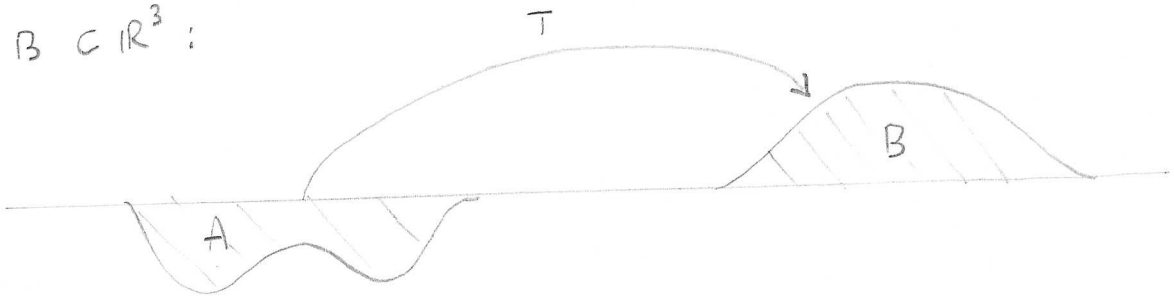
Teoria del Trasporto Ottimo ed Elementi di Analisi Superiore

Appunti del Corso 2022-23

Roberto Monti

Teoria del Trasporto ottimo. Introduzione storica, problemi e applicazioni

1) Nel 1781 Monge si pone il problema di trasportare in modo ottimale un quantitativo di materiale $A \subset \mathbb{R}^3$ preso in una cava in un sito di costruzione $B \subset \mathbb{R}^3$:

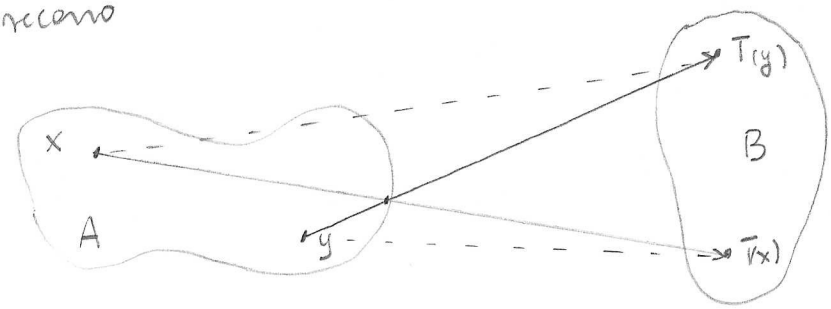


L'incognita è una funzione $T : A \to B$ che preserva il volume. Per Monge il costo da minimizzare è

$$C(T) = \int_A \underbrace{|X - T(x)|}_{\text{"distanza fra } x \text{ e } Tx\text{"}} dx .$$

Vedremo che questo problema è difficile, in ultima analisi perché $x \mapsto |x| = \left(\sum_{i=1}^{3} x_i^2\right)^{1/2}$ non è strettamente convessa.

Un'osservazione semplice è che il trasporto deve avvenire lungo raggi di trasporto, i segmenti $[x, Tx]$, che non si intersecano



Questa T non è ottimale.

2) Nel 1939 Kantorovic riformula il problema di Monge in termini di "programmazione lineare" (→ Ottimizzazione discreta). Uno dei problemi considerati da Kantorovic è il seguente. Ci sono N siti produttivi $x_1 \dots x_N$ (ad esempio in $\mathbb{R}^2$) ed M centri di distribuzione/consumo $y_1 \dots y_M$. Il sito x_i produce un quantitativo p_i di prodotto, mentre y_j ne riceve un quantitativo q_j. Dovrà essere: $p_1 + \dots + p_N = q_1 + \dots + q_M$ (= 1 ad esempio)
Sia $c(x,y)$ il costo per trasportare un'"unità" di prodotto da x ed y.
Kantorovic vuole trovare il "piano di trasporto" che minimizza il costo totale.
È chiaro che se $M \neq N$ (e $\{p_1 \dots p_N\} \neq \{q_1 \dots q_M\}$) la soluzione non può essere una mappa da $\{x_1 \dots x_N\}$ a $\{y_1 \dots y_N\}$. Il prodotto può venir spezzato e distribuito fra vari $y_1 \dots y_M$. Occorre rispettare la legge additiva: Kantorovic ha l'idea di riformulare il problema in termini di teoria della misura.
Siano X, Y due insiemi con misure unitarie μ e ν. Un "accoppiamento" di μ e ν è una misura π sul prodotto $X \times Y$ tale che

$$\mu(A) = \pi(A \times Y) \quad \text{per } A \subset X$$
$$\nu(B) = \pi(X \times B) \quad \text{per } B \subset Y.$$

Il problema di Kantorovic è di minimizzare

$$C(\pi) = \int_{X \times Y} c(x,y)\, d\pi,$$

su tutti gli accoppiamenti.

Kantorovic collega il suo problema con quello di Monge, considera via il caso discreto che quello continuo. Riconosce anche il "problema duale", che vedremo. Si occupa anche del problema detto "determinazione dei prezzi", Poi si dedica all'informatica teorica. Ottiene il Nobel per l'economia nel 1975.

3) Nel 1976 Sudakov affronta il problema dell'esistenza di soluzione per il Problema di Monge. Incollando le soluzioni nel caso di dimensione $n=1$, Sudakov "costruisce" una soluzione del problema.
La dimostrazione contiene un "gap" finito nel 2003 da Ambrosio - Pratelli.
Una prima dimostrazione alternative dell'esistenza è del 1999, Evans - Gangbo.

4) Nel 1991 Brenier fa un passo in avanti importante sul problema di Monge con costo quadratico:

$$C(T) = \int_{\mathbb{R}^n} |X - T(x)|^2 \, d\mu$$

μ misura su $\mathbb{R}^n$ finita, uniteria e assolutamente continua.
La mappa T ottimale è il gradiente di una funzione convessa, $T = \nabla \varphi$.
Il punto di partenza di Brenier è la Fluidodinamica.
Il problema è di provare l'esistenza di soluzioni $v : \mathbb{R} \times \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ per l'equazione di Eulero

$$\begin{cases} v_t(t,x) + v \cdot \nabla_x v(t,x) = -\nabla_x p(t,x) & \text{in } A \subset \mathbb{R}^n,\ t>0 \\ \text{div}_x(v(t,x)) = 0 \end{cases}$$

p = pressione

v = velocità

fluido incomprimibile

con condizioni iniziali al $t=0$ e condizioni al bordo.

In generale questo problema non ha soluzioni regolari $v \in C^1$ in tempo e spazio.

Brenier usa la mappa ottimale $T = \nabla\varphi$ per costruire soluzioni "generalizzate" di questo problema.

5) Nel 1996 Gangbo e McCann estendono la teoria del trasporto ottimo al caso in cui $\mathbb{R}^n$ è sostituito da una varietà Riemanniana. L'estensione getta una nuova luce sul Tensore di Curvatura di Ricci.

Nel 2006 (Sturm) e nel 2009 (Lott-Villani) nasce una "teoria delle curvature" in spazi metrici con tecniche di trasporto ottimo.

