Partial Differential Equations and Monge–Kantorovich Mass Transfer

Lawrence C. Evans*
Department of Mathematics
University of California, Berkeley

September, 2001 version

*Supported in part by NSF Grant DMS-94-24342. This paper appeared in Current Developments in Mathematics 1997, ed. by S. T. Yau

In Memory of
Frederick J. Almgren, Jr.
and
Eugene Fabes

1 Introduction

These notes are a survey documenting an interesting recent trend within the calculus of variations, the rise of differential equations techniques for Monge–Kantorovich type optimal mass transfer problems. I will discuss in some detail a number of recent papers on various aspects of this general subject, describing newly found applications in the calculus of variations itself and in physics. An important theme will be the rather different analytic and geometric tools for, and physical interpretations of, Monge–Kantorovich problems with a uniformly convex cost density (here exemplified by $c(x,y) = \frac{1}{2}|x-y|^2$) versus those problems with a nonuniformly convex cost (exemplified by $c(x,y) = |x-y|$). We will as well study as applications several physical processes evolving in time, for which we can identify optimal Monge–Kantorovich mass transferences on "fast" time scales.

The current text corrects some minor errors in earlier versions, improves the exposition a bit, and adds a few more references. The interested reader may wish to consult as well the lecture notes of Urbas [U1] and of Ambrosio [Am] for more.

1.1 Optimal mass transfer

The original transport problem, proposed by Monge in the 1780's, asks how best to move a pile of soil or rubble ("déblais") to an excavation or fill ("remblais"), with the least amount of work. In modern parlance, we are given two nonnegative Radon measures $\mu^\pm$ on $\mathbb{R}^n$, satisfying the overall *mass balance* condition

$$\mu^+(\mathbb{R}^n) = \mu^-(\mathbb{R}^n) < \infty, \tag{1.1}$$

and we consider the class of measurable, one-to-one mappings $\mathbf{s} : \mathbb{R}^n \to \mathbb{R}^n$ which rearrange μ^+ into μ^-:

$$\mathbf{s}_\#(\mu^+) = \mu^-. \tag{1.2}$$

In other words, we require

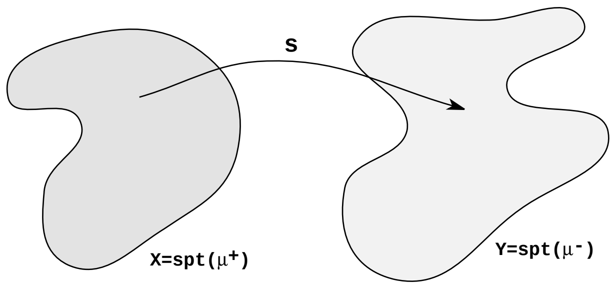
$$\int_X h(\mathbf{s}(x))\, d\mu^+(x) = \int_Y h(y)\, d\mu^-(y) \tag{1.3}$$

for all continuous functions h, where $X = \operatorname{spt}(\mu^+)$, $Y = \operatorname{spt}(\mu^-)$. We denote by $\mathcal{A}$ the admissible class of mappings $\mathbf{s}$ as above, satisfying (1.2), (1.3)

Given also is the *work* or *cost density* function

$$c : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty);$$

so that $c(x,y)$ records the work required to move a unit mass from the position $x \in \mathbb{R}^n$ to a new position $y \in \mathbb{R}^n$. (In Monge's original problem $c(x,y) = |x-y|$; that is, the work is simply proportional to the distance moved.)



The total *work* corresponding to a mass rearrangement plan $\mathbf{s} \in \mathcal{A}$ is thus

$$I[\mathbf{s}] := \int_{\mathbb{R}^n} c(x, \mathbf{s}(x))\, d\mu^+(x). \tag{1.4}$$

Our problem is therefore to find and characterize an *optimal mass transfer* $\mathbf{s}^* \in \mathcal{A}$ which minimizes the work:

$$I[\mathbf{s}^*] = \min_{\mathbf{s} \in \mathcal{A}} I[\mathbf{s}]. \tag{1.5}$$

In other words, we wish to construct a one-to-one mapping $\mathbf{s}^* : \mathbb{R}^n \to \mathbb{R}^n$ which pushes the measure μ^+ onto μ^- and, among all such mappings, minimizes $I[\cdot]$. We will later see that a really remarkable array of interesting mathematical and physical interpretations follow.

This is even now, over two hundred years later, a difficult mathematical problem, owing mostly to the highly nonlinear structure of the constraint. For instance, if $\mu^\pm$ have smooth densities $f^\pm$, that is, if

$$d\mu^+ = f^+ dx,\ d\mu^- = f^- dy, \tag{1.6}$$

then (1.2) reads

$$f^+(x) = f^-(\mathbf{s}(x)) \det(D\mathbf{s}(x)) \quad (x \in X), \tag{1.7}$$

where we write $\mathbf{s} = (s^1, \dots, s^n)$ and

$$D\mathbf{s} = \begin{pmatrix} s^1_{x_1} & \dots & s^1_{x_n} \\ & \ddots & \\ s^n_{x_1} & \dots & s^n_{x_n} \end{pmatrix}_{n \times n} = \text{ Jacobian matrix of the mapping } \mathbf{s}.$$

It is not at all apparent offhand that there exists any mapping, much less an optimal mapping, satisfying this constraint. Additionally, if $\{\mathbf{s}_k\}_{k=1}^{\infty} \subset \mathcal{A}$ is a minimizing sequence,

$$I[\mathbf{s}_k] \to \inf_{\mathbf{s} \in \mathcal{A}} I[\mathbf{s}],$$

an obvious guess is that we can somehow pass to a subsequence $\{\mathbf{s}_{k_j}\}_{j=1}^{\infty} \subset \{\mathbf{s}_k\}_{k=1}^{\infty}$, which in turn converges to an optimal mass allocation plan $\mathbf{s}^*$:

$$\mathbf{s}_{k_j} \to \mathbf{s}^*.$$

However, there is no clear way to extract such a subsequence, converging in any reasonable sense to a limit. The direct methods of the calculus of variations fail spectacularly, as there are no terms creating any sort of compactness built into the work functional $I[\cdot]$. In particular, $I[\cdot]$ does not involve the gradient of $\mathbf{s}$ at all and so is not coercive on any Sobolev space. And yet, as we will momentarily see, precisely this feature opens up the problem to methods of linear programming.

1.2 Relaxation, duality

Kantorovich in the 1940's [K1],[K2] (see also [R]) resolved certain of these problems by introducing:

(i) a "relaxed" variant of Monge's original mass allocation problem

and, more importantly,

(ii) a dual variational principle.

The idea behind (i) is, remarkably, to transform (1.5) into a linear problem. The trick is firstly to introduce the class

$$\mathcal{M} := \left\{ \text{Radon probability measures } \mu \text{ on } \mathbb{R}^n \times \mathbb{R}^n \mid \ \text{proj}_x \mu = \mu^+, \ \ \text{proj}_y \mu = \mu^- \right\} \tag{1.8}$$

of measures on the product space $\mathbb{R}^n \times \mathbb{R}^n$, whose projections on the first n coordinates and the last n coordinates are, respectively, μ^+, μ^-. Given then $\mu \in \mathcal{M}$, we define the *relaxed cost* functional

$$J[\mu] := \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y)\, d\mu(x, y). \tag{1.9}$$

The point of course is that if we have a mapping $\mathbf{s} \in \mathcal{A}$, then the induced measure

$$\mu(E) := \mu^+\{x \in \mathbb{R}^n \mid (x, \mathbf{s}(x)) \in E\} \qquad (E \subset \mathbb{R}^n \times \mathbb{R}^n, \ E \text{ Borel}) \tag{1.10}$$

belongs to $\mathcal{M}$. Furthermore the new functional (1.9) is *linear* in μ and so, under appropriate assumptions on the cost c, simple compactness arguments assert the existence of at least one optimal measure $\mu^* \in \mathcal{M}$, satisfying

$$J[\mu^*] = \min_{\mu \in \mathcal{M}} J[\mu]. \tag{1.11}$$

Such a measure μ^* need not, however, be generated by any one-to-one mapping $\mathbf{s} \in \mathcal{A}$, and consequently the foregoing construction allows us only to establish the existence of a "weak" or "generalized" solution of Monge's original problem. We will several times later return to the central problem of fashioning some sort of "strong" solution, which actually corresponds to a mapping.

Of even greater importance for our purposes was Kantorovich's additional introduction of a *dual problem*. The best way to motivate this is by analogy with the finite dimensional case. Suppose that then c_{ij}, μ_i^+, μ_j^- $(i = 1, \dots, n;\ j = 1, \dots, m)$ are given nonnegative numbers, satisfying the balance condition

$$\sum_{i=1}^{n} \mu_i^+ = \sum_{j=1}^{m} \mu_j^-,$$

and we wish to find μ_{ij}^* $(i = 1, \dots, n;\ j = 1, \dots, m)$ to

$$\text{minimize} \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} \mu_{ij}, \tag{1.12}$$

subject to the constraints

$$\sum_{j=1}^{m} \mu_{ij} = \mu_i^+,\ \sum_{i=1}^{n} \mu_{ij} = \mu_j^-,\ \mu_{ij} \geq 0 \qquad (i = 1, \dots, n;\ j = 1, \dots, m). \tag{1.13}$$

This is clearly the discrete analogue of (1.8), (1.9), (1.11). As explained in the appendix (§13) the discrete linear programming dual problem to (1.12) is to find $u = (u_1, \dots, u_n) \in \mathbb{R}^n$, $v = (v_1, \dots, v_m) \in \mathbb{R}^m$ so as to

$$\text{maximize} \qquad \sum_{i=1}^{n} u_i \mu_i^+ + \sum_{j=1}^{m} v_j \mu_j^-, \tag{1.14}$$

subject to the inequalities

$$u_i + v_j \leq c_{ij} \quad (i = 1, \dots, n;\ j = 1, \dots, m). \tag{1.15}$$

We can now by analogy guess the dual variational principle to (1.11). For this, we introduce a continuous variant of (1.15) by defining

$$\mathcal{L} := \{(u, v) \mid u, v : \mathbb{R}^n \to \mathbb{R} \text{ continuous},\ u(x) + v(y) \leq c(x, y) \ \ (x, y \in \mathbb{R}^n)\}. \tag{1.16}$$

Likewise, we introduce the continuous analogue of (1.14) by setting

$$K[u, v] := \int_{\mathbb{R}^n} u(x)\, d\mu^+(x) + \int_{\mathbb{R}^n} v(y)\, d\mu^-(y). \tag{1.17}$$

Consequently our *dual problem* is to find an optimal pair $(u^*, v^*) \in \mathcal{L}$ such that

$$K[u^*, v^*] = \max_{(u,v)\in\mathcal{L}} K[u, v]. \tag{1.18}$$

In summary then, the transformation of Monge's original mass allocation problem (1.5) into the dual problem (1.18) presents us with a rather different vantage point: rather than struggling to construct an optimal mapping $\mathbf{s}^* \in \mathcal{A}$ satisfying a highly nonlinear constraint, we are now confronted with the task of finding an optimal pair $(u^*, v^*) \in \mathcal{L}$. And this, as we will see later, is really easy. The mathematical structure of the dual problem provides precisely what was missing for the original problem, enough compactness to construct a minimizer as some sort of limit of a minimizing sequence.

And yet this of course is not the story's end. Kantorovich's methods of first relaxing and then dualizing have brought us into a realm where routine mathematical tools work, but have also taken us far away from the original issue: namely, how do we actually fashion an optimal allocation plan $\mathbf{s}^*$?

We devote much of the remainder of the paper to answering this question, in two most important cases of the uniformly convex cost density:

$$c(x, y) = \frac{1}{2}|x - y|^2 \quad (x, y \in \mathbb{R}^n), \tag{1.19}$$

and the nonuniformly convex cost density

$$c(x, y) = |x - y| \quad (x, y \in \mathbb{R}^n). \tag{1.20}$$

These "L^2" and "L^1" theories are rich in mathematical structure, and serve as archetypes for other models. Observe carefully the very different geometric consequences: in the first case the graph of the mapping $y \mapsto c(x, y)$ contains no straight lines, and in the second case it does. The latter degeneracy will create interesting problems.

Remark. I make no attempt in this paper to survey the vast literature on Monge–Kantorovich methods in probability and statistics, a nice summary of which may be found in Rachev [R]. □

Part I: Cost $= \frac{1}{2}$(Distance)^2

2 Heuristics

For this and the next five sections we take the quadratic cost density

$$c(x,y) := \frac{1}{2}|x-y|^2 \qquad (x,y \in \mathbb{R}^n), \tag{2.1}$$

$|\cdot|$ denoting the usual Euclidean norm. We hereafter wish to understand if and how we can construct an optimal mass transfer plan $\mathbf{s}^*$ solving (1.5), where now

$$I[\mathbf{s}] := \frac{1}{2}\int_{\mathbb{R}^n} |x - \mathbf{s}(x)|^2 \, d\mu^+(x) \tag{2.2}$$

for $\mathbf{s}$ in $\mathcal{A}$, the admissible class of measurable, one-to-one maps of $\mathbb{R}^n$ which push forward the measure μ^+ to μ^-. The following techniques were largely pioneered by Y. Brenier in [B2].

2.1 Geometry of optimal transport

We begin with some informal insights, our goal being to understand, for the moment without proofs, what information about an optimal mapping we can extract directly from the original and dual variational principles. In other words, how can we exploit the very fact that a given mapping minimizes the work functional $I[\cdot]$ (among all other mappings in $\mathcal{A}$), to understand its precise geometric properties?

So now let us assume that $\mathbf{s}^* \in \mathcal{A}$ minimizes the work functional (2.2), among all other mappings $\mathbf{s} \in \mathcal{A}$. Fix a positive integer m, take distinct points $\{x_k\}_{k=1}^m \subset X = \text{ spt } (\mu^+)$, and assume we can find small disjoint balls

$$E_k := B(x_k, r_k) \quad (k = 1, \dots, m), \tag{2.3}$$

and radii $\{r_k\}_{k=1}^m$, adjusted so that

$$\mu^+(E_1) = \cdots = \mu^+(E_m) = \varepsilon. \tag{2.4}$$

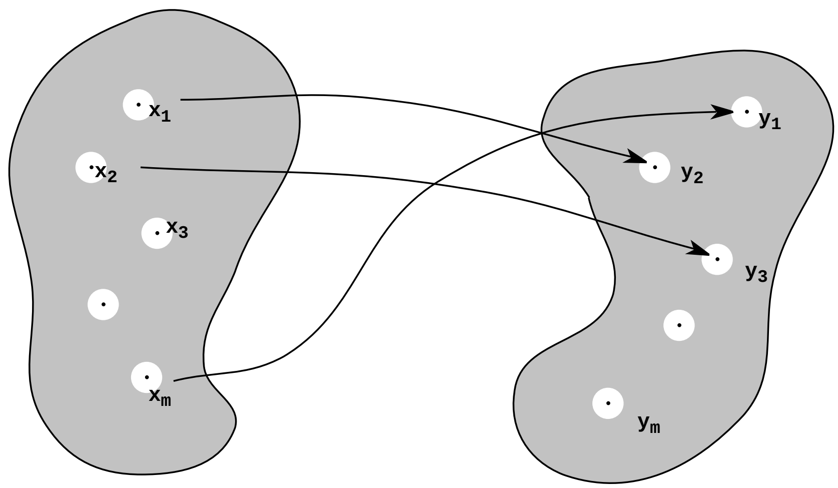
Next write $y_k := \mathbf{s}^*(x_k)$, $F_k := \mathbf{s}^*(E_k)$. Since $\mathbf{s}^*$ pushes μ^+ to μ^-, we have

$$\mu^-(F_1) = \cdots = \mu^-(F_m) = \varepsilon. \tag{2.5}$$

We construct another mapping $\mathbf{s} \in \mathcal{A}$ by cyclically permuting the images of the balls $\{E_k\}_{k=1}^m$. That is, we design $\mathbf{s} \in \mathcal{A}$ so that

$$\begin{cases} \mathbf{s}(x_k) = y_{k+1}, \ \mathbf{s}(E_k) = F_{k+1} \quad (k = 1, \dots, m) \\ \mathbf{s} \equiv \mathbf{s}^* \text{ on } X - \displaystyle\bigcup_{k=1}^m E_k, \end{cases} \tag{2.6}$$

where $y_{m+1} := y_1$, $F_{m+1} := F_1$.



Then, since $\mathbf{s}^*$ is a minimizer,

$$I[\mathbf{s}^*] \leq I[\mathbf{s}]. \tag{2.7}$$

Remembering (2.2), we deduce

$$\sum_{k=1}^{m} \int_{E_k} |x - \mathbf{s}^*(x)|^2 \, d\mu^+(x) \leq \sum_{k=1}^{m} \int_{E_k} |x - \mathbf{s}(x)|^2 \, d\mu^+(x).$$

Since both $\mathbf{s}^*, \mathbf{s}$ push μ^+ to μ^-, we can further simplify and then divide by ε:

$$\sum_{k=1}^{m} ⨍_{E_k} x \cdot (\mathbf{s}(x) - \mathbf{s}^*(x)) \, d\mu^+(x) \leq 0, \tag{2.8}$$

the slash through the integral denoting an average. Now send $\varepsilon \to 0$. Assuming that the mapping $\mathbf{s}^*$ and the measure μ^+ are well-behaved, we deduce from (2.8) that

$$\sum_{k=1}^{m} x_k \cdot (y_{k+1} - y_k) \leq 0 \tag{2.9}$$

for $y_k = \mathbf{s}^*(x_k)$, $y_{m+1} = y_1$, $k = 1, 2, \ldots, m$.

