Optimal transportation problem

Alexander Kolesnikov

Higher School of Economics

Bonn, Hausdorff Research Institute, 2013

Discrete transportation problem

Formulation of the problem

There are n plants $P_1, \cdots, P_n$ producing certain commodity and m markets $M_1, \cdots, M_m$. Plant P_i posesses an amount s_i of this commodity and market M_j must receive the amount r_j of this commodity. The transportation of one unit of the commodity from P_i to M_j costs c_{ij}. How to transport the required quantity of the commodity at the lowest cost?

Linear programming formulation

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{ij} c_{ij} \to \min$$

under the constraints

$$\pi_{ij} \geq 0, \ \sum_{j} \pi_{ij} \leq s_i, \sum_{i} \pi_{ij} \geq r_j.$$

Discrete transportation problem

Formulation of the problem

There are n plants $P_1, \cdots, P_n$ producing certain commodity and m markets $M_1, \cdots, M_m$. Plant P_i posesses an amount s_i of this commodity and market M_j must receive the amount r_j of this commodity. The transportation of one unit of the commodity from P_i to M_j costs c_{ij}. How to transport the required quantity of the commodity at the lowest cost?

Linear programming formulation

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \pi_{ij} c_{ij} \to \min$$

under the constraints

$$\pi_{ij} \geq 0, \ \sum_{j} \pi_{ij} \leq s_i, \sum_{i} \pi_{ij} \geq r_j.$$

Balance condition

Our constraints imply the **balance condition**

$$\sum_i s_i \geq \sum_j r_j.$$

The problem is feasible if and only if this condition is satisfied. The feasible problem can be always reduced to the case

$$\sum_i s_i = \sum_j r_j.$$

To this end we introduce a **dummy market** M_d with the property $c_{id} = 0$ and $r_d = \sum_i s_i - \sum_j r_j$. Clearly, in this case

$$\sum_j \pi_{ij} = s_i, \ \sum_i \pi_{ij} = r_j$$

General linear programming duality (L. Kantorovich)

Primal problem

$$\langle c, x\rangle \to \max, \quad Ax \le b,\ x \ge 0,$$

where $c, x \in \mathbb{R}^k$, $b \in \mathbb{R}^l$, $A : \mathbb{R}^k \to \mathbb{R}^l$ is linear.

Dual problem

$$\langle b, \lambda\rangle \to \min, \quad A^T\lambda \ge c,\ \lambda \ge 0,$$

where $\lambda \in \mathbb{R}^l$.

Theorem

1. *$\langle c, x\rangle \le \langle b, \lambda\rangle$ for all feasible x, λ.*
2. *$\langle c, \hat{x}\rangle = \langle b, \hat{\lambda}\rangle$ if and only if $\hat{x}, \hat{\lambda}$ are solutions to these problems.*

General linear programming duality (L. Kantorovich)

Primal problem

$$\langle c, x\rangle \to \max, \quad Ax \le b,\ x \ge 0,$$

where $c, x \in \mathbb{R}^k$, $b \in \mathbb{R}^l$, $A : \mathbb{R}^k \to \mathbb{R}^l$ is linear.

Dual problem

$$\langle b, \lambda\rangle \to \min, \quad A^T\lambda \ge c,\ \lambda \ge 0,$$

where $\lambda \in \mathbb{R}^l$.

Theorem

1. *$\langle c, x\rangle \le \langle b, \lambda\rangle$ for all feasible x, λ.*
2. *$\langle c, \hat{x}\rangle = \langle b, \hat{\lambda}\rangle$ if and only if $\hat{x}, \hat{\lambda}$ are solutions to these problems.*

General linear programming duality (L. Kantorovich)

Primal problem

$$\langle c, x \rangle \to \max, \quad Ax \le b, \ x \ge 0,$$

where $c, x \in \mathbb{R}^k$, $b \in \mathbb{R}^l$, $A : \mathbb{R}^k \to \mathbb{R}^l$ is linear.

Dual problem

$$\langle b, \lambda \rangle \to \min, \quad A^T \lambda \ge c, \ \lambda \ge 0,$$

where $\lambda \in \mathbb{R}^l$.

Theorem

1. *$\langle c, x \rangle \le \langle b, \lambda \rangle$ for all feasible x, λ.*
2. *$\langle c, \hat{x} \rangle = \langle b, \hat{\lambda} \rangle$ if and only if $\hat{x}, \hat{\lambda}$ are solutions to these problems.*

General linear programming duality (L. Kantorovich)

Primal problem

$$\langle c, x \rangle \to \max, \quad Ax \le b, \ x \ge 0,$$

where $c, x \in \mathbb{R}^k$, $b \in \mathbb{R}^l$, $A : \mathbb{R}^k \to \mathbb{R}^l$ is linear.

Dual problem

$$\langle b, \lambda \rangle \to \min, \quad A^T \lambda \ge c, \ \lambda \ge 0,$$

where $\lambda \in \mathbb{R}^l$.

Theorem

1. $\langle c, x \rangle \le \langle b, \lambda \rangle$ *for all feasible* x, λ.
2. $\langle c, \hat{x} \rangle = \langle b, \hat{\lambda} \rangle$ *if and only if* $\hat{x}, \hat{\lambda}$ *are solutions to these problems.*

General linear programming duality (L. Kantorovich)

Primal problem

$$\langle c, x\rangle \to \max, \quad Ax \le b,\ x \ge 0,$$

where $c, x \in \mathbb{R}^k$, $b \in \mathbb{R}^l$, $A : \mathbb{R}^k \to \mathbb{R}^l$ is linear.

Dual problem

$$\langle b, \lambda\rangle \to \min, \quad A^T\lambda \ge c,\ \lambda \ge 0,$$

where $\lambda \in \mathbb{R}^l$.

Theorem

1. $\langle c, x\rangle \le \langle b, \lambda\rangle$ *for all feasible* x, λ.
2. $\langle c, \hat{x}\rangle = \langle b, \hat{\lambda}\rangle$ *if and only if* $\hat{x}, \hat{\lambda}$ *are solutions to these problems.*

General linear programming duality (L. Kantorovich)

Primal problem

$$\langle c, x\rangle \to \max, \quad Ax \le b,\ x \ge 0,$$

where $c, x \in \mathbb{R}^k$, $b \in \mathbb{R}^l$, $A : \mathbb{R}^k \to \mathbb{R}^l$ is linear.

Dual problem

$$\langle b, \lambda\rangle \to \min, \quad A^T\lambda \ge c,\ \lambda \ge 0,$$

where $\lambda \in \mathbb{R}^l$.

Theorem

1. $\langle c, x\rangle \le \langle b, \lambda\rangle$ *for all feasible* x, λ.
2. $\langle c, \hat{x}\rangle = \langle b, \hat{\lambda}\rangle$ *if and only if* $\hat{x}, \hat{\lambda}$ *are solutions to these problems.*

About the proof

The first part of this duality statement is easy. The second part is more delicate and can be viewed as a corollary of the saddle point theorem (minmax=maxmin) for convex programming.

Dual transportation problem

$$\sum_i s_i u_i + \sum_j r_j v_j \to \max$$

under constraints

$$u_i + v_j \leq c_{ij}, \quad v_j \geq 0, \; u_i \leq 0.$$

Measures and potentials

It is always helpful to interpret the solution $\{\pi_{ij}\}$ to the primal transportation problem as a **measure** and the solution $\{v_j, u_i\}$ to the dual transportation problem as a couple of **potentials**.

About the proof

The first part of this duality statement is easy. The second part is more delicate and can be viewed as a corollary of the saddle point theorem (minmax=maxmin) for convex programming.

Dual transportation problem

$$\sum_i s_i u_i + \sum_j r_j v_j \to \max$$

under constraints

$$u_i + v_j \leq c_{ij}, \quad v_j \geq 0, \; u_i \leq 0.$$

Measures and potentials

It is always helpful to interpret the solution $\{\pi_{ij}\}$ to the primal transportation problem as a **measure** and the solution $\{v_j, u_i\}$ to the dual transportation problem as a couple of **potentials**.

About the proof

The first part of this duality statement is easy. The second part is more delicate and can be viewed as a corollary of the saddle point theorem (minmax=maxmin) for convex programming.

Dual transportation problem

$$\sum_i s_i u_i + \sum_j r_j v_j \to \max$$

under constraints

$$u_i + v_j \leq c_{ij}, \quad v_j \geq 0, \; u_i \leq 0.$$

Measures and potentials

It is always helpful to interpret the solution $\{\pi_{ij}\}$ to the primal transportation problem as a **measure** and the solution $\{v_j, u_i\}$ to the dual transportation problem as a couple of **potentials**.