6) A partire dagli anni 2000 varie diseguaglianze funzionali e geometriche vengono (ri-)dimostrate con Tecniche di trasporto ottimo. Noi vedremo una dimostrazione della diseguaglianza di Brunn-Minkowski:

$$|A+B|^{\frac{1}{n}} \geq |A|^{\frac{1}{n}} + |B|^{\frac{1}{n}} \quad \text{con } A, B \subset \mathbb{R}^n.$$

misura di Lebesgue

Il concetto chiave è la "displacement-convexity" di funzionali lungo le geodetiche che collegano due misure (geodetiche di Wasserstein).

7) Altre applicazioni della teoria del trasporto ottimo riguardano: sistemi dinamici (Teoria di Mather), meteorologia (sistema quantistico), trasporto su reti e pianificazione urbana, idrologia.

Commentiamo il lavoro di cosmologia "A reconstruction of the initial conditions of the Universe by optimal mass transportation" di Frisch e altri apparso su Nature nel 2002.

Consideriamo la distribuzione di massa nell'Universo ad un tempo iniziale (variabile q) ed al tempo attuale (variabile x). Queste distribuzioni sono date.
Gli autori dicono che la mappa $q \mapsto x$ può essere approssimata dal (gradiente di) un potenziale):

$$x = \nabla_q \Phi(q).$$

Gli autori introducono l'ipotesi: Φ è convessa.
La trasformazione inversa è del tipo

$$q = \nabla_x \Theta(x)$$

dove Θ è la trasformata di Legendre-Fenchel di Φ
La conservazione della massa è (Equazione di Monge-Ampère)

$$\det(\mathrm{Hess}\,\Theta(x)) = \rho(x)/\rho_0$$

dove $\rho : \mathbb{R}^3 \to \mathbb{R}^+$ è la distribuzione attuale di massa e ρ_0 distribuzione iniziale ($\sim$ costante).

La mappa $x \longmapsto T(x) = \nabla_x \Phi$ è la soluzione del problema di minimo

$$C(T) = \int_{\mathbb{R}^3} |x - Tx|^2 \rho_0(x)\, dx .$$

Gli autori affermano che questa "ricostruzione" si accorda bene con altre simulazioni ed osservazioni.

Aggiungiamo che la Teoria del trasporto ottimo propone anche una ricostruzione dell'evoluzione nei tempi intermedi.

8) A partire dal 1990, Caffarelli ha sviluppato una importante teoria della regolarità per le soluzioni generalizzate dell'equazione di Monge-Ampère

$$\det\big(H(\varphi(x))\big) = f(x) \quad , \quad x \in A$$

H = matrice Hessiana di φ

La teoria è rilevante in trasporto ottimo.
Ci sono risultati recenti in questo ambito di A. Figalli.

A reconstruction of the initial conditions of the Universe by optimal mass transportation

Uriel Frisch*, Sabino Matarrese†, Roya Mohayaee‡* & Andrei Sobolevski§*

* CNRS, UMR 6529, Observatoire de la Côte d'Azur, BP 4229, 06304 Nice Cedex 4, France
† Dipartimento di Fisica "G. Galilei" and INFN, Sezione di Padova, via Marzolo 8, 35131-Padova, Italy
‡ Dipartimento di Fisica, Universitá Degli Studi di Roma "La Sapienza", P.le A. Moro 5, 00185-Roma, Italy
§ Department of Physics, M.V. Lomonossov University 119899-Moscow, Russia

Nature| VOL 417| 16 MAY 2002

Reconstructing the density fluctuations in the early Universe that evolved into the distribution of galaxies we see today is a challenge of modern cosmology.[1] An accurate reconstruction would allow us to test cosmological models by simulating the evolution starting from the reconstructed state and comparing it to the observations. Several reconstruction techniques have been proposed,[2-9] but they all suffer from lack of uniqueness because the velocities of galaxies are usually not known. Here we show that reconstruction can be reduced to a well-determined problem of optimisation, and present a specific algorithm that provides excellent agreement when tested against data from N-body simulations. By applying our algorithm to the new redshift surveys now under way,[10] we will be able to recover reliably the properties of the primeval fluctuation field of the local Universe and to determine accurately the peculiar velocities (deviations from the Hubble expansion) and the true positions of many more galaxies than is feasible by any other method.

Starting from the available data on the galaxy distribution, can we trace back in time and map to its initial locations the highly structured distribution of mass in the Universe (Fig. 1)? Here we show that, with a suitable hypothesis, the knowledge of both the present non-uniform distribution of mass and of its primordial quasi-uniform distribution uniquely determines the *inverse Lagrangian map*, defined as the transformation from present (Eulerian) positions $\mathbf{x}$ to the respective initial (Lagrangian) positions $\mathbf{q}$.

We first consider the direct Lagrangian map $\mathbf{q} \mapsto \mathbf{x}$, which can be approximately written in terms of a potential as $\mathbf{x} = \nabla_{\mathbf{q}}\Phi(\mathbf{q})$ at those scales where nonlinearity stays moderate.[11] This is supported by numerical N-body simulations showing good agreement with a very simple potential approximation, due to Zel'dovich,[12] which assumes that the particles move on straight trajectories. Even better agreement is obtained with a refinement, the second-order Lagrangian perturbation method,[13-16] also known to be potential.

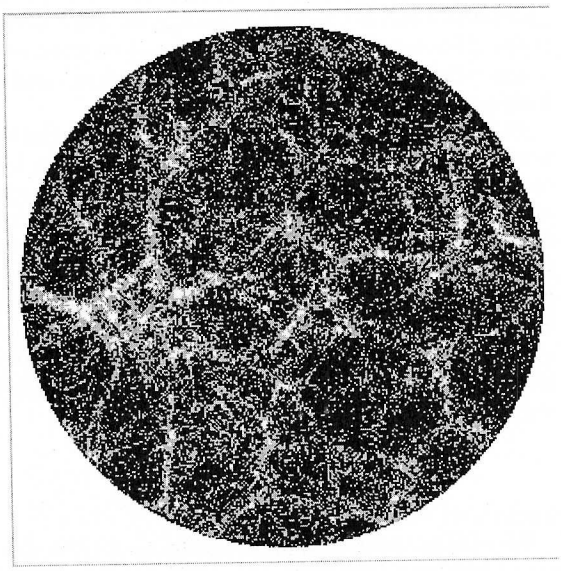


FIG. 1. N-body simulation output (present epoch) used for testing our reconstruction method. In the standard model of structure formation, the distribution of matter in the Universe is believed to have emerged from a very smooth initial state: tiny irregularities of the gravitational potential, which we can still observe as temperature fluctuations of the cosmic microwave background, gave rise to density fluctuations, which grew under their self-gravity, developing a rich and coherent pattern of structures. Most of the mass is in the form of cold dark matter; the luminous matter (galaxies) can be assumed to trace – up to some form of bias – the underlying dark matter. Shown here is a projection onto the x-y plane of a 10% slice of the simulation box of size $200h^{-1}$ Mpc. The model, ΛCDM, uses cold dark matter with cosmological constant and the following parameters: Hubble constant $h = 0.65$, density parameters $\Omega_\Lambda = 0.7$ and $\Omega_{\rm m} = 0.3$, normalization factor $\sigma_8 = 0.99$. Points are highlighted in yellow when reconstruction fails by more than $6\,h^{-1}$ Mpc, which happens mostly in high-density regions.