In the terminology of convex analysis, (2.9) asserts that the graph $\{(x, \mathbf{s}^*(x)) \mid x \in X\} \subset \mathbb{R}^n \times \mathbb{R}^n$ is *cyclically monotone*. This is an interesting deduction in light of an important theorem of Rockafeller [Rk], asserting that *a cyclically monotone subset of* $\mathbb{R}^n \times \mathbb{R}^n$ *lies in the subdifferential of a convex mapping of* $\mathbb{R}^n$ into $\mathbb{R}$. In other words,

$$\mathbf{s}^* \subset \partial\phi^*, \tag{2.10}$$

for some convex function ϕ^*, in the sense of possibly multivalued graphs on $\mathbb{R}^n \times \mathbb{R}^n$. Moreover, a convex function is differentiable a.e., and so

$$\mathbf{s}^* = D\phi^* \quad \text{a.e. in } X, \tag{2.11}$$

where D denotes the gradient.

We have come upon an important deduction: *an optimal mass allocation plan is the gradient of a convex potential* ϕ^*. (Cf. McCann [MC2], etc.)

2.2 Lagrange multipliers

In view of its importance we provide next an alternative, but still strictly formal, analytic derivation of (2.11).

For this, we assume the measures $\mu^{\pm}$ have smooth densities, $d\mu^+ = f^+dx$, $d\mu^- = f^-dy$, and introduce the *augmented work functional*

$$\tilde{I}[\mathbf{s}] := \int_{\mathbb{R}^n} \frac{1}{2}|x - \mathbf{s}(x)|^2 f^+(x) + \lambda(x)[f^-(\mathbf{s}(x))\det(D\mathbf{s}(x)) - f^+(x)]dx, \tag{2.12}$$

where the function λ is the *Lagrange multiplier* corresponding to the pointwise constraint that $\mathbf{s}_{\#}(\mu^+) = \mu^-$ (that is, $f^+ = f^-(\mathbf{s})\det(D\mathbf{s})$). Computing the first variation, we find for $k = 1, \dots, m$:

$$(\lambda f^-(\mathbf{s}^*)(\mathrm{cof} D\mathbf{s}^*)^k_i)_{x_i} = (s^{*k} - x_k)f^+ + \lambda f^-_{y_k}(\mathbf{s}^*)\det(D\mathbf{s}^*). \tag{2.13}$$

Here $\mathrm{cof}D\mathbf{s}^*$ is the cofactor matrix of $D\mathbf{s}^*$; that is, the $(k,i)^{th}$ entry of $\mathrm{cof}D\mathbf{s}^*$ is $(-1)^{k+i}$ times the $(k,i)^{th}$ minor of the matrix $D\mathbf{s}^*$.

Standard matrix identities assert $(\mathrm{cof}D\mathbf{s}^*)^k_{i,x_i} = 0$, $s^{*l}_{x_i}(\mathrm{cof}Ds^*)^k_i = \delta_{kl}(\det D\mathbf{s}^*)$, and $\mathbf{s}^{*k}_{x_j}(\mathrm{cof}D\mathbf{s}^*)^k_i = \delta_{ij}(\det D\mathbf{s}^*)$. We employ these equalities to simplify (2.13), and thereby discover after some calculations

$$\lambda_{x_i} f^-(\mathbf{s}^*)(\mathrm{cof}D\mathbf{s}^*)^k_i = (s^{*k} - x_k)f^+. \tag{2.14}$$

Now multiply by $s^{*k}_{x_j}$ and sum on k, to deduce:

$$\lambda_{x_j} = (s^{*k} - x_k)s^{*k}_{x_j}.$$

But then

$$\left(\lambda - \frac{|\mathbf{s}^* - x|^2}{2} + \frac{|x|^2}{2}\right)_{x_j} = s^{*j} \qquad (j = 1, \dots, n),$$

and so (2.11) again follows, for the potential $\phi^* := \lambda - \frac{|\mathbf{s}^*-x|^2}{2} + \frac{|x|^2}{2}$.

3 Optimal mass transport, polar factorization

3.1 Solution of dual problem

The foregoing heuristics done with, we turn next to the task of proving rigorously the existence of an optimal mass allocation plan. We expect $\mathbf{s}^* = D\phi^*$ almost everywhere for some convex potential ϕ^*, and the task is now really to deduce this. We will do so from the Kantorovich *dual variational principle* (1.16)– (1.18) introduced in §1. We hereafter concentrate on the situation

$$d\mu^+ = f^+ dx, \; d\mu^- = f^- dy, \tag{3.1}$$

where $f^\pm$ are bounded, nonnegative functions with compact support, satisfying the *mass balance* condition

$$\int_X f^+(x)dx = \int_Y f^-(y)\, dy \tag{3.2}$$

where, as always, $X :=$ spt (f^+), $Y :=$ spt (f^-). For the case at hand, the dual problem is to find (u^*, v^*) so as to *maximize*

$$K[u,v] := \int_X u(x) f^+(x)\, dx + \int_Y v(y) f^-(y)\, dy, \tag{3.3}$$

subject to the constraint

$$u(x) + v(y) \leq \frac{1}{2}|x-y|^2 \qquad (x \in X, \; y \in Y). \tag{3.4}$$

We wish to take up tools from convex analysis, and for this must first change variables:

$$\begin{cases} \phi(x) := \frac{1}{2}|x|^2 - u(x) & (x \in X) \\ \psi(y) := \frac{1}{2}|y|^2 - v(y) & (y \in Y). \end{cases} \tag{3.5}$$

Note that now (3.4) says

$$\phi(x) + \psi(y) \geq x \cdot y \qquad (x \in X, \; y \in Y), \tag{3.6}$$

and so the variational problem is then to *minimize*

$$L[\phi, \psi] := \int_X \phi(x) f^+(x)\, dx + \int_Y \psi(y) f^-(y)\, dy, \tag{3.7}$$

subject to the constraint (3.6).

Lemma 3.1 (i) *There exist* (ϕ^*, ψ^*) *solving this minimization problem.*

(ii) *Furthermore,* (ϕ^*, ψ^*) *are dual convex functions, in the sense that*

$$\begin{cases} \phi^*(x) = \max_{y \in Y}(x \cdot y - \psi^*(y)) & (x \in X) \\ \psi^*(y) = \max_{x \in X}(x \cdot y - \phi^*(x)) & (y \in Y). \end{cases} \tag{3.8}$$

Proof. 1. If ϕ, ψ satisfy (3.6), then

$$\phi(x) \geq \max_{y \in Y}(x \cdot y - \psi(y)) =: \hat{\phi}(x) \tag{3.9}$$

and

$$\hat{\phi}(x) + \psi(y) \geq x \cdot y \qquad (x \in X,\ y \in Y). \tag{3.10}$$

Consequently

$$\psi(y) \geq \max_{x \in X}(x \cdot y - \hat{\phi}(x)) =: \hat{\psi}(y) \tag{3.11}$$

and

$$\hat{\phi}(x) + \hat{\psi}(y) \geq x \cdot y \qquad (x \in X,\ y \in Y). \tag{3.12}$$

As $\psi \geq \hat{\psi}$, (3.9) implies

$$\max_{y \in Y}(x \cdot y - \hat{\psi}(y)) \geq \hat{\phi}(x).$$

This and (3.12) say

$$\hat{\phi}(x) = \max_{y \in Y}(x \cdot y - \hat{\psi}(y)). \tag{3.13}$$

Since $f^{\pm} \geq 0$ and $\psi \geq \hat{\psi}$, $\phi \geq \hat{\phi}$, we see that $L[\hat{\phi}, \hat{\psi}] \leq L[\phi, \psi]$.

2. Consequently in seeking for minimizers of L we may restrict attention to convex dual pairs $(\hat{\phi}, \hat{\psi})$, as above. Such functions are uniformly Lipschitz continuous, and so, after adding or subtracting constants, we can extract a uniformly convergent subsequence from any minimizing sequence for L. We thereby secure an optimal, convex dual pair. □

3.2 Existence of optimal mass transfer plan

Let us now regard (3.8) as defining $\phi^*(x), \psi^*(y)$ for all $x, y \in \mathbb{R}^n$. Then $\phi^*, \psi^* : \mathbb{R}^n \to \mathbb{R}$ are convex, and consequently differentiable a.e. We demonstrate next that

$$\mathbf{s}^*(x) := D\phi^*(x) \qquad (\text{a.e. } x \in X) \tag{3.14}$$

solves the mass allocation problem.

Theorem 3.1 *Define* $\mathbf{s}^*$ *by (3.14). Then*

(i) $\mathbf{s}^* : X \to Y$ *is essentially one-to-one and onto.*

(ii) $\int_X h(\mathbf{s}^*(x))\,d\mu^+(x) = \int_Y h(y)\,d\mu^-(y)$ *for each* $h \in C(Y)$.

(iii) *Lastly,*

$$\frac{1}{2}\int_X |x - \mathbf{s}^*(x)|^2\,d\mu^+(x) \le \frac{1}{2}\int_X |x - \mathbf{s}(x)|^2\,d\mu^+(x)$$

for all $\mathbf{s} : X \to Y$ *such that* $\mathbf{s}_\#(\mu^+) = \mu^-$.

Proof. 1. From the max-representation function (3.8) we see that $\mathbf{s}^* = D\phi^* \in Y$ a.e.

2. Fix $\tau > 0$, and define the variations

$$\begin{cases} \psi_\tau(y) := \psi^*(y) + \tau h(y) & (y \in Y) \\ \phi_\tau(x) := \max_{y\in Y}(x \cdot y - \psi_\tau(y)) & (x \in X). \end{cases} \tag{3.15}$$

Then

$$\phi_\tau(x) + \psi_\tau(y) \ge x \cdot y \qquad (x \in X,\ y \in Y) \tag{3.16}$$

and so

$$L[\phi^*, \psi^*] \le L[\phi_\tau, \psi_\tau] =: i(\tau).$$

As the mapping $\tau \mapsto i(\tau)$ thus has a minimum at $\tau = 0$,

$$\begin{aligned} 0 &\le \frac{1}{\tau}(L[\phi_\tau, \psi_\tau] - L[\phi^*, \psi^*]) \\ &= \int_X \left[\frac{\phi_\tau(x) - \phi^*(x)}{\tau}\right] f^+(x)\,dx + \int_Y h(y) f^-(y)\,dy. \end{aligned} \tag{3.17}$$

Now $\left|\frac{\phi_\tau - \phi^*}{\tau}\right| \le \|h\|_{L^\infty}$. Furthermore if we take $y_\tau \in Y$ so that

$$\phi_\tau(x) = x \cdot y_\tau - \psi_\tau(y_\tau), \tag{3.18}$$

then

$$\phi_\tau(x) - \phi^*(x) = x \cdot y_\tau - \psi^*(y_\tau) - \tau h(y_\tau) - \phi^*(x) \le -\tau h(y_\tau). \tag{3.19}$$

On the other hand, if we select $y \in Y$ such that

$$\phi^*(x) = x \cdot y - \psi^*(y), \tag{3.20}$$

then

$$\phi_\tau(x) - \phi^*(x) \ge x \cdot y - \psi^*(y) - \tau h(y) - \phi^*(x) = -\tau h(y). \tag{3.21}$$

Thus

$$- h(y) \leq \frac{\phi_\tau(x) - \phi^*(x)}{\tau} \leq -h(y_\tau). \tag{3.22}$$

If we take a point $x \in X$ where $\mathbf{s}^*(x) := D\phi^*(x)$ exists, then (3.20) implies $y = \mathbf{s}^*(x)$. Furthermore as $\tau \to 0$, $y_\tau \to \mathbf{s}^*(x)$. Thus (3.17), (3.22) and the Dominated Convergence Theorem imply

$$\int_X h(\mathbf{s}^*(x)) f^+(x)\, dx \leq \int_Y h(y) f^-(y)\, dy.$$

Replacing h by $-h$, we deduce that equality holds. This is statement (ii).

3. Now take $\mathbf{s}$ to be any admissible mapping. Then

$$\int_X \psi^*(\mathbf{s}(x)) f^+(x)\, dx = \int_Y \psi^*(y) f^-(y)\, dy.$$

Since $\phi^*(x) + \psi^*(y) \geq x \cdot y$, with equality for $y = \mathbf{s}^*(x)$, we consequently have

$$\begin{aligned} 0 &\geq \textstyle\int_X [x \cdot (\mathbf{s}(x) - \mathbf{s}^*(x)) - \phi^*(x) + \phi^*(x)] f^+(x)\, dx \\ &= \textstyle\int_X [x \cdot (\mathbf{s}(x) - \mathbf{s}^*(x))] f^+(x)\, dx. \end{aligned}$$

This implies assertion (iii) of the Theorem: $\mathbf{s}^*$ is optimal. □

This proof follows Gangbo [G] and Caffarelli [C4]. Another neat approach is due to McCann [MC2]; he approximates the measures $\mu^\pm$ by point masses, solves the resulting discrete linear programming problem, and the passes to limits. See also Gangbo–McCann [G-M1], [G-M2], McCann [MC4]. An interesting related work is Wolfson [W].

3.3 Polar factorization of vector fields

Assume next that $U \subset \mathbb{R}^n$ is open, bounded, with $|\partial U| = 0$, and that $\mathbf{r} : U \to \mathbb{R}^n$ is a bounded measurable mapping satisfying the *nondegeneracy condition*

$$\left\{ \begin{array}{l} |\mathbf{r}^{-1}(N)| = 0 \text{ for each bounded Borel} \\ \text{set } N \subset \mathbb{R}^n, \text{ with } |N| = 0. \end{array} \right. \tag{3.23}$$

Define the modified functional

$$\hat{L}[\phi, \psi] := \int_U \phi(\mathbf{r}(y)) + \psi(y)\, dy, \tag{3.24}$$

which we propose to minimize among functions (ϕ, ψ) satisfying

$$\phi(x) + \psi(y) \geq x \cdot y. \tag{3.25}$$

Let (ϕ^*, ψ^*) solve this problem. Then taking variations as in the previous proof, we deduce

$$\int_U h(y)\,dy = \int_U h(D\phi^*(\mathbf{r}(y)))\,dy$$

for all $h \in C(U)$. Define now

$$\mathbf{s}^*(x) := D\phi^*(\mathbf{r}(x)).$$

Then $\mathbf{s}^*$ is measure preserving and

$$\mathbf{r}(x) = D\psi^*(\mathbf{s}^*(x)). \tag{3.26}$$

This is the *polar factorization* of the vector field $\mathbf{r}$, as *the composition of the gradient of a convex function and a measure preserving mapping.* □

This remarkable polar factorization was established by Brenier [B1], [B2], and the proof later simplified by Gangbo [G]. (Brenier was motivated by problems in fluid mechanics, which we will not discuss in this paper: see [B3].)

Remark. We can regard (3.26) as a nonlinear generalization of the Helmholtz decomposition of a vector field into the sum of a gradient and a divergence-free field. To see this formally, let $\mathbf{a}$ be a given vector field and write

$$\mathbf{r_0} := id,\ \mathbf{s_0} := id,\ \psi_0 := \frac{|x|^2}{2}.$$

Set

$$\mathbf{r}(\tau) := \mathbf{r_0} + \tau\mathbf{a} \tag{3.27}$$

for small $|\tau|$. The polar factorization gives

$$\mathbf{r}(\tau) := D\psi(\tau) \circ \mathbf{s}(\tau), \tag{3.28}$$

where $\psi(\tau)$ is convex, $\mathbf{s}(\tau)$ is measure preserving, and we drop the superscript *. Next put

$$\psi(\tau) = \psi_0 + \tau b(\tau),\ \mathbf{s}(\tau) = \mathbf{s_0} + \tau\mathbf{c}(\tau) \tag{3.29}$$

into (3.28), differentiate with respect to τ, and set $\tau = 0$. We deduce

$$\mathbf{a} = Db + \mathbf{c}, \tag{3.30}$$

for $b = b(0)$, $\mathbf{c} = \mathbf{c}(0)$. This is a Helmholtz decomposition of $\mathbf{a}$, since

$$1 = \det(D\mathbf{s}(\tau)) = 1 + \tau \mathrm{div}\ \mathbf{c} + O(\tau^2) \tag{3.31}$$

implies div $\mathbf{c} = 0$. □

4 Regularity

4.1 Solving the Monge–Ampere equation

The mass allocation problem solved in §3 can be interpreted as providing a weak or generalized solution to the Monge–Ampere PDE mapping problem:

$$\begin{cases} \text{(a)} \quad f^-(D\phi(x))\text{det}D^2\phi(x) = f^+(x) \quad (x \in X) \\ \text{(b)} \quad D\phi \text{ maps } X \text{ into } Y. \end{cases} \tag{4.1}$$

Now and hereafter we omit the superscript *. Recall that we interpret (4.1)(a) to mean

$$\int_X h(D\phi(x))f^+(x)\,dx = \int_Y h(y)f^-(y)\,dy \tag{4.2}$$

for each continuous function h.

