that

$$(\nabla u)|_{\bar{U}} \le C \operatorname{osc} u|_U.$$

This gives us a sort of "energy inequality" that controls a positive quantity of D^2u by the oscillation of u:

$$\int_{\bar{U}} \det D^2u \le C(\bar{U}, U)(\operatorname{osc} u)^n.$$

Invariances

The Monge-Ampere equation is invariant of course, under the the standard families of transformations:

a) Rigid motions, R:
$$\det D^2 u(Rx) = f(Rx),$$
b) Translations:
$$\det D^2 u(x+v) = f(x+v),$$
c) Quadratic dialations:
$$\det D^2 \frac{1}{t^2} u(tx) = f(tx).$$

But also

d) Monge-Ampere is invariant under any affine transformation A, of determinant one:
$$\det D^2 u(Ax) = f(Ax)\ .$$

If f is, for instance, in one of the following classes:

a) f constant,
b) f close to constant ($|f-1| \leq \varepsilon$),
c) f bounded away from zero and infinity ($0 < \frac{1}{\sigma} \leq f \leq \sigma$),

any of the transformations above gives a new u in the *same class* of solutions.

For instance, if u is a solution of
$$\det D^2 u = 1$$
then, $u(\varepsilon x, \frac{1}{\varepsilon} y)$ is also a solution of the *same equation*. But this has dramatically "deformed" the graph of u. It is then almost unavoidable that there are singular solutions (Pogorelov).

In fact, for $n \geq 3$, one can construct convex solutions u that contain a line their graph and are not differentiable in the direction transversal to that line, solutions of
$$\det D^2 u = f(x)$$
with f a smooth positive function.

Fortunately, this geometry can only be inherited from the boundary of the domain.

Theorem 0.1. *If in the domain* $U \subset R^n$

a) $\frac{1}{\sigma} \leq \det D^2 u \leq \sigma$,
b) $u \geq 0$,
c) The set $\Gamma = \{u = 0\}$ *is not a point, then* Γ *is generated as "convex combinations" of its boundary points*
$$\Gamma = \textit{ convex envelope of } \ \Gamma \cap \partial U\ .$$

A corollary of this theorem is that

a) If we can "cut a slice" of the graph of u, with a hyperplane $l(x)$ so that the support S of $(u-l)^-$ is compactly contained in U, then u is, inside S, both $C^{1,\alpha}$ regular and also $C^{1,\alpha}$- strictly convex, i.e., separates from any of its supporting planes with polynomial growth.

This is the equivalent of De Giorgi's and Krylov-Safanov result (remember that the C^α theorems were applied to the *derivatives* of the solutions of the non-linear equations under consideration).

Note that by an affine transformation and a dilation we can always renormalize the support of the "slice" S to be equivalent to the unit ball of R^n: $B_1 \subset S \subset B_n$.

After this normalization, it is possible to reproduce for u all the classical estimates we had for the Laplacian:

a) (Calderon-Zygmund). If f is close to constant ($|f-1| < \varepsilon$), then $D^2u \in L^p(B_{1/2})$ ($p = p(\varepsilon)$ goes to infinity when ε goes to zero).
b) If $f \in C^{k,\alpha}$ (has up to k derivatives Hölder continuous) then $u \in C^{k+2,\alpha}$ (all second derivatives of u are $C^{k,\alpha}$.

Note that f plays, for Monge-Ampere, simultaneously the role of "right hand side" and "coefficients" due to the structure of its non-linearity.

The Monge-Ampere equation and optimal transportation (the Monge problem)

The Monge-Ampere equation has many applications, not only in geometry, but also in applied areas: optimal design of antenna arrays, vision, statistical mechanics, front formation in meteorology, financial mathematics.

Many of these applications are related to optimal transportation and the Wasserstein metric between probability distributions. In the discrete case, optimal transportation consists of the following.

We are given two sets of k points in R^n: $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_k$, and want to map the X's onto the Y's, i.e., we look at all one-to-one functions $Y(X_j)$. But we want to do so, minimizing some transportation costs

$$\mathcal{C} = \sum_j C\Big(Y(X_j) - X_j\Big) .$$

For our discussion $C(X-Y) = \frac{1}{2}|X-Y|^2$. It is easy to see that the minimizing map must be the gradient (subdifferential) of a convex potential φ.

In the continuous case, instead of having k-points we have two probability densities, $f(X)\,dX$ and $g(Y)\,dY$ and we want to consider those (admissible) maps $Y(X)$ that "push forward" f to g.

Heuristically that means that in the change of variable formula, we can substitute

$$g(Y(X)) \det D_X Y(X) \text{“="} f(X).$$

A weak formulation, substitutes the map $Y(X)$, by a joint probability density $\nu(X,Y)$ with marginals $f(X)\,dX$ and $g(Y)\,dY$, i.e.,

$$f(X_0) = \int d_Y \nu(X_0, Y),$$
$$g(Y_0) = \int d_X \nu(X, Y_0).$$

(We don't ask the "map" to be one-to-one any more, the image of X_0 may now spread among "many Y's".

Among all such ν, we want to maximize correlation

$$\mathcal{K} = \int \langle X, Y \rangle d\nu(X,Y)$$

or minimize cost

$$\mathcal{C} = \int \frac{1}{2}|X - Y|^2\, d\nu(X,Y),$$

$\sqrt{\mathcal{C}}$ defines a metric, the Wasserstein metric among probability densities.