In our "reconstruction hypothesis", we furthermore assume the convexity of the potential $\Phi(\mathbf{q})$, a consequence of which is the absence of multi-streaming: for almost any Eulerian position, there is a single Lagrangian antecedent. As is well-known, the Zel'dovich approximation leads to caustics and to multi-streaming. This can be overcome in various ways, for example by a modification known as the adhesion model, an equation of viscous pressureless gas dynamics.[17,18] The latter, which

leads to shocks rather than caustics, is known to have a convex potential[19] and to be in better agreement with N-body simulations. Suppression or reduction of multi-streaming requires a mechanism of momentum exchange, such as viscosity, between neighbouring streams having different velocities. This is a common phenomenon in ordinary fluids, such as the flow of air or water in our natural environment. Dark matter is however essentially collisionless and the usual mechanism for generating viscosity (discovered by Maxwell) does not operate, so that a non-collisional mechanism involving a small-scale gravitational instability must be invoked.

Our reconstruction hypothesis implies that the initial positions can be obtained from the present ones by another gradient map: $\mathbf{q} = \nabla_{\mathbf{x}}\Theta(\mathbf{x})$, where Θ is a convex potential related to Φ by a Legendre–Fenchel transform (see Methods). We denote by ρ_0 the initial mass density (which can be treated as uniform) and by $\rho(\mathbf{x})$ the final one. Mass conservation implies $\rho_0 d^3q = \rho(\mathbf{x})d^3x$. Thus, the ratio of final to initial density is the Jacobian of the inverse Lagrangian map. This can be written as the following Monge–Ampère equation[20] for the unknown potential Θ

$$\det\left(\nabla_{x_i}\nabla_{x_j}\Theta(\mathbf{x})\right) = \rho(\mathbf{x})/\rho_0, \qquad (1)$$

where 'det' stands for determinant.

We emphasize that no information about the dynamics of matter other than the reconstruction hypothesis is needed for our method, whose degree of success depends crucially on how well this hypothesis is satisfied. Exact reconstruction is obtained, for example, for the Zel'dovich approximation (before particle trajectories cross) and for adhesion-model dynamics (at arbitrary times).

We note that our Monge–Ampère equation for self-gravitating matter may be viewed as a nonlinear generalisation of a Poisson equation (used for reconstruction in ref. 4), to which it reduces if particles have moved very little from their initial positions.

It has been discovered recently that the map generated by the solution to the Monge-Ampère equation (1) is the (unique) solution to an optimisation problem[21] (see also refs 22 and 23). This is the 'mass transportation' problem of Monge and Kantorovich[24,25] in which one seeks the map $\mathbf{x} \mapsto \mathbf{q}$ that minimises the quadratic 'cost' function

$$I = \int_{\mathbf{q}} \rho_0 |\mathbf{x}-\mathbf{q}|^2 \, d^3q = \int_{\mathbf{x}} \rho(\mathbf{x})|\mathbf{x}-\mathbf{q}|^2 \, d^3x. \qquad (2)$$

Note that $\mathbf{x} = \mathbf{q}$ is forbidden: as the initial and final density fields ρ_0 and $\rho(\mathbf{x})$ are prescribed there is a constraint on Jacobian of the map (see Methods).

Next, we take into account that information on the mass distribution is provided in the form of N discrete particles both in simulations and when handling observational data from galaxy surveys. The cost minimisation then becomes what is known in optimisation theory as the assignment problem: find the unique one-to-one pairing of a set of N initial points $\mathbf{q}_j$'s and N final points $\mathbf{x}_i$'s that minimises $I_{\rm discr} = \sum_{i=1}^{N} |\mathbf{x}_i - \mathbf{q}_{j(i)}|^2$. An immediate consequence is that, for any subset of k pairs of initial and final points ($2 \le k \le N$), the contribution of these points to the cost function should not decrease under arbitrary permutations of initial points. This property is known to be equivalent to having a Lagrangian map that is the gradient of a convex function.[26]

If we restrict ourselves to interchanging just pairs ($k = 2$), the map is said to be monotonic, a condition not equivalent to minimisation of the cost function (except in one dimension). In ref. 8, a method of reconstruction called the Path Interchange Zel'dovich Approximation (PIZA) is introduced which uses the same quadratic cost function (obtained by applying a minimum-action argument within the framework of the Zel'dovich approximation). In PIZA, a randomly chosen tentative correspondence between initial and final points is successively improved by swapping *pairs* of initial particles whenever this decreases the cost function. Eventually, a monotonic map is obtained which usually does not minimise the cost. This explains the non-uniqueness of PIZA reconstruction (also noticed in ref. 8).

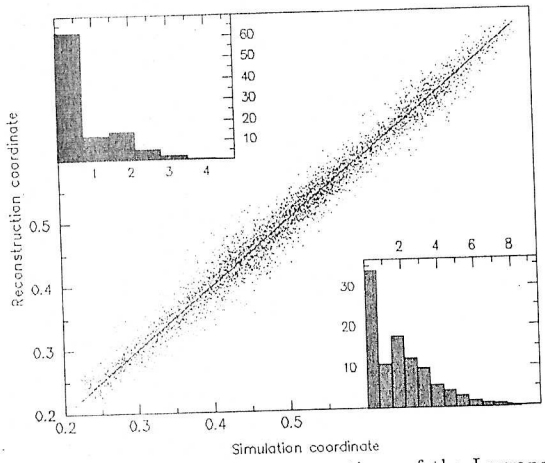


FIG. 2. Tests of MAK reconstructions of the Lagrangian positions, using the data shown in Fig. 1. The dots near the diagonal are a scatter plot of reconstructed initial points vs. simulation initial points for the coarsest $6.25\,h^{-1}$ Mpc grid with 17,178 points. The scatter diagram uses a 'quasi-periodic projection' coordinate $\tilde{q} \equiv (q_1 + q_2\sqrt{2} + q_3\sqrt{3})/(1+\sqrt{2}+\sqrt{3})$, which guarantees a one-to-one correspondence between $\tilde{q}$-values and points on the regular Lagrangian grid. The upper left inset is a histogram (by percentage) of distances in reconstruction mesh units between such points; the first slightly darker bin (whose width was taken to be slightly less than one mesh) corresponds to perfect reconstruction (thereby allowing a good determination of the peculiar velocities of galaxies); the lower right inset is a similar histogram for reconstruction on a finer $3.12\,h^{-1}$ Mpc grid using 19,187 points. With the $6.25h^{-1}$ Mpc grid, 62% of the 17,178 points are assigned perfectly and about 75% are within not more than one mesh. With the $3.12h^{-1}$ Mpc grid, we have 34% of exact reconstruction out of 19,187 points. On further refinement of the mesh by a factor two, this degrades to 14%.