There are, however, subtleties here. First of all, recall that since ϕ is convex, we can interpret $D^2\phi$ as a matrix of signed measures, which have absolutely continuous and singular parts with respect to Lebesgue measure:

$$d[D^2\phi] = [D^2\phi]_{ac}\,dx + d[D^2\phi]_s \tag{4.3}$$

(see for instance [E-G2]). We say ϕ is a solution of

$$f^-(D\phi)\text{det}(D^2\phi) = f^+ \tag{4.4}$$

in the *sense of Alexandrov* if the identity (4.2) holds, and additionally

$$d[D^2\phi]_s = 0, \ \phi \text{ is strictly convex.} \tag{4.5}$$

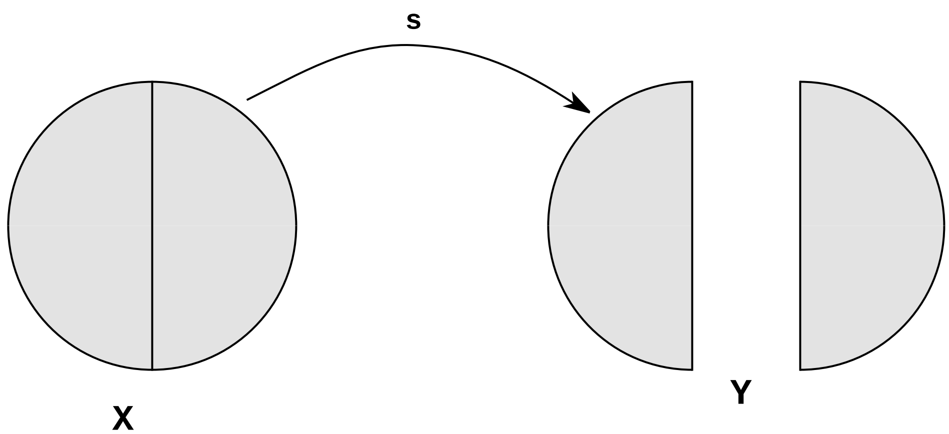
4.2 Examples

But, as noted by L. Caffarelli [C5], the solution constructed in §3 need *not* solve (4.4) in the Alexandrov sense. Consider first of all the situation that $n = 2$ and X is the unit disk $B(0,1)$. Then $\mathbf{s} = D\phi$, for

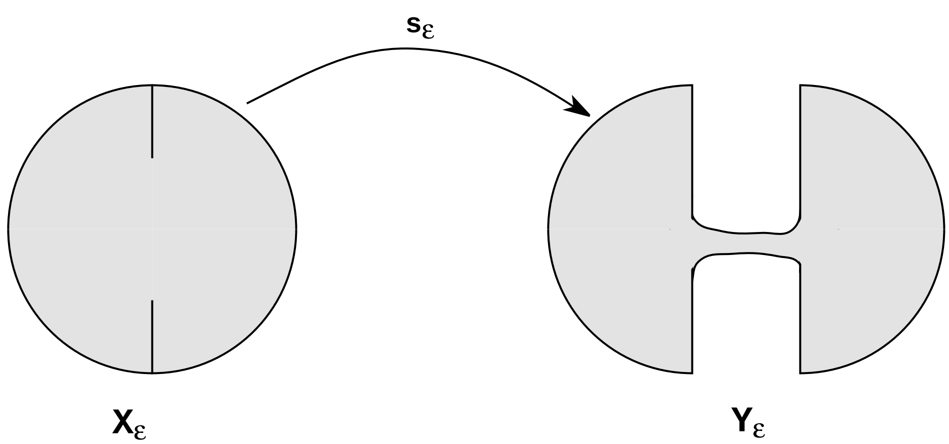
$$\phi(x_1, x_2) = |x_1| + \frac{1}{2}(x_1^2 + x_2^2),$$

optimally rearranges $\mu^+ = \chi_X dx$ into $\mu^- = \chi_Y\,dy$, where

$$\begin{aligned} Y \quad = \quad & \{(x_1, x_2) \mid 0 \le (x_1 - 1)^2 + x_2^2, \ x_1 \ge 1\} \\ & \cup \{(x_1, x_n) \mid 0 \le (x_1 + 1)^2 + x_2^2, \ x_1 \le -1\}. \end{aligned}$$



In this case $D^2\phi$ has a singular part concentrated along $\{x_1 = 0\}$. Caffarelli shows also by a perturbation argument that if we replace Y by a connected set Y_ε, as drawn, with $|Y_\epsilon| = |Y| = |X|$, the optimal mapping $\mathbf{s}_\varepsilon$ still has a singularity.



4.3 Interior regularity for convex targets

On the other hand Caffarelli also demonstrated that if the target Y is convex, the optimal mapping $\mathbf{s} = D\phi$ is indeed regular. More precisely, assume that X, Y are bounded, connected, open sets in $\mathbb{R}^n$ and

$$f^+ : X \to (0, \infty), \;\; f^- : Y \to (0, \infty)$$

are bounded above and below, away from zero.

Theorem 4.1 *Assume Y is convex.*

(i) *Then an optimal mapping* $\mathbf{s} = D\phi$ *solves the Monge–Ampere equation*

$$f^-(D\phi)\det(D^2\phi) = f^+$$

in the Alexandrov sense, and ϕ *is strictly convex.*

(ii)*In addition,* $\phi \in C^{1,\alpha}_{\text{loc}}(X)$ *for some* $0 < \alpha < 1$.

(iii) *If* f^+, f^- *are continuous, then*

$$\phi \in W^{2,p}_{\text{loc}}(X) \qquad \textit{for each } 1 \leq p < \infty.$$

(iv) *Finally if* $f^+ \in C^\beta(X)$, $f^- \in C^\beta(Y)$ *for some* $0 < \beta < 1$, *then*

$$\phi \in C^{2,\alpha}_{\text{loc}}(X) \qquad \textit{for each } 0 < \alpha < \beta.$$

The idea of the proof is to show that $\phi : \mathbb{R}^n \to \mathbb{R}$ is an Alexandrov solution of

$$\lambda\chi_X \leq \det D^2\phi \leq \Lambda\chi_X \quad \text{ in } \mathbb{R}^n$$

for constants $0 < \lambda \leq \Lambda$, and to apply then the deep regularity theory developed in [C1], [C2], [C3].

4.4 Boundary regularity for convex domain and target

Theorem 4.2 *Assume both* X, Y *are convex.*

(i) *Then* $\phi \in C^{1,\alpha}(\bar{X})$, $\psi \in C^{1,\alpha}(\bar{Y})$ *for some* $0 < \alpha < 1$.

(ii) *If, in addition,* $\partial X, \partial Y$ *are smooth, then* $\phi \in C^{2,\alpha}(\bar{X})$, $\psi \in C^{2,\alpha}(\bar{Y})$.

These assertions have been proved independently by Urbas [U2] and Caffarelli [C6], [C7], under slightly different hypotheses and using quite different techniques.

5 Application: Nonlinear interpolation

The next three sections discuss several new and interesting applications of the ideas set forth above in §2-4.

We begin with a clever procedure, due to R. McCann [MC1], which resolves a uniqueness problem in the calculus of variations. We consider a model for the *equilibrium configuration of an interacting gas*. Take

$$\mathcal{P} := \{\rho : \mathbb{R}^n \to \mathbb{R} \mid \rho \in L^1(\mathbb{R}^n),\ \rho \geq 0 \text{ a.e.}, \int_{\mathbb{R}^n} \rho\, dx = 1\} \tag{5.1}$$

to be the collection of mass densities, and to each $\rho \in \mathcal{P}$ associate the *energy*

$$E(\rho) := \int_{\mathbb{R}^n} \rho^\gamma(x)\, dx + \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \rho(x) V(x-y) \rho(y)\, dxdy. \tag{5.2}$$

The first term

$$E_1(\rho) := \int_{\mathbb{R}^n} \rho^\gamma(x)\, dx$$

represents the *internal energy*, where $\gamma > 1$ is a constant. The second expression

$$E_2(\rho) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \rho(x) V(x-y) \rho(y)\, dxdy$$

corresponds to the *potential energy* owing to interaction, where $V : \mathbb{R}^n \to \mathbb{R}$ is nonnegative and strictly convex.

Suppose now $\rho_0, \rho_1 \in \mathcal{P}$ are both minimizers:

$$E(\rho_0) = E(\rho_1) = \min_{\rho \in \mathcal{P}} E(\rho). \tag{5.3}$$

A basic question concerns the uniqueness of these minimizers, up to translation invariance. Now the usual procedure would be to take the linear interpolation of ρ_0, ρ_1 and to ask if the mapping $t \mapsto E((1-t)\rho_1 + t\rho_0)$ is convex on $[0, 1]$. This is however false in general for the example at hand. McCann instead employs the ideas of §2–3 to build a sort of "nonlinear interpolation" between ρ_0, ρ_1. His idea is to take the convex potential ϕ such that

$$\mathbf{s}_\#(\rho_0) = \rho_1, \tag{5.4}$$

where

$$\mathbf{s} = D\phi. \tag{5.5}$$

Define then

$$\rho_t := ((1-t)id + t\mathbf{s})_\# \rho_0 \quad (0 \le t \le 1). \tag{5.6}$$

This nonlinear interpolation between ρ_0 and ρ_1 in effect locates a realm of convexity for E.

Theorem 5.1 *The mapping $t \mapsto E(\rho_t)$ is convex on $[0, 1]$, and is strictly convex unless ρ_1 is a translate of ρ_0.*

Proof. 1. We first compute

$$
\begin{aligned}
E_2(\rho_t) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \rho_t(x)V(x-y)\rho_t(y)\,dxdy \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \rho_0(x)V((1-t)(x-y)+t(\mathbf{s}(x)-\mathbf{s}(y))\rho_0(y)\,dxdy.
\end{aligned} \tag{5.7}
$$

As V is convex, the mapping $t \mapsto E_2(\rho_t)$ is convex.

2. Next, (5.6) implies

$$
\begin{aligned}
E_1(\rho_t) &= \int_{\mathbb{R}^n} \rho_t^\gamma(x)\,dx \\
&= \int_{\mathbb{R}^n} \left[\frac{\rho_0}{\det[(1-t)I+tD\mathbf{s}]}\right]^\gamma \det[(1-t)I+tD\mathbf{s}]\,dx.
\end{aligned} \tag{5.8}
$$

Now the matrix

$$A_t := (1-t)I + tD\mathbf{s} = (1-t)I + tD^2\phi$$

is positive definite for $0 \le t < 1$ and a.e. x. (In particular, where $d[D^2\phi]_s = 0$.) Now if A is a positive-definite, symmetric matrix, we have

$$(\det A)^{1/n} = \frac{1}{n} \inf_{\substack{B>0 \\ \det B = 1}} (A:B),$$

and thus $A \mapsto (\det A)^{1/n}$ is concave for positive definite matrices. Consequently

$$\alpha(t) := (\det[(1-t)I+tD\mathbf{s}])^{1/n} \text{ is concave on } [0,1]. \tag{5.9}$$

Now set

$$\beta(t) := \alpha(t)^{n(1-\gamma)} \quad (0 \le t \le 1).$$

Then

$$
\begin{aligned}
\beta''(t) &= n(1-\gamma)(n-n\gamma-1)\alpha(t)^{n(1-\gamma)-2}(\alpha')^2 \\
&\quad + n(1-\gamma)\alpha(t)^{n(1-\gamma)-1}\alpha'' \ge 0,
\end{aligned}
$$

since $\gamma > 1$. Thus

$$t \mapsto \beta(t) \text{ is convex on } [0,1]. \tag{5.10}$$

Return to (5.8), which we rewrite to read

$$E_1(\rho_t) = \int_{\mathbb{R}} \rho_0\beta(t)\,dx.$$

It follows that $t \mapsto E_1(\rho_t)$ is convex.

3. To show a minimizer is unique up to translation, we go back to (5.7). As V is strictly convex, the mapping $t \mapsto E_2(\rho_t)$ is strictly convex, unless $x - y = \mathbf{s}(x) - \mathbf{s}(y)$ ρ_0 a.e. Now such strict convexity would violate the fact that ρ_0, ρ_1 are minimizers. Consequently

$$x - \mathbf{s}(x) \text{ is independent of } x, \quad \rho_0 \text{ a.e.,}$$

and so ρ_1 is a translate of ρ_0. □

A related technique for understanding the equilibrium shape of crystals is presented in McCann [MC3]. See also Barthe [Ba] for the derivation of various inequalities using mass transport methods.

6 Application: Time-step minimization and nonlinear diffusion

An interesting emerging theme in much current research is the interplay between Monge–Kantorovich mass transform problems and partial differential equations involving time. A nice example, based upon Otto [O1] and Jordan–Kinderlehrer–Otto [J-K-O1], [J-K-O2], concerns a time-step minimization approximation to nonlinear diffusion equation.

Preparatory to describing this procedure, let us first define for the functions $f^+, f^- \in \mathcal{P}$ the *Wasserstein distance*

$$d(f^+, f^-)^2 := \inf \left\{ \frac{1}{2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x - y|^2 \, d\mu(x, y) \right\}, \tag{6.1}$$

the infimum taken over all nonnegative Radon measure μ whose projections are $d\mu^+ = f^+ dx$, $d\mu^- = f^- dy$. In view of §1, $d(f^+, f^-)^2$ is least cost of the Monge–Kantorovich mass reallocation of μ^+ to μ^-, for $c(x, y) = \frac{1}{2}|x - y|^2$.

6.1 Discrete time approximation

We next initiate a *time stepping procedure*, by first taking a small step size $h > 0$ and an initial profile $g \in \mathcal{P}$. We set $u^0 = g$ and then inductively define $\{u^k\}_{k=1}^\infty \subset \mathcal{P}$ by taking $u^{k+1} \in \mathcal{P}$ to minimize the functional

$$N_k(v) := \frac{d(v, u^k)^2}{h} + \int_{\mathbb{R}^n} \beta(v) \, dx \tag{6.2}$$

among all $v \in \mathcal{P}$. Here $\beta : \mathbb{R}^n \to \mathbb{R}$ is a given convex function, with superlinear growth. We suppose

$$\int_{\mathbb{R}^n} \beta(u^0) \, dx < \infty. \tag{6.3}$$

Convexity and weak convergence arguments show that there exists $u^{k+1} \in \mathcal{P}$ satisfying

$$N_k(u^{k+1}) = \min_{v \in \mathcal{P}} N_k(v) \qquad (k = 0, \dots). \tag{6.4}$$

We can envision (6.2), (6.4) as a discrete, dissipation evolution, in which at step $k+1$ the updated density u^{k+1} strikes a balance between minimizing the potential energy $\int_{\mathbb{R}^n} \beta(v)\, dx$ and the distance $\frac{d(v,u^k)^2}{2h}$ to the previous density at step k.

Taking $v = u^k$ on the right-hand side of (6.3), we have

$$\frac{d(u^{k+1}, u^k)^2}{2h} + \int_{\mathbb{R}^n} \beta(u^{k+1})\, dx \leq \int_{\mathbb{R}^n} \beta(u^k)\, dx;$$

and so

$$\frac{1}{2h} \sum_{k=1}^{\infty} d(u^{k+1}, u^k)^2,\ \max_{k \geq 1} \int_{\mathbb{R}^n} \beta(u^{k+1})\, dx \leq \int_{\mathbb{R}^n} \beta(u^0)\, dx < \infty. \tag{6.5}$$

Next, define $u_h : [0, \infty) \to P$ by setting $t_k = hk$, $u_h(t_k) = u^k$, and taking u_h to be linear on each time interval $[t_k, t_{k+1}]$ $(k = 0, 1, \dots)$.

The basic question is this: what happens when the step size h goes to 0?

6.2 Euler–Lagrange equation

To gain some insight, we follow Otto [O1] to compute the first variation of the minimization principle (6.4). For this, let us define

$$\alpha(z) := \beta'(z)z - \beta(z) \qquad (z \in \mathbb{R}) \tag{6.6}$$

and also write

$$d(u^{k+1}, u^k)^2 = \frac{1}{2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x - y|^2\, d\mu_{k+1}(x, y), \tag{6.7}$$

where

$$\text{proj}_x \mu_{k+1} = u^{k+1} dx,\ \text{proj}_y \mu_{k+1} = u^k dy. \tag{6.8}$$

In other words, the measure μ_{k+1} solves the relaxed problem.

Lemma 6.1 *Let $\boldsymbol{\xi} \in C_c^\infty(\mathbb{R}^n; \mathbb{R}^n)$ be a smooth, compactly supported vector field. Then*

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{(x - y)}{h} \cdot \boldsymbol{\xi}(x)\, d\mu_{k+1}(x, y) = \int_{\mathbb{R}^n} \alpha(u^{k+1}) \text{div}\, \boldsymbol{\xi}\, dx. \tag{6.9}$$

Proof. 1. We employ $\boldsymbol{\xi}$ to generate a domain variation, as follows. First solve the ODE

$$\begin{cases} \dot{\boldsymbol{\Phi}} = \boldsymbol{\xi}(\boldsymbol{\Phi}) & \left(\dot{} = \frac{d}{d\tau}\right) \\ \boldsymbol{\Phi}(0) = x, \end{cases} \tag{6.10}$$

and then write $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\tau, x)$ to display the dependence of $\boldsymbol{\Phi}$ on the parameter τ and the initial point x. Then for each τ, the mapping $x \mapsto \Phi(\tau, x)$ is a diffeomorphism. Thus for small τ, we can implicitly define

$$u_\tau : \mathbb{R}^n \to \mathbb{R}$$

by the formula

$$(\det D\boldsymbol{\Phi}(x,\tau))u_\tau(\boldsymbol{\Phi}(\tau, x)) = u^{k+1}(x). \tag{6.11}$$

Clearly then

$$\int_{\mathbb{R}^n} u_\tau \, dx = \int_{\mathbb{R}^n} u^{k+1} \, dx = 1.$$

Thus $u_\tau \in \mathcal{P}$, and so

$$i(\tau) := N_k(u_\tau) \text{ has a minimum at } \tau = 0. \tag{6.12}$$

2. Now if $z = \Phi(\tau, x)$, then

$$\int_{\mathbb{R}^n} \beta(u_\tau) dz = \int_{\mathbb{R}^n} \beta(u^{k+1}(\det D\boldsymbol{\Phi})^{-1}) \det D\boldsymbol{\Phi} \, dx.$$

Since $\frac{d}{dt}(\det D\boldsymbol{\Phi}) =$ div $\boldsymbol{\xi}$ at $\tau = 0$, we compute

$$\frac{d}{dt} \int_{\mathbb{R}^n} \beta(u_\tau) dz|_{\tau=0} = -\int_{\mathbb{R}^n} \alpha(u^{k+1}) \operatorname{div} \boldsymbol{\xi} \, dx, \tag{6.13}$$

owing to (6.6).

