The Monge-Ampère Equation and Optimal Transportation, an elementary review

Luis Caffarelli

Department of Math, UT Austin, Austin, TX 78712
caffarel@fireant.ma.utexas.edu

1 Optimal Transportation

Optimal transportation can be better described in the discrete case:

We are given "goods" sitting at k different locations, x_j, in $\mathbb{R}^n$, and we want to transport them to k new locations y_j.

We do not care which goods go to which point, and transporting them from x_i to y_j incurs a cost $C(x_i - y_j)$ (think of $C_p(x-y) = \frac{1}{p}|x-y|^p$).

We want to choose a delivery scheme $y(x)$ that would minimize the total cost:

$$\mathcal{J}(y) = \sum_j C\big(y(x_j) - x_j\big) \ ,$$

among all admissible transportation plans $y(x)$.

Of course, everything being finite, such a problem has a solution $y_0(x)$,

$$\mathcal{J}(y_0(x)) \leq \min \sum C(y_0(x_j) - x_j) \tag{1}$$

Clearly, this imposes some geometric condition on the map. For instance, suppose that $C = C_p$ (and in particular rotationally invariant).

If we take two points x_1, x_2 and their images $y_0(x_1)$, $y(x_2)$ we may wonder what does it mean to switch them (that would increase cost). We can, for instance, take a system of coordinates where $x_1 = 0$, $x_2 = \lambda e_1$. Then, $y_0(x_1)$, $y_0(x_2)$ can be rotated with respect to this axis to make the configuration coplanar without changing cost.

This reduces the question to a problem in the plane and we see that for each position $y_0(x_1) = \alpha e_1 + \beta e_2$, $y_0(x_2)$ is forced to stay in some predetermined region above $y_0 = Ry_0$. That is, the map has to have some monotonicity.

For instance, in the case $p = 1$ (the usual Euclidean distance) we see that the vectors from x_i to y_{0_i} should not cross. For $p = 2$, instead the map $y_0(x)$ has to be monotone, i.e.,

$$\langle y(x_1) - y_(x_2), x_1 - x_2 \rangle \geq 0 \ .$$

L. Ambrosio et al.: LNM 1813, L.A. Caffarelli and S. Salsa (Eds.), pp. 1–10, 2003.
© Springer-Verlag Berlin Heidelberg 2003

2 The continuous case:

In the continuous case, instead of having two finite families of locations, we are given two "goods densities", $f(x)$ and $g(y)$; that is we have our goods "spreaded" through $\mathbb{R}^n$ with density $f(x)$ and we want to reorganize them, so they become "spreaded" with density $g(y)$. (There is an obvious compatibility condition $\int f = \int g$.)

So, heuristically, we are looking for a map, $y_0(x)$, that will reorganize f into g, and that, among all "admissible" maps will minimize the transportation cost

$$\mathcal{J}(y) = \int C(y-x)f(x)\,dx$$

Admissible means that for any set A, the total f-mass of A, be identical to the g-mass of its image, or infinitesimally (if such a thing were allowed)

$$g\big(y(x)\big)\det D_x y = f(x)\ .$$

A weak way of expressing admissibility is that, for any continuous test function $\eta(y)$

$$\int \eta(y)g(y)\,dy \text{ ``=''} \int \eta\big(y(x)\big)g\big(y(x)\big)\det y_x(x)\,dx$$

be actually equal to

$$\int \eta\big(y(x)\big)f(x)\,dx\ .$$

That is, we "allow" ourself to replace the formal term

$$g\big(y(x)\big)\det D_x y \text{ by } f(x)\ .$$

A still "weaker" formulation of the problem can be proposed if we agree that it is not necessary to require that whatever is located at the point x has to be mapped to a single point y, but that we may "spread" it around and vice versa that the necessary "quota" at y may be filled by a combination of x's.

In this case we could make a "table", $h(x,y)$ of how much of $f(x)\,dx$ goes into $g(y)\,dy$ and the "shipping plan" becomes a joint probability density $h(x,y)$, with marginals

$$f(x) = \int h(x,y)\,dy$$
$$g(y) = \int h(x,y)\,dx$$

Our minimization problem is, then, minimize

$$\int C(x-y)h(x,y)\,dx\,dy$$

among all such h.