There are, however, known deterministic strategies for the assignment problem which give the correct unique solution; their complexity (dependence on N of the number of operations needed) is close to N^3 for arbitrary cost functions, but can be sharply reduced when the cost function is quadratic. Combining the organisation of data taken from Hénon's mechanical analogue machine[27] for solving the assignment problem with the dual simplex method of Balinski,[28] we have designed an algorithm which gives the optimal assignement for about 20,000 particles in a few hours of CPU on a fast Alpha machine. For historical reasons we call our approach Monge–Ampère–Kantorovich or MAK (see Methods). Details of the algorithms will be given elsewhere; we merely note that, when working with catalogues of several hundred thousands of galaxies expected within a few years, a direct application of the assignment algorithm in its present state could require unreasonable computational resources. A mixed strategy can however be used, in which the assignment problem is solved on a coarse grid while, on smaller scales, the Monge–Ampère equation (1) is solved by a relaxation technique (adapted from ref. 23).

We tested the MAK reconstruction on data obtained by a cosmological N-body simulation with 128^3 particles, using the HYDRA code[29] (Fig. 1). Reconstruction was performed on three 32^3 grids with (comoving) meshes given by $\Delta x = 6.25\,h^{-1}$ Mpc, $\Delta x/2$ and $\Delta x/4$, where h is the Hubble constant in units of 100 km s^{-1} Mpc^{-1}. In comoving coordinates, the typical displacement of our mass elements over one Hubble time is about ten h^{-1} Mpc. We discarded those points that, at the end of the simulation (present epoch), were not within a sphere containing about 20,000 points, a number comparable to that of currently available all-sky galaxy redshift catalogues. As the simulation assumes periodic boundary conditions, we also took into account periodicity when calculating the distance between pairs of points. The MAK reconstructions were used to generate a scatter diagram and various histograms allowing comparisons of simulation and reconstructed Lagrangian points (Fig. 2). The results demonstrate the essentially potential character of the Lagrangian map above $\sim 6\,h^{-1}$ Mpc (within the given cosmological model).

We also performed PIZA reconstructions on the coarsest grid and obtained typically 30–40% exactly reconstructed points, but severe non-uniqueness: for two different seeds of the random generator only about half of the exactly reconstructed points were the same.

When reconstructing from observational data, in redshift space (Fig. 3), the galaxies appear displaced radially (as seen by the observer) by an amount proportional to the radial component of the peculiar velocity. We thus performed another reconstruction, with an accordingly modified cost function, that led to somewhat degraded results (Fig. 3) but at the same time provided an approximate determination of peculiar velocities. More accurate determination of peculiar velocities can be done using second-order Lagrangian perturbation theory. The effect of the catalogue selection function can be handled by standard techniques; for instance one can assign each galaxy a 'mass' inversely proportional to the catalog selection function[9,10].

What is the smallest length scale at which an optimisation algorithm such as MAK can be expected to give a unique and reliable reconstruction? The key ingredient here is the *simultaneous* knowledge of the initial and present mass density fields. MAK-type reconstruction (with a suitable cost based on the equation of a self-gravitating fluid) should therefore be possible down to scales comparable to the thickness of collapsed structures, below which the hydrodynamical description ceases to be meaningful.

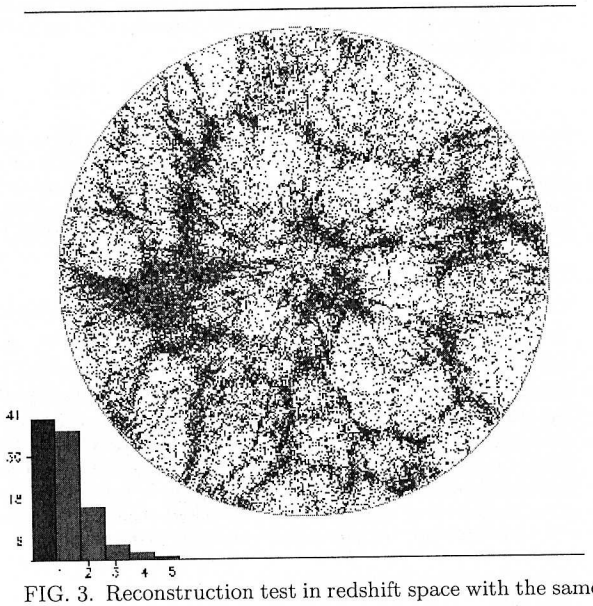


FIG. 3. Reconstruction test in redshift space with the same data as for the real-space reconstruction tested in the upper left histogram of Fig. 2. The circular redshift map (violet points) corresponds to the same real-space slice as displayed in Fig. 1. The observer is taken to be at the center of the simulation box. Points used for reconstruction within the displayed slice are highlighted in red. Reconstruction is performed by the MAK algorithm with a different cost function, obtained (as in ref. 8) by assuming that the peculiar velocities $\mathbf{v}$ can be estimated by the Zel'dovich approximation: $\boldsymbol{v} = f(\mathbf{x} - \mathbf{q})$, where $f \approx \Omega_m^{0.6} \approx 0.49$. Note that we now have 43% of exactly reconstructed points, included among the 60% which are within not more than $6.25\,h^{-1}$ Mpc from their correct positions.

The fact that MAK guarantees a unique solution and that our present reconstruction hypothesis proved to be very faithful down to $6.5\,h^{-1}$ Mpc makes our method very promising for the analysis of galaxy redshift surveys.[10] Reconstruction of the primordial positions and velocities of matter will allow us to test the Gaussian nature of the primordial perturbations and the self-consistency of cosmological hypotheses, such as the choice of the global cosmological parameters and the assumed biasing scheme.

By obtaining a point-by-point reconstruction of the specific realisation that describes the observed patch of our Universe, we can distinguish between universal properties and the influence of the large-scale environment on the galaxy formation process. Moreover, reconstruction will open a new window not only onto the past but also into the present Universe: it would enable us to make a first-time determination of the peculiar velocities of a very large number of galaxies, using their positions in redshift catalogues.

Methods

Monge–Ampère equation

The Lagrangian map $\mathbf{q} \mapsto \mathbf{x}$ is taken to be the gradient of a convex potential $\Phi(\mathbf{q})$; therefore its inverse $\mathbf{x} \mapsto \mathbf{q}$ also has a potential representation $\mathbf{q} = \nabla_{\mathbf{x}}\Theta(\mathbf{x})$, where $\Theta(\mathbf{x})$ is again a convex function; the two potentials are Legendre–Fenchel transforms of each other (see ref. [30], pp. 61–65):

$$\Theta(\mathbf{x}) = \max_{\mathbf{q}} \left[\mathbf{q}\cdot\mathbf{x} - \Phi(\mathbf{q})\right]\,;\ \Phi(\mathbf{q}) = \max_{\mathbf{x}} \left[\mathbf{x}\cdot\mathbf{q} - \Theta(\mathbf{x})\right]. \qquad (3)$$

The potential Θ satisfies the Monge–Ampère equation (1), written for the first time by Ampère[20] by exploiting the property of the Legendre transformation. Note that within the more restricted framework of the Zel'dovich approximation, Θ differs just by a quadratic additive term from the Eulerian velocity potential.[19]

Quadratic cost function

To show that the quadratic cost minimisation leads to the Monge–Ampère equation, we define the displacement field $\boldsymbol{\xi}(\mathbf{x}) \equiv \mathbf{x} - \mathbf{q}(\mathbf{x})$ and perform a variation $\delta\boldsymbol{\xi}(\mathbf{x})$ to obtain, to lowest order, the variation of the cost function $\delta I = \int_{\mathbf{x}} 2\,\boldsymbol{\xi}(\mathbf{x}) \cdot (\rho(\mathbf{x})\,\delta\boldsymbol{\xi})\,d^3x$. The condition that the Eulerian density remains unchanged, which constrains the variation, is expressed as $\nabla_{\mathbf{x}} \cdot (\rho(\mathbf{x})\delta\boldsymbol{\xi}(\mathbf{x})) = 0$. By a simple Lagrange multiplier argument, this implies that $\boldsymbol{\xi}$ must be a gradient of some function of $\mathbf{x}$; thus, $\mathbf{q} = \mathbf{x} - \boldsymbol{\xi} = \nabla_{\mathbf{x}}\Theta(\mathbf{x})$. Furthermore, should Θ be non-convex and thus lead to multi-streaming, this would prevent the Lagrangian map from being optimal.