3. Next define $\mu_\tau \in \mathcal{M}_k$ so that

$$\frac{1}{2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |z - y|^2 \, d\mu_\tau(z, y) = \frac{1}{2} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |\Phi(\tau, x) - y|^2 \, d\mu_{k+1}(x, y).$$

Then for $\tau > 0$:

$$\frac{1}{\tau}(d(u_\tau, u^k)^2 - d(u^{k+1}, u^k)^2) \leq \frac{1}{2\tau} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (|\boldsymbol{\Phi}(\tau, x) - y|^2 - |x - y|^2) \, d\mu_{k+1}(x, y).$$

Let $\tau \to 0^+$,

$$\frac{1}{h} \limsup_{\tau \to 0^+} \left[\frac{d(u_\tau, u^k)^2 - d(u^{k+1}, u^k)^2}{\tau} \right] \leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{(x - y)}{h} \cdot \boldsymbol{\xi} \, d\mu_{k+1}.$$

Replacing $\boldsymbol{\xi}$ by $-\boldsymbol{\xi}$ and recalling (6.4), (6.12), we obtain (6.9). □

6.3 Convergence

Let us suppose now that as $h \to 0$,

$$u_h \to u \text{ strongly in } L^1_{\text{loc}}(\mathbb{R}^n \times (0,\infty)). \tag{6.14}$$

Theorem 6.1 *The function u is a weak solution of the nonlinear diffusion problem*

$$\begin{cases} u_t = \Delta\alpha(u) & \textit{in } \mathbb{R}^n \times (0,\infty) \\ u = g & \textit{on } \mathbb{R}^n \times \{t=0\}. \end{cases} \tag{6.15}$$

Notice from (6.6) that

$$\alpha'(z) = \beta''(z)z \geq 0 \quad (z \geq 0),$$

and so this nonlinear PDE is parabolic, corresponding to a nonlinear diffusion.

Proof. Let us fix $\zeta \in C_c^\infty(\mathbb{R}^n)$ and take $\boldsymbol{\xi} := D\zeta$. Then

$$\begin{aligned} \left| \int_{\mathbb{R}^n} u^{k+1}(x)\zeta(x)\,dx - \int_{\mathbb{R}^n} u^k(y)\zeta(y)\,dy \right. &- \left. \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (x-y)\cdot D\zeta(x)\,d\mu_{k+1}(x,y) \right| \\ &= \left| \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} [\zeta(x) - \zeta(y) - (x-y)\cdot D\zeta(x)]\,d\mu_{k+1} \right| \\ &\leq C\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |x-y|^2\,d\mu_{k+1} = Cd(u^{k+1},u^k)^2 \end{aligned}$$

for some constant C. Consequently if $\phi \in C_c^\infty(\mathbb{R}^n \times [0,\infty))$, the Lemma lets us estimate

$$\begin{aligned} \left| -\int_0^\infty \int_{\mathbb{R}^n} u_n \frac{(\phi(\cdot,t+h) - \phi(\cdot,t))}{h}\,dxdt \right. &- \left. \int_h^\infty \int_{\mathbb{R}^n} \alpha(u_h)\Delta\phi\,dxdt - \frac{1}{h}\int_0^h \int_{\mathbb{R}^n} g\phi\,dxdt \right| \\ &\leq C\sum_{k=1}^\infty d(u^{k+1},u^k)^2 \leq Ch. \end{aligned}$$

Owing then to (6.5),

$$\int_0^\infty \int_{\mathbb{R}^n} -u\phi_t - \alpha(u)\Delta\phi\,dxdt - \int_{\mathbb{R}^n} g\phi(\cdot,0)\,dx = 0.$$

That this identity hold for all ϕ as above means u is a weak solution of the nonlinear diffusion PDE. □

Remark. An interesting example is

$$\beta(z) = z\log z \quad (z > 0),$$

in which case the term

$$\int_{\mathbb{R}^n} \beta(v)dx = \int_{\mathbb{R}^n} v \log v \, dx$$

corresponds to *entropy*, and

$$\alpha(z) = \beta'(z)z - \beta(z) = z.$$

So the usual linear heat equation results from this time-stepping procedure. It is best, however, to continue to regard each u^k as a density, which at each stage balances the Wasserstein distance from the previous density against the entropy production. As explained in Jordan–Kinderlehrer–Otto [J-K-O1], we can therefore envision the individual approximations u^k as something like "Gibbs states", in a sort of local equilibrium. □

7 Application: Semigeostrophic models in meteorology

In this section we explain the remarkable connections between Monge–Kan-torovich theory and the so-called semigeostrophic equations from meteorology, a model for large scale, stratified atmospheric flows with front formation. The most accessible reference for mathematicians is Cullen–Norbury–Purser [C-N-P] (but see also [C-P1],[C-P2],[C-P3],[C-S].)

7.1 The PDE in physical variables

In this system of PDE there are seven unknowns:

$$\begin{cases} \mathbf{v}_g &= (v_g^1, v_g^2, 0) = \text{ \textit{geostrophic} wind velocity} \\ \mathbf{v}_a &= (v_a^1, v_a^2, v_a^3) = \text{ \textit{ageostrophic} wind velocity} \\ p &= \textit{pressure} \\ \theta &= \textit{potential temperature}, \end{cases} \tag{7.1}$$

defined in $X \times [0, \infty)$, X denoting a fixed region in $\mathbb{R}^3$. We also set

$$\mathbf{v} := \mathbf{v}_g + \mathbf{v}_a = \text{ total wind velocity.}$$

Write $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, $D_x = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right)$, and define the *convective derivative*

$$\frac{D_x}{Dt} := \frac{\partial}{\partial t} + \mathbf{v} \cdot D_x = \frac{\partial}{\partial t} + v^1 \frac{\partial}{\partial x_1} + v^2 \frac{\partial}{\partial x_2} + v^3 \frac{\partial}{\partial x_3}. \tag{7.2}$$

The *semigeostrophic equations* are then these seven equations:

$$\begin{cases} \text{(i)} & \dfrac{D_x v_g^1}{Dt} - v_a^2 = 0, \ \dfrac{D_x v_g^2}{Dt} + v_a^1 = 0 \\ \text{(ii)} & \dfrac{D_x \theta}{Dt} = 0 \\ \text{(iii)} & \operatorname{div}(\mathbf{v}) = 0 \\ \text{(iv)} & D_x p = (v_g^2, -v_g^1, \theta), \end{cases} \tag{7.3}$$

where we have set all physical parameters to 1. The equations (i) represent a simplification of the full Euler equations in regimes where centrifugal forces dominate, and (ii) is the passive transport of the potential θ with the flow. The equality (iii) of course means incompressibility. The first two components of (iv) embody the definition of the geostrophic wind $\mathbf{v}_g$ and the last is the definition of θ. Observe that v_a^1, v_a^2 are defined by (i) and that v_a^3 enters the system of PDE only through (iii).

We couple (7.3) with appropriate initial conditions and the boundary condition

$$\mathbf{v} \cdot \boldsymbol{\nu} = 0 \qquad \text{on } \partial U, \tag{7.4}$$

$\boldsymbol{\nu}$ being the outward unit normal along ∂U.

7.2 The PDE in dual variables

The system (7.3) is quite complicated, but its structure clarifies under an appropriate change of variable. We hereafter introduce new functions

$$\mathbf{y} = (y^1, y^2, y^3) := (x_1 + v_g^2, x_2 - v_g^1, \theta) \tag{7.5}$$

and

$$s := \det(D_x \mathbf{y}). \tag{7.6}$$

Then (7.3)(i),(ii) transform to read

$$\frac{D_x \mathbf{y}}{Dt} = \mathbf{v}_g, \tag{7.7}$$

and

$$\frac{D_x s}{Dt} = 0. \tag{7.8}$$

We have in mind next to change to new independent variables $y = (y_1, y_2, y_3)$. To do so introduce

$$\phi := \frac{1}{2}(x_1^2 + x_2^2) + p. \tag{7.9}$$

Then

$$\mathbf{y} = D_x \phi. \tag{7.10}$$

Consequently (7.6) says

$$s = \det(D_x^2 \phi). \tag{7.11}$$

We assume ϕ is convex, and write ψ for its convex dual, $\psi = \psi(y)$. Then

$$\mathbf{x} = D_y \psi, \tag{7.12}$$

for $\mathbf{x} = (x^1, x^2, x^3)$. If we write

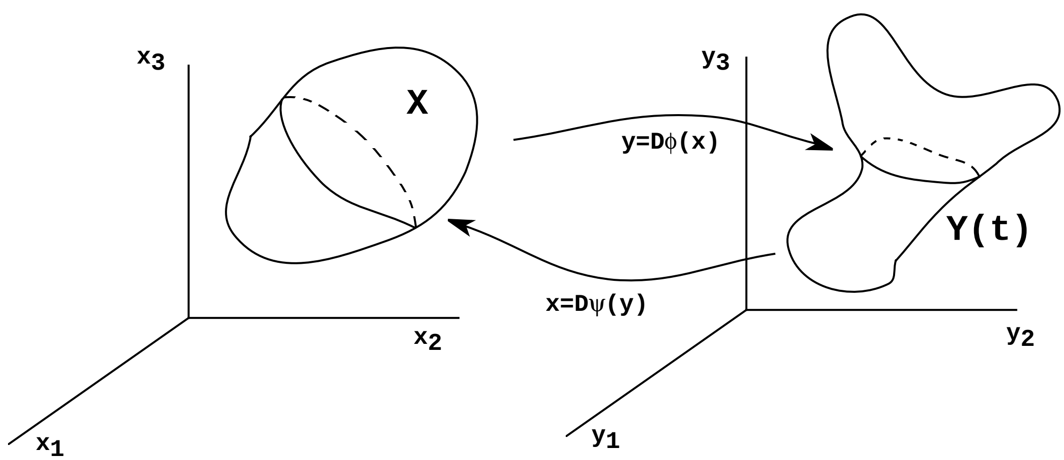
$$r := \det(D_y \mathbf{x}) = \det(D_y^2 \psi), \tag{7.13}$$

then using (7.7), (7.8) we can compute

$$\frac{D_y r}{Dt} = 0, \tag{7.14}$$

where

$$\frac{D_y}{Dt} := \frac{\partial}{\partial t} + \mathbf{v}_g \cdot D_y = \frac{\partial}{\partial t} + v_g^1 \frac{\partial}{\partial y_1} + v_g^2 \frac{\partial}{\partial y_2}. \tag{7.15}$$



Now let $Y(t)$ denote the range of $D_x \phi : X \to \mathbb{R}^3$, at time t. Then we can in summary rewrite (7.13), (7.14) to read

$$\begin{cases} \text{(i)} & r_t + \operatorname{div}(r\mathbf{w}) = 0 \\ \text{(ii)} & \mathbf{w} = (\psi_{y_2} - y_2, -(\psi_{y_1} - y_1), 0) \quad \text{in } Y(t) \times [0, \infty) \\ \text{(iii)} & \det(D^2 \psi) = r, \end{cases} \tag{7.16}$$

where we have set $\mathbf{w} := \mathbf{v}_g$. The additional requirement is

$$D\psi \in X. \tag{7.17}$$

The system (7.16) is a sort of time-dependent Monge–Ampere equation involving the moving free boundary $\partial Y(t)$. Perhaps more interestingly, (7.16) is an obvious variant of the vorticity formulation of the two-dimensional Euler equations

$$\begin{cases} \text{(i)} & \omega_t + \operatorname{div}(\omega \mathbf{v}) = 0 \\ \text{(ii)} & \mathbf{v} = (-\psi_{x_2}, \psi_{x_1}) \quad \text{in } \mathbb{R}^2 \times [0, \infty). \\ \text{(iii)} & \Delta\psi = \omega, \end{cases} \tag{7.18}$$

where ψ is the stream function, and ω the scalar vorticity. Roughly speaking, the system (7.18) is a linearization of (7.16).

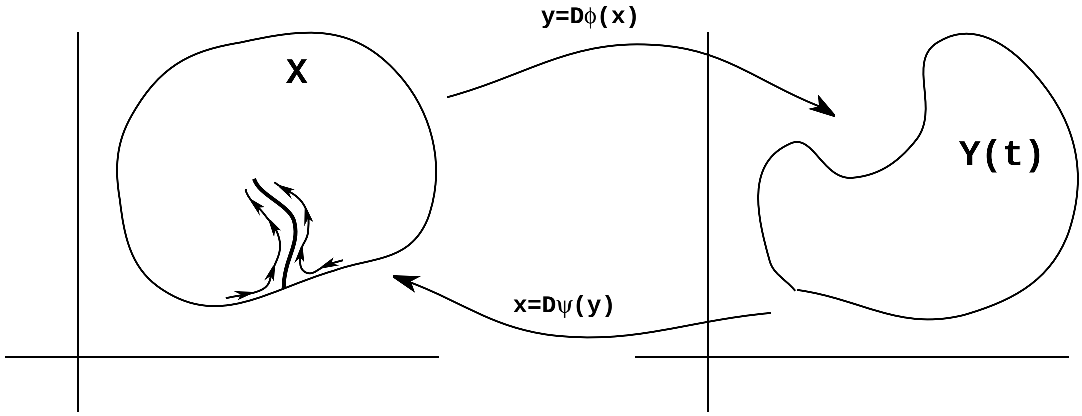
F. Otto [O2] and Benamou–Brenier [Be-B1] have shown the existence of a weak solution of (7.16), making use of the pair (ϕ, ψ) in the sense of the Monge–Kantorovich theory discussed above. One interpretation is that while $r, \mathbf{w}$ evolve on an "order-one" time scale, there is an optimal Monge–Kantorovich rearrangement of air parcels on a "fast" time scale. See also Brenier [B4].

7.3 Frontogenesis

We can informally interpret the dynamics (7.16) as supporting the onset of "fronts", i.e. surfaces of discontinuity in the velocity and temperature.

To understand this, suppose $X, Y(0)$ are uniformly convex, and for simplicity take $r \equiv 1$. Then (7.16)(i) is trivial and we can understand (7.16)(ii), (iii) as a law of evolution of $\partial Y(t)$. Suppose for heuristic purposes that $\partial Y(t)$ remains smooth. Then owing to the regularity theory (§4) $\psi(\cdot, t)$ will be smooth for each $t \geq 0$. But since $Y(t)$ is changing shape, it is presumably possible that $Y(t)$ is no longer convex at some sufficiently large time t.

Then the regularity theory from §4 no longer applies to $\phi(\cdot, t)$. We may consequently expect that although $\psi(\cdot, t)$ remains smooth, $\phi(\cdot, t)$ will not.



Far from being a defect, the advent of singularities of $\phi(\cdot, t)$ in the physical variables $x = (x_1, x_2, x_3)$ is a definite advantage in the model. As discussed in Cullen [C], the meteorologists wish to model how fronts arise in large scale weather patterns. Tracking these fronts, that is, thin regions across which there are large variations in wind and temperature, is a central goal, and the semigeostrophic equations provide a plausible mechanism for frontogenesis. An interesting physical rationale occurs in [C], where the author points out that such discontinuities are *contact discontinuities* in the sense of fluid mechanics: this means that the air parcels move parallel to, and not across, them. Obviously in regions of rapid temperature and wind change the approximations which transform the full Euler equations into the semigeostrophic PDE are very suspect. But, since the proposed fronts are contact discontinuities, most of the air mass stays away from such regions, and so the various approximations should overall be pretty good.

There are extremely interesting mathematical problems here, which have only in small part been studied. There are likewise many issues concerning computing these flows: see Benamou [Be1],[Be2] and Benamou–Brenier [Be-B2] for this.

Part II: Cost = Distance

8 Heuristics

For the rest of this paper we turn attention to the nonuniformly convex cost density

$$c(x, y) = |x - y| \qquad (x, y \in \mathbb{R}^n). \tag{8.1}$$

The task now is to find an optimal mass transfer plan $\mathbf{s}^*$ solving Monge's original problem (1.5), where

$$I[\mathbf{s}] := \int_{\mathbb{R}^n} |x - \mathbf{s}(x)| \, d\mu^+(x) \tag{8.2}$$

for $\mathbf{s} \in \mathcal{A}$, the admissible class of measurable, one-to-one maps of $\mathbb{R}^n$ which push μ^+ to μ^-.

It will turn out that the nonuniform convexity of the cost density (8.1) defeats any simple attempt to modify the techniques described before in §2–7. We will therefore first of all need some new insights, to help us sort out the structure of optimal mass allocations.

8.1 Geometry of optimal transport

As in §2 we start with heuristics, our intention being to read off useful information from a "twist variation".

