A Benamou-Brenier approach to branched transport

Lorenzo Brasco

Dipartimento di Matematica e Applicazioni “R. Caccioppoli”
Università di Napoli “Federico II”

May 14, 2009
Some results of this talk are contained in

Outline

1. Branched transport: introduction and models
2. An Eulerian point of view on branched transport
3. The variational setting
4. Equivalences with other models
Some notations

- $\Omega \subset \mathbb{R}^N$ compact and convex
- $\mathcal{P}(\Omega) =$ Borel probability measures over $\Omega$
- $\mathcal{M}(\Omega; \mathbb{R}^N) =$ $\mathbb{R}^N$-valued Radon measures over $\Omega$
- $w_p = p-$Wasserstein distance
  
  $$w_p(\rho_0, \rho_1) = \min \left\{ \left( \int_{\Omega \times \Omega} |x - y|^p \, d\gamma(x, y) \right)^{1/p} : \gamma \in \Pi(\rho_0, \rho_1) \right\}$$

- $\mathcal{W}_p(\Omega) =$ $p-$Wasserstein space over $\Omega$, i.e. $\mathcal{P}(\Omega)$ equipped with $w_p$

- $|\mu'_t|_{w_p} = \lim_{h \to 0} \frac{w_p(\mu_{t+h}, \mu_t)}{|h|}$ metric derivative

- $\alpha =$ exponent between 0 and 1
Branched transport: what’s this?

Transport problems where the cost has a **subadditive** dependence on the mass, i.e. moving a mass $m$ for a distance $\ell$ costs

$$\varphi(m) \ell,$$

with $\varphi(m_1 + m_2) < \varphi(m_1) + \varphi(m_2) \implies \text{total cost} = \sum \varphi(m) \ell$

**typical choice**  $\varphi(t) = t^\alpha$, $\alpha \in [0, 1]$

Due to concavity, **grouping the mass** during the transport could lower the total cost $\implies$ typical optimal structures are **tree-shaped**
Remark

Many natural and artificial transportation systems satisfy this **cost saving requirement** (root systems in a tree, blood vessels...)

---

*Note: The PDF contains a watermark and a disclaimer regarding licensing.*
Example: a power supply station

- $\rho_0 = \delta_{x_0}$ power supply station
- $\rho_1 = \sum_{i=1}^{k} m_i \delta_{x_i}$ houses ($\sum_{i=1}^{k} m_i = 1$)

Comment

It is better to construct an optimal network of wires (right) to save cost; this is not possible by looking at Monge-Kantorovich (left).
Some models: Gilbert’s weighted oriented graphs

This is only suitable for **discrete measures**

\[
\rho_0 = \sum_{i=1}^{k} a_i \delta_{x_i} \in \mathcal{P}(\Omega) \quad \text{and} \quad \rho_1 = \sum_{j=1}^{m} b_j \delta_{y_j} \in \mathcal{P}(\Omega)
\]

**Transport path between** $\rho_0$ **and** $\rho_1$

- **weighted oriented graph** consisting of:
  - $\{v_s\}_{s \in V}$ vertices (comprising $x_i$ **sources** and $y_j$ **sinks**)
  - $\{e_h\}_{h \in H}$ edges
  - $\{\overrightarrow{\tau_h}\}_{h \in H}$ orientations of the edges
  - $\{m_h\}_{h \in H}$ weights (i.e. transiting mass on the edge $e_h$)

  + **Kirchhoff’s Law** for circuits
**Interior vertices**

\[
\begin{align*}
m_1 + m_2 + m_3 &= m_4
\end{align*}
\]

**“Boundary” vertices**

\[
\begin{align*}
m_1 + m_2 &= a_i \\
m_1 + m_2 &= b_j
\end{align*}
\]

**Total cost**

\[
M_\alpha(g) = \sum_{h \in H} m_h^\alpha \mathcal{H}^1(e_h) \quad (\text{Gilbert-Steiner energy})
\]
Some models: Xia’s transport path model I

Idea: for the discrete case...