History of mass transportation

Monge[24] posed the following problem: how to optimally move material from one place to another, knowing only its initial and final spatial distributions, the cost being a prescribed function of the distance travelled by 'molecules' of material (a linear function in Monge's original work). Kantorovich[25] showed that Monge's query was an instance of the linear programming problem and developed for it a theory which found numerous applications in economics and applied mathematics.

1. Narayanan, V.K. & Croft, R.A. Recovering the primordial density fluctuations: a comparison of methods. *Astrophys. J.* **515**, 471–486 (1999).
2. Peebles, P.J.E. Tracing galaxy orbits back in time. *Astrophys. J.* **344**, L53–L56 (1989).
3. Weinberg, D.H. Reconstructing primordial density fluctuations-I. Method. *Mon. Not. R. Astron. Soc.* **254**, 315–342 (1992).
4. Nusser, A. & Dekel, A. Tracing large-scale fluctuations back in time. *Astrophys. J.* **391**, 443–452 (1992).
5. Croft, R.A. & Gaztañaga, E. Reconstruction of cosmological density and velocity fields in the Lagrangian Zel'dovich approximation. *Mon. Not. R. Astron. Soc.* **285**, 793–805 (1997).
6. Nusser, A. & Branchini, E. On the least action principle in cosmology. *Mon. Not. R. Astron. Soc.* **313**, 587–595 (2000).
7. Goldberg, D.M. & Spergel, D.N. Using perturbative least action to recover cosmological initial conditions. *Astrophys. J.* **544**, 21–29 (2000).
8. Valentine, H., Saunders, W. & Taylor, A. Reconstructing the IRAS point source catalog redshift survey with a generalized PIZA. *Mon. Not. R. Astron. Soc.* **319**, L13–L17 (2000).
9. Branchini, E., Eldar, A., Nusser, A. Peculiar velocity reconstruction with fast action method: tests on mock redshift surveys. *Mon. Not. R. Astron. Soc.*, submitted. Preprint astro-ph/0110618 (2001).
10. Frieman, J.A. & Szalay, A.S. Large-scale structure: entering the precision era. *Phys. Reports* **333–334**, 215–232 (2000).
11. Bertschinger, E. & Dekel, A., Recovering the full velocity and density fields from large-scale redshift-distance samples. *Astrophys. J.* **336**, L5–L8 (1989).
12. Zel'dovich, Ya.B. Gravitational instability: an approximate theory for large density perturbations. *Astron.& Astrophys.* **5**, 84–89 (1970).
13. Moutarde, F., Alimi, J.-M., Bouchet, F.R., Pellat, R. & Ramani, A. Precollapse scale invariance in gravitational instability. *Astrophys. J.* **382**, 377–381 (1991).
14. Buchert, T. Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies and the Zel'dovich approximation. *Mon. Not. R. Astron. Soc.* **254**, 729–737 (1992).
15. Munshi, D., Sahni, V. & Starobinsky, A. Nonlinear approximations to gravitational instability: a comparison in the quasi-linear regime. *Astrophys. J.* **436**, 517–527 (1994).
16. Catelan, P. Lagrangian dynamics in non-flat universes and nonlinear gravitational evolution. *Mon. Not. R. Astron. Soc.* **276**, 115–124 (1995).
17. Gurbatov, S. & Saichev, A.I. Probability distribution and spectra of potential hydrodynamic turbulence. *Radiophys. Quant. Electr.* **27**, 303–313 (1984).
18. Shandarin, S.F. & Zel'dovich, Ya.B. The large-scale structure of the universe: turbulence, intermittency, structures in a self-gravitating medium *Rev. Mod. Phys.* **61**, 185–220 (1989).
19. Vergassola, M., Dubrulle, B., Frisch, U. & Noullez, A. Burgers' equation, Devil's staircases and the mass distribution for large-scale structures *Astron.& Astrophys.* **289**, 325–356 (1994).
20. Ampère, A.-M. Mémoire concernant ... l'intégration des équations aux différentielles partielles du premier et du second ordre. *Journal de L'École Royale Polytechnique* **11**, 1–188 (1820).
21. Brenier, Y. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris* **305**, 805–808 (1987).
22. Gangbo, W. & McCann, R.J. The geometry of optimal transportation. *Acta Mathematica* **177**, 113–161 (1996).
23. Benamou, J.-D. & Brenier, Y. The optimal time-continuous mass transport problem and its augmented Lagrangian numerical resolution. *Numer. Math.*, **84**, 375–393 (2000). (www.inria.fr/rrrt/rr-3356.html).
24. Monge, G. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Science de Paris*, 666–704 (1781).
25. Kantorovich, L. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)* **37**, 199–201 (1942).
26. Rockafellar, R.T. *Convex Analysis* (Princeton University Press, 1970).
27. Hénon, M. A mechanical model for the transportation problem. *C. R. Acad. Sci. Paris.* **321**, 741–745 (1995). A more detailed version, including optimisation algorithms, is available at www.obs-nice.fr/etc7/henon.pdf
28. Balinski, M.L. A competitive (dual) simplex method for the assignment problem. *Mathematical Programming* **34**, 125–141 (1986).
29. Couchman, H.M.P., Thomas, P.A. & Pearce, F.R. Hydra: an Adaptive-Mesh Implementation of P^3M-SPH. *Astrophys. J.* **452**, 797–813 (1995).
30. Arnold, V.I. *Mathematical Methods of Classical Mechanics* (Springer, Berlin, 1978).

Acknowledgements

Special thanks are due to E. Branchini (observational and conceptual aspects), to Y. Brenier (mathematical aspects) and to M. Hénon (algorithmic aspects and the handling of spatial periodicity and of scatter plots). We also thank J. Bec, H. Frisch, B. Gladman, L. Moscardini, A. Noullez, C. Porciani, M. Rees, E. Spiegel, A. Starobinsky and P. Thomas for comments. This work was supported by the BQR program of Observatoire de la Côte d'Azur, by the TMR program of the European Union (UF, RM), by MIUR (SM), by the French Ministry of Education, the McDonnel Foundation, the Russian RFBR and INTAS (AS).

Formula del cambio di variabile

Siano $(X, \mathcal{A})$ e $(Y, \mathcal{B})$ due spazi misurabili, ovvero $\mathcal{A} \subset \mathcal{P}(X)$ e $\mathcal{B} \subset \mathcal{P}(Y)$ sono σ-algebre. Una funzione $T : X \to Y$ si dice misurabile se $T^{-1}(B) \in \mathcal{A}$ per ogni $B \in \mathcal{B}$.