Duality

In what follows we always assume that $\sum_i s_i = \sum_j r_j$.
Example is taken from **[Ferg]**. Find solution to the following transportation problem

	2	**9**	**4**	**5**
5	4	7	11	3
7	2	5	6	4
8	1	3	4	8

We claim that the following transportation plan is optimal

	2	**9**	**4**	**5**
5	0	0	0	5
7	2	1	4	0
8	0	8	0	0

Proof.

Find the dual problem. Verify that

$$v = (2, 5, 6, 4), \quad u = (-1, 0, -2)$$

are feasible for the dual problem and both problems have the same value

$$\sum_{i,j} \pi_{ij} c_{ij} = 2 * 2 + 1 * 5 + 8 * 3 + 4 * 6 + 5 * 3 = 72$$

$$\sum_i s_i u_i + \sum_j r_j v_j = -1 * 5 - 2 * 8 + 2 * 2 + 5 * 9 + 6 * 4 + 4 * 5 = 72.$$

□

Cyclical monotonicity

Let us denote $I = \{1, \cdots, n\}$, $J = \{1, \cdots, m\}$.

Definition

Given a cost function

$$c : I \times J \to \mathbb{R}, \ c(i, j) = c_{ij},$$

a set $S \subset I \times J$ is called called cyclically monotone if for any $\{(i_1, j_1), \cdots, (i_k, j_k)\} \subset S$ one has

$$c(i_1, j_1) + c(i_2, j_2) + \cdots + c(i_k, j_k) \\ \leq c(i_1, j_2) + c(i_2, j_3) + \cdots + c(i_k, j_{k+1}),$$

where $j_{k+1} = j_1$.

Cyclical monotonicity

Let us denote $I = \{1, \cdots, n\}$, $J = \{1, \cdots, m\}$.

Definition

Given a cost function

$$c : I \times J \to \mathbb{R}, \; c(i,j) = c_{ij},$$

a set $S \subset I \times J$ is called called cyclically monotone if for any $\{(i_1, j_1), \cdots, (i_k, j_k)\} \subset S$ one has

$$\begin{aligned} c(i_1, j_1) + c(i_2, j_2) + \cdots + c(i_k, j_k) \\ \leq c(i_1, j_2) + c(i_2, j_3) + \cdots + c(i_k, j_{k+1}), \end{aligned}$$

where $j_{k+1} = j_1$.

Cyclical monotonicity of solutions

Support of any solution is cyclically monotone

Theorem

Assume that $\{\pi_{ij}\}$ is a solution to the optimal transportation problem. Then the set $S = \{\pi_{ij} \neq 0\} \subset A$ is cyclically monotone.

Proof: Assume that S contains a non cyclically monotone sequence $\{(i_1, j_1), \cdots, (i_k, j_k)\}$. Set

$$\varepsilon = \min_k \pi_{i_k j_k}.$$

Construct a new transportation plan

$$\pi'_{ij} = \pi_{ij} - \sum_{l=1}^{k} \varepsilon \delta_{ii_l} \delta_{jj_l} + \sum_{l=1}^{k} \varepsilon \delta_{ii_l} \delta_{jj_{l+1}}.$$

Cyclical monotonicity of solutions

Support of any solution is cyclically monotone

Theorem

Assume that $\{\pi_{ij}\}$ is a solution to the optimal transportation problem. Then the set $S = \{\pi_{ij} \neq 0\} \subset A$ is cyclically monotone.

Proof: Assume that S contains a non cyclically monotone sequence $\{(i_1, j_1), \cdots, (i_k, j_k)\}$. Set

$$\varepsilon = \min_k \pi_{i_k j_k}.$$

Construct a new transportation plan

$$\pi'_{ij} = \pi_{ij} - \sum_{l=1}^{k} \varepsilon \delta_{i i_l} \delta_{j j_l} + \sum_{l=1}^{k} \varepsilon \delta_{i i_l} \delta_{j j_{l+1}}.$$

Cyclical monotonicity of solutions

Support of any solution is cyclically monotone

Theorem

Assume that $\{\pi_{ij}\}$ *is a solution to the optimal transportation problem. Then the set* $S = \{\pi_{ij} \neq 0\} \subset A$ *is cyclically monotone.*

Proof: Assume that S contains a non cyclically monotone sequence $\{(i_1, j_1), \cdots, (i_k, j_k)\}$. Set

$$\varepsilon = \min_k \pi_{i_k j_k}.$$

Construct a new transportation plan

$$\pi'_{ij} = \pi_{ij} - \sum_{l=1}^{k} \varepsilon \delta_{ii_l} \delta_{jj_l} + \sum_{l=1}^{k} \varepsilon \delta_{ii_l} \delta_{jj_{l+1}}.$$

This means that we remove a mass which gives a value ε to any atom (i_l, j_l) and add to the plan another mass which gives a value ε to any atom (i_l, j_{l+1}). Clearly, 1) the new transportation plan π'_{ij} is feasible, 2) the new transportation plan is better than the previous one, because

$$\sum_{i,j} \pi'_{ij} c_{ij} - \sum_{i,j} \pi_{ij} c_{ij} = -\sum_{l=1}^{k} c_{i_l j_l} + \sum_{l=1}^{k} c_{i_l j_{l+1}} < 0.$$

The last inequality follows from the definition of cyclically monotone sets. We get a contradiction.

Remark

The cyclical monotonicity is equivalent to the following property:

$$\sum_k c(i_k, j_k) \leq \sum_k c(i_k, \sigma(j_k)),$$

where σ is any permutation. This follows from the fact that any finite permutation can be decomposed in cycles.

Rockafellar's theorem

Definition

A set $S = \{(x_i, y_i)\} \subset I \times J$ is included in the c-superdifferential of a function $u : I \to \mathbb{R}$ iff for every $x \in I$

$$u(x) \leq u(x_i) + c(x, y_i) - c(x_i, y_i)$$

Theorem

A set S is c-cyclically monotone if and only if it is included in the c-subdifferential of a function u

Proof. Assume that S is included in the c-subdifferential of u. Then for every sequence $\{(i_1, j_1), \cdots, (i_k, j_k)\}$ one has

$$u(i_{l+1}) \leq u(i_l) + c(i_{l+1}, j_l) - c(i_l, j_l).$$

Taking the sum over l we get the desired inequality.

Rockafellar's theorem

Definition

A set $S = \{(x_i, y_i)\} \subset I \times J$ is included in the c-superdifferential of a function $u : I \to \mathbb{R}$ iff for every $x \in I$

$$u(x) \leq u(x_i) + c(x, y_i) - c(x_i, y_i)$$

Theorem

A set S is c-cyclically monotone if and only if it is included in the c-subdifferential of a function u

Proof. Assume that S is included in the c-subdifferential of u. Then for every sequence $\{(i_1, j_1), \cdots, (i_k, j_k)\}$ one has

$$u(i_{l+1}) \leq u(i_l) + c(i_{l+1}, j_l) - c(i_l, j_l).$$

Taking the sum over l we get the desired inequality.

Rockafellar's theorem

Definition

A set $S = \{(x_i, y_i)\} \subset I \times J$ is included in the c-superdifferential of a function $u : I \to \mathbb{R}$ iff for every $x \in I$

$$u(x) \leq u(x_i) + c(x, y_i) - c(x_i, y_i)$$

Theorem

A set S is c-cyclically monotone if and only if it is included in the c-subdifferential of a function u

Proof. Assume that S is included in the c-subdifferential of u. Then for every sequence $\{(i_1, j_1), \cdots, (i_k, j_k)\}$ one has

$$u(i_{l+1}) \leq u(i_l) + c(i_{l+1}, j_l) - c(i_l, j_l).$$

Taking the sum over l we get the desired inequality.

Assume that S is cyclically monotone. Fix any $(x_0, y_0) \in S$ and set

$$u(x) = \min_{\Gamma}\Big[c(x, y_r) + \sum_{k=1}^{r} c(x_k, y_{k-1}) - \sum_{k=1}^{r} c(x_k, y_k) - c(x_0, y_0)\Big],$$

where the minimum is taken over all the sequences

$$\Gamma = \{(x_1, y_1), \cdots (x_r, y_r)\},\ (x_i, y_i) \in S,\ r \in \mathbb{N}$$

with the agreement that $y_r = y_0$.
Let us note that this minimum exists. Indeed, this follows from the cyclical monotonicity: $\sum_{k=1}^{r} c(x_k, y_{k+1}) - \sum_{k=1}^{r} c(x_k, y_k) \geq 0$ implies that

$$u(x) \geq \min_{\Gamma}\Big[c(x, y_m) - c(x_0, y_0)\Big] > -\infty.$$

Finally, take $a, b \in I$, $(a, b) \in S$. There exists $\Gamma_a = \{(x_1, y_1), \cdots, (x_r, y_r)\}$ such that

$$u(a) = \Big[c(a, y_r) + \sum_{k=1}^{r} c(x_k, y_{k-1}) - \sum_{k=1}^{r} c(x_k, y_k) - c(x_0, y_0)\Big].$$

Set $\Gamma = \{(x_1, y_1), \cdots, (x_r, y_r), (a, b)\}$.Then for every x

$$\begin{aligned} u(x) &\leq c(x, b) + \sum_{k=1}^{r+1} c(x_k, y_{k-1}) - \sum_{k=1}^{r+1} c(x_k, y_k) - c(x_0, y_0) \\ &= c(x, b) + u(a) - c(a, b). \end{aligned}$$

This completes the proof.