- \( g \leadsto \phi_g \) vector measure \( \langle \phi_g, \vec{\phi} \rangle = \sum_{h \in H} m_h \int_{e_h} \vec{\phi} \cdot \vec{\tau}_h \, d\mathcal{H}^1 \)
- Kirchhoff’s Law \( \leadsto \text{div} \phi_g = \rho_0 - \rho_1 \)

...for the general case

\( \phi \) transport path between \( \rho_0 \) and \( \rho_1 \) if \( \exists \{g_n, \rho^n_0, \rho^n_1\} \}_{n \in \mathbb{N}} \) s.t.

\[ \phi_{g_n} \rightharpoonup \phi, \quad \rho^n_i \rightharpoonup \rho_i, \quad i = 0, 1 \]

Total cost

\[ M^*_\alpha := \text{relaxation of } M_\alpha \]

\[ M^*_\alpha(\phi) = \begin{cases} \int_{\Sigma} m(x)^\alpha \, d\mathcal{H}^1(x), & \text{if } \phi = m \vec{\tau} \mathcal{H}^1 \cup \Sigma, \\ +\infty, & \text{otherwise} \end{cases} \]
The variational setting

Some models: Xia’s transport path model II

**Theorem (Xia, Morel-Santambrogio)**

Let \( \alpha \in (1 - 1/N, 1] \) and \( \rho_0, \rho_1 \in \mathcal{P}(\Omega) \), then

\[
d_\alpha(\rho_0, \rho_1) := \min \left\{ M_\alpha^*(\phi) : \text{div} \phi = \rho_0 - \rho_1 \right\} < +\infty
\]

Moreover \( d_\alpha \) defines a **distance** on \( \mathcal{P}(\Omega) \), **equivalent** to \( w_1 \) (and thus to any \( w_p \), with \( 1 \leq p < \infty \))

\[
w_1(\rho_0, \rho_1) \leq d_\alpha(\rho_0, \rho_1) \leq C w_1(\rho_0, \rho_1)^{N(\alpha-1)+1}
\]

**Remark**

- the exponent \( N(\alpha - 1) + 1 \) can not be improved
- the lower bound is not optimal, actually we have \( w_1/\alpha \leq d_\alpha \)
  (Devillanova-Solimini)
Some models: a Lagrangian approach I

Transportation is described through $Q$ probability measures on Lipschitz paths (parametrized on $[0, 1]$, let us say)

Constraints

$$(e_0)_\# Q = \rho_0, \ (e_1)_\# Q = \rho_1$$

(where $e_t(\sigma) = \sigma(t)$ evaluation at $t$)

Multiplicity (i.e. “transiting mass”)

$$[x]_Q = Q(\{\tilde{\sigma} : x \in \tilde{\sigma}([0, 1])\}) \leq 1$$

Energy (Bernot-Caselles-Morel)

$$E_\alpha(Q) = \int_{Lip([0,1];\Omega)} \int_0^1 [\sigma(t)]_{Q}^{\alpha-1} |\sigma'(t)| \ dt \ dQ(\sigma)$$
If $E_\alpha(Q) < +\infty$ and $Q$ gives full mass to injective curves...

Gilbert-Steiner energy, again!

$$E_\alpha(Q) = \int_\Omega [x]^\alpha_Q \, d\mathcal{H}^1(x)$$

Theorem (Bernot-Caselles-Morel)

For every $\rho_0, \rho_1$, this Lagrangian model is equivalent to Xia’s one (i.e. same optimal structures, different description of the same energy)

There exist other Lagrangian models (Maddalena-Morel-Solimini\(^a\), Bernot-Figalli) that we are neglecting, differing for the definition of the multiplicity: the one chosen here is not local in time

\(^a\)This was actually the first!
Aim of the talk

We want to present a model for branched transport of the type

**Energy**

\[ G(\mu, v) = \int_0^1 G_\alpha(\mu_t, v_t) \, dt \quad \text{with} \quad t \mapsto \mu_t \text{ curve in } \mathcal{P}(\Omega) \]

\[ t \mapsto v_t \text{ velocity field} \]

**Constraints: the continuity equation**

\[
\begin{align*}
\partial_t \mu_t + \text{div}_x(v_t \mu_t) &= 0 \quad \text{in } \Omega, \\
\mu_0 &= \rho_0, \quad \mu_1 = \rho_1
\end{align*}
\]