So assume that $\mathbf{s}^* \in \mathcal{A}$ minimizes (8.2). Fix then $m \geq 2$, select any points $\{x_k\}_{k=1}^m \subset X = \text{ spt } (\mu^+)$, and small balls $E_k := B(x_k, r_k)$, the radii selected so that $\mu^+(E_1) = \cdots = \mu^+(E_m) = \varepsilon$. Next set $y_k = \mathbf{s}^*(x_k)$, $F_k := \mathbf{s}^*(E_k)$, and finally fix an integer $l \geq 1$. Similarly now to the construction in §2, we build $\mathbf{s} \in \mathcal{A}$ so that

$$\begin{cases} \mathbf{s}(x_k) = y_{k+l}, \ \mathbf{s}(E_k) = F_{k+l} \\ \mathbf{s} \equiv \mathbf{s}^* \text{ on } X - \displaystyle\bigcup_{k=1}^m E_k, \end{cases}$$

where we compute the subscripts mod m. As in §2 we deduce from the inequality $I[\mathbf{s}^*] \leq I[\mathbf{s}]$ that

$$\sum_{k=1}^m |x_k - y_k| \leq \sum_{k=1}^m |x_k - y_{k+l}|. \tag{8.3}$$

We as follows draw a geometric deduction from (8.3). Take any closed curve C lying in $X = \text{spt } (\mu^+)$, with constant speed parameterization $\{\mathbf{r}(t) \mid 0 \leq t \leq 1\}$, $\mathbf{r}(0) = \mathbf{r}(1)$. Let $x_k := \mathbf{r}\left(\frac{k}{m}\right)$ be m equally spaced points along C.

Fix $\tau > 0$. Then using (8.3) and letting $m \to \infty$, $\frac{l}{m} \to \tau$, we deduce:

$$\begin{aligned} \int_0^1 |\mathbf{r}(t) - \mathbf{s}^*(\mathbf{r}(t))| dt \ &\leq \ \int_0^1 |\mathbf{r}(t) - \mathbf{s}^*(\mathbf{r}(t+\tau))| dt \\ &= \ \int_0^1 |\mathbf{r}(t-\tau) - \mathbf{s}^*(\mathbf{r}(t))| dt =: i(\tau). \end{aligned}$$

Hence $i(\cdot)$ has a minimum at $\tau = 0$, and so

$$0 = i'(0) = -\int_0^1 \frac{[\mathbf{r}(t) - \mathbf{s}^*(\mathbf{r}(t))] \cdot \mathbf{r}'(t)}{|\mathbf{r}(t) - \mathbf{s}^*(\mathbf{r}(t))|} dt.$$

This is to say,

$$\int_C \boldsymbol{\xi} \cdot d\mathbf{r} = 0, \tag{8.4}$$

for the vector field

$$\boldsymbol{\xi}(x) := \frac{\mathbf{s}^*(x) - x}{|\mathbf{s}^*(x) - x|}.$$

As (8.4) holds for all closed curves C, we conclude that $\boldsymbol{\xi}$ is a gradient:

$$\frac{\mathbf{s}^*(x) - x}{|\mathbf{s}^*(x) - x|} = -Du^*(x) \tag{8.5}$$

for some scalar mapping $u^* : \mathbb{R}^n \to \mathbb{R}$, which solves the *eikonal equation*

$$|Du^*| = 1 \tag{8.6}$$

on $X = \text{spt}(\mu^+)$. In other words, the direction $\mathbf{s}^*$ maps each point x is given by the gradient of a potential:

$$\mathbf{s}^*(x) = x - d^*(x)Du^*(x).$$

Note very carefully however that this deduction provides absolutely no information about the distance $d^*(x) := |\mathbf{s}^*(x) - x|$.

Remark. The foregoing insights were first discovered by Monge, based upon completely different, geometric arguments concerning developable surfaces. See Monge [M], Dupin [D]. Monge's and his students' investigations of this problem lead to many of their foundational discoveries in differential geometry: see for instance Struik [St]. □

8.2 Lagrange multipliers

It is interesting to see also a formal, analytic derivation of (8.5), (8.6). The following computations are essentially those of Appell [A], from the turn of the century.

We assume for this that $d\mu^+ = f^+dx$, $d\mu^- = f^-dy$ for smooth densities $f^\pm$. The *augmented work functional* is

$$\tilde{I}[\mathbf{s}] := \int_{\mathbb{R}^r} |x - \mathbf{s}(x)|f^+(x) + \lambda(x)[f^-(\mathbf{s}(x))\text{det}(D\mathbf{s}(x)) - f^+(x)]\, dx,$$

where λ is the *Lagrange multiplier* for the constraint $f^+ = f^-(\mathbf{s})\text{det}(D\mathbf{s})$. The first variation is

$$(\lambda f^-(\mathbf{s}^*)(\text{cof}D\mathbf{s}^*)_i^k)_{x_i} = \frac{s^{*k} - x_k}{|\mathbf{s}^* - x|} f^+(x) + \lambda f^-_{y_k}(\mathbf{s}^*)\text{det}Ds^*.$$

Simplifying as in §2, we deduce

$$\lambda x_j = \frac{s^{*k} - x_k}{|\mathbf{s}^* - x|} s^{*k}_{x_j} \quad (j = 1, \dots, n).$$

Next define u^* by

$$u^*(\mathbf{s}^*(x)) = -\lambda(x).$$

Then

$$u^*_{y_k} s^{*k}_{x_j} = -\lambda x_j = \frac{s^{*k} - x_k}{|\mathbf{s}^* - x|} s^{*k}_{x_j}.$$

As $D\mathbf{s}^*$ is invertible, we see

$$Du^*(s^*(x)) = -\frac{s^* - x}{|s^* - x|}.$$

But $Du^*(\mathbf{s}^*(x)) = Du^*(x)$, and so (8.5), (8.6) again follow.

Remark. It is instructive to compare and contrast the heuristics of this section with those in §2. The argument based upon the "twist variation" in §8.1 is more general than that using cyclic monotonicity in §2.1, and adapts without trouble to a general cost density $c(x, y)$. The conclusion is then that

$$D_x c(x, \mathbf{s}^*(x)) = Du^*(x) \text{ for some scalar potential function } u^*. \tag{8.7}$$

Now *if* we can invert (8.7), to solve for $\mathbf{s}^*$ in terms of Du^*, we thereby derive a structural formula for an optimal mapping. Much of the interest in (8.1) is precisely that this inversion is not possible if c is not uniformly convex. □

9 Optimal mass transport

9.1 Solution of dual problem

Our intention now is to transform the foregoing heuristic calculations into a proof of the existence of an optimal mapping $\mathbf{s}^*$, where we expect

$$\mathbf{s}^*(x) = x - d^*(x) Du^*(x) \tag{9.1}$$

for some potential u^*, with $|Du^*(x)| = 1$ and so $d^*(x) = |\mathbf{s}^*(x) - x|$.

We turn to the case

$$d\mu^+ = f^+ dx, \; d\mu^- = f^- dy \tag{9.2}$$

where $f^\pm$ are bounded, nonnegative functions with compact support, satisfying the *mass balance* compatibility condition

$$\int_X f^+(x)\, dx = \int_Y f^-(y)\, dy \tag{9.3}$$

where $X = \operatorname{spt}(f^+)$, $Y = \operatorname{spt}(f^-)$.

We further remember the dual problem (§1), which asks us to find (u^*, v^*) to *maximize*

$$K[u,v] := \int_X u(x)f^+(x)\,dx + \int_Y v(y)f^-(y)\,dy \tag{9.4}$$

subject to

$$u(x) + v(y) \leq |x-y| \qquad (x \in X,\ y \in Y). \tag{9.5}$$

Lemma 9.1 (i) *There exist (u^*, v^*) solving this maximization problem.*
(ii) *Furthermore, we can take*

$$v^* = -u^*, \tag{9.6}$$

where $u^* : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous, with

$$|u^*(x) - u^*(y)| \leq |x-y| \qquad (x, y \in \mathbb{R}^n). \tag{9.7}$$

Proof. 1. If u, v satisfy (9.5), then

$$u(x) \leq \inf_{y\in Y}(|x-y| - v(y)) =: \hat{u}(x) \tag{9.8}$$

and

$$\hat{u}(x) + v(y) \leq |x-y| \qquad (x \in X,\ y \in Y).$$

Therefore

$$v(y) \leq \min_{x\in X}(|x-y| - \hat{u}(x)) =: \hat{v}(y) \tag{9.9}$$

and

$$\hat{u}(x) + \hat{v}(y) \leq |x-y| \qquad (x \in X,\ y \in Y). \tag{9.10}$$

Furthermore, since $\hat{v} \geq v$, (9.8) implies

$$\hat{u}(x) \geq \min_{y\in Y}(|x-y| - \hat{v}(y));$$

and so (9.10) implies

$$\hat{u}(x) = \min_{y\in Y}(|x-y| - \hat{v}(y)). \tag{9.11}$$

Since $f^\pm \geq 0$ and $\hat{u} \geq u$, $\hat{v} \geq v$ we see that $K[u,v] \leq K[\hat{u}, \hat{v}]$.

2. Thus in seeking to maximize K we may restrict attention to "dual" pairs $(\hat{u}, \hat{v})$, as above. But then

$$\hat{u} + \hat{v} = 0 \text{ on } X \cap Y. \tag{9.12}$$

To see this, take $z \in X \cap Y$ (if $X \cap Y \neq \emptyset$), and suppose

$$\hat{u}(z) + \hat{v}(z) < 0. \tag{9.13}$$

Take $x \in X$, $y \in Y$ so that

$$\begin{cases} \hat{u}(z) = |z - y| - \hat{v}(y) \\ \hat{v}(z) = |x - z| - \hat{u}(x). \end{cases}$$

Rearrange and add:

$$\begin{aligned} |x - z| + |z - y| &= \hat{u}(x) + \hat{v}(y) + \hat{u}(z) + \hat{v}(z) \\ &< \hat{u}(x) + \hat{v}(y) \text{ by (9.13)} \\ &\leq |x - y| \text{ by (9.10)}. \end{aligned}$$

This contradiction shows $\hat{u} + \hat{v} \geq 0$ on $X \cap Y$. As (9.10) clearly implies $\hat{u} + \hat{v} \leq 0$ on $X \cap Y$, assertion (9.12) follows.

3. In view of (9.12) let us extend the definition of $\hat{u}$ by setting

$$\hat{u} := -\hat{v} \qquad \text{on } Y.$$

Then (9.10) reads

$$\hat{u}(x) - \hat{u}(y) \leq |x - y| \qquad (x \in X,\ y \in Y).$$

Finally it is an exercise to check that we can extend $\hat{u}$ to all of $\mathbb{R}^n$, so that

$$|\hat{u}(x) - \hat{u}(y)| \leq |x - y| \qquad (x, y \in \mathbb{R}^n).$$

Our problem is thus to maximize

$$K[u] := \int_{\mathbb{R}^n} u(f^+ - f^-)\, dz, \tag{9.14}$$

subject to the Lipschitz constraint

$$|u(x) - u(y)| \leq |x - y| \qquad (x, y \in \mathbb{R}^n). \tag{9.15}$$

This problem clearly admits a solution u^*, and we define $v^* := -u^*$ to obtain a pair solving the original dual problem. □

Note that the Lipschitz condition implies that Du^* exists a.e. in $\mathbb{R}^n$, with

$$|Du^*| \leq 1 \quad \text{a.e.}\ .$$

We further expect

$$|Du^*| = 1 \quad \text{a.e. on } X \cup Y.$$

9.2 Existence of optimal mass transfer plan

We wish next to employ u^* to build an optimal mass allocation mapping $\mathbf{s}^*$. This is not so easy as in the uniformly convex case that $c(x,y) = \frac{1}{2}|x-y|^2$, discussed in §3. The central problem is that although we expect $\mathbf{s}^*$ to have the structure (9.1) there is still an unknown, namely the *distance* $d^*(x) = |\mathbf{s}^*(x) - x|$ that the point x should move: $Du^*(x)$ tells us only the *direction*.

This problem was solved by Sudakov [Su] using rather subtle measure theoretic techniques (but see also the comments in Ambrosio [Am].). In keeping with the overall theme of this paper, we will here discuss instead an alternative, differential-equations-based procedure, from [E-G1]. Our argument introduces some useful ideas, but is really complicated. Far better proofs have been recently found by Trudinger–Wang [T-W] and [C-F-M] Caffarelli–Feldman–McCann.

To repeat, the basic issue is that we must somehow extract the missing information about the distance $d^*(x) = |\mathbf{s}^*(x) - x|$ from the variational problem (9.14), (9.15). It is convenient in doing so to introduce some standard notion from convex analysis. Let us set

$$\mathbb{K} := \{u \in L^2(\mathbb{R}^n) \mid |Du| \leq 1 \text{ a.e.}\} \tag{9.16}$$

and

$$I_\infty[u] := \begin{cases} 0 & \text{if } u \in \mathbb{K} \\ +\infty & \text{otherwise.} \end{cases} \tag{9.17}$$

Then u^* minimizes $K[\cdot]$ over $\mathbb{K}$. The corresponding Euler Lagrange equation is

$$f^+ - f^- \in \partial I_\infty[u^*]; \tag{9.18}$$

that is,

$$I_\infty[v] \geq I_\infty[u^*] + (f^+ - f^-, v - u^*)_{L^2}$$

for all $v \in L^2(\mathbb{R}^n)$.

We need firstly to convert (9.18) into more concrete form:

Lemma 9.2 *Assume additionally that $f^\pm$ are Lipschitz continuous.*

(i) *Then there exists a nonnegative L^∞ function a such that*

$$-\operatorname{div}(aDu^*) = f^+ - f^- \quad \text{in } \mathbb{R}^n. \tag{9.19}$$

(ii) *Furthermore*

$$|Du| = 1 \quad \text{a.e. on the set } \{a > 0\}.$$

We call a the *transport density*. The PDE (9.19) looks linear, but is not: *the function a is the Lagrange multiplier corresponding to the constraint $|Du^*| \leq 1$*, and is a highly nonlinear and nonlocal function of u^*. On the other hand once u^* is known, (9.19) can be thought of as a linear, first-order PDE for a.

Outline of Proof. Take $n+1 \leq p < \infty$. We approximate by the quasilinear PDE

$$-\operatorname{div}(|Du_p|^{p-2} Du_p) = f^+ - f^-, \tag{9.20}$$

which corresponds to the problem of maximizing

$$K_p[u] := \int_{\mathbb{R}^n} u(f^+ - f^-) - \frac{1}{p}|Du|^p \, dz.$$

A maximum principle argument shows that

$$\sup_p |u_p|, |Du_p|, |Du_p|^p \leq C < \infty$$

for some constant C. (Cf. Bhattacharya–DiBenedetto–Manfredi [B-B-M].)

It follows that there exists a sequence $p_k \to \infty$ such that

$$\begin{cases} u_{p_k} \to u^* & \text{locally uniformity} \\ Du_{p_k} \to Du^* & \text{boundedly, a.e.} \\ |Du_{p_k}|^{p-2} \rightharpoonup a & \text{weakly } * \text{ in } L^\infty. \end{cases}$$

Then passing to limits in (9.20) we obtain (9.19). □

As noted above, a is the Lagrange multiplier from the constraint $|Du^*| \leq 1$. It turns out furthermore that *a in fact "contains" the missing information as to the distance $d^*(x)$*.

The recipe is to build $\mathbf{s}^*$ by solving a flow problem involving Du^*, a, etc. So fix $x \in X$ and consider the ODE

$$\begin{cases} \dot{\mathbf{z}}(t) = \mathbf{b}(\mathbf{z}(t), t) & (0 \leq t \leq 1) \\ \mathbf{z}(0) = x, \end{cases} \tag{9.21}$$

for the time-varying vector field

$$\mathbf{b}(z, t) := \frac{-a(z) Du^*(z)}{t f^-(z) + (1-t) f^+(z)}. \tag{9.22}$$

Ignore for the moment that a, Du^* are not smooth (or even continuous in general) and that we may be dividing by zero in (9.22). Proceeding formally then, let us write

$$\mathbf{s}^*(x) = z(1), \tag{9.23}$$

the time-one map of the flow.

Theorem 9.1 *Define* $\mathbf{s}^*$ *by* (9.23). *Then*

(i) $s^* : X \to Y$ *is essentially one-to-one and onto.*

(ii) $\int_X h(\mathbf{s}^*(x))\, d\mu^+(x) = \int_Y h(y)\, d\mu^-(y)$ *for each* $h \in C(Y)$.

(iii) *Lastly,*

$$\int_X |x - \mathbf{s}^*(x)|\, d\mu^+(x) \leq \int_X |\mathbf{x} - \mathbf{s}(x)|\, d\mu^+(x)$$

for all $\mathbf{s} : X \to Y$ *such that* $\mathbf{s}_{\#}(\mu^+) = \mu^-$.

Outline of Proof. 1. Write

$$\mathbf{s}(t,x) := z(t) \qquad (0 \leq t \leq 1),$$

$$J(z,t) := \det D\mathbf{s}(z,t).$$

Then $J_t = (\text{div } \mathbf{b})J$; and so, following Dacorogna–Moser [D-M], we may compute

$$\begin{aligned}
&\frac{\partial}{\partial t}[(tf^-(\mathbf{s}(z,t)) + (1-t)f^+(\mathbf{s}(z,t)))J(z,t)] \\
&= (f^- - f^+)J + (tDf^- \cdot \mathbf{s}_t + (1-t)Df^+ \cdot \mathbf{s}_t)J \\
&\qquad + (tf^- + (1-t)f^+)J_t \\
&= [(f^- - f^+) + (tDf^- \cdot \mathbf{b} + (1-t)Df^+ \cdot \mathbf{b}) \\
&\qquad + (tf^- + (1-t)f^+)\text{div } \mathbf{b}]J.
\end{aligned} \tag{9.24}$$

But in view of (9.19), (9.22):

$$\begin{aligned}
\text{div } \mathbf{b} &= \frac{f^+ - f^-}{tf^- + (1-t)f^+} + \frac{(tDf^- + (1-t)Df^+) \cdot (aDu^*)}{(tf^- + (1-t)f^+)^2} \\
&= \frac{f^+ - f^- - (tDf^- + (1-t)Df^+) \cdot \mathbf{b}}{tf^- + (1-t)f^+}.
\end{aligned}$$

Consequently the last term in (9.24) is zero, and hence

$$f^-(\mathbf{s}^*)\det D\mathbf{s}^* = f^+.$$

This confirms assertion (ii).