The original Rockafellar's theorem has been proved in the general (continuous) case for the quadratic cost function $c(x, y) = |x - y|^2$.

Rockafellar's theorem for optimal transportation plan

Let π_{ij} be a solution to some optimal transportation problem and $S = \{(i, j) : \pi_{ij} \neq 0\}$. By the duality theorem

$$\sum_{i \in I, j \in J} c_{ij} \pi_{ij} = \sum_{i \in J} r_i u_i + \sum_{j \in J} s_j v_j.$$

In the other hand $u_i + v_j \le c_{ij}$ for every $i \in I$, $j \in J$. This implies that $u_i + v_j = c_{ij}$ for every $(i, j) \in S$. Indeed, otherwise

$$\sum_{i \in I, j \in J} c_{ij} \pi_{ij} > \sum_{i \in I, j \in J} \pi_{ij}(u_i + v_j) = \sum_{i \in J} r_i u_i + \sum_{j \in J} s_j v_j.$$

The original Rockafellar's theorem has been proved in the general (continuous) case for the quadratic cost function $c(x, y) = |x - y|^2$.

Rockafellar's theorem for optimal transportation plan

Let π_{ij} be a solution to some optimal transportation problem and $S = \{(i, j) : \pi_{ij} \neq 0\}$. By the duality theorem

$$\sum_{i \in I, j \in J} c_{ij} \pi_{ij} = \sum_{i \in J} r_i u_i + \sum_{j \in J} s_j v_j .$$

In the other hand $u_i + v_j \leq c_{ij}$ for every $i \in I$, $j \in J$. This implies that $u_i + v_j = c_{ij}$ for every $(i, j) \in S$. Indeed, otherwise

$$\sum_{i \in I, j \in J} c_{ij} \pi_{ij} > \sum_{i \in I, j \in J} \pi_{ij} (u_i + v_j) = \sum_{i \in J} r_i u_i + \sum_{j \in J} s_j v_j .$$

Take $(i, j) \in S$. One has

$$u_i + v_j = c_{ij}$$

and

$$u_k + v_j \leq c_{kj}$$

for any other k. Hence

$$u_k - u_i \leq c_{kj} - c_{ij}.$$

We get the following variant of the Rockafellar theorem. **The support of the optimal transportation plan is contained in the superdifferential of the solution u to the dual transportation problem.**

Relation to other linear programming problems

Algorithms: as a linear programming problem the transportation problem can be solved by some standard algorithm (e.g. simplex method). But there are special algorithms for this problem (see **[Ferg]**).

Shortest path problem and negative cycles

Let us consider the directed graph with nodes (i, j) and "distance" $a_{ij} = c(x_j, y_i) - c(x_i, y_i)$ (the number a_{ij} is allowed to be negative). The sequence $\{i_1, i_2, \cdots, i_n\}$ is called negative cycle if

$$\sum_{k=1}^{n} a_{i_k i_{k+1}} < 0, \quad i_{n+1} = i_1.$$

It is clear that **absence of negative cycles** $\iff$ **cyclical monotonicity**.

Relation to other linear programming problems

Algorithms: as a linear programming problem the transportation problem can be solved by some standard algorithm (e.g. simplex method). But there are special algorithms for this problem (see **[Ferg]**).

Shortest path problem and negative cycles

Let us consider the directed graph with nodes (i, j) and "distance" $a_{ij} = c(x_j, y_i) - c(x_i, y_i)$ (the number a_{ij} is allowed to be negative). The sequence $\{i_1, i_2, \cdots, i_n\}$ is called negative cycle if

$$\sum_{k=1}^{n} a_{i_k i_{k+1}} < 0, \quad i_{n+1} = i_1.$$

It is clear that **absence of negative cycles** $\Longleftrightarrow$ **cyclical monotonicity**.

Warshall-Floyd algorithm

The absence of negative cycles in a graph is a necesary and sufficent condition for existence of a **shortest path** joining any two points x_i, x_j. The **Warshall-Floyd algorithm** finds this path or reveals a negative cycle. Thus it can be used to check whether our set is cyclically monotone.

Let $d_{ij}^{(r)}$ be the shortest path from i to j which uses exactly r edges. Then $d_{ij}^{(1)} = a_{ij}$ and the minimal path length d equals to $d_{ij}^{(n-1)}$.

Bellman dynamical programming principle

$$d_{ij}^{(r)} = \min_{1 \le k \le n} (d_{ik}^{(r-1)} + a_{kj}).$$

Beautiful oservation (see **[M-St]**): assume we are given operations

$$x \oplus y = \min(x, y), \quad \text{tropical computation}$$

$$x \odot y = x + y, \quad \text{tropical multiplication.}$$

Then the previous relation can be written in the form

$$d_{ij}^{(r)} = d_{i1}^{(r-1)} \odot a_{1j} \oplus d_{i2}^{(r-1)} \odot a_{2j} \cdots d_{in}^{(r-1)} \odot a_{nj}.$$

Equivalently

$$d^{(r)} = d^{(r-1)} c \Longrightarrow d^{(r)} = a^r,$$

thus the matrix $(d_{ij}^{(r)})$ is the tropical r-th power of (c_{ij}).

Afriat's theorem

Definition

A finite set $D = \{(x_i, y_i)\} \subset \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is called rationalizable if there exists a function $u : \mathbb{R}^n_+ \to \mathbb{R}$ (**utility function**) satisfying the following property

$$\langle x, y_i \rangle < \langle x_i, y_i \rangle \implies u(x) < u(x_i).$$

Interpreting $\{x_i\}$ as distributions of goods and $\{y_i\}$ as distributions of prices the existence of u means that every x_i is chosen in order to maximize utility for given prices and fixed income.

Afriat's theorem

Definition

A finite set $D = \{(x_i, y_i)\} \subset \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is called rationalizable if there exists a function $u : \mathbb{R}^n_+ \to \mathbb{R}$ (**utility function**) satisfying the following property

$$\langle x, y_i \rangle < \langle x_i, y_i \rangle \implies u(x) < u(x_i).$$

Interpreting $\{x_i\}$ as distributions of goods and $\{y_i\}$ as distributions of prices the existence of u means that every x_i is chosen in order to maximize utility for given prices and fixed income.

Afriat's theorem

Definition

A finite set $D = \{(x_i, y_i)\} \subset \mathbb{R}^n_+ \times \mathbb{R}^n_+$ is called rationalizable if there exists a function $u : \mathbb{R}^n_+ \to \mathbb{R}$ (**utility function**) satisfying the following property

$$\langle x, y_i \rangle < \langle x_i, y_i \rangle \implies u(x) < u(x_i).$$

Interpreting $\{x_i\}$ as distributions of goods and $\{y_i\}$ as distributions of prices the existence of u means that every x_i is chosen in order to maximize utility for given prices and fixed income.

Problem. What are the necessary and sufficient conditions for existence of u? The answer is given by a theorem of S. Afriat: a set is rationalizable if and only if it satisfies the **strong axiom of revealed preferences (SARP)** .
Assume, in addition, that u must be homogenious: $u(\lambda x) = \lambda u(x)$.
This special case of the Afriat's theorem was studied by H. Varian.

Lemma
D is rationalizable with a positive homogenious u if and only if

$$\frac{u(x_i)}{\langle y_i, x_i \rangle} \geq \frac{u(x)}{\langle y_i, x \rangle}, \quad \forall x \in \mathbb{R}^n_+. \tag{1}$$

Proof. (1) $\implies$ D is rationalizable (obvious).

Problem. What are the necessary and sufficient conditions for existence of u? The answer is given by a theorem of S. Afriat: a set is rationalizable if and only if it satisfies the **strong axiom of revealed preferences (SARP)** .
Assume, in addition, that u must be homogenious: $u(\lambda x) = \lambda u(x)$.
This special case of the Afriat's theorem was studied by H. Varian.

Lemma

D is rationalizable with a positive homogenious u if and only if

$$\frac{u(x_i)}{\langle y_i, x_i \rangle} \geq \frac{u(x)}{\langle y_i, x \rangle}, \quad \forall x \in \mathbb{R}^n_+. \tag{1}$$

Proof. (1) $\implies$ D is rationalizable (obvious).