Confronted with the problem one is of course tempted to look at the two particular cases

$$C(x-y) = |x-y| \quad \text{or} \quad |x-y|^2 \ .$$

The case $|x-y|$ is the original Monge problem. The quadratic case arises in many applications, particularly in fluid dynamics. Let us start by discussing the case $C(\xi) = |\xi|^2$:

The first observation or calculation that comes naturally to mind is that if $y_0(x)$ is the minimizing map, any "permutation" of k images would increase cost.

This can be better expressed in the discrete case, and we may hope it will lead us to an "Euler like" equation for this variational problem.

If π is a permutation of the x_j's it simply reads

$$\sum |y_0(x_j) - x_j|^2 \leq \sum |y_0(x_{\pi_j}) - x_j|^2$$

after some simplification

$$\sum \langle y_0(x_j), x_j \rangle \geq \sum \langle y_0(x_{\pi_j}), x_j \rangle \ .$$

This condition is called "cyclical" monotonicity of the map (in the case of two points x_1, x_2, it is the classical monotonicity condition of the map $\langle y(x_2) - y(x_1), x_2 - x_1 \rangle \geq 0$) and a theorem of Rockafeller asserts that the map $y(x)$ is "cyclically" monotone if and only if it is the subdifferential of a convex function $\varphi(x)$, what we would call a "convex potential" in the spirit of fluid dynamics.

To understand the meaning of $\varphi(x)$, we should go to the dual problem, that is the "shippers" point of view:

3 The dual problem:

Suppose that a shipping company wants to bid for the full transportation business. It has to charge each initiation point x_i, an amount $\mu(x_i)$ and any arrival point y_j an amount $\nu(y_j)$.

But it is constrained to charge $\mu(x_i) + \nu(y_j) \leq C(x_i - y_j)$.

If not x_i and y_j would leave the coalition and find another shipper. So the shipper wants to maximize

$$\sum \mu(x_i) + \sum \nu(y_i)$$

with the constraint that

$$\mu(x_i) + \nu(y_j) \leq C(x_i - y_j) \ .$$

In the continuous case, this becomes:
Maximize
$$\mathcal{K}(\mu,\nu) = \int \mu(x)f(x)\,dx + \int \nu(y)g(y)\,dy$$
with the constraint that
$$\mu(x) + \nu(y) \leq C(x-y)\ .$$
In principle, we would like to try and maximize this quantity $\mathcal{K}(\mu,\nu)$, for a reasonable family of admissible functions, say continuous.

But we note that given an admissible pair (μ,ν), we can find a better one μ, ν^*, by replacing $\nu(y)$ by
$$\nu^*(y) = \inf_x C(x-y) - \mu(x)$$
Indeed, ν^* is again admissible and $\nu^*(y) \geq \nu(y)$ for any y.

Similarly we can change μ to μ^*.

Thus the minimization process can be done in a much better class of functions, that we can call C-concave, that are of the form
$$\mu = \inf_{y\in S} C(x-y) + \nu(y)$$
(Note that if $C(x-y) = (x-y)^2$
$$-\varphi(x) = \mu(x) - |x|^2 = \inf_{y\in S} -2\langle x,y\rangle + |y|^2 - \nu(y)$$
Thus $\varphi(x)$ and $\varphi(y) = |y|^2 - \nu(y)$ are regular convex functions).

4 Existence and Uniqueness:

Since, for x, y varying in a bounded set the family of functions $C(x-y)$ is (equi) Lipschitz, it is not hard to pass to the limit and obtain a maximizing pair:

Theorem 4.1.

a) *There exists a unique maximizing pair μ_0, ν_0.*
b) *For any $x \in \Omega$, there exists at least a $y(x)$, for which*
$$C(x-y) = \mu(x) + \nu(y)\ .$$

Further, if C is strictly convex and smooth ($C^{1,\alpha}$), on can prove

Theorem 4.2.

a) *$y(x)$ is unique a.e. and the map $x \to y(x)$ is the unique optimal transportation.*

b) $y(x)$ is also defined by

$$\nabla C(x-y) = \nabla \nu(x)$$

or

$$y = x + \big(\nabla E\big)\big(\nabla \nu(x)\big)$$

where (∇E) is the gradient of the Legendre transform of C, (for instance, for $C = \frac{1}{p}|x|^p$, $E = \frac{1}{q} + |x|^q$ $(\frac{1}{p} + \frac{1}{q} = 1)$)

Theorem 1 is not hard, all we have to remember is that the minimization process was originally done in the space of continuous functions, that we are allowed to make just continuous perturbations of μ, ν, not necessarily C-concave ones, and hence if for some x, $\nu(x) + \mu(y) < C(x-y)$ for all $y \in \bar{\Omega}_2$, we can increase a little bit ν near x, and keep it admissible.