2. We next verify $\mathbf{s}^*$ is optimal. For this, take $\mathbf{s}$ to be any admissible mapping.

Then

$$\begin{aligned}\int_X |x - \mathbf{s}^*(x)|\, d\mu^+(x) &= \int_X [u^*(x) - u^*(\mathbf{s}^*(x))] f^+(x)\, dx \\ &= \int_X u^* f^+ - \int_Y u^* f^-\, dy \\ &= \int_X [u^*(x) - u^*(\mathbf{s}(x))] f^+(x)\, dx \\ &\leq \int_X |x - \mathbf{s}(x)| f^+(x)\, dx.\end{aligned}$$

□

This "Theorem" should of course really be in quotes, since the "proof" just outlined is purely formal. In reality neither a nor Du^* nor $f^\pm$ are smooth enough to justify these computations. A careful proof is available in [E-G1]: the full details are extremely complicated, involving a very careful smoothing of u^*. The proof requires as well the additional conditions that $\partial X, \partial Y$ are nice and $X \cap Y = \emptyset$, although these requirements are presumably not really necessary.

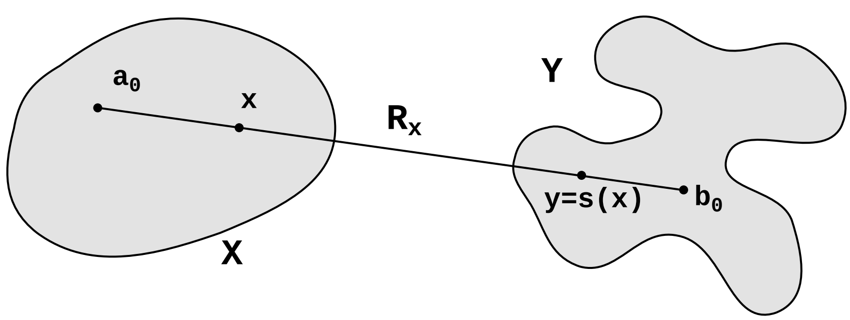
Remark. A nice paper by Cellina and Perrotta [C-P] discusses somewhat related issues. Jensen [Je] considers the subtle problem of what happens to the approximations u_p within the regions $\{f^+ \equiv f^- \equiv 0\}$, in the limit $p \to \infty$.

More on the connections with optimal flow problems may be found in Iri [I] and Strang [S1], [S2]. □

9.3 Detailed mass balance, transport density

For reference later we record here some properties of the optimal potential u^* and transport density a, proved in [E-G1].

First a.e. point $x \in X$ lies in a unique maximal line segment R_x along which u^* decreases linearly at rate one. We call R_x the *transport ray* through x. The idea is that we move the point x "downhill" along R_x to $y = \mathbf{s}^*(x)$, and the ODE (9.21), (9.22) tells us how far to go.



Next we note that these transport rays subdivide X and Y into subregions of equal μ^+, μ^- measure.

Lemma 9.3 *Let $E \subset \mathbb{R}^n$ be a measurable set with the property that for each point $x \in E$, the full transport ray R_x also lies in E. Then*

$$\int_{X \cap E} f^+ \, dx = \int_{Y \cap E} f^- \, dy. \tag{9.25}$$

We call (9.25) the *detailed mass balance* relation. It asserts that the line segments along which u^* changes linearly with slope one naturally partition X and Y into subregions of equal masses. This must be so if our transport scheme (9.23) is to work.

Outline of Proof. We provide an interesting, but purely heuristic, derivation of a somewhat weaker statement. The trick is to employ a Hamilton–Jacobi PDE to generate a variation for the maximization principle (9.14), (9.15).

For this, take $H : \mathbb{R}^n \to \mathbb{R}$ to be any smooth Hamiltonian. We solve then the initial-value problem:

$$\begin{cases} w_t + H(Dw) = 0 & \text{in } \mathbb{R}^n \times \{t > 0\} \\ w = u^* & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{9.26}$$

Now the mapping $t \to w(\cdot, t)$ is a contraction in the sup-norm, and therefore for each time $t > 0$:

$$\|Dw(\cdot, t)\|_{L^\infty} \leq \|Du^*(\cdot, t)\|_{L^\infty} \leq 1.$$

Hence $w(\cdot, t)$ is a valid competitor in the variational principle. Since u^* solves (9.14), (9.15) and $w(\cdot, 0) = u^*$, it follows that

$$i(t) \leq i(0) \quad (t \geq 0),$$

where

$$i(t) := \int_{\mathbb{R}^n} w(\cdot, t)(f^+ - f^-) dz.$$

Hence $i'(0) \leq 0$. In view of (9.26) therefore,

$$\int_{\mathbb{R}^n} H(Du^*)(f^+ - f^-) dz = -i'(0) \geq 0.$$

Replacing H by $-H$, we conclude that

$$\int_{\mathbb{R}^n} H(Du^*)(f^+ - f^-) dz = 0 \tag{9.27}$$

for *all* smooth $H : \mathbb{R}^n \to \mathbb{R}$.

Taking H to approximate χ_A, where $A \subset S^{n-1}$, we deduce from (9.27) that the detailed mass balance holds for the particular set $E := \{z \mid Du^*(z) \in A\}$. □

This proof is not really rigorous as u^*, w are not smooth enough to justify the stated computations. See e.g. [E-G1] for a careful proof.

It is also useful to understand how smooth the transport density is along the transport rays, and we also need to check that the ODE flow (9.21) does not "overshoot" the endpoints.

Lemma 9.4 (i) *For a.e. x, the transport density a, restricted to R_x, is locally Lipschitz continuous.*

(ii) *Furthermore,*

$$\lim_{\substack{z \to a_0, b_0 \\ z \in R_x}} a(z) = 0,$$

where a_0, b_0 are the endpoints of R_x.

The first calculations in the literature related to assertion (ii) seem to be those of Janfalk [J].

Remark. The argument discussed above introduces various ideas useful in the following applications, but is *very* complicated in detail. Far superior, and much shorter, proofs have been independently found by Trudinger–Wang [T-W] and [C-F-M] Caffarelli–Feldman–McCann. A clear proof is also available in Ambrosio [Am]. □

10 Application: Shape optimization

The ensuing three sections describe some applications extending the ideas set forth in §8,9.

As a first application, we discuss the following shape optimization problem. Suppose we are given two nonnegative Radon measures $\mu^{\pm}$ on $\mathbb{R}^n$, with $\mu^+(\mathbb{R}^n) = \mu^-(\mathbb{R}^n)$. Think of μ^+ as giving the density of a given electric charge. We interpret $\mathbb{R}^n$ as an insulating medium, into which we place a fixed amount of some conducting material, whose conductivity (= inverse resistivity) is described by a nonnegative measure σ. We imagine then the resulting steady current flow within $\mathbb{R}^n$ from μ^+ to μ^-, and ask if we can optimize the placement of the conducting material so as to *minimize the heating* induced by the flow.

To be more precise, consider the admissible class

$$\mathcal{S} = \{\text{nonnegative Radon measures } \sigma \text{ on } \mathbb{R}^n \mid \sigma(\mathbb{R}^n) = 1\}$$

and think of each $\sigma \in \mathcal{S}$ as describing how to arrange a unit quantity of conducting material within $\mathbb{R}^n$. Corresponding to each $\sigma \in \mathcal{S}$ and scalar function $v \in C_c^\infty(\mathbb{R}^n)$ we define

$$E(\sigma, v) := \frac{1}{2} \int_{\mathbb{R}^n} |Dv|^2 \, d\sigma - \int_{\mathbb{R}^n} v \, d(\mu^+ - \mu^-). \tag{10.1}$$

Then

$$E(\sigma) := \inf\{E(\sigma, v) \mid v \in C_c^\infty(\mathbb{R}^n)\} \tag{10.2}$$

represents the negative of the Joule heating (= energy dissipation) corresponding to the given conductivity. If there exists $v \in C_c^\infty(\mathbb{R}^n)$ giving the minimum, then

$$-\operatorname{div}(\sigma Dv) = \mu^+ - \mu^- \quad \text{in } \mathbb{R}^n; \tag{10.3}$$

that is,

$$\int_{\mathbb{R}^n} Dv \cdot Dw \, d\sigma = \int_{\mathbb{R}^n} w \, d(\mu^+ - \mu^-)$$

for each $w \in C_c^\infty(\mathbb{R}^n)$. We can interpret

$$\begin{cases} v = \text{ electrostatic potential, } \mathbf{e} = -Dv = \text{electric field,} \\ \mathbf{j} = \sigma \mathbf{e} = \text{ current density (by Ohm's law).} \end{cases}$$

We now ask: Can we find $\sigma^* \in \mathcal{S}$ to *maximize* $E(\sigma)$? In other words, is there an optimal way to arrange the given amount of conductivity material so as to *minimize* the heating?

Following Bouchitte–Buttazzo–Seppechere [B-B-S], we introduce the related Monge–Kantorovich dual problem of maximizing

$$\int_{\mathbb{R}^n} u \, d(\mu^+ - \mu^-) \tag{10.4}$$

among all $u : \mathbb{R}^n \to \mathbb{R}$, with

$$|u(x) - u(y)| \leq 1 \quad (x, y \in \mathbb{R}^n). \tag{10.5}$$

Lemma 10.1 (i) *There exists* u^* *solving* (10.4), (10.5).

(ii) *Furthermore, there exists* $\alpha^* \in \mathcal{A}$ *such that*

$$\begin{cases} -\operatorname{div}(\alpha^* Du^*) = \mu^+ - \mu^- \\ |Du^*| = 1 \quad \alpha^* \ a.e. \end{cases} \tag{10.6}$$

This is clearly a generalization of our work in §9. Note carefully that since α^* is merely a Radon measure, and thus may be in part singular with respect to n-dimensional Lebesgue measure, care is needed in interpreting the PDE in (10.6): see [B-B-S].

Now set

$$\begin{cases} \sigma^* := (\alpha^*(\mathbb{R}^n))^{-1} \alpha^* \\ v^* := \alpha^*(\mathbb{R}^n) u^* \end{cases} \tag{10.7}$$

Then $\sigma^* \in \mathcal{S}$ and

$$-\operatorname{div}(\sigma^* Dv^*) = \mu^+ - \mu^- \quad \text{in } \mathbb{R}^n. \tag{10.8}$$

We claim now that

$$E(\sigma^*) = \max_{\sigma \in \mathcal{S}} E(\sigma). \tag{10.9}$$

For this we invoke another duality principle, namely

$$E(\sigma) = \sup \left\{ -\frac{1}{2} \int_{\mathbb{R}^n} |\mathbf{e}|^2 \, d\sigma \mid \operatorname{div}(\sigma \mathbf{e}) = \mu^+ - \mu^- \right\}. \tag{10.10}$$

Now if u satisfies (10.5) and $\operatorname{div}(\sigma \mathbf{e}) = \mu^+ - \mu^-$, then

$$\begin{aligned}
\int_{\mathbb{R}^n} u \, d(\mu^+ - \mu^-) &= -\int_{\mathbb{R}^n} Du \cdot \mathbf{e} \, d\sigma \\
&\leq \left(\int_{\mathbb{R}^n} |Du|^2 \, d\sigma \right)^{1/2} \left(\int_{\mathbb{R}^n} |\mathbf{e}|^2 \, d\sigma \right)^{1/2} \\
&\leq \left(\int_{\mathbb{R}^n} |\mathbf{e}|^2 \, d\sigma \right)^{1/2},
\end{aligned}$$

since $|Du| \leq 1$ and $\sigma(\mathbb{R}^n) = 1$. Taking the suprema over all u, we deduce

$$-\frac{1}{2} \int_{\mathbb{R}^n} |\mathbf{e}|^2 \, d\sigma \leq -\frac{1}{2} \alpha^*(\mathbb{R}^n)^2.$$

In light of (10.10), then

$$E(\sigma) \leq -\frac{1}{2} \alpha^*(\mathbb{R}^n). \tag{10.11}$$

But since $|Du^*| = 1$ σ^* a.e., we compute for $\mathbf{e}^* := -Dv^*$ that

$$E(\sigma^*) \geq -\frac{1}{2} \int_{\mathbb{R}^n} |\mathbf{e}^*|^2 \, d\sigma^* = -\frac{1}{2} \alpha^*(\mathbb{R}^n)^2.$$

According then to (10.11), σ^* is optimal.

Remark. Results strongly related to these are to be found in earlier work of Iri [I] and Strang [S1], [S2]. □

11 Application: Sandpile models

As a completely different class of applications, we next introduce some physical "sandpile" models evolving in time, for which we can identify a Monge–Kantorovich mass transfer mechanism on a "fast" time scale. In the following examples we regard u as the height function of our sandpiles: the constraint

$$|Du| \leq 1 \tag{11.1}$$

is everywhere imposed, and has the physical meaning that the sand cannot remain in equilibrium if the slope anywhere exceeds the angle of repose $\pi/4$. We will later reinterpret (11.1) as the Monge–Kantorovich constraint: the interplay of these interpretations animates much of the following. The exposition is based in part upon the interesting papers of Prigozhin [P1], [P2], [P3], and also upon [E-F-G], [A-E-W], [E-R].

11.1 Growing sandpiles

As a simple preliminary model, suppose that the function $f \geq 0$ is a source term, representing the rate that sand is added to a sandpile, whose initial height is zero. We then have $u_t = f$ in any region where the constraint (11.1) is active. But if adding more sand at some location would break the constraint, we may imagine the newly added sand particles "to roll downhill instantly", coming to rest at new sites where their addition maintains the constraint.

We propose as a model for the resulting evolution:

$$\begin{cases} f - u_t \in \partial I_\infty[u] & (t > 0) \\ u = 0 & (t = 0), \end{cases} \tag{11.2}$$

the functional $I_\infty[\cdot]$ defined in §9. The interpretation is that at each moment of time, the mass $d\mu^+ = f^+(\cdot, t)dx$ is instantly and optimally transported downhill by the potential $u(\cdot, t)$ into the mass $d\mu^- = u_t(\cdot, t)dy$. In other words *the height function $u(\cdot, t)$ of the sandpile is deemed also to be the potential generating the Monge–Kantorovich reallocation of f^+dx to u_tdy*. This requirement forces the dynamics (11.2).

Example: Interacting sandcones. Consider, for example, the case that mass is added only at fixed sites:

$$f = \sum_{k=1}^{m} f_k(t)\delta_{d_k}, \tag{11.3}$$

where $f_k > 0$. In this case we expect the height function u to be the union of interacting *sandcones*:

$$u(x, t) = \max\{0, z_1(t) - |x - d_1|, \ldots, z_m(t) - |x - d_m|\}. \tag{11.4}$$

Owing to conservation of mass, we expect the cone heights $\mathbf{z}(t) = (z_1(t), \dots, z_m(t))$ to solve the coupled system of ODE

$$\begin{cases} \dot{z}_k(t) = \dfrac{f_k(t)}{|D_k(t)|} \quad (t \geq 0) \\ z_k(0) = 0 \end{cases} \tag{11.5}$$

for $k = 1, \dots, m$, where $|D_k(t)|$ denotes the measure of the set $D_k(t) \subset \mathbb{R}^n$ on which the k-th cone determines u:

$$D_k(t) = \{x \in \mathbb{R}^n \mid z_k(t) - |x - d_k| > 0,\ z_l(t) - |x - d_l| \text{ for } l \neq k\}. \tag{11.6}$$

These ODE were originally derived by Aronsson [AR1].

Let us confirm that (11.4)–(11.6) give the solution of (11.2), in the case of point sources (11.3). To check (11.2) we must show at each time $t \geq 0$ that

$$I_\infty[u] + (f - u_t, v - u)_{L^2} \leq I_\infty[v] \tag{11.7}$$

for all $v \in L^2(\mathbb{R}^n)$. Now $I_\infty[u(\cdot, t)] = 0$, and if $I_\infty[v] = +\infty$, then (11.7) is obvious. So we may as well assume $I_\infty[v] = 0$; that is,

$$|Dv| \leq 1 \text{ a.e.} \tag{11.8}$$

The problem now is to show that

$$(f - u_t(\cdot, t), v - u(\cdot, t))_{L^2} \leq 0. \tag{11.9}$$

Owing to (11.3) and (11.4), the term on the left means

$$\sum_{k=1}^{m} f_k(t)(v(d_k) - u(d_k, t)) - \sum_{k=1}^{\infty} \int_{D_k(t)} \dot{z}_k(t)(v(x) - u(x, t))\, dx.$$

Consequently the ODE (11.5) means that we must show

$$\sum_{k=1}^{m} f_k(t) \fint_{D_k(t)} v(d_k) - v(x)\, dx \leq \sum_{k=1}^{m} f_k(t) \fint_{D_k(t)} u(d_k, t) - u(x, t)\, dx, \tag{11.10}$$

the slash through the integral signs denoting average. But (11.10) is easy: in light of (11.8) and (11.4), we have

$$v(d_k) - v(x) \leq |d_k - x| = u(d_k, t) - u(x, t)$$

on $D_k(t)$. □

11.2 Collapsing sandpiles

The dynamics (11.2) model "surface flows" for sandpiles: once a sand grain is added, rolls downhill and comes to rest, it never again moves. It is therefore interesting to modify this model to allow for "avalanches".