Problem. What are the necessary and sufficient conditions for existence of u? The answer is given by a theorem of S. Afriat: a set is rationalizable if and only if it satisfies the **strong axiom of revealed preferences (SARP)** .
Assume, in addition, that u must be homogenious: $u(\lambda x) = \lambda u(x)$.
This special case of the Afriat's theorem was studied by H. Varian.

Lemma

D is rationalizable with a positive homogenious u if and only if

$$\frac{u(x_i)}{\langle y_i, x_i \rangle} \geq \frac{u(x)}{\langle y_i, x \rangle}, \quad \forall x \in \mathbb{R}^n_+. \tag{1}$$

Proof. (1) $\Longrightarrow$ D is rationalizable (obvious).

D is rationalizable $\implies$ (1). Fix $x_i, x \in \mathbb{R}^d_+$ and find $\lambda > 0$ such that $\langle y_i, \lambda \cdot x\rangle = \langle y_i, x_i\rangle$. Since D is rationalizable, one has $u(\lambda' \cdot x) < u(x_i)$ for every $\lambda' < \lambda$. Since u is homogeneous and $\lambda = \frac{\langle y_i, x_i\rangle}{\langle y_i, x\rangle}$ we immediately get (1). The proof is complete.

Inequality (1) implies that $\log u$ is included in the superdifferential of the cost function $c(x, y) = \log\langle x, y\rangle$. This implies that any optimal transportation plan for this cost function is supported by a rationalizable set and the corresponding homogenious utility function can be found by solving the dual transportation problem. We will see that the inverse is also true (see details in **[KKN]**).

D is rationalizable $\implies$ (1). Fix $x_i, x \in \mathbb{R}^d_+$ and find $\lambda > 0$ such that $\langle y_i, \lambda \cdot x\rangle = \langle y_i, x_i\rangle$. Since D is rationalizable, one has $u(\lambda' \cdot x) < u(x_i)$ for every $\lambda' < \lambda$. Since u is homogeneous and $\lambda = \frac{\langle y_i, x_i\rangle}{\langle y_i, x\rangle}$ we immediately get (1). The proof is complete.

Inequality (1) implies that $\log u$ is included in the superdifferential of the cost function $c(x, y) = \log\langle x, y\rangle$. This implies that any optimal transportation plan for this cost function is supported by a rationalizable set and the corresponding homogenious utility function can be found by solving the dual transportation problem. We will see that the inverse is also true (see details in **[KKN]**).

General transportation (Monge-Kantorovich) problem

Kantorovich problem (L. Kantorovich, 1942)

We define separately **Kantorovich problem** and **Monge problem**.
Assume that we are given
1) Spaces X and Y (Polish, Suslin, completely regular $\cdots$)
2) Borel probability measures: μ on X and ν on Y
3) "Cost function" $c(x, y) : X \times Y \to \mathbb{R}$

Let us denote by $\Pi(\mu, \nu)$ the space of probability measures on the product $(X \times Y)$ giving μ and ν under projecting to X and Y:

$$\sigma \in \Pi(\mu, \nu) \iff \mathrm{Pr}_X \sigma = \mu, \ \mathrm{Pr}_Y \sigma = \nu.$$

Problem. Find a probability measure $\pi \in \Pi(\mu, \nu)$ such that

$$\int c(x, y) \, d\pi = \min_{\sigma \in \Pi(\mu, \nu)} \int c(x, y) \, d\sigma.$$

General transportation (Monge-Kantorovich) problem

Kantorovich problem (L. Kantorovich, 1942)

We define separately **Kantorovich problem** and **Monge problem**.
Assume that we are given

1) Spaces X and Y (Polish, Suslin, completely regular $\cdots$)
2) Borel probability measures: μ on X and ν on Y
3) "Cost function" $c(x, y) : X \times Y \to \mathbb{R}$

Let us denote by $\Pi(\mu, \nu)$ the space of probability measures on the product $(X \times Y)$ giving μ and ν under projecting to X and Y:

$$\sigma \in \Pi(\mu, \nu) \iff \mathrm{Pr}_X \sigma = \mu, \ \mathrm{Pr}_Y \sigma = \nu.$$

Problem. Find a probability measure $\pi \in \Pi(\mu, \nu)$ such that

$$\int c(x, y) \, d\pi = \min_{\sigma \in \Pi(\mu,\nu)} \int c(x, y) \, d\sigma.$$

Clearly, the discrete transportation problem is a particular case (X, Y are discrete spaces) of the Kantorovich problem.

Existence

The Kantorovich problem has a solution under quite general assumptions. The following theorem is applicable in most of the applications.

Theorem

For any Borel probability measures μ and ν on Polish spaces and a lower semicontinuous function $c \geq 0$ the Kantorovich problem has a solution.

Clearly, the discrete transportation problem is a particular case (X, Y are discrete spaces) of the Kantorovich problem.

Existence

The Kantorovich problem has a solution under quite general assumptions. The following theorem is applicable in most of the applications.

Theorem

For any Borel probability measures μ and ν on Polish spaces and a lower semicontinuous function $c \geq 0$ the Kantorovich problem has a solution.

Clearly, the discrete transportation problem is a particular case (X, Y are discrete spaces) of the Kantorovich problem.

Existence

The Kantorovich problem has a solution under quite general assumptions. The following theorem is applicable in most of the applications.

Theorem

For any Borel probability measures μ and ν on Polish spaces and a lower semicontinuous function $c \geq 0$ the Kantorovich problem has a solution.

Sketch of the proof: Consider a sequence of measures $\{\pi_n\} \in \Pi(\mu, \nu)$ such that the sequence $\{\int c(x,y)\pi_n\}$ tends to $\inf_{\sigma \in \Pi(\mu,\nu)} \int c(x,y)\ d\sigma$. Note that $\Pi(\mu,\nu)$ is a tight set of measures. By the Prokhorov's compactness theorem one can extract a weakly convergent subsequence $\pi_n \to \pi$. It follows from the lower continuity of c and the properties of weak convergence that

$$\liminf_k \int c(x,y) d\pi_n \geq \int c(x,y) d\pi.$$

Hence $\int c(x,y) d\pi = \inf_{\sigma \in \Pi(\mu,\nu)} \int c(x,y)\ d\sigma$.

Counterexample to existence

If $c(x, y)$ is not lower-semicontinuous, the problem may have no solutions (Example 1.3 of **[BGMS]**).
Let μ be uniform measure on $[0, 1]$, ν be uniform measure on $[1, 2]$, $c = (x - y)^2$. We will see soon that the optimal transport plan π is the uniform measure on the set $\{[t, t + 1], 0 \leq t \leq 1\}$.
Hence $W_c(\mu, \nu) = \int c(x, y)d\pi = 1$.
Set

$$\tilde{c} = \begin{cases} 2, & y = x + 1 \\ (x - y)^2, & y \neq x + 1. \end{cases}$$

It is impossible to find a plan π such that $\int \tilde{c}(x, y)d\pi = 1$, but for every $\varepsilon > 0$ there is a plan π_ε such that $\int \tilde{c}(x, y)d\pi_\varepsilon \leq 1 + \varepsilon$.

Properties of solutions

Metric on the space of measures

Let $X = Y$ be a metric space with metric $d(x,y)$. An important case of the Kantorovich problem

$$c(x,y) = d^p(x,y), \; p \geq 1.$$

Theorem

For every $p \geq 1$ the functional

$$W_p(\mu,\nu) = \min_{\sigma \in \Pi(\mu,\nu)} \Big(\int d^p(x,y) \, d\sigma \Big)^{\frac{1}{p}} = \Big(\int d^p(x,y) \, d\pi \Big)^{\frac{1}{p}}$$

defines a distance of the space of probability measures on X with a finite p-th monent (π is the minimizer in the Kantorovich problem).

The space of measures with W_p-distance will be denoted by $\mathcal{W}_p$.

Properties of solutions

Metric on the space of measures

Let $X = Y$ be a metric space with metric $d(x,y)$. An important case of the Kantorovich problem

$$c(x,y) = d^p(x,y), \ p \geq 1.$$

Theorem

For every $p \geq 1$ the functional

$$W_p(\mu,\nu) = \min_{\sigma \in \Pi(\mu,\nu)} \Big(\int d^p(x,y) \ d\sigma \Big)^{\frac{1}{p}} = \Big(\int d^p(x,y) \ d\pi \Big)^{\frac{1}{p}}$$

defines a distance of the space of probability measures on X with a finite p-th monent (π is the minimizer in the Kantorovich problem).

The space of measures with W_p-distance will be denoted by $\mathcal{W}_p$.

Properties of solutions

Metric on the space of measures

Let $X = Y$ be a metric space with metric $d(x, y)$. An important case of the Kantorovich problem

$$c(x, y) = d^p(x, y), \ \ p \geq 1.$$

Theorem

For every $p \geq 1$ the functional

$$W_p(\mu, \nu) = \min_{\sigma \in \Pi(\mu,\nu)} \Big(\int d^p(x, y) \ d\sigma \Big)^{\frac{1}{p}} = \Big(\int d^p(x, y) \ d\pi \Big)^{\frac{1}{p}}$$

defines a distance of the space of probability measures on X with a finite p-th monent (π is the minimizer in the Kantorovich problem).