Sia ora $(X, \mathcal{A}, \mu)$ uno spazio di misura e sia $T : X \to Y$ misurabile. La misura push-forward $T_{\#}\mu$ su $(Y, \mathcal{B})$ si definisce ponendo

$$T_{\#}\mu(B) = \mu(T^{-1}(B)) \tag{1}$$

per ogni $B \in \mathcal{B}$. È facile verificare che $T_{\#}\mu$ è effettivamente una misura σ-additiva su $\mathcal{B}$.

Teorema Nella situazione precedente, sia $f \in L^1(Y; T_{\#}\mu)$. Allora $f \circ T \in L^1(X; \mu)$ e inoltre vale la formula del cambio di variabile

$$\int_Y f(y)\, d T_{\#}\mu(y) = \int_X f(T(x))\, d\mu(x). \tag{2}$$

Formalmente si pone $y = T(x)$ e si trasforma $d T_{\#}\mu(y) = d\mu(x)$.

Dim. Quando $f = \chi_B$ con $B \in \mathcal{B}$, la formula diventa

$$\begin{aligned} T_{\#}\mu(B) &= \mu(\{x \in X : T(x) \in B\}) \\ &= \mu(T^{-1}(B)) \end{aligned} \tag{3}$$

che è la (1).

Quando $f \geq 0$ è misurabile della forma

$$f(y) = \sum_{n=1}^{\infty} \frac{1}{n} \chi_{B_n}(y) \tag{4}$$

con $B_n \in \mathcal{B}$, la formula (2) segue da (3) e dal Teorema della convergenza monotona.

Quando $f \in L^1(Y)$ si decompone $f = f^+ - f^-$ con $f^+, f^- \geq 0$ misurabili (integrabili) e si applica il ragionamento precedente separatamente ad f^+, f^-.

□

Quando $X = Y = \mathbb{R}^n$ la formula (2) ha una versione differenziabile (Analisi 2).

Sia $A \subset \mathbb{R}^n$ un insieme aperto e sia $T \in C^1(A; T(A))$ un diffeomorfismo di classe C^1. Sia $\mu = \mathcal{L}^n \llcorner A$ la misura di Lebesgue su A. Vogliamo calcolare la misura $T_\# \mu$ su $T(A) \subset \mathbb{R}^n$. Sappiamo che per $B \subset T(A)$ aperto ed $f \in L^1(A; \mu)$

$$\int_{T^{-1}(B)} f(x)\, dx = \int_B f(T^{-1}(y)) \, |\det \mathrm{J} T^{-1}(y)| \, dy \tag{5}$$

e con $f = \chi_{T^{-1}(B)}$ si ottiene

$$T_\# \mu(B) = \mu(T^{-1}(B)) = \int_B |\det \mathrm{J} T^{-1}(y)| \, dy$$

e quindi $T_\#(\mathcal{L}^n \llcorner A) = |\det \mathrm{J} T^{-1}| \, \mathcal{L}^n \llcorner T(A)$.

(6)

Dalla formula diretta con $B(x_0,r) \subset A$ palla

$$(7) \qquad \mathcal{L}^n(T(B(x_0,r))) = \int_{B(x_0,r)} |\det \mathcal{J}T(x)|\, dx$$

Dividendo per $\mathcal{L}^n(B(x_0,r))$ e con il limite per $r \to 0^+$ si scopre che

$$(8) \qquad |\det \mathcal{J}T(x)| = \lim_{r\to 0^+} \frac{\mathcal{L}^n(TB(x,r))}{\mathcal{L}^n(B(x,r))}, \quad x \in A.$$

Quando T è (un diffeomorfismo) di classe C^1 il limite esiste e la formula (8) è un teorema.

Commento Può succedere che il limite in (8) esista per qualche (ogni) $x \in A$ senza che T sia di classe C^1 (ad esempio, T è solo un omeomorfismo). La (8) diventa la definizione "debole" di $|\det \mathcal{J}T(x)|$. Questo è interessante perché stiamo definendo in modo debole un oggetto nonlineare.

Esercizio 1 Quando $X = \mathbb{R}^n$, $\mathcal{A} = \mathcal{M} = \{$ boreliani di $\mathbb{R}^n\}$ ed $Y = \mathbb{R}$, raccordare la definizione di $T : \mathbb{R}^n \to \mathbb{R}$ misurabile data sopra con quella appresa in Analisi 2

Esercizio 2 X spazio metrico, μ misura di Borel finita, $B \subset X$ Boreliano.

1) $\mu(B) = \sup\{\mu(C) : C \subset B,\ C$ chiuso$\}$, Vero | Falso?

2) $\mu(B) = \sup\{\mu(K) : K \subset B,\ K$ compatto$\}$, Vero | Falso?

Problema di Monge

Siano X, Y due spazi topologici e siano μ misura di Borel su X e ν misura di Borel su Y tali che

$$\mu(X) = 1 = \nu(Y)$$

"misure unitarie", "misure di probabilità".

Introduciamo l'insieme delle mappe di trasporto

$$T(\mu,\nu) = \{ T : X \to Y \mid T \text{ è di Borel e } T_{\#}\mu = \nu \}.$$

Osservazione Se $\mu = \delta_x$, delta di Dirac in $x \in X$ fissato e $\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$ con $y_1, y_2 \in Y$ e $y_1 \neq y_2$ allora

$$T(\mu,\nu) = \emptyset.$$

Fissiamo una funzione costo $c : X \times Y \longrightarrow [0,\infty)$ continua. Se $T : X \to Y$ è di Borel allora $\bar{T} := Id \times T : X \to X \times Y$ è di Borel. Dunque se $A \subset \mathbb{R}$ è aperto

$$(c \circ \bar{T})^{-1}(A) = \underbrace{\bar{T}^{-1}\big(\underbrace{c^{-1}(A)}_{\text{aperto}}\big)}_{\text{boreliano}},$$

ovvero $x \mapsto c(x, T(x))$ è Boreliana.

Definiamo il funzionale costo $C : T(\mu,\nu) \to [0,\infty]$

$$C(T) = \int_X c(x, T(x))\, d\mu ,$$

e consideriamo il Problema di Monge

$$\min \{ C(T) : T \in T(\mu,\nu) \}.$$

Le domande sono:

1) Esistenza di soluzioni. Non c'è in generale.
2) Unicità di soluzioni. Non c'è in generale.
3) Struttura delle soluzioni.
4) Regolarità delle soluzioni.

La Teoria arriva in "profondità" in questi casi:

1) $X = Y = \mathbb{R}^n$ e $c(x,y) = \phi(|x-y|)$ con ϕ convessa.
Il caso migliore è $c(x,y) = \frac{1}{2}|x-y|^2$.
Si deve richiedere $\mu \ll \mathcal{L}^n$

2) $X = Y = M$ Varietà Riemanniana con $c(x,y) = \phi(d(x,y))$, ϕ convessa e $d(x,y)$ = distanza Riemanniana. Se M è completa: più facile.