In view of the approximation in §9 of the term $\partial I_\infty[u]$ by the p-Laplacian operator $-\text{div}(|Du|^{p-2}Du)$, we propose now to investigate the limiting behavior as $p \to \infty$ of the quasilinear parabolic problem

$$\begin{cases} u_{p,t} - \text{div}(|Du_p|^{p-2}Du_p) = 0 & \text{in } \mathbb{R}^n \times (0, \infty) \\ u_p = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{11.11}$$

Here $g : \mathbb{R}^n \to \mathbb{R}$ is the given initial height of a sandpile, satisfying the *instability condition*

$$L := \sup_{\mathbb{R}^n} |Dg| > 1. \tag{11.12}$$

For large p, the PDE in (11.11) supports very fast diffusion in regions where $|Du_p| > 1$ and very slow diffusion in regions where $|Du_p| < 1$. We expect therefore that for $p >> 1$ the solutions u_p of (11.11) will rapidly rearrange its mass to achieve the stability condition. As the initial profile is unstable, we expect there to be a short time period during which there is rapid mass flow, followed by a time in which u_p is practically unchanging.

Indeed, simple estimates suffice to show that there exists a function $u = u(x)$ with

$$|Du| \leq 1 \qquad \text{in } \mathbb{R}^n \tag{11.13}$$

such that

$$u_{p_k} \to u \qquad \text{uniformly on compact subsets of } \mathbb{R}^n \times (0, \infty) \tag{11.14}$$

for some sequence $p_k \to \infty$. We call u the *collapsed profile* and our problem is to understand how the initial, unstable profile g rearranges itself into u.

To understand the mapping $g \mapsto u$, we rescale the PDE (11.11) to stretch out the initial time layer of rapid mass motion. For this, set

$$v_p(x, t) := t u_p \left(x, \frac{t^{p-1}}{p-1} \right) \qquad (x \in \mathbb{R}^n, \ t > 0) \tag{11.15}$$

and write $\tau := L^{-1}$. It turns out then that

$$v_{p_k} \to v \qquad \text{uniformly on } \mathbb{R}^n \times [\tau, 1],$$

where

$$\begin{cases} \dfrac{v}{t} - v_t \in \partial I_\infty[v] & (\tau \leq t \leq 1) \\ v = h & (t = \tau) \end{cases} \tag{11.16}$$

for $h := \frac{1}{L}g$. Furthermore the collapsed profile is

$$u = v(\cdot, 1). \tag{11.17}$$

In summary, *our procedure for calculating the collapse $g \mapsto u$ is to define h as above, solve the evolution (11.16) and then set $u := v(\cdot, 1)$.*

Now (11.16) is interpreted, as above, as an evolution in which the mass $d\mu^+ = \frac{v}{t}dx$ is instantly and optimally rearranged by the potential v into $d\mu^- = v_t dy$. Again we have a Monge–Kantorovich mass reallocation occurring on a fast time scale, which thereby generates the dynamics (11.16).

Example: Collapse of a convex cone. As an application let us take Γ_τ to be the boundary of an open convex region $U_\tau \subset \mathbb{R}^2$. Assume

$$g(x) := \begin{cases} L \operatorname{dist}(x, \Gamma_\tau) & x \in U_\tau \\ 0 & \text{otherwise} \end{cases}$$

is the height of the initial, unstable sandpile. How does it collapse into equilibrium? Following the procedure above, we look to the evolution

$$\begin{cases} \dfrac{v}{t} - v_t \in \partial I_\infty[v] & (\tau \le t \le 1) \\ v = h & (t = \tau), \end{cases} \tag{11.18}$$

for

$$h(x) := \begin{cases} \operatorname{dist}(x, \Gamma_\tau) & x \in U_\tau \\ 0 & \text{otherwise.} \end{cases}$$

We next guess that our solution v has the form

$$v(x,t) = \begin{cases} \operatorname{dist}(x, \Gamma_t) & x \in U_t \\ 0 & \text{otherwise,} \end{cases}$$

where Γ_t is the boundary of an open set U_t. In other words, we conjecture that v is the distance function to a moving family of surfaces $\{\Gamma_t\}_{\tau \le t \le 1}$.

We next employ techniques from Monge–Kantorovich theory to derive a geometric law of motion for the surfaces $\{\Gamma_t\}_{\tau \le t \le 1}$, which describe the moving edge of the collapsing sandpile. We hereafter write for each point $y \in \Gamma_t$,

$$\begin{cases} \gamma = \gamma(y,t) = \text{radius of the largest disk within} \\ U_t \text{ which touches } \Gamma_t \text{ at } y. \end{cases}$$

Take a point $x \in U_t$ with a unique closest point $y \in \Gamma_t$. Let $\kappa :=$ curvature of Γ_t at y and $R := \frac{1}{\kappa}$.

We may assume the segment $[x, y]$ is vertical and $y = (0, R)$. Set

$$A_\varepsilon := \{z \mid \theta(z, e_2) < \varepsilon,\ R - \gamma \le |z| \le R\},$$

where $\theta(z, e_2)$ denotes the angle between z and $e_2 = (0, 1)$. Now we interpret (11.18) as saying that the mass $d\mu^+ = \frac{v}{t} dx$ is transferred in the direction $-Dv$, to $d\mu^- = v_t dy$. This transfer, restricted to the set A_ε, forces the (approximate) *detailed mass balance* relation

$$\int_{A_\varepsilon} \frac{v}{t}\, dx \approx \int_{A_\varepsilon} v_t\, dy, \tag{11.19}$$

according to Lemma 9.3. Now divide both sides by $|A_\varepsilon|$ and send $\varepsilon \to 0$. A calculation shows

$$\lim_{\varepsilon \to 0} ⨍_{A_\varepsilon} \frac{v}{t}\, dx = \frac{\gamma}{3t} \left(\frac{3 - 2\kappa\gamma}{2 - \kappa\gamma} \right)$$

and, since $v_t = V (=$ outward normal velocity of Γ_t at $y)$ along the ray through y,

$$\lim_{\varepsilon \to 0} ⨍_{A_\varepsilon} v_t\, dy = V.$$

We derive therefore the *nonlocal geometric law of motion*

$$V = \frac{\gamma}{3t} \left(\frac{3 - 2\kappa\gamma}{2 - \kappa\gamma} \right) \quad \text{on } \Gamma_t \tag{11.20}$$

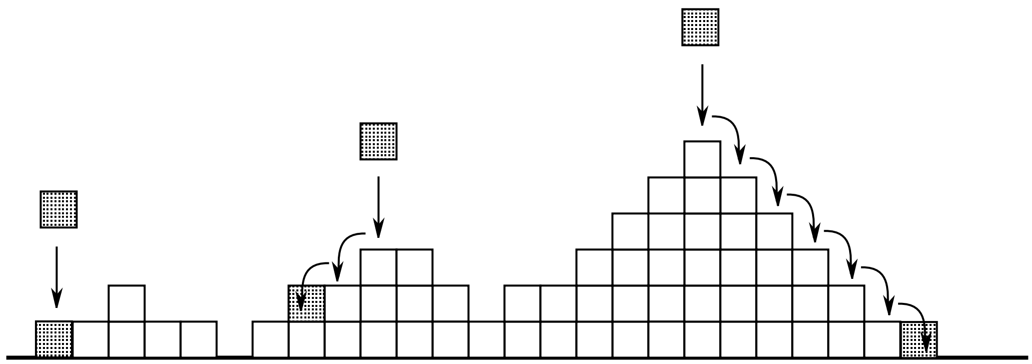
for the moving surfaces $\{\Gamma_t\}_{\tau \le t \le 1}$. (See Feldman [F] for a rigorous analysis of (11.20).)

11.3 A stochastic model

A variant of our growing sandpile evolution, discussed in §11.1, arises a rescaled continuum limit of the following stochastic model. Consider the lattice $\mathbb{Z}^2 \subseteq \mathbb{R}^2$ as subdividing the plane into unit squares. We introduce a discrete model for a "sandpile", as a stack of unit cubes resting on the plane, each column of cubes above a unit square. At each moment the configuration must be *stable*, which means that the heights of any two adjacent columns of cubes can differ by at most one. (A given column has four adjacent columns, in the coordinate directions.)

We image additional cubes being added randomly to the top of columns. If the new pile is stable, the new cube remains in place. Otherwise it "falls downhill", until it reaches a stable position.

What happens in the scaled continuum limit, when we take more and more, smaller and smaller cubes added faster and faster?



We make precise our model, generalizing to $\mathbb{R}^n$. Let us write $\mathbf{i} = (i_1, \dots, i_n)$ for a typical site in $\mathbb{Z}^n$ and say two sites $\mathbf{i}, \mathbf{j}$ are *adjacent*, written $\mathbf{i} \sim \mathbf{j}$, if

$$\max_{1 \le k \le n} |i_k - j_k| = 1.$$

A (*stable*) *configuration* is a mapping $\eta : \mathbb{Z}^n \to \mathbb{Z}$ such that

$$\begin{cases} |\eta(\mathbf{i}) - \eta(\mathbf{j})| \le 1 \text{ if } \mathbf{i} \sim \mathbf{j} \\ \text{and } \eta \text{ has bounded support.} \end{cases}$$

The *state space* S is the collection of all configurations.

We introduce as follows a Markov process on S. Given $\eta \in S$ and $\mathbf{i} \in \mathbb{Z}^n$, write

$$\begin{aligned} \Gamma(\mathbf{i}, \eta) \ := \ \{\mathbf{j} \in \mathbb{Z}^n \mid & \text{ there exist sites } \mathbf{i} = \mathbf{i}_1 \sim \mathbf{i}_2 \sim \cdots \sim \mathbf{i}_m = \mathbf{j} \\ & \text{with } \eta(\mathbf{i}_{l+1}) = \eta(\mathbf{i}_l) - 1 \ (l = 1, \dots, m-1), \\ & \eta(\mathbf{k}) \ne \eta(\mathbf{j}) - 1 \text{ for all } \mathbf{k} \sim \mathbf{j}\}. \end{aligned} \tag{11.21}$$

Thus $\Gamma(\mathbf{i}, \eta)$ is the set of sites $\mathbf{j}$ at which a cube newly added to $\mathbf{i}$ can come to rest. Assign to each $\mathbf{j} \in \Gamma(\mathbf{i}, \eta)$ a number

$$0 \le p(\mathbf{i}, \mathbf{j}, \eta) \le 1$$

such that

$$\sum_{\mathbf{j} \in \Gamma(\mathbf{i}, \eta)} p(\mathbf{i}, \mathbf{j}, \eta) = 1, \tag{11.22}$$

and think of $p(\mathbf{i}, \mathbf{j}, \eta)$ as the probability that a cube added to $\mathbf{i}$ will fall downhill and come to rest at $\mathbf{j}$. Finally set

$$c(\mathbf{j}, \eta, t) := \sum_{\mathbf{i} : \mathbf{j} \in \Gamma(\mathbf{i}, \eta)} p(\mathbf{i}, \mathbf{j}, \eta) f\left(\frac{\mathbf{i}}{\mathbf{N}}, \frac{t}{N}\right) \tag{11.23}$$

where $f : \mathbb{R}^n \times [0, \infty) \to [0, \infty)$ is a given function, the *source density*. Then $c(\mathbf{j}, \eta, t)$ is the rate at which cubes are coming to rest at $\mathbf{j}$.

We introduce the *infinitesimal generator* $\mathcal{L}_t$ of our Markov process by taking any $F : S \to \mathbb{R}$ and defining then

$$(\mathcal{L}_t F)(\eta) := \sum_{\mathbf{j} \in \mathbb{Z}^n} c(\mathbf{j}, \eta, t)(F(\eta^{\mathbf{j}}) - F(\eta)), \tag{11.24}$$

where

$$\eta^{\mathbf{j}}(\mathbf{i}) := \begin{cases} \eta(\mathbf{i}) + 1 & \text{if } \mathbf{i} = \mathbf{j} \\ \eta(\mathbf{i}) & \text{if } \mathbf{i} \neq \mathbf{j}. \end{cases}$$

The formula (11.24) encodes the foregoing probabilistic interpretation.

Let $\{\eta(\cdot, t)\}_{t \geq 0}$ denote the inhomogeneous Markov process on S generated by $\{\mathcal{L}_t\}_{t \geq 0}$, with $\eta(\cdot, 0) \equiv 0$. Thus $\eta(\mathbf{i}, t)$ is the (random) height of the pile of cubes at site $\mathbf{i}$, time $t \geq 0$.

We intend now to rescale, taking now cubes of side $\frac{1}{N}$ added on a time scale multiplied by N. We construct therefore the rescaled process $\frac{1}{N}\eta([xN], tN)$, where $x \in \mathbb{R}^n$, $t \geq 0$, and inquire what happens as $N \to \infty$.

Theorem 11.1 *For each $t \geq 0$, we have*

$$E\left(\sup_{x \in \mathbb{R}^n} \left| \frac{1}{N}\eta([xN], tN) - u(x, t) \right| \right) \to 0 \tag{11.25}$$

as $N \to \infty$, where u is the unique solution of the evolution

$$\begin{cases} f - u_t \in \partial \hat{I}[u] & (t = 0) \\ u = 0 & (t = 0). \end{cases} \tag{11.26}$$

Here

$$\hat{I}[v] := \begin{cases} 0 & \text{if } v \in \hat{\mathbb{K}} \\ +\infty & \text{otherwise} \end{cases}$$

for

$$\hat{\mathbb{K}} := \{v \in L^2(\mathbb{R}^n) \mid v \text{ is Lipschitz, } |v_{x_i}| \leq 1 \text{ a.e. } (i = 1, \dots, n)\}.$$

Consequently, the continuum dynamics (11.26), while similar to (11.2), differ by "remembering" the anisotropic structure of the lattice.

The proof in [E-R] is too complicated to reproduce here: the key new idea is a combinatorial lemma asserting that if η, ξ are any two configurations in S, then

$$\sum_{\mathbf{j}} c(\mathbf{j}, \eta, t)(\eta(\mathbf{j}) - \xi(\mathbf{j})) \leq \sum_{\mathbf{i}} f\left(\frac{\mathbf{i}}{\mathbf{N}}, \frac{t}{N}\right)(\eta(\mathbf{i}) - \xi(\mathbf{i})). \tag{11.27}$$

Since the term c, describing the rate at which cubes come to rest at a given site, is a sort of rescaled, discrete analogue of u_t, (11.27) is a "microscopic" analogue of the "macroscopic" inequality

$$\int_{\mathbb{R}^n} (f - u_t)(v - u)dx \le 0 \quad (v \in \hat{\mathbb{K}}).$$

But this is just, as we have seen before, another way of writing the evolution (11.26).

A further interpretation of (11.26) is that the mass $d\mu^+ = fdx$ is instantly rearranged into $d\mu^- = u_t dy$, so as to minimize the cost

$$\int_{\mathbb{R}^n} c(x, \mathbf{s}(x)) f^+(x)\, dx$$

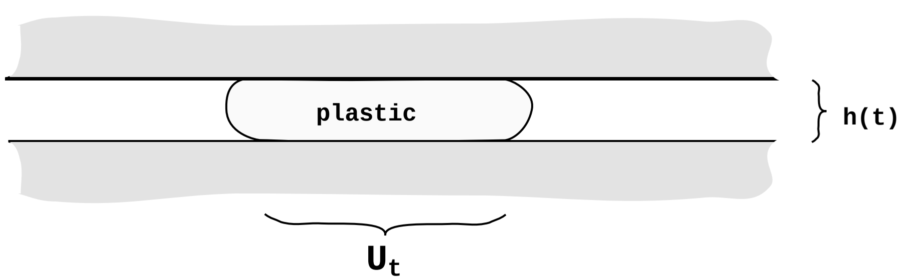
for the l^1-distance

$$c(x,y) = |x_1 - y_1| + \cdots + |x_n - y_n|.$$

In the case of point sources, $f = \sum f_k(t)\delta_{d_k}$, the dynamics (11.26) correspond to growing, interacting *pyramids* of sand.

12 Application: Compression molding

The utterly different physical situation of *compression molding* gives rise to similar mathematics. In this setting we consider an incompressible plastic material being squeezed between two horizontal plates.



Assume the lower plate is fixed, and the plate separation at time t is $h(t)$, with $\dot{h}(t) \le 0$, $h(T) = 0$. Let U_t denote the (approximate) projection onto $\mathbb{R}^2$ of the region occupied by the plastic at time t and write $\Gamma_t := \partial U_t$. As $t \to T$, the region U_t expands to fill the entire plane.

Following [AR2], [A-E], we introduce a highly simplified model of the physics, with the goal of tracking the air-plastic interface Γ_t for times $0 \le t < T$. After a number of simplifying assumptions for this highly viscous flow, the relevant PDE becomes

$$\operatorname{div}(|Dp|^{\frac{1}{\sigma}-1} Dp) = \frac{\dot{h}}{h} \qquad \text{in } U_t \tag{12.1}$$

where $0 \le t < T$, $p =$ pressure, and σ is a small constant arising in a power-law constitutive relation for the highly non-Newtonian flow.

We can rescale in time to convert to the case $T = \infty$, $\dot{h}/h \equiv -1$, in which case (12.1) reads

$$-\operatorname{div}(|Dp|^{\frac{1}{\sigma}-1} Dp) = 1 \qquad \text{in } U_t \tag{12.2}$$

for $t \ge 0$. In addition, we have the boundary conditions

$$\begin{cases} p = 0 & \text{on } \Gamma_t \\ V = |Dp|^{1/\sigma} & \text{on } \Gamma_t, \end{cases} \tag{12.3}$$

where $V =$ outward normal velocity of Γ_t.