The space of measures with W_p-distance will be denoted by $\mathcal{W}_p$.

Sketch of the proof: All the axioms are obvious except of the triangle inequality. Given three measures μ_1, μ_2, μ_3 let $\pi_{1,2}$ and $\pi_{2,3}$ be optimal plans for the pairs (μ_1, μ_2) and (μ_2, μ_3).
Step 1 ("gluing lemma"). Find a probability measure η on $X \times X \times X$, whose projection to the product of the first two factors is $\pi_{1,2}$, and the projection to the product of the last two factors is the measure $\pi_{2,3}$.
Step 2. The projections of η to the first and third factors are μ_1 and μ_3, whence we obtain

$$W_p(\mu_1, \mu_3) \leq \left(\int d(x_1, x_3)^p \;\; \eta(dx_1 dx_2 dx_3) \right)^{1/p} \leq$$

$$\leq \left(\int [d(x_1, x_2) + d(x_2, x_3)]^p \;\; \eta(dx_1 dx_2 dx_3) \right)^{1/p} \leq$$

$$\leq \left(\int d(x_1, x_2)^p \;\; \eta(dx_1 dx_2 dx_3) \right)^{1/p} + \left(\int d(x_2, x_3)^p \;\; \eta(dx_1 dx_2 dx_3) \right)^{1/p}.$$

The right-hand side coincides with
$\|d\|_{L^p(\pi_{1,2})} + \|d\|_{L^p(\pi_{2,3})} = W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)$.

Application: ergodicity of Markov chains

Consider a finite Markov chain M endowed with a metric $d(x, y)$ and transition probability operator $P : P(M) \to P(M)$. Following **[Olivier]** let us call a number k "**curvature**" of M (or "contraction coefficient" of M) if for every $x, y \in M$ one has

$$W_1(P\delta_x, P\delta_y) \leq (1-k)W_1(Px, Py) = (1-k)d(x, y). \quad (2)$$

If (2) holds, then it can be easily shown that that for every couple of probability measures μ, ν one has

$$W_1(P\mu, P\nu) \leq (1-k)W_1(\mu, \nu).$$

In particular, if ν is the stationary distribution, then for every initial μ the measures $P^n\mu$ converges to ν exponentially fast

$$W_1(P^n\mu, \nu) \leq (1-k)^n W_1(\mu, \nu).$$

There are plenty of examples, where (2) holds.

Optimality $\Longrightarrow$ cyclical monotonicity

We consider the Kantorovich problem with the cost function $c(x, y)$.

Definition

Recall that a topological support S of a Borel probability measure μ is the smallest closed set of full measure.

Theorem

Let X, Y be complete separable metric spaces, let a function $c \geq 0$ be continuous on $X \times Y$ and suppose that there are finite functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$ such that $c(x, y) \leq a(x) + b(y)$. Then the topological support of π is c-cyclically monotone.

Optimality $\Longrightarrow$ cyclical monotonicity

We consider the Kantorovich problem with the cost function $c(x, y)$.

Definition

Recall that a topological support S of a Borel probability measure μ is the smallest closed set of full measure.

Theorem

Let X, Y be complete separable metric spaces, let a function $c \geq 0$ be continuous on $X \times Y$ and suppose that there are finite functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$ such that $c(x, y) \leq a(x) + b(y)$. Then the topological support of π is c-cyclically monotone.

Optimality $\Longrightarrow$ cyclical monotonicity

We consider the Kantorovich problem with the cost function $c(x, y)$.

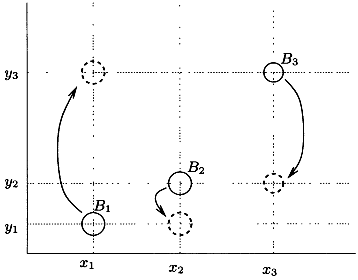
Definition

Recall that a topological support S of a Borel probability measure μ is the smallest closed set of full measure.

Theorem

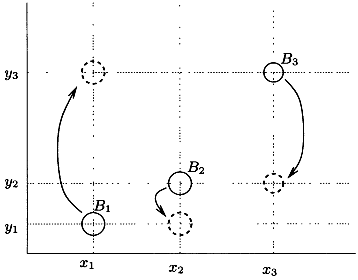
Let X, Y be complete separable metric spaces, let a function $c \geq 0$ be continuous on $X \times Y$ and suppose that there are finite functions $a \in L^1(\mu)$ and $b \in L^1(\nu)$ such that $c(x, y) \leq a(x) + b(y)$. Then the topological support of π is c-cyclically monotone.

The theorem goes back to **[Kn-Sm]** (quadratic case). More general statement: **[Ruesch]**, **[Amb-Pra]** $\cdots$). The proof is similar to the proof in the discrete case. Iinstead of permutation of points we apply permutation of small balls (see **[G-Mc]**). **Picture**: cyclical permutation of small balls (taken from **[Vill]**).



It is also possible to prove existence of a cyclically monotone solutions constructing a discrete approximation sequences.

The theorem goes back to **[Kn-Sm]** (quadratic case). More general statement: **[Ruesch]**, **[Amb-Pra]** $\cdots$). The proof is similar to the proof in the discrete case. Iinstead of permutation of points we apply permutation of small balls (see **[G-Mc]**). **Picture**: cyclical permutation of small balls (taken from **[Vill]**).



It is also possible to prove existence of a cyclically monotone solutions constructing a discrete approximation sequences.

Cyclical monotonicity $\Longrightarrow$ optimality

Theorem

Let X, Y be complete separable metric spaces and $c(x, y)$ be finite measurable cost function. Let $\pi \in \Pi(\mu, \nu)$ admits a set Γ which is c-cyclically monotone and $\pi(\Gamma) = 1$. Then π is optimal.

The proof is non-trivial. The theorem does not hold for cost functions (Example 3.5 in **[Amb-Pra]**) which can take $+\infty$ value. The statement was verified by Pratelli (08) (quadratic cost function), Schachermayer and Teichmann (09) (lower semicontinuous cost function). Further generalizations: **[BGMS]**.

Cyclical monotonicity $\Longrightarrow$ optimality

Theorem

Let X, Y be complete separable metric spaces and $c(x, y)$ be finite measurable cost function. Let $\pi \in \Pi(\mu, \nu)$ admits a set Γ which is c-cyclically monotone and $\pi(\Gamma) = 1$. Then π is optimal.

The proof is non-trivial. The theorem does not hold for cost functions (Example 3.5 in **[Amb-Pra]**) which can take $+\infty$ value. The statement was verified by Pratelli (08) (quadratic cost function), Schachermayer and Teichmann (09) (lower semicontinuous cost function). Further generalizations: **[BGMS]**.

Cyclical monotonicity and ergodic theory

It was shown by M. Beiglböck **[Beig]** that this theorem follows from the Birkhoff ergodic theorem.

Sketch of the proof:

Assume that $\pi, \tilde{\pi} \in \Pi(\mu, \nu)$ and $\pi(\Gamma) = 1$, where Γ is c-cyclically monotone. We want to show that $\int c(x, y) d\tilde{\pi} \geq \int c(x, y) d\pi$. Consider $Z = (X \times Y)^{\mathbb{N}}$. Denote by σ the Bernoulli shift:

$$\sigma : (x_i, y_i)_{i=1}^{\infty} \to (x_{i+1}, y_{i+1})_{i=1}^{\infty}.$$

Construct a measure m on Z which is σ-invariant: $m \circ \sigma^{-1} = \sigma$ such that

$$m \circ P^{-1}_{(x_1, y_1)} = \pi, \quad m \circ P^{-1}_{(x_1, y_2)} = \tilde{\pi}.$$

One has

$$\int c(x, y) d\tilde{\pi} - \int c(x, y) d\pi = \int f dm,$$

where $f = c(P_{(x_1, y_2)}) - c(P_{(x_1, y_1)})$. It suffices to show that $\int f dm > 0$.

Using ergodic theorem one gets that there exists f^* such that

$$f^* = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} f(\sigma^k)$$

and $\int f^* dm = \int f dm$. For the sake of simplicity let us assume that c is bounded. It follows immediately from the cyclical monotonicity property that

$$f* = \lim_n \frac{1}{n} \sum_{k=0}^{n-1} (c(x_k, y_{k+1}) - c(x_k, y_k)) \geq 0.$$

The proof is compelete.