Esempio di non unicità

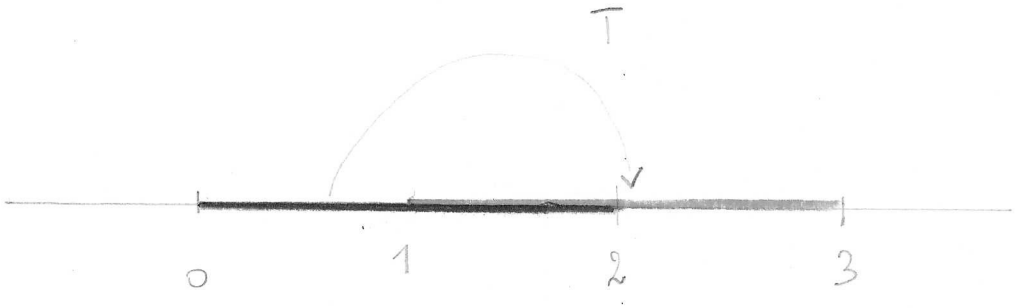
Su $\mathbb{R}$ consideriamo le due misure

$$\mu = \frac{1}{2}\,\mathcal{L}^1 \llcorner [0,2] \qquad e \qquad \nu = \frac{1}{2}\,\mathcal{L}^1 \llcorner [1,3]$$

e il corrispondente problema di Monge

$$\min \left\{ \int_{\mathbb{R}} |x - T(x)|\, d\mu \;:\; T \in \mathcal{T}(\mu,\nu) \right\}$$

Stiamo considerando la funzione costo = distanza $c(x,y) = |x-y|$.



Osserviamo che per ogni $T \in \mathcal{T}(\mu,\nu)$ si ha:

$$\int_{\mathbb{R}} |x - T(x)|\, d\mu \geq \int_{\mathbb{R}} (T(x) - x)\, d\mu =$$

$$= \int_{\mathbb{R}} T(x)\, d\mu - \int_{\mathbb{R}} x\, d\mu$$

$$= \int_{\mathbb{R}} y\, d\nu - \int_{\mathbb{R}} x\, d\mu$$

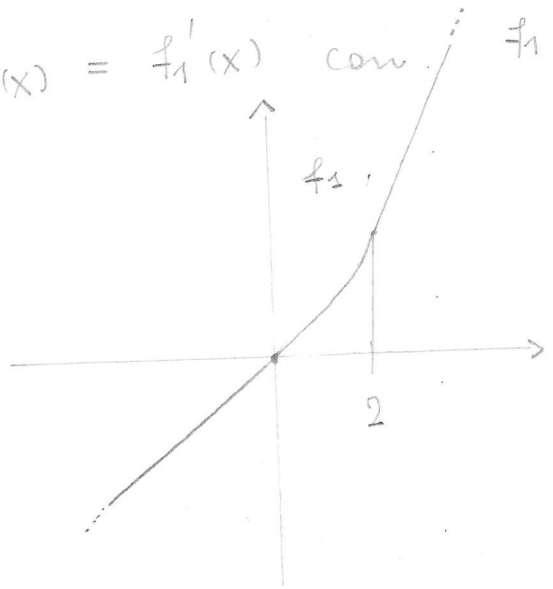
$$= \frac{1}{2}\int_1^3 y\, dy - \frac{1}{2}\int_0^2 x\, dx$$

$$= \frac{1}{4}\left[y^2\right]_1^3 - \frac{1}{4}\left[x^2\right]_0^2 = \frac{1}{4}(9 - 1 - 4) = 1.$$

Un minimo dovrebbe essere dato dal rearrangiamento monotono $T_1 : \mathbb{R} \to \mathbb{R}$

$$T_1(x) = \begin{cases} 1 & \text{se } x \leq 0 \\ 1+x & \text{se } x \in [0,2] \\ 3 & \text{se } x \geq 2 . \end{cases}$$

Abbiamo $T_1(x) = f_1'(x)$ con $f_1 : \mathbb{R} \to \mathbb{R}$ convessa.



Calcoliamo

$$\int_{\mathbb{R}} |X - T_1(x)| \, d\mu = \frac{1}{2} \int_0^2 1 \, dx = 1$$

e dunque T_1 è un minimo.

Consideriamo ora la funzione $T_2 : \mathbb{R} \to \mathbb{R}$ che lascia fisso $[1,2]$ e sposta $(0,1)$ in $(2,3)$

$$T_2(x) = \begin{cases} x+2 & \text{per } x \in (0,1) \\ x & \text{per } x \in [1,2] \end{cases}$$

Su $\mathbb{R} \setminus (0,2]$ definiamo T_2 a piacere.

Calcoliamo

$$\int_{\mathbb{R}} |x - T_2(x)| \, d\mu = \frac{1}{2} \int_0^1 |x - (x+2)| \, dx + \frac{1}{2} \int_1^2 |x - x| \, dx$$

$$= 1 .$$

Dunque anche T_2 è un minimo. Su $(0,2)$ abbiamo $T_2(x) = f_2'(x)$ con

$$f_2(x) = \begin{cases} \frac{x^2}{2} + 2x & \text{se } x \in (0,1) \\ \frac{x^2}{2} + 2 & \text{se } x \in (1,2] \end{cases}$$

non derivabile per $x = 1$, non conviene.

Le mappe minime lasciano fina = la parte comune = delle misure quando il conto non è (strettamente) convesso.

□

Esercizio Siano μ, ν come sopra. Per $p > 0$ studiare il problema di minimo

$$\min \left\{ \int_{\mathbb{R}} |x - Tx|^p \, d\mu : T \in \mathcal{T}(\mu, \nu) \right\}.$$

Idea: per $p > 1$ vince T_1, per $0 < p < 1$ non vince T_1.

Problema di Kantorovic

Siano X, Y due spazi topologici con misure di probabilità μ e ν.

Consideriamo il prodotto cartesiano $X \times Y$ e le due proiezioni sui fattori

$$pr^1 : X \times Y \to X, \quad pr^1(x,y) = x,$$
$$pr^2 : X \times Y \to Y, \quad pr^2(x,y) = y.$$

Sono continue e dunque di Borel.

Introduciamo l'insieme dei piani di trasporto

$$\Pi(\mu,\nu) = \{ \pi \mid \pi \text{ misura di Borel (unitarie) su } X\times Y \text{ tale che } pr^1_\# \pi = \mu \text{ e } pr^2_\# \pi = \nu \}$$

$\neq \emptyset$ sempre

Sia $c : X \times Y \to [0,\infty)$ un "costo" continuo, consideriamo il funzionale di costo $I : \Pi(\mu,\nu) \to [0,\infty]$

$$I(\pi) = \int_{X\times Y} c(x,y)\, d\pi,$$

e il problema di Kantorovic

$$\min \{ I(\pi) \,;\, \pi \in (\mu,\nu) \}.$$

Le domande sono:

1) Provare l'esistenza di soluzioni. C'è, molto in generale.
2) Trovare condizioni necessarie e/o sufficienti che "descrivono" i piani di trasporto ottimali.

L'esistenza si riesce a provare quando X e Y sono spazi topologici "polacchi".

Definizione Uno spazio topologico è "Polish" se è metrizzabile in modo completo e separabile.

Noi vedremo il caso $X = Y = \mathbb{R}^n$.

Le condizioni necessarie/sufficienti che si trovano sono chiare solo in casi speciali.

Osservazione L'insieme

$$T(\mu,\nu) \subset \{ T : X \to Y \mid T \text{ di Borel} \}$$

è privo di una qualsiasi struttura lineare.