**Remark**

This is **Eulerian** and **dynamical**, i.e. an optimal \( \mu \) provides the evolution in time of the branched transport with its **velocity field** \( v \), not just the optimal ramified structure.
The Benamou-Brenier formula I

First of all, recall the dynamical formulation for $w_p$ ($p > 1$)

Benamou-Brenier [Numer. Math. 84 (2000)]

$$w_p(\rho_0, \rho_1) = \min \left\{ \int_0^1 \int_\Omega |v_t(x)|^p \, d\mu_t(x) \, dt : \partial_t \mu_t + \text{div}_x(v_t \mu_t) = 0 , \mu_0 = \rho_0, \mu_1 = \rho_1 \right\}$$

Important

It can be reformulated as a **convex optimization** + **linear constraints**, introducing

$$\phi_t := v_t \cdot \mu_t \ (\text{momentum}) \implies |v_t|^p \mu_t = |\phi_t|^p \mu_t^{1-p} \text{ convex}$$

Thanks to the Disintegration Theorem...

$(\mu, \phi)$ can be thought as measures on $[0, 1] \times \Omega$ disintegrating as

$$\mu = \int \mu_t \, dt \text{ and } \phi = \int \phi_t \, dt$$
The Benamou-Brenier formula II

The functional can be rewritten as follows

$$f_p(x, y) = \begin{cases} 
|y|^p x^{1-p}, & \text{if } x > 0, y \in \mathbb{R}^N, \\
0, & \text{if } x = 0, y = 0, \\
+\infty, & \text{otherwise}
\end{cases}$$

is jointly convex and 1–homogeneous

Benamou-Brenier functional

$$\mathcal{F}_p(\mu, \phi) = \int_{[0,1] \times \Omega} f_p \left( \frac{d\mu}{dm}, \frac{d\phi}{dm} \right) dm$$

Comment

$$\mathcal{F}_p \text{ l.s.c. and does not depend on the choice of } m$$

$$w_p(\rho_0, \rho_1) = \min \left\{ \mathcal{F}_p(\mu, \phi) : \partial_t \mu + \text{div}_x \phi = \delta_0 \otimes \rho_0 - \delta_1 \otimes \rho_1 \right\}$$
Remark

By its very definition

$$\mathcal{F}_p(\mu, \phi) < +\infty \implies \phi \ll \mu$$

and in this case

$$\mathcal{F}_p(\mu, \phi) = \int_{[0,1] \times \Omega} \left| \frac{d\phi}{d\mu} \right|^p d\mu$$

If moreover $\mu = \int \mu_t \, dt$, then $\phi = \int \phi_t \, dt$ with $\phi_t = v_t \cdot \mu_t$ and

$$\mathcal{F}_p(\mu, \phi) = \int_0^1 \int_\Omega \left| \frac{d\phi_t}{d\mu_t} \right|^p d\mu_t \, dt = \int_0^1 \int_\Omega |v_t|^p d\mu_t \, dt$$
We consider the **local** and **l.s.c. functional on measures**

\[ g_\alpha(\lambda) = \begin{cases} \int_\Omega |\lambda(\{x\})|^\alpha \, d\#(x), & \text{if } \lambda \text{ is atomic} \\ +\infty, & \text{otherwise} \end{cases} \]

**Energy?**

For \( \mu = \int \mu_t \, dt \) and \( \phi = \int \phi_t \, dt \) with \( \phi_t \ll \mu_t \)

\[ G_\alpha(\mu, \phi) = \int_0^1 g_\alpha \left( \left| \frac{d\phi_t}{d\mu_t} \right|^{1/\alpha} \mu_t \right) \, dt = \int_0^1 g_\alpha \left( |v_t|^{1/\alpha} \mu_t \right) \, dt \]

**This is a Gilbert-Steiner energy!**

\[ G_\alpha(\mu, \phi) = \int_0^1 \sum_{k \in \mathbb{N}} |v_t(x_{k,t})| \mu_t(\{x_{k,t}\})^\alpha \, dt \]
A possible variant for branched transport: setting

\[ \mathcal{D} = \text{admissible pairs } (\mu, \phi) \]

\[ \mu \in C([0, 1]; \mathcal{P}(\Omega)) \]

\[ \phi \in L^1([0, 1]; \mathcal{M}(\Omega; \mathbb{R}^N)) \]