Cyclical monotonicity: concluding remarks

- It is known that cyclical monotonicity for n points does not imply cyclical monotonicity for any number of points (Asplund's counterexample).
- (Krauss) Vector field u is 2-cyclically monotone if and only if

$$u(x) \in \partial_2 H(x, x)$$

where H is a concave-convex antisymmetric Hamiltonian on $\mathbb{R}^d \times \mathbb{R}^d$

- Ghoussoub and Galishon **[GG]** charaterized N-cyclically monotone vector fields as solutions of a particular mass transportation problem.

Cyclical monotonicity: concluding remarks

- It is known that cyclical monotonicity for n points does not imply cyclical monotonicity for any number of points (Asplund's counterexample).
- (Krauss) Vector field u is 2-cyclically monotone if and only if

$$u(x) \in \partial_2 H(x, x)$$

where H is a concave-convex antisymmetric Hamiltonian on $\mathbb{R}^d \times \mathbb{R}^d$

- Ghoussoub and Galishon **[GG]** charaterized N-cyclically monotone vector fields as solutions of a particular mass transportation problem.

Cyclical monotonicity: concluding remarks

- It is known that cyclical monotonicity for n points does not imply cyclical monotonicity for any number of points (Asplund's counterexample).
- (Krauss) Vector field u is 2-cyclically monotone if and only if

$$u(x) \in \partial_2 H(x, x)$$

  where H is a concave-convex antisymmetric Hamiltonian on $\mathbb{R}^d \times \mathbb{R}^d$
- Ghoussoub and Galishon **[GG]** charaterized N-cyclically monotone vector fields as solutions of a particular mass transportation problem.

One-dimensional case

For $X = Y = \mathbb{R}$ and $c = (x-y)^2$ (actually for every $c = h(x-y)$, where h is convex) it is easy to check that the cyclical monotonicity of a set $\Gamma \subset \mathbb{R} \times \mathbb{R}$ is equivalent to the usual monotonicity. More precisely, if $(x_1, y_1), (x_2, y_2) \in \Gamma$ and $x_2 > x_1$, then $y_2 \geq y_1$.
Conclusion: **any quadratic optimal transportation plan for measures on $\mathbb{R}$ is carried by a graph of some increasing function**.

This observation allows to describe exactly the solutions to the one-dimensional Kantorovich problem. If $\mu = f\ dx$, $\nu = g\ dx$ are probability measures, then π is supported on $\{(x, T(x)),\ x \in \mathbb{R}\}$, where T is defined by the relation

$$\int_{-\infty}^{T(x)} g(t)\ dt = \int_{-\infty}^{x} f(t)\ dt.$$

Theorem

Let μ and ν be two probability measures on $\mathbb{R}$, with respective cumulative distribution functions F and G. Let π be the probability measure on $\mathbb{R}^2$ with joint two-dimensional cumulative distribution function $\min(F(x), H(y))$ Then π is optimal in the Kantorovich transportation problem between μ and ν for any convex cost function $c(x - y)$. Moreover, the value of the optimal transportation cost is

$$\int_0^1 c(F^{-1}(t) - G^{-1}(t))\, dt.$$

Optimal transportation mappings

What happens in the multidimensional case? Assume that $c = \frac{1}{2}(x-y)^2$. According to the Rockafellar theorem and cyclical monotonicity of the support of π there exists u such that for π-almost all (x, y) and every z

$$u(z) - u(x) \le \frac{1}{2}(z-y)^2 - \frac{1}{2}(x-y)^2.$$

Set $\Phi(x) = \frac{1}{2}x^2 - \varphi(x)$.
Then this inequality implies

$$\Phi(z) \ge \Phi(x) + \langle y, z-y \rangle.$$

This means that

- Φ is convex
- y belongs to the subdifferential of Φ at the point x

Optimal transportation mappings

What happens in the multidimensional case? Assume that $c = \frac{1}{2}(x-y)^2$. According to the Rockafellar theorem and cyclical monotonicity of the support of π there exists u such that for π-almost all (x, y) and every z

$$u(z) - u(x) \leq \frac{1}{2}(z-y)^2 - \frac{1}{2}(x-y)^2.$$

Set $\Phi(x) = \frac{1}{2}x^2 - \varphi(x)$.
Then this inequality implies

$$\Phi(z) \geq \Phi(x) + \langle y, z-y \rangle.$$

This means that

- Φ is convex
- y belongs to the subdifferential of Φ at the point x

Optimal transportation mappings

What happens in the multidimensional case? Assume that $c = \frac{1}{2}(x-y)^2$. According to the Rockafellar theorem and cyclical monotonicity of the support of π there exists u such that for π-almost all (x, y) and every z

$$u(z) - u(x) \le \frac{1}{2}(z-y)^2 - \frac{1}{2}(x-y)^2.$$

Set $\Phi(x) = \frac{1}{2}x^2 - \varphi(x)$.
Then this inequality implies

$$\Phi(z) \ge \Phi(x) + \langle y, z-y \rangle.$$

This means that

- Φ is convex
- y belongs to the subdifferential of Φ at the point x

Assume that μ has a density with respect to Lebesgue measure $\mu = \rho dx$. It is known (and easy to show) the subdifferential of a convex function Φ contains a unique element for μ-almost all x. Thus our transportation plan is supported by a graph of the mapping $T : x \to \nabla\Phi(x)$. The measure ν is the image of μ under T. This means that

$$\nu(A) = \mu(T^{-1}(A))$$

for any Borel set A.
Here is a more precise statement of this result (Knott–Smith, Brenier, McCann $\cdots$).

Theorem

Let μ and ν be probability measures on $\mathbb{R}^d$ and let $\mu(A) = 0$ for every set A of Hausdorff dimension not greater than $d-1$. Then there exists a Borel mapping $T\colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\nu = \mu \circ T^{-1}$ and $T = \nabla\Phi$ for some convex function Φ.

Assume that μ has a density with respect to Lebesgue measure $\mu = \rho dx$. It is known (and easy to show) the subdifferential of a convex function Φ contains a unique element for μ-almost all x. Thus our transportation plan is supported by a graph of the mapping $T : x \to \nabla\Phi(x)$. The measure ν is the image of μ under T. This means that

$$\nu(A) = \mu(T^{-1}(A))$$

for any Borel set A.
Here is a more precise statement of this result (Knott–Smith, Brenier, McCann $\cdots$).

Theorem

Let μ and ν be probability measures on $\mathbb{R}^d$ and let $\mu(A) = 0$ for every set A of Hausdorff dimension not greater than $d - 1$. Then there exists a Borel mapping $T\colon \mathbb{R}^d \to \mathbb{R}^d$ such that $\nu = \mu \circ T^{-1}$ and $T = \nabla\Phi$ for some convex function Φ.

Duality

The existence of a **convex** transportational potential Φ can be extracted from the continuous version of the Kantorovich duality theorem.
The functional

$$K(\pi) \to \int c(x,y)d\pi, \quad \pi \in \Pi(\mu,\nu)$$

is "dual" to

$$J(u(x),v(y)) \to \int u d\mu + \int v d\nu, \quad u(x)+v(y) \le c(x,y).$$

which is defined on **measurable integrable functions**.
Remark: it is not a rigorous statement because the space of measure is dual to a space of functions under quite restrictive assumptions.

Theorem

Assume that X, Y are complete, separable, metric spaces, $c(x, y)$ is lower semicontinuous and $c(x, y) \leq a(x) + b(y)$ with some functions $a \in L^1(\mu)$, $b \in L^1(\nu)$. Then

$$\inf_{\pi \in \Pi(\mu,\nu)} K(\pi) = \sup_{u(x)+v(y) \leq c(x,y)} J(\mu, \nu)$$

Moreover, inf *and* sup *are attained on some π and (u, v) respectively.*

1. The proof (see **[Vill]**) is based on a special case of the minimax principle applied to the spaces of bounded continuous functions and measures.
2. For general cost function c different cases are possible: $\sup \neq \inf$; $\sup = \inf$, but inf is not attained etc.

Theorem

Assume that X, Y are complete, separable, metric spaces, $c(x, y)$ is lower semicontinuous and $c(x, y) \leq a(x) + b(y)$ with some functions $a \in L^1(\mu)$, $b \in L^1(\nu)$. Then

$$\inf_{\pi \in \Pi(\mu,\nu)} K(\pi) = \sup_{u(x)+v(y)\leq c(x,y)} J(\mu, \nu)$$

Moreover, inf *and* sup *are attained on some π and (u, v) respectively.*

1. The proof (see **[Vill]**) is based on a special case of the minimax principle applied to the spaces of bounded continuous functions and measures.
2. For general cost function c different cases are possible: $\sup \neq \inf$; $\sup = \inf$, but inf is not attained etc.