Theorem 2 is more delicate. The main ideas are due to Brenier:

To show that $y(x)$ is admissible, that is, according to our definition, that

$$\int \eta(y) g(y)\, dy = \int \eta\big(y(x)\big) f(x)\, dx$$

corresponds basically to the Euler equation of the variational problem:

We perturb $\nu(y)$ to

$$\nu_\varepsilon = \nu(y) + \varepsilon \eta(y)$$

and $\mu(x) = \inf_y C(x-y) - \nu(y)$ to

$$\mu_\varepsilon(x) = \inf_y C(x-y) - \nu_\varepsilon(y)$$

to keep the pair μ_ε, ν_ε admissible. Thus by maximality

$$\int (\nu - \nu_\varepsilon)(y) g(y)\, dy + \int (\mu - \mu_\varepsilon)(x) f(x)\, dx = \varepsilon[\mathrm{I} + \mathrm{II}] \geq 0$$

where

$$\mathrm{I} = -\int \eta(y) g(y)\, dy$$

and

$$\mathrm{II} = \int \frac{\mu - \mu_\varepsilon}{\varepsilon}(x) f(x)\, dx$$

But by definition

$$\left| \frac{\mu - \mu_\varepsilon}{\varepsilon} \right| \leq \sup \eta < C\ .$$

When ε goes to zero, by dominated convergence II_ε converges to the integral of the a.e. limit of $\frac{\mu - \mu_\varepsilon}{\varepsilon} f(x)$.

But, if the $y(x)$ for which the infimum in

$$\mu(x) = \inf C(x-y) - \nu(y)$$

is attained is unique, then $\frac{\mu - \mu_\varepsilon}{\varepsilon}(x)$ converges to $\eta\big(y(x)\big)$.

Once we know that $y(x)$ is admissible, if $z(x)$ is any other admissible map, we write

$$\begin{aligned}\int C\big(z(x)-x\big)f(x)\,dx &\geq \int \big[\nu\big(z(x)\big)+\mu(x)\big]f(x)\,dx \\ &= \int \nu(y)g(y)\,dy+\mu(x)f(x)\,dx\end{aligned}$$

(since z is admissible) while

$$\begin{aligned}\int C\big(y(x)-x\big)f(x) &= \int \big[\nu\big(y(x)\big)+\mu(x)\big]f(x) \\ &= \text{(by definition of } y) \\ &= \int \nu(y)g(y)\,dy+\int \mu(x)f(x)\,dx\end{aligned}$$

(since y is also admissible).

5 The potential equation:

We find ourselves now in a very good position: Our mapping problem (usually a relatively hard one) has been reduced to study a single function, the potential $\mu(x)$.

The variational process made out of μ the potential of an admissible map.

That is, heuristically

$$g(y)\det D\big(y(x)\big)=f(x)\ .$$

Replacing $y(x)$ by its formula

$$y=x+\nabla E(-\nabla\mu)$$

we obtain

$$g\big(x+\nabla(E(-\nabla\mu(x)))\big)\cdot\det\big(I+D(\nabla E(-\nabla\mu))\big)=f(x)$$

In the case in which

$$C(x)=\frac{1}{2}|x|^2=E(x)$$

writing $x=\nabla(\frac{1}{2}|x|^2)$ and $\varphi(x)=\frac{1}{2}|x|^2-\mu(x)$ the equation becomes

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

That is, φ is convex and it satisfies (formally) a Monge-Ampère type equation.

In the general case, always computing formally, we get

$$g(\cdots)\det\big[I+E_{ij}(-\nabla\mu)D_{ij}\mu\big]=f$$

Or if we multiply by C_{ij}, that happens to be the inverse matrix to E_{ij}, we get

$$g^*\det\big(D_{ij}\mu+C_{ij}(\nabla\mu)\big)=f$$

6 Some remarks on the structure of the equation

This last equation has a structure similar to that of the Monge-Ampère equation in a Riemannian manifold but instead of having a first order term with the structure $a_{ij}D_i\mu D_j\mu$, has the more complex one

$$E_{ij}(\nabla\mu)$$

resembling a Finsler metric term.

In confronting this equation two issues arise.

The first one, in which sense are the equations satisfied, and the second one, even if the equations are satisfied in the best of all possible senses, what can we expect of the structure of the set of solutions.