Dynamical branched energy

\[ G_\alpha(\mu_t, \phi_t) = \begin{cases} 
\int_\Omega |v_t(x)| \mu_t(\{x\})^\alpha \ d\#(x) & \text{if } \phi_t = v_t \cdot \mu_t, \\
+\infty & \text{if } \phi_t \ll \mu_t 
\end{cases} \]

\[ G_\alpha(\mu, \phi) = \int_0^1 G_\alpha(\mu_t, \phi_t) \ dt, \ (\mu, \phi) \in \mathcal{D} \]

Important remark

\[ G_\alpha(\mu, \phi) < +\infty \iff \mu_t \text{ atomic } \forall t \]

\[ \Rightarrow \phi \ll \mu \text{ and } \mu_t \text{ atomic on } \{|v_t(x)| > 0\} \]
Main result

Theorem (B.-Buttazzo-Santambrogio)

For every $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$, the minimization problem

$$B_\alpha(\rho_0, \rho_1) = \min_{(\mu, \phi) \in \mathcal{D}} \{ G_\alpha(\mu, \phi) : \mu_0 = \rho_0, \mu_1 = \rho_1 \}$$

admits a solution

 Remark 1

The proof uses Direct Methods...l.s.c.? coercivity? As always, it is a matter of choosing the right topology

 Remark 2

Observe that the problem is not convex, but rather concave
Choice of the topology

Proposal: pointwise convergence

What about “$\mu^n_t \rightharpoonup \mu_t$ for every $t$ and $\phi^n_t \rightharpoonup \phi_t$ for a.e. $t$”? 

Answer: NO

Good for l.s.c. (you simply apply Fatou Lemma, because $G_\alpha$ is l.s.c.) but not so good for coercivity (how can we infer compactness from $G_\alpha \leq C$?)

Choice: weak topology

$$(\mu^n, \phi^n) \rightharpoonup (\mu, \phi)$$ (as measures on $[0, 1] \times \Omega$$)
The basic inequalities

If \((\mu, \phi) \in \mathcal{D}\) such that \(\phi \ll \mu\) and \(\phi_t = v_t \cdot \mu_t\)

\((B.I.)_1\)

\[
G_\alpha(\mu_t, \phi_t) = \sum_i \mu_t(\{x_i\})^\alpha |v_t(x_i)| = \sum_i \left( \mu_t(\{x_i\}) |v_t(x_i)|^{1/\alpha} \right)^\alpha \\
\geq \left( \sum_i \mu_t(\{x_i\}) |v_t(x_i)|^{1/\alpha} \right)^\alpha = \|v_t\|_{L^{1/\alpha}(\mu_t)} \geq |\mu'_t|_{W^{1/\alpha}}
\]

\((B.I.)_2\)

\[
G_\alpha(\mu, \phi) = \int_0^1 G_\alpha(\mu_t, \phi_t) \, dt \geq \int_0^1 |\phi_t|(\Omega) \, dt = |\phi|([0, 1] \times \Omega)
\]

Remark

\[
\sup_t G_\alpha \leq C \implies |\phi|([0, 1] \times \Omega) \leq C \text{ and } \mu_t \text{ Lipschitz in } W^{1/\alpha}_{1/\alpha}
\]
Proof of the main result 1

Stage 1 – Extraction of a subsequence

- \{ (\mu^n, \phi^n) \} \subset \mathcal{D} minimizing sequence
- we can assume \( G_\alpha(\mu^n, \phi^n) \leq C \) for every \( n \)
- \( G_\alpha \) 1–homogeneous w.r.t. \( \nu_t \) (i.e. reparametrization invariant)
- \((\mu^n, \phi^n) \leadsto (\tilde{\mu}^n, \tilde{\phi}^n)\), with \( \tilde{\mu}^n_s = \mu^n_{t(s)} \) and \( \tilde{\phi}^n_s = t'(s) \cdot \phi^n_{t(s)} \)
- choose \( t \) s.t. \( G_\alpha(\tilde{\mu}^n_s, \tilde{\phi}^n_s) \equiv G_\alpha(\tilde{\mu}^n, \tilde{\phi}^n) = G_\alpha(\mu^n, \phi^n) \leq C \)
- \( \implies \tilde{\mu}^n \rightharpoonup \mu \) and \( \tilde{\phi}^n \rightharpoonup \phi \) (thanks to (B.I.)_1 and (B.I.)_2)
Proof of the main result II