Ad esempio, $T \mapsto |\det \nabla T|$ con $T : \mathbb{R}^n \to \mathbb{R}^n$ diffeomorfismo è nonlineare.

Al contrario, l'insieme

$$\Pi(\mu,\nu) \subset \{ \pi \mid \text{misure di Borel su } X \times Y \text{ unitarie} \}$$

è convesso.

Mappe di trasporto e piani di trasporto

Siano μ, ν due misure di Borel in $\mathbb{R}^n$, $\mu(\mathbb{R}^n) = \nu(\mathbb{R}^n) = 1$. Sia poi $T \in T(\mu,\nu)$ una mappa di Borel tale che $T_\# \mu = \nu$. Sul prodotto $\mathbb{R}^n \times \mathbb{R}^n$ possiamo definire la misura di Borel

$$\pi = (Id \times T)_\# \mu .$$

Controlliamo che $pr^1_\# \pi = \mu$ e $pr^2_\# \pi = \nu$.

1) Dato $B \subset \mathbb{R}^n$ di Borel

$$\begin{aligned} pr^1_\# \pi (B) &= \pi (B \times \mathbb{R}^n) \\ &= (Id \times T)_\# \mu (B \times \mathbb{R}^n) \\ &= \mu(\{x \in \mathbb{R}^n : x \in B \text{ e } T(x) \in \mathbb{R}^n\}) \\ &= \mu(B). \end{aligned}$$

2) $$\begin{aligned} pr^2_\# \pi (B) &= \pi (\mathbb{R}^n \times B) \\ &= (Id \times T)_\# \mu (\mathbb{R}^n \times B) \\ &= \mu(\{x \in \mathbb{R}^n : T(x) \in B\}) \\ &= T_\# \mu (B) \\ &= \nu(B). \end{aligned}$$

Ora consideriamo una misura di Borel π (unitaria) su $\mathbb{R}^n \times \mathbb{R}^n$. Supponiamo che esista $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ con queste proprietà:

i) $\Gamma = \mathrm{gr}(T)$ con $T : \mathbb{R}^n \to \mathbb{R}^n$ (a priori no regolarità)

ii) Γ è π-misurabile (Borel)

iii) $\pi(\mathbb{R}^n \times \mathbb{R}^n \setminus \Gamma) = 0$, ovvero π è concentrata su Γ.

Vogliamo provare che:

1) T è "essenzialmente" di Borel

2) $\pi = (\mathrm{Id} \times T)_{\#} \mu$ con $\mu = \mathrm{pr}^1_{\#} \pi$

3) Detta $\nu = \mathrm{pr}^2_{\#} \pi$, si ha $T_{\#} \mu = \nu$.

Regolarità di μ in $\mathbb{R}^n$

Esiste $\Gamma_1 \subset \Gamma$ di Borel tale $\mu(\Gamma \setminus \Gamma_1) = 0$.

Esiste $K_1 \subset \Gamma_1$ compatto tale che $\mu(\Gamma_1 \setminus K_1) < \frac{1}{2}$

Detto $H_1 = \mathrm{pr}^1 K_1$ avremo $T : H_1 \to \mathbb{R}^n$ con grafico $K_1 = \mathrm{Id} \times T(H_1) \subset \mathbb{R}^n \times \mathbb{R}^n$ compatto. Quindi T è continua da H_1 in $\mathbb{R}^n$ e quindi di Borel.

Per induzione, dato $\Gamma_m = \Gamma_{m-1} \setminus K_{m-1}$ esiste un compatto $K_m \subset \Gamma_m$ tale che $\mu(\Gamma_m \setminus K_m) < \frac{1}{2^m}$ e $T : H_m = \mathrm{pr}^1 K_m \to \mathbb{R}^n$ è continua.

Avremo $\mu\left(\Gamma \setminus \bigcup_{m=1}^{\infty} K_m\right) = 0$ e $T : \bigcup_{m=1}^{\infty} H_m \to \mathbb{R}^n$ continua ristretta ad ogni H_m. Quindi T è di Borel. Questo prova 1). (Fuori da $\bigcup_{m=1}^{\infty} H_m$, T definita contando).

Verifichiamo 2). Sia $B \subset \mathbb{R}^n \times \mathbb{R}^n$ di Borel. Allora:

$$
\begin{aligned}
(\mathrm{Id} \times T)_{\#} \mu(B) &= \mu(\{x \in \mathbb{R}^n : (x, Tx) \in B\}) \\
&= \mathrm{pr}^1_{\#} \pi(\{x \in \mathbb{R}^n : (x, Tx) \in B\}) \\
&= \pi(\{x \in \mathbb{R}^n : (x, Tx) \in B\} \times \mathbb{R}^n) \\
&\overset{\text{usiamo iii)}}{=} \pi(\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : (x, Tx) \in B,\ y = Tx\}) \\
&= \pi(B).
\end{aligned}
$$

Verifichiamo 3). Sia $B \subset \mathbb{R}^n$ di Borel. Allora

$$\begin{aligned} T_{\#}\mu(B) &= \mu(\{x \in \mathbb{R}^n : T(x) \in B\}) \\ &= pr^1_{\#}\pi(\{x \in \mathbb{R}^n : T(x) \in B\}) \\ &= \pi(\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : T(x) \in B\}) \\ &\quad \text{Usiamo iii)} \\ &= \pi(\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : y = T(x) \in B\}) \\ &= \pi(\mathbb{R}^n \times B) \\ &= pr^2_{\#}\pi(B) = \nu(B). \end{aligned}$$

□

<u>Teorema</u> Siano μ e ν due misure di Borel unitarie su $\mathbb{R}^n$.

Siano poi:

$$\inf_M = \inf\left\{ \int_{\mathbb{R}^n} c(x, Tx)\, d\mu \;:\; T \in \mathcal{T}(\mu,\nu) \right\}$$

e

$$\inf_K = \inf\left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y)\, d\pi \;:\; \pi \in \Pi(\mu,\nu) \right\},$$

dove $c : \mathbb{R}^n \times \mathbb{R}^n \xrightarrow{\text{cont.}} [0,\infty)$ è una funzione costo.

Allora $\inf_K \leq \inf_M$ e se $\inf_K$ è un minimo realizzato da una $\pi \in \Pi(\mu,\nu)$ concentrata su $gr(T)$ per una $T : \mathbb{R}^n \to \mathbb{R}^n$, allora $\inf_M = \inf_K$ e sono minimi.

Dim. Sia $T \in \mathcal{T}(\mu,\nu)$ e definiamo $\pi = (Id \times T)_{\#}\mu$. Allora $\pi \in \Pi(\mu,\nu)$ e per la formula di cambio di variabile:

$$\int_{\mathbb{R}^n \times \mathbb{R}^m} c(x,y)\, d\pi = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x,y)\, d(Id \times T)_{\#}\mu$$

$$= \int_{\mathbb{R}^n} c(x, T(x))\, d\mu .$$

Questo prova che $\inf_K \leq \inf_M$.

Se π realizza il minimo di Kantorovic e π è concentrata su $gr(T)$, allora $\pi = (Id \times T)_{\#}\mu$ e $T \in \mathcal{T}(\mu,\nu)$. Il claim segue.

□