Let us answer first the second question: what can we expect about "nice" solutions of an equation of the form

$$\det\left(D^2\mu + C_{ij}(\nabla\mu)\right) = h$$

The left hand side satisfies heuristically a comparison principle (Given μ_1, μ_2 two solutions, the difference "should not" have a maximum in the interior) and it is translation invariant. Therefore, directional derivative of μ should have a "maximum principle".

Once could explore, thus, monotonicity properties of the optimal map, that is, what properties of f and g imply $y(x) \succ x$ in some order, $\succ$, in the spirit of $[C]$.

On the other hand, in the spirit of the regularity theory for fully nonlinear equations, one may ask if there are second derivative estimates. Here one should draw a parallel with solutions of divergence type equations coming from the calculus of variations. Indeed

$$\operatorname{div} D\nabla E(\nabla\mu) \tag{a}$$

very much resembles a linearization of

$$\det\left(I + \varepsilon D(\nabla E(\nabla\mu))\right) \tag{b}$$

and in general (a) possess no second derivative estimates.

So we feel that one should not expect, in general, second derivative estimates for (b) unless E is very special.

This is a very serious obstacle to a regularity theory for μ, since the linearized equation involves, as usual, second derivatives of μ in its coefficients.

About the first question, in what sense is the equation satisfied, we also have a serious difficulty. Let's consider the simple quadratic case, in which the equation is simple Monge-Ampère:

$$\det D^2\varphi \ldots$$

and φ is convex.

Since convex functions are in principle only Lipschitz, in order to make sense of $\det D^2\varphi$, the natural approach has been, precisely to look at the gradient map

$$\nabla\varphi : \Omega \to \mathbb{R}^n$$

and give an interpretation of $\det D^2$ as the Jacobian of such a map, that is the ratio between the volume of the image of a set and the volume of the set

$$\frac{|\nabla\varphi(S)|}{|S|}$$

The problem is that $\nabla\varphi(S)$ can be thought in the L^∞ sense (i.e. $\nabla\varphi \in L^\infty_{\text{loc}}$ and thus is defined a.e.) or in the maximal monotone map, i.e. $\nabla\varphi(S)$ is the set of gradients of all supporting planes to the convex function φ, on the set S.

The difference is clear in the case of $\varphi(x) = |x|$ in just one variable.

In the a.e. sense $\nabla\varphi = \pm 1$ and hence $D_{xx}\varphi = 0$. In the maximal monotone map $\nabla\varphi(0)$ is the full interval $[-1, 1]$ and thus $D_{xx}\varphi$ is the expected Dirac's δ at the origin.

Of course, this second definition is the correct one from almost any point of view, but unfortunately optimal transportation does not care much about sets of measure zero:

$$\varphi(x) = |x| + \frac{1}{2}|x|^2$$

is the optimal transportation from say $\Omega_1 = [-1, 1]$ to $\Omega_2 = [-2, -1] \cup [1, 2]$, with densities $f(y) = g(y) = 1$ but unfortunately $D^2\varphi$ has the extra density δ_0 at the origin.

Always in the particular case of $C(x) = \frac{1}{2}|x|^2$, the natural geometric condition to impose on Ω_2 to avoid this difficulty is simple: Ω_2 must be convex.

Indeed it is easy to see that the difference between $\det D^2\varphi$ (a.e.) and $\det D^2\varphi$ (max. mon.) consists of a singular measure, whose image by the gradient map is always contained in the (closure of) the convex envelope of the image of the regular part.

Thus, if Ω_2 is convex, and since we have the compatibility condition that $\int_{\Omega_2} f(x) = \int_{\Omega_2} g(y)$, any extra singular measure has nowhere to go inside Ω_2.

Once we know that φ satisfies the equation in the maximal monotone sense (Alexandrov sense) there exists a reasonable local regularity theory that asserts that the map $\nabla\varphi$ is, as expected, "one derivative better" than $f(x)$, $g(y)$.

For general cost functions is is not clear how to proceed both to find the appropriate condition on Ω_2, and how to develop at least the first steps on a local regularity theory, asserting for instance that the map is continuous.

Part of the difficulty is that convexity is simultaneously a local (infinitesimal) and global condition, that is $D^2\varphi \geq 0$ or the graph of φ stays below the segment joining two points, while in the general case that does not seem to be the case (with $-C(x-y)$ replacing linear functions and $-\big[D^2\mu + E_{ij}(\nabla\mu)\big]$ replacing $D^2\varphi$.