Stage 2 – Admissibility of the limit

- clearly $\mu = \int \mu_t \, dt$ (uniform limit of continuous curves)
- to show that $\phi = \int \phi_t \, dt$, we use l.s.c. of Benamou-Brenier functional

$$F_{1/\alpha}(\mu, \phi) \leq \liminf_{n \to \infty} F_{1/\alpha}(\tilde{\mu}^n, \tilde{\phi}^n) \overset{(B.I.)_1}{\leq} C$$

$\implies \phi \ll \mu$ and $\phi = \int \phi_t \, dt$

- $(\mu, \phi)$ still solves the continuity equation $\implies (\mu, \phi) \in \mathcal{D}$
- $\mu_0 = \rho_0$ and $\mu_1 = \rho_1$
Proof of the main result III (conclusion)

Stage 3 – l.s.c. along a minimizing sequence

- remember that $\tilde{\phi}^n = \tilde{v}^n \cdot \tilde{\mu}^n$ and $G_\alpha(\tilde{\mu}^n, \tilde{\phi}^n) \leq C$
- define $m^n = \int \sum_i |\tilde{v}_t^i(x_i, t)| \tilde{\mu}_t^i(\{x_i, t\})^\alpha \delta_{x_i,t} \, dt \in \mathcal{M}([0, 1] \times \Omega)$
- $m^n([0, 1] \times \Omega) = G_\alpha(\tilde{\mu}^n, \tilde{\phi}^n) \leq C$
- $\implies m^n \rightharpoonup m$ and $m = \int m_t \, dt$
- $m^n([0, 1] \times \Omega) \to m([0, 1] \times \Omega) \implies G_\alpha(\tilde{\mu}^n, \tilde{\phi}^n) \to \int_0^1 m_t(\Omega) \, dt$
- show $m_t(\Omega) \geq G_\alpha(\mu_t, \phi_t)$ (a little bit delicate)
- $\implies G_\alpha(\mu, \phi) \leq \liminf_{n \to \infty} G_\alpha(\tilde{\mu}^n, \tilde{\phi}^n) = \inf G_\alpha$
Equivalences with other models

Theorem (B.-Buttazzo-Santambrogio)

$$\mathfrak{B}_\alpha(\rho_0, \rho_1) = \min \{ E_\alpha(Q) : (e_i)_\# Q = \rho_i \} = d_\alpha(\rho_0, \rho_1)$$

As always, we have equivalence of the problems, not just equality of the minima.

Recall that

$$E_\alpha(Q) = \int_{\text{Lip}([0,1]:\Omega)} \int_0^1 [\sigma(t)]_Q^{\alpha-1} |\sigma'(t)| \, dt \, dQ(\sigma)$$

Remark

In order to compare the two models, we need to switch from curves of measures to measures on curves (and back!)
Some preliminary comments

Alert!
- Transiting mass in our model $\Rightarrow \mu_t(\{x\})$ (local in space/time)
- Transiting mass in $E_\alpha$ model $\Rightarrow [x]Q$ (not local in time)

We will need the following

Theorem (Superposition principle (AGS, Theorem 8.2.1))

Let $(\mu, \nu)$ solve the continuity equation, with $\|\nu_t\|_{L^p(\mu_t)}^{P}$ integrable in time. Then $\mu_t = (e_t)_\sharp Q$ with $Q$ concentrated on solutions of the ODE $\sigma'(t) = \nu_t(\sigma(t))$

Comment

This is a probabilistic version of the method of characteristics
Sketch of the proof: \( \mathcal{B}_\alpha(\rho_0, \rho_1) \geq d_\alpha(\rho_0, \rho_1) \)