Now experimental results in the engineering literature suggest that for real materials the evolution of Γ_t does not much depend upon the exact choice of σ, so long as σ is small: see, for instance, [F-F-T]. We therefore propose to study the asymptotic limit $\sigma \to 0$. For this, we first change notation, and rewrite (12.2), (12.3) as

$$\begin{cases} -\operatorname{div}(|Du_p|^{p-2} Du_p) = 1 & \text{in } U_t \\ u_p = 0, \; V = |Du_p|^{p-2} & \text{on } \Gamma_t. \end{cases} \tag{12.4}$$

When $p \to \infty$, we expect in light of the calculations in §9 that $u_p \to u$, where

$$\begin{cases} -\operatorname{div}(aDu) = 1 & \text{in } U_t \\ u = 0, \; V = a & \text{on } \Gamma_t \end{cases} \tag{12.5}$$

for some nonnegative function a. The physical interpretations are now

$$\begin{cases} u = \text{pressure}, \; \mathbf{v} = -aDu = \text{velocity}, \\ a = \text{speed}. \end{cases}$$

We can further reinterpret (12.5) as the evolution

$$\begin{cases} w - w_t \in \partial I_\infty[u] \\ u \in \beta(w), \end{cases} \qquad (t > 0) \tag{12.6}$$

where β is the multivalued mapping

$$\beta(x) = \begin{cases} [0,\infty) & \text{if } \; x \geq 1 \\ 0 & \text{if } \; 0 \leq x \leq 1 \\ (-\infty, 0] & \text{if } \; x \leq 0. \end{cases}$$

We interpret $w = \chi_{U_t}$, and so the initial condition for (12.6) is

$$w = \chi_{U_0} \quad (t = 0). \tag{12.7}$$

The Monge–Kantorovich interpretation is that at each moment the measure $d\mu^+ = \chi_{U_t} dx$ is being instantly and optimally rearranged to $d\mu^- = V$ times (n-1)-dimensional Hausdorff measure $\mathcal{H}^{n-1}$ restricted to Γ_t. These dynamics in turn determine the velocity V of Γ_t.

It is not hard to check that

$$u(x,t) = \begin{cases} \operatorname{dist}(x, \Gamma_t) & x \in U_t \\ 0 & \text{otherwise;} \end{cases}$$

so that the pressure in this asymptotic model is just the distance to the boundary.

Using detailed mass balance arguments somewhat like those in §11 we can further show that

$$V = \gamma\left(1 - \frac{\kappa\gamma}{2}\right) \quad \text{ on } \Gamma_t \tag{12.8}$$

where γ = radius of largest disk within U_t touching Γ_t at y and κ = curvature. Feldman [F] has rigorously analyzed this geometric flow that characterizes the spread of the plastic.

Appendix

13 Finite-dimensional linear programming

To motivate some functional analytic considerations in §1, we record here some facts about linear programming. Good references are Bertsimas–Tsitsiklis [B-T], Ekeland–Turnbull [E-T] or Papadimitriou–Steiglitz [P-S].

Notation. If $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$, we write

$$x \geq 0$$

to mean $x_k \geq 0$ $(k = 1, \ldots, N)$. □

Assume we are given $c \in \mathbb{R}^N$, $b \in \mathbb{R}^M$, and an $M \times N$ matrix A.

The *primal linear programming problem* is to find $x^* \in \mathbb{R}^N$ so as to

$$
\text{(P)} \qquad \begin{cases} \text{minimize } c \cdot x, \text{ subject to the} \\ \text{constraints } Ax = b,\ x \geq 0. \end{cases}
$$

The *dual problem* is then to find $y^* \in \mathbb{R}^M$ so as to

$$
\text{(D)} \qquad \begin{cases} \text{maximize } b \cdot y, \text{ subject to the} \\ \text{constraints } A^T y \leq c. \end{cases}
$$

Assume $x^* \in \mathbb{R}^N$ solves (P), $y^* \in \mathbb{R}^M$ solves (D). We may then regard y^* as the Lagrange multiplier for the constraints in (P), and, likewise, x^* as the Lagrange multiplier for (D). Furthermore we have the saddle point relation:

$$
c \cdot x^* = b \cdot y^*; \tag{13.1}
$$

that is,

$$
\min\{c \cdot x \mid Ax = b,\ x \geq 0\} = \max\{b \cdot y \mid A^T y \leq c\}. \tag{13.2}
$$

An example. Suppose we are given nonnegative numbers c_{ij}, μ_i^+, μ_j^- $(i = 1, \dots, n;\ j = 1, \dots, m)$, with

$$
\sum_{i=1}^{n} \mu_i^+ = \sum_{j=1}^{m} \mu_j^-,
$$

and we are asked to find μ_{ij}^* $(i = 1, \dots, n;\ j = 1, \dots, m)$, so as to

$$
\text{minimize } \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} \mu_{ij}, \tag{13.3}
$$

subject to the constraints

$$
\sum_{j=1}^{m} \mu_{ij} = \mu_i^+,\ \sum_{i=1}^{n} \mu_{ij} = \mu_j^-,\ \mu_{ij} \geq 0\ (i = 1, \dots, n;\ j = 1, \dots, m). \tag{13.4}
$$

Papadimitriou and Steiglitz call this the *Hitchcock problem*. It has the form (P), with

$$
\begin{cases} N = nm,\ M = n + m,\ x = (\mu_{11}, \mu_{12}, \dots, \mu_{1m}, \mu_{21}, \dots, \mu_{nm}), \\ c = (c_{11}, \dots, c_{1m}, c_{21}, \dots, c_{nm}),\ b = (\mu_1^+, \dots, \mu_n^+, \mu_1^-, \dots, \mu_m^-), \end{cases}
$$

and A the $(n+m) \times nm$ matrix

$$\begin{array}{l} n \text{ rows} \left\{ \begin{array}{l} \\ \\ \\ \end{array} \right. \\ \\ m \text{ rows} \left\{ \begin{array}{l} \\ \\ \\ \end{array} \right. \end{array} \begin{pmatrix} \mathbb{1} & 0 & \dots & 0 \\ 0 & \mathbb{1} & \dots & 0 \\ 0 & 0 & \dots & \mathbb{1} \\ \dots & \dots & \dots & \dots \\ e_1 & e_1 & \dots & e_1 \\ e_2 & e_2 & \dots & e_2 \\ e_m & e_m & \dots & e_m \end{pmatrix}$$

Above the middle horizontal dotted line, each entry is a row vector in $\mathbb{R}^m$ and $\mathbb{1} = (1, \dots, 1)$, $0 = (0, 0, \dots, 0)$. Below the dotted line each entry is a row vector in $\mathbb{R}^n$ and $e_i = (0, 0, \dots, 1, \dots, 0)$, the 1 in the i-th slot $(i = 1, \dots, n)$.

Now write $y = (u_1, \dots, u_n, v_1, \dots, v_m) \in \mathbb{R}^{n+m}$. Employing the explicit form of the matrix A, we translate the dual problem (D) for the case at hand to read

$$\text{maximize} \quad \sum_{i=1}^{n} u_i \mu_i^+ + \sum_{j=1}^{m} v_j \mu_j^-, \tag{13.5}$$

subject to the constraints

$$u_i + v_j \le c_{ij} \quad (i = 1, \dots, n;\ j = 1, \dots, m). \tag{13.6}$$

The Monge–Kantorovich mass transfer problem is a continuum version of (13.3), (13.4) and its dual problem a continuum version of (13.5), (13.6): see §1. □

References

[A] P. Appell, Le probleme geometrique des déblais et remblais, Memor. des Sciences Math., Acad. de Sciences de Paris, Gauthier Villars **27** (1928), 1–34.

[Am] L. Ambrosio,*Lecture Notes on Optimal Transport Problems*, to appear.

[AR1] G. Aronsson, A mathematical model in sand mechanics, SIAM J. Applied Math, **22** (1972), 437–458.

[AR2] G. Aronsson, Asymptotic solutions of a compression molding problem, preprint, Department of Mathematics, Linkoping University.

[A-E] G. Aronsson and L. C. Evans, An asymptotic model for compression molding, to appear.

[A-E-W] G. Aronsson, L. C. Evans and Y. Wu, Fast/slow diffusion and growing sandpiles, J. Differential Equations **131** (1996), 304–335.

[Ba] F. Barthe, Inégalités fonctionnelles et géométriques obtenues par transport des mesures, Thesis, Université de Marne–la–Vallée, 1997.

[Be1] J–D. Benamou, Transformations conservant la mesure, mécanique des fluides incompressibles et modèle semi-géostrophique en météorologie, Thesis, Université de Paris IX, 1992.

[Be2] J–D. Benamou, A domain decomposition method for the polar factorization of vector–valued mappings, SIAM J Numerical Analysis **32** (1995), 1808–1838.

[Be-B1] J–D. Benamou and Y. Brenier, Weak existence for the semi–geostrophic equations formulated as a coupled Monge–Ampere/transport problem, preprint.

[Be-B2] J–D. Benamou and Y. Brenier, The optimal time–continuous mass transport problem and its augmented Lagrangian numerical resolution, preprint.

[B-T] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific (1997)

[B-B-M] T. Bhattacharya, E. DiBenedetto, and J. Manfredi, Limits at $p \to \infty$ of $\Delta_p u = f$ and related extremal problems, Rend. Sem. Mat. Univ. Pol. Torino, Fascicolo Speciale (1989).

[B-B-S] G. Bouchitte, G. Buttazzo and P. Seppechere, Shape optimization solutions via Monge–Kantorovich equation, CRAS **324** (1997), 1185–1191.

[B1] Y. Brenier, Décomposition polaire et réarrangement monotone des champs de vecteurs, CRAS **305** (1987), 805–808.

[B2] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, Comm. Pure Appl. Math. **44** (1991), 375–417.

[B3] Y. Brenier, The dual least action problem for an ideal, compressible fluid, Arch. Rat. Mech. Analysis **122** (1993), 323–351.

[B4] Y. Brenier, A geometric presentation of the semi–geostrophic equations, preprint.

[C1] L. Caffarelli, A localization property of viscosity solutions of the Monge–Ampere equaqtion, Annals of Math. **131** (1990), 129–134.

[C2] L. Caffarelli, Interior $W^{2,p}$ estimates for solutions of the Monge–Ampere equation, Annals of Math. **131** (1990), 135–150.

[C3] L. Caffarelli, Some regularity properties of solutions to the Monge–Ampere equation, Comm. in Pure Appl. Math. **44** (1991), 965–969.

[C4] L. Caffarelli, Allocation maps with general cost functions, in *Partial Differential Equations with Applications* (ed. by Talenti), 1996, Dekker.

[C5] L. Caffarelli, The regularity of mappings with a convex potential, J. Amer. Math. Soc. **5** (1992), 99–104.

[C6] L. Caffarelli, Boundary regularity of maps with convex potentials, Comm. Pure Appl. Math. **45** (1992), 1141–1151.

[C7] L. Caffarelli, Boundary regularity of maps with convex potentials II, Annals of Math. **144** (1996), 453–496.

[C-F-M] L. Caffarelli, M. Feldman and R. McCann, Constructing optimal maps for Monge's transport problem as the limit of strictly convex costs, to appear

[C-P] A. Cellina and S. Perrotta, On the validity of the Euler–Lagrange equations, preprint.

[C-S] S. Chynoweth and M. J. Sewell, Dual variables in semigeostrophic theory, Proc. Royal Soc. London A **424** (1989), 155–186.

[C] M. J. P. Cullen, Solutions to a model of a front forced by deformation, Quart. J. Royal Meteor. Soc. **109** (1983), 565–573.

[C-N-P] M. J. P. Cullen, J. Norbury, and R. J. Purser, Generalized Lagrangian solutions for atmospheric and oceanic flows, SIAM J. Appl. Math. **51** (1991), 20–31.

[C-P1] M. J. P. Cullen and R. J. Purser, An extended Lagrangian theory of semigeostrophic frontogenesis, J. of the Atmospheric Sciences **41** (1984), 1477–1497.

[C-P2] M. J. P. Cullen and R. J. Purser, A duality principle in semigeostgrophic theory, J. of the Atmospheric Sciences **44** (1987), 3449–3468.

[C-P3] M. J. P. Cullen and R. J. Purser, Properties of the Lagrangian semigeostrophic equations, J. of the Atmospheric Sciences **46** (1989), 2684–2697.

[D-M] B. Dacorogna and J. Moser, On a partial differential equation involving the Jacobian determinant, Ann. Inst. H. Poincare **7** (1990), 1–26.

[D] C. Dupin, *Applications de Geometrie et de Mechanique*, Bachelier, Paris, 1822.

[E-T] I. Ekeland and T. Turnbull, *Infinite-dimensional optimization and convexity*, Chicago Lectures in Mathematics, Univ. of Chicago Press, 1983.

[E-G1] L. C. Evans and W. Gangbo, Differential equations methods in the Monge–Kantorovich mass transfer problem, Memoirs American Math. Society, to appear.

[E-G2] L. C. Evans and R. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, 1992. (An errata sheet for the first printing of this book is available through the math.berkeley.edu website.)

[E-F-G] L. C. Evans, M. Feldman and R. Gariepy, Fast/slow diffusion and collapsing sandpiles, J. Differential Equations **137** (1997), 166–209.

[E-R] L. C. Evans and F. Rezakhanlou, A stochastic model for sandpiles and its continuum limit, to appear.

[F] M. Feldman, Variational evolution problems and nonlocal geometric motion, forthcoming.

[F-F-T] F. Folgar, C.-C. Lee and C. L. Tucker, Simulation of compression molding for fiber-reinforced thermosetting polymers, Trans. ASME J. of Eng. for Industry **106** (1984), 114–125.

[G] W. Gangbo, An elementary proof of the polar factorization of vector-valued functions, Arch. Rat. Math. Analysis **128** (1994), 381–399.

[G-M1] W. Gangbo and R. McCann, Optimal maps in Monge's mass transport problem, C.R. Acad. Sci. Paris **321** (1995), 1653–1658.

[G-M2] W. Gangbo and R. McCann, The geometry of optimal transport, Acta Mathematica **177**, (1996), 113–161.

[I] M. Iri, Theory of flows in continua as approximation to flows in networks, in *Survey of Math Programming* (ed. by Prekopa), North-Holland, 1979.

[J] U. Janfalk, On certain problems concerning the p-Laplace operator, Linköping Studies in Science and Technology, Dissertation #326, Linköping University, Sweden, 1993.

[J-K-O1] R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker–Planck equation, preprint.

[J-K-O2] R. Jordan, D. Kinderlehrer, and F. Otto, The route to stability through Fokker–Planck dynamics, Proc. First US-China Conference on Differential Equations and Applications.

[Je] R. Jensen, Uniqueness of Lipschitz extensions minimizing the sup-norm of the gradient, Arch. Rat. Mech. Analysis **123** (1993), 51–74.

[K1] L. V. Kantorovich, On the transfer of masses, Dokl. Akad. Nauk. SSSR **37** (1942), 227–229 (Russian).

[K2] L. V. Kantorovich, On a problem of Monge, Uspekhi Mat. Nauk. **3** (1948), 225–226.

[MC1] R. McCann, A convexity theory of interacting gases, Advances in Mathematics **128** (1997), 153–179.

[MC2] R. McCann, Existence and uniqueness of monotone measure-preserving maps, Duke Math. J. **80** (1995), 309–323.

[MC3] R. McCann, Equilibrium shapes for planar crystals in an external field, preprint.

[MC4] R. McCann, Exact solutions to the transportation problem on the line, preprint.

[M] G. Monge, *Memoire sur la Theorie des Déblais et des Remblais*, Histoire de l'Acad. des Sciences de Paris, 1781.

[O1] F. Otto, Dynamics of labyrinthine pattern formation in magnetic fields, preprint.

[O2] F. Otto, letter to LCE.

[P-S] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization*, Dover Press (2001).

[P1] L. Prigozhin, A variational problem of bulk solid mechanics and free-surface segratation, Chemical Eng. Sci. **78** (1993), 3647–3656.

[P2] L. Prigozhin, Sandpiles and river networks: extended systems with nonlocal interactions, Phys. Rev E **49** (1994), 1161–1167.

[P3] L. Prigozhin, Variational model of sandpile growth, European J. Appl. Math. **7** (1996), 225–235.

[R] S. T. Rachev, The Monge–Kantorovich mass transference problem and its stochastic applications, Theory of Prob. and Appl. **29** (1984), 647–676.

[Rk] T. Rockafeller, Characterization of the subdifferentials of convex functions, Pacific Journal of Math. **17** (1966), 497–510.

[Ro] V. A. Rohlin, On the fundamental ideas of measure theory, American Math. Society, Translations in Math. **71** (1952).

[S1] G. Strang, Maximal flow through a domain, Math. Programming **26** (1983), 123–143.

[S2] G. Strang, L^1 and L^∞ approximations of vector fields in the plane, in *Lecture Notes in Num. Appl. Analysis* **5** (1982), 273–288.

[St] D. Struik, *Lectures on Classical Differential Geometry* (2nd ed.), Dover, 1988.

[Su] V. N. Sudakov, Geometric problems in the theory of infinite-dimensional probability distributions, Proceedings of Steklov Institute **141** (1979), 1–178.

[T-W] N. Trudinger and X. J. Wang, On the Monge mass transfer problem, to appear in Calculus of Variations and PDE.

[U1] J. Urbas, Lecture notes on mass transfer problems, U of Bonn.

[U2] J. Urbas, On the second boundary value problem for equations of Monge–Ampere type, J. Reine Angew. Math. **487** (1997).

[W] J. Wolfson, Minimal Lagrangian diffeomorphisms and the Monge–Ampere equation, J. Differential Geom. **46** (1997), 335–373.