Under mild hypothesis, we have the

Theorem 0.2. *The unique optimal ν_0 concentrates in a graph (is actually a one-to-one map, $Y(X)$). Further $Y(X)$ is the subdifferential of a convex potential φ, i.e., $Y(X) = \nabla\varphi$. Heuristically, then, φ must satisfy the Monge-Ampere equation*

$$g(\nabla\varphi)\det D^2\varphi = f(X).$$

For several reasons, the weak theory does not apply in general, but one can still prove, for instance:

Theorem 0.3. *If f and g never vanish or if the supports of f and g are convex sets, the map $Y(X)$ is "one derivative better" than f and g.*

Some applications and current issues

a) It was pointed out by Otto, that the Wasserstein metric can be used to describe the evolution of several of the classical "diffusion" equations: heat equation, porous media, lubrication.

The idea is that a diffusion process for one equation with conservation of mass, consists of the balance of two factors: trying to minimize distance between consecutive distributions ($u(x,t_k)$ and $u(x,t_{k+1})$), plus trying to flatten or smooth (diffuse), $u(x,t_{k+1})$.

This fact has allowed to prove rates of decay to equilibrium in many of the classical equations, as well as a number of new phenomena. The fine relations between the discrete and continuous problems is an evolving issue (rate of convergence, regularity of the discrete problems, etc.).

b) Another family of problems, coming both from geometry and optimal transportation concerns the study of several issues on solutions of Monge-Ampere equations in periodic or random media.

b$_1$) Liouville type theorems: We start with a theorem of Calabi of Liouville type: Given a global convex solution of Monge-Ampere equation, $\det D^2u = 1$, u must be a quadratic polynomial. Suppose now that instead of RHS equal to one, we have a general RHS, $f(x)$. Given a global solution, to discover its behavior at infinity we may try to "shrink it" through quadratic transformations:

$$u_\varepsilon = \varepsilon^2 u\left(\frac{x}{\varepsilon}\right) \ , \quad \text{satisfies} \quad \det D^2 u_\varepsilon = \left(\frac{x}{\varepsilon}\right) \ .$$

Suppose now that f averages out at infinity, for instance f is periodic. Then due to the "divergence structure" of Monge-Ampere u_ε should converge to a quadratic polynomial.

Theorem 0.4. *Given a RHS $f(x)$, periodic, with average $⨍ f = a$*

i) Given any quadratic polynomial P with $\det D^2P = a$, there exists a unique periodic function w, such that

$$\det D^2(P + w) = f(x)$$

(w is a "corrector" in homogenization language).

ii) Conversely (Liouville type theorem): Given a global solution u, it must be of the form $P + w$.

What are the implications for homogenization? What can we say if $f(X, u, \nabla u)$ is periodic in X and u? What can we say if $f_\omega(x)$ is random in X?

b$_2$) Vorticity transport: (2 dimensions) Again in the periodic context we seek a "vorticity density", $\rho(X,t)$ periodic in X. At each time t, ρ generates a periodic "stream function", $\psi(X,t)$ by the equation

$$\det(I + D^2\psi) = \rho \ .$$

In turn, ψ generates a periodic velocity field $v = -(\psi_y, \psi_x)$ that transports ρ:

$$\rho_t + \operatorname{div}(v\rho) = 0 \ .$$

Given some initial data $\rho_0(x)$, what can we say about ρ?

If ρ_0 is a vorticity patch, $\rho_0(x) = 1 + \chi_\Omega$, does it stay that way?

If we choose ρ_0, ψ_0 so that $\rho_0 = F(\psi_0)$, that is $\det I + D^2\psi_0 = F(\psi_0)$, we have a stationary vorticity array, i.e., $\rho(X,t) \equiv \rho_0$.

What can we say, in parallel to the classic theory of rotating fluids, or plasma, where det is substituted by $\Delta\psi$?

c) Another area of research relates to optimal transportation as a natural "map" between probability densities. It has been shown that optimal transportation explains naturally interpolation properties of densities (of Brunn Minkowski type), monotonicity properties (like correlation inequalities that express in which way the probability density, g, is shifted in some cone of directions with respect to f), and concentration properties of g versus f (in which sense for instance, a log concave perturbation of a Gaussian is more concentrated than a Gaussian).

Of particular interest would be to understand optimal transportation as dimension goes to infinity. Since convex potentials are very stable objects, this would provide, under some circumstances, an "infinite dimensional" change of variables formula between probability densities.

d) Finally, one of my favorite problems is to understand the geometry of optimal transportation in the case in which the cost function $C(X-Y)$ is still strictly convex, but not quadratic. In that case, the optimal map is still related to a potential that satisfies

$$\det(I + D(F_j(\nabla\psi))) = \cdots$$

where F_j is now the gradient of the convex conjugate to C.

At this point, we have come full circle and we are now in a higher hierarchy, in a sort of Lagrangian version of the Euler-Lagrange equation from the calculus of variations.

In fact if we put an epsilon in front of D and linearize,

$$\det(I+\varepsilon D(F_j(\nabla\psi))) = 1+\varepsilon \ \mathrm{Trace}(D(F_j(\nabla\psi)))+O(\varepsilon^2) = 1+\varepsilon \ \mathrm{div} \ F_j(\nabla\psi) + O(\varepsilon^2).$$

Bibliographical references can be found in the books of J. Gilbarg-N. Trudinger, L.C. Evans and L.A. Caffarelli-X. Cabre for nonlinear PDE's; T. Aubin, I. Bakelman and C. Gutierrez for the Monge-Ampere equation, and the recent surveys by L. Ambrosio and C. Villani for optimal transportation.