**Step 1**

\((\mu, \phi)\) optimal \(\implies\) \(\phi = v \cdot \mu\) and \(\int_0^1 \|v_t\|_{L^{1/\alpha}(\mu_t)} \, dt \leq \mathcal{B}_\alpha(\rho_0, \rho_1)\)

**Step 2 - superposition principle**

\(\exists Q \text{ s.t. } \mu_t = (e_t)^\# Q\) and \(\sigma'(t) = v_t(\sigma(t))\) for \(Q-\text{a.e. } \sigma\)

**Step 3 - comparison of the multiplicities**

\(\mu_t = (e_t)^\# Q \implies [x]_Q \geq Q(\{\tilde{\sigma} : \tilde{\sigma}(t) = x\}) = \mu_t(\{x\})\)

\[
\int [\sigma(t)]_Q^{\alpha-1} |\sigma'(t)| \, dQ(\sigma) \overset{\text{Step 2}}{=} \int [x]_Q^{\alpha-1} |v_t(x)| \, d\mu_t(x) \\
\overset{\text{Step 3}}{\leq} \int \mu_t(\{x\})^{\alpha-1} |v_t(x)| \, d\mu_t(x)
\]
Sketch of the proof: \( \mathcal{B}_\alpha(\rho_0, \rho_1) \leq d_\alpha(\rho_0, \rho_1) \)

**Step 0 – approximation**
Approximate \((\rho_0, \rho_1)\) with \((\rho^n_0, \rho^n_1)\) (finite sums of Dirac masses) s.t. \(d_\alpha(\rho^n_0, \rho^n_1) \to d_\alpha(\rho_0, \rho_1)\)

**Remark: why approximation?**
\(\exists Q\) optimal s.t.
\[ [\sigma(t)]_Q = Q(\{\tilde{\sigma}(t) = \sigma(t)\}) \]
the mass is **synchronized**
(this is true if \(\rho_0\) is finitely atomic)

**Step 1 – curve in \(\mathcal{P}(\Omega)\)**
\(\mu_t := (e_t)_\#Q\) and disintegrate \(Q = \int Q^t_x \, d\mu_t(x)\) (i.e. \(Q^t_x\) is concentrated on \(\{\sigma : \sigma(t) = x\}\))
Sketch of the proof: $\mathcal{B}_\alpha(\rho_0, \rho_1) \leq d_\alpha(\rho_0, \rho_1)$

**Step 2 – velocity field**

$$v_t(x) := \int_{\{\sigma : \sigma(t) = x\}} \sigma'(t) \, dQ_t^x(\sigma) \text{ (average velocity)}$$

**Step 3**

$$(\mu, v \cdot \mu) \in D \text{ and } G_\alpha(\mu, v \cdot \mu) \leq E_\alpha(Q) = d_\alpha(\rho_0^n, \rho_1^n), \text{ with } \mu_0 = \rho_0^n \text{ and } \mu_1 = \rho_1^n$$

**Step 4**

Putting all together, we have

$$\mathcal{B}_\alpha(\rho_0, \rho_1) \leq \liminf_{n \to \infty} \mathcal{B}_\alpha(\rho_0^n, \rho_1^n) \leq \lim_{n \to \infty} d_\alpha(\rho_0^n, \rho_1^n) = d_\alpha(\rho_0, \rho_1)$$
A final remark: comparison of $d_\alpha$ and $w_{1/\alpha}$

Taking $(\mu, \phi)$ optimal for $\mathcal{B}_\alpha(\rho_0, \rho_1)$

$$\int_0^1 |\mu'_t|w_{1/\alpha} \, dt \leq \mathcal{B}_\alpha(\rho_0, \rho_1) \quad \text{equivalence} \quad d_\alpha(\rho_0, \rho_1)$$

i.e. we have another proof of

$$w_{1/\alpha}(\rho_0, \rho_1) \leq d_\alpha(\rho_0, \rho_1)$$

Remark

$d_\alpha$ and $w_{1/\alpha}$ have exactly the same scaling

$$d_\alpha = \sum m^\alpha \ell \quad w_{1/\alpha} = \left(\sum m \ell^{1/\alpha}\right)^\alpha$$
Further readings

- Standard reference on branched transport

- Other models employing curves in Wasserstein spaces (but avoiding the use of the continuity equation) have been studied