This becomes evident for instance with the issue of $\nabla\mu$ (a.e.) versus $\nabla\mu$ (maximal monotone).

Since $\mu = \inf_y C(x-y) - \nu(y)$ one may suggest that $\nabla\mu$ (maximal monotone) should be the $\nabla\mu(x)$ coming from all supporting cost functions $C(x-y)$ at the point x. But a cost function $C(x-y)$ that supports μ locally near x does not necessarily support it globally.

A possible way out is to study these cases for which both $f(x)$, $g(y)$ do not vanish, thus $\Omega_1 = \Omega_2 = \mathbb{R}^n$ or periodic problems ($f(x)$, $g(y)$ periodic) where hopefully, whatever singular part one should add to the a.e. definition of the map, has nowhere to go again.

Finally, some remarks on the case $C(x-y) = |x-y|$, the classic Monge problem.

In this case, strict convexity is lost and therefore ∇C is not anymore a "nice" invertible map from $\mathbb{R}^n$ to $\mathbb{R}^n$.

So if one tries to reconstruct the map from μ, ν the solution pair of the dual problem, $\nabla\mu$ only gives us the general direction of the optimal transportation, i.e.

$$\nabla\mu(x) = \frac{y-x}{|y-x|}$$

but not the distance.

In fact, the optimal allocation is, in general, not unique: If we want to transport, in one dimension the segment $[-2,-1] = I_0$ onto $[1,2] = I_1$, the cost function $|y-x|$ becomes simply $y-x$, and from the change of variable formula, we see that any measure preserving transformation $y(x)$ is a minimizer.

So, for instance we can split I_0, I_1 into a bunch of little intervals and map each other more or less arbitrarily and still have a minimizer.

On the other hand, in more than one dimension, optimal maps have a strong geometric restriction: If $x_1 \to x_2$ and $y_1 \to y_2$, the segments $[x_1, y_1]$, $[x_2, y_2]$ cannot intersect. (Remember the discrete case.)

Further, if x_2 lies in the interior of $[x_1, y_1]$, y_2 must be aligned to $[x_1, y_1]$.

Otherwise transportation can be improved.

Thus, transportation is aligned along "transportation rays". That is, if x_1 goes to y_1, all points in $\Omega_1 \cap [x_1, y_1]$ are mapped to the line containing $[x_1, y_1]$.

This is easy to visualize if for instance Ω_1 is contained in $\{x : x_n < 0\} = \mathbb{R}^n_-$, and Ω_2 in $\{x : x_n > 0\} = \mathbb{R}^n_+$.

Then transportation occurs along rays going from left to right.

There are several geometrical quantities one can look at to try to understand what is going on.

For instance, if Ω_1 is a long vertical rectangle and Ω_2 a horizontal one, one can see that in general, we cannot expect a "nice, clean" foliation of Ω_1, Ω_2 by transport rays, and that the domains must split in patches each one foliated by these "transportation" rays.

Another observation, from the definition of μ, ν as

$$\mu(x) = \inf_{y \in S} C(x-y) - \nu(y)$$

we see that if we define the two auxiliary functions

$$h(z) = \inf_{y \in \Omega_2} |z-y| - \nu(y)$$

and

$$g(z) = \sup_{x \in \Omega_1} \mu(x) - |z-x|$$

then $h(z) = g(z)$ along transportation rays, where both are linear.

Since along $x_n = 0$, h is $C^{1,1}$ by above (quasiconcave) and g $C^{1,1}$ by below (quasiconvex). The direction $\mathcal{T}(x)$ of transportation rays is Lipschitz along $\{x_n = 0\}$.

Two possible ideas to construct a solution are then to write an equation for the potential and the infinitesimal transportation along these rays, given by the mass balance (this is the approach of Evans-Gangbo) or to pass to the limit on a strictly ε-approximation of the limiting problem.

This last approach was worked out by Trudinger and Wang, and Feldman, McCann and the author, and was recently completed by L. Ambrosio and A. Pratelli by incorporating higher order T-convergence arguments to obtain strong convergence of the map.

Bibliographical Remarks: There is extensive work in this area.

Instead of attempting a long list of references let me mention a few names whose work is addressed. General issues of existence and regularity, that can be easily traced in the MathSciNet: L. Ambrosio, Y. Bremier, L.C. Evans, M. Feldman, W. Gangbo, R. McCann, N. Trudinger, J. Urbas, X.J. Wang.