Theorem

Assume that X, Y are complete, separable, metric spaces, $c(x, y)$ is lower semicontinuous and $c(x, y) \leq a(x) + b(y)$ with some functions $a \in L^1(\mu)$, $b \in L^1(\nu)$. Then

$$\inf_{\pi \in \Pi(\mu,\nu)} K(\pi) = \sup_{u(x)+v(y) \leq c(x,y)} J(\mu, \nu)$$

Moreover, $\inf$ *and* $\sup$ *are attained on some π and (u, v) respectively.*

1. The proof (see **[Vill]**) is based on a special case of the minimax principle applied to the spaces of bounded continuous functions and measures.
2. For general cost function c different cases are possible: $\sup \neq \inf$; $\sup = \inf$, but $\inf$ is not attained etc.

Solution (u, v) of the dual problem

Definition

Function f is called c-concave, if it has the representation

$$f(x) = g^c(x) = \inf_y \{c(x, y) - g(y)\}.$$

The functions u, v have the following properties (after redefining (if necessary) on a set of π-measure zero).

- u and v are c-concave
- they satisfy $u(x) + v(y) \leq c(x, y)$ for every (x, y)
- $u(x) + v(y) = c(x, y)$ for π-almost (x, y)
- (**General Legendre duality**) $u^c = v, v^c = u$.

Solution (u, v) of the dual problem

Definition

Function f is called c-concave, if it has the representation

$$f(x) = g^c(x) = \inf_y \{c(x, y) - g(y)\}.$$

The functions u, v have the following properties (after redefining (if necessary) on a set of π-measure zero).

- u and v are c-concave
- they satisfy $u(x) + v(y) \leq c(x, y)$ for every (x, y)
- $u(x) + v(y) = c(x, y)$ for π-almost (x, y)
- (**General Legendre duality**) $u^c = v, v^c = u$.

Solution (u, v) of the dual problem

Definition

Function f is called c-concave, if it has the representation

$$f(x) = g^c(x) = \inf_y \{c(x, y) - g(y)\}.$$

The functions u, v have the following properties (after redefining (if necessary) on a set of π-measure zero).

- u and v are c-concave
- they satisfy $u(x) + v(y) \leq c(x, y)$ for every (x, y)
- $u(x) + v(y) = c(x, y)$ for π-almost (x, y)
- (**General Legendre duality**) $u^c = v, v^c = u$.

Solution (u, v) of the dual problem

Definition

Function f is called c-concave, if it has the representation

$$f(x) = g^c(x) = \inf_y \{c(x, y) - g(y)\}.$$

The functions u, v have the following properties (after redefining (if necessary) on a set of π-measure zero).

- u and v are c-concave
- they satisfy $u(x) + v(y) \leq c(x, y)$ for every (x, y)
- $u(x) + v(y) = c(x, y)$ for π-almost (x, y)
- (**General Legendre duality**) $u^c = v, v^c = u$.

Solution (u, v) of the dual problem

Definition

Function f is called c-concave, if it has the representation

$$f(x) = g^c(x) = \inf_y \{c(x, y) - g(y)\}.$$

The functions u, v have the following properties (after redefining (if necessary) on a set of π-measure zero).

- u and v are c-concave
- they satisfy $u(x) + v(y) \leq c(x, y)$ for every (x, y)
- $u(x) + v(y) = c(x, y)$ for π-almost (x, y)
- (**General Legendre duality**) $u^c = v, v^c = u$.

Solution (u, v) of the dual problem

Definition

Function f is called c-concave, if it has the representation

$$f(x) = g^c(x) = \inf_y \{c(x, y) - g(y)\}.$$

The functions u, v have the following properties (after redefining (if necessary) on a set of π-measure zero).

- u and v are c-concave
- they satisfy $u(x) + v(y) \leq c(x, y)$ for every (x, y)
- $u(x) + v(y) = c(x, y)$ for π-almost (x, y)
- **(General Legendre duality)** $u^c = v, v^c = u$.

Solution (u, v) of the dual problem

Definition

Function f is called c-concave, if it has the representation

$$f(x) = g^c(x) = \inf_y \{c(x, y) - g(y)\}.$$

The functions u, v have the following properties (after redefining (if necessary) on a set of π-measure zero).

- u and v are c-concave
- they satisfy $u(x) + v(y) \leq c(x, y)$ for every (x, y)
- $u(x) + v(y) = c(x, y)$ for π-almost (x, y)
- (**General Legendre duality**) $u^c = v, v^c = u$.

Monge problem (G. Monge, 1781)

The Monge problem can be described as follows (in the general form this formulation of the problem is due to A.M. Vershik).
Assume that we are given
1) Spaces X and Y (Polish, Suslin, completely regular $\cdots$)
2) Borel probability measures: μ on X and ν on Y
3) Cost function $c(x, y) : X \times Y \to \mathbb{R}$.
Find a Borel mapping $T\colon\ X \to Y$, taking the measure μ into ν and minimizing the expression

$$\int_X c(x, T(x))\, \mu(dx)$$

among all such mappings. The condition that μ is taken to ν means that

$$\mu \circ T^{-1}(B) := \mu(T^{-1}(B)) = \nu(B),$$

If some $T \in T(\mu, \nu)$ gives a minimum, then this T is called an **optimal mapping** of the measure μ to ν or an **optimal transportation**.

Counterexamples: Monge versus Kantorovich

Let us denote by $K(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int c \; d\pi$ the value of the Kantorovich problem.
(i) Let $X = Y = [-1, 1]$, $\mu = \delta_0$, $\nu = 2^{-1}(\delta_{-1} + \delta_1)$, $h(x, y) = |x - y|^2$. Then μ cannot be transformed into ν at all, but the half-sum of the Dirac measures at the points $(0, -1)$ and $(0, 1)$ serves as a unique solution to the Kantorovich problem.

(ii) Let $X = Y = [-1, 1]^2$, $\mu = \lambda \otimes \delta_0$, $\nu = 2^{-1}(\lambda \otimes \delta_{-1} + \lambda \otimes \delta_1)$, let λ be the normalized Lebesgue measure on $[-1, 1]$, and let $h(x, y) = |x - y|^2$, where $| \cdot |$ is the usual norm in $\mathbb{R}^2$. Then both measures have no atoms, the Kantorovich problem has a solution and $K(\mu, \nu) = 1$, but the Monge problem has no solutions.

(i) If a measure σ on $[-1,1]^2$ has projections μ and ν, then it is concentrated on the intersection of the interval $\{0\}\times[-1,1]$ with the union of the intervals $[-1,1]\times\{-1\}$ and $[-1,1]\times\{1\}$, i.e., is a combination of the Dirac measures at $(0,-1)$ and $(0,1)$, whence it is seen that it must be their half-sum. Therefore, here $\Pi(\mu,\nu)$ consists of a single element.

(ii) Here $\Pi(\mu,\nu)$ contains many measures, but the quantity $K(\mu,\nu)$ is easily found, since minimizing the integral of the function

$$|x-y|^2=(x_1-y_1)^2+(x_2-y_2)^2$$

with respect to measures $\sigma\in\Pi(\mu,\nu)$ we obtain that the integral of the first summand equals 1 due to the fact that $x_1=0$ σ-a.e. and $y_1^2=1$ σ-a.e., so that the minimum equal to 1 is attained precisely when $x_2=y_2$ σ-a.e. An optimal plan is the distribution of the random vector $(0,\eta,\xi,\eta)$, where ξ and η are independent random variables such that ξ assumes the values -1 and 1 with probabilities $1/2$ and η has the uniform distribution in $[-1,1]$.

If we suppose that the Monge problem has a solution $T = (T_1, T_2)$, then we obtain that $|T_1(x)| = 1$ μ-a.e., so that the integral of $|x - T(x)|^2 = (x_1 - T_1(x))^2 + (x_2 - T_2(x))^2$ with respect to the measure μ is equal to the integral of $1 + (x_2 - T_2(x))^2$ with respect to μ. On account of the equality $K(\mu, \nu) = 1$ this means that $T_2(x) = x_2$ μ-a.e., so that $T_1(x) = \xi(x_2)$ for a.e. x_2 is either 1 or -1. Then the measure $\mu \circ T^{-1}$ will be concentrated on the union of the sets $A \times \{-1\}$ and $B \times \{1\}$, where A and B are disjoint measurable sets with $A \cup B = [-1, 1]$, therefore, it cannot coincide with the measure ν.

Existence and uniqueness

For what kind of cost finctions the Monge problem has a solution? Typical example: $X = Y = \mathbb{R}^d$, $c(x, y) = |x - y|^p$, $p \geq 1$.

Consider cost function $c(x - y)$, where c is strictly convex, superlinear, and $|c(x - y)| \leq a(x) + b(x)$, $a \in L^1(\mu), b \in L^1(\nu)$. There exists a unique solution T to the Monge problem. The unique solution to the Kantorovich problem is supported on the graph $(x, T(x))$. T has the form

$$T(x) = x - \nabla c^*(\nabla \varphi(x)),$$

Here c^* is the Legendre transform of c: $c^*(t) = \inf_s \big(st - c(s)\big)$. The potential φ is c-concave, i.e. there exists a function ψ such that

$$\varphi(x) = \inf_{y \in \mathbb{R}^d} \big(c(x - y) - \psi(y)\big).$$

Existence and uniqueness

For what kind of cost finctions the Monge problem has a solution? Typical example: $X = Y = \mathbb{R}^d$, $c(x, y) = |x - y|^p$, $p \geq 1$.

Consider cost function $c(x - y)$, where c is strictly convex, superlinear, and $|c(x - y)| \leq a(x) + b(x)$, $a \in L^1(\mu), b \in L^1(\nu)$. There exists a unique solution T to the Monge problem. The unique solution to the Kantorovich problem is supported on the graph $(x, T(x))$. T has the form

$$T(x) = x - \nabla c^*(\nabla \varphi(x)),$$

Here c^* is the Legendre transform of c: $c^*(t) = \inf_s \big(st - c(s)\big)$. The potential φ is c-concave, i.e. there exists a function ψ such that

$$\varphi(x) = \inf_{y \in \mathbb{R}^d} \big(c(x - y) - \psi(y)\big).$$

Scheme of the proof:

1. There exists a solution to the corresponding Kantorovich problem.
2. It is supported by a c-cyclical monotone set S.
3. S is concentrated on the graph of a mapping (via Rockafellar's theorem).
4. If T_1, T_2 are two optimal mappings, then $T_1 = T_2$ μ-almost everywhere.

We have already discussed 1)-3). How to prove 4)?

Scheme of the proof:

1. There exists a solution to the corresponding Kantorovich problem.
2. It is supported by a c-cyclical monotone set S.
3. S is concentrated on the graph of a mapping (via Rockafellar's theorem).
4. If T_1, T_2 are two optimal mappings, then $T_1 = T_2$ μ-almost everywhere.

We have already discussed 1)-3). How to prove 4)?

Scheme of the proof:

1. There exists a solution to the corresponding Kantorovich problem.
2. It is supported by a c-cyclical monotone set S.
3. S is concentrated on the graph of a mapping (via Rockafellar's theorem).
4. If T_1, T_2 are two optimal mappings, then $T_1 = T_2$ μ-almost everywhere.

We have already discussed 1)-3). How to prove 4)?

Scheme of the proof:

1. There exists a solution to the corresponding Kantorovich problem.
2. It is supported by a c-cyclical monotone set S.
3. S is concentrated on the graph of a mapping (via Rockafellar's theorem).
4. If T_1, T_2 are two optimal mappings, then $T_1 = T_2$ μ-almost everywhere.

We have already discussed 1)-3). How to prove 4)?

Scheme of the proof:

1. There exists a solution to the corresponding Kantorovich problem.
2. It is supported by a c-cyclical monotone set S.
3. S is concentrated on the graph of a mapping (via Rockafellar's theorem).
4. If T_1, T_2 are two optimal mappings, then $T_1 = T_2$ μ-almost everywhere.

We have already discussed 1)-3). How to prove 4)?

Scheme of the proof:

1. There exists a solution to the corresponding Kantorovich problem.
2. It is supported by a c-cyclical monotone set S.
3. S is concentrated on the graph of a mapping (via Rockafellar's theorem).
4. If T_1, T_2 are two optimal mappings, then $T_1 = T_2$ μ-almost everywhere.

We have already discussed 1)-3). How to prove 4)?

Let π_1 and π_2 be solutions to the Kantorovich problem, $c = |x - y|^2$, and $T_1 = \nabla\varphi_1$, $T_2 = \nabla\varphi_2$ be the corresponding optimal mappings. One has $\varphi_i(x) + \varphi_i^*(y) \le \langle x, y\rangle$ and $\varphi_i(x) + \varphi_i^*(y) = \langle x, y\rangle$ for π_i-almost all (x, y). In follows from the optimality of π_1, π_2 that $\int \langle x, y\rangle \ d\pi_1 = \int \langle x, y\rangle \ d\pi_2$. One has

$$\int \langle x, \nabla\varphi_1\rangle d\mu = \int \langle x, y\rangle \ d\pi_1 = \int \langle x, y\rangle \ d\pi_2$$

$$= \int \varphi_2 \ d\mu + \int \varphi_2^* \ d\nu = \int \Big(\varphi_2 + \varphi_2^*(\nabla\varphi_1)\Big) \ d\mu.$$

Since $\varphi_2 + \varphi_2^*(\nabla\varphi_1) \ge \langle x, \nabla\varphi_1\rangle$ (Fenchel inequality), one has $\varphi_2 + \varphi_2^*(\nabla\varphi_1) = \langle x, \nabla\varphi_1\rangle$. This is possible if and only if $\nabla\varphi_1 \in \partial\varphi_2$. Hence $\nabla\varphi_1 = \nabla\varphi_2$ μ-a.e.

Distance case. Monge problem. Kantorovich-Rubinstein theorem.

$$c(x,y) = |x-y|, \; X = Y = \mathbb{R}^d.$$

This is the classical Monge problem! Solution exists but **not unique**.

Example: μ is uniform measure on $[0,1]$, ν is uniform measure on $[1/2, 3/2]$. Then $T_1(x) = x + 1/2$ and T_2 which is identical on $[1/2, 1]$ and $T_2(x) = x + 1$ on $[0, 1/2]$ are both optimal. Optimality of T_1 follows from the one-dimensional description of the optimal mappings. It can be verified by direct computations that T_1 and T_2 have the same cost.

The existence theorem is very hard. First proof: V. Sudakov (1974). Some gaps were revealed later (see **Amb-Pra**). Assume that $\mu = f\ dx$, $\nu = g\ dx$. Then the transportational potential φ solves

$$-\mathrm{div}(a\nabla\varphi) = f - g, \quad |\nabla\varphi| = 1.$$

The function $a \geq 0$ is called transportational density. Optimal mapping: move in direction $\nabla\varphi$ by a distance a.

Distance case. Dual problem.

It follows from the relations
$u(x) = \inf_y(|x-y| - v(y)), v(y) = \inf_x(|x-y| - u(x))$ that

1. u, v are 1-Lipschitz
2. $u = -v$, because $|x-y| \geq v(x) - v(y) \Longrightarrow$ $\inf_y(|x-y| - v(y))$ is attained at $y = x \Longrightarrow u(x) = -v(x)$.

Thus in the distance case the dual problem takes the following form:

$$\int u(\mu - \nu) \to \sup, \ \|u\|_{\mathrm{Lip}} \leq 1.$$

Corollary: the distance W_1 is a **norm** on the space of probability measures. It is invariant under addition of a mass $W_1(\mu, \nu) = W_1(\mu + \sigma, \nu + \sigma)$.

Distance case. Dual problem.

It follows from the relations
$u(x) = \inf_y(|x-y| - v(y)), v(y) = \inf_x(|x-y| - u(x))$ that

1. u, v are 1-Lipschitz
2. $u = -v$, because $|x-y| \geq v(x) - v(y) \Longrightarrow$
   $\inf_y(|x-y| - v(y))$ is attained at $y = x \Longrightarrow u(x) = -v(x)$.

Thus in the distance case the dual problem takes the following form:

$$\int u(\mu - \nu) \to \sup, \ \|u\|_{\mathrm{Lip}} \leq 1.$$

Corollary: the distance W_1 is a **norm** on the space of probability measures. It is invariant under addition of a mass
$W_1(\mu, \nu) = W_1(\mu + \sigma, \nu + \sigma)$.

Distance case. Dual problem.

It follows from the relations
$u(x) = \inf_y(|x-y| - v(y)), v(y) = \inf_x(|x-y| - u(x))$ that

1. u, v are 1-Lipschitz
2. $u = -v$, because $|x-y| \geq v(x) - v(y) \Longrightarrow$
   $\inf_y(|x-y| - v(y))$ is attained at $y = x \Longrightarrow u(x) = -v(x)$.

Thus in the distance case the dual problem takes the following form:

$$\int u(\mu - \nu) \to \sup, \ \|u\|_{\mathrm{Lip}} \leq 1.$$

Corollary: the distance W_1 is a **norm** on the space of probability measures. It is invariant under addition of a mass
$W_1(\mu, \nu) = W_1(\mu + \sigma, \nu + \sigma)$.

Distance case. Dual problem.

It follows from the relations
$u(x) = \inf_y(|x-y| - v(y)), v(y) = \inf_x(|x-y| - u(x))$ that

1. u, v are 1-Lipschitz
2. $u = -v$, because $|x-y| \geq v(x) - v(y) \Longrightarrow$
   $\inf_y(|x-y| - v(y))$ is attained at $y = x \Longrightarrow u(x) = -v(x)$.

Thus in the distance case the dual problem takes the following form:

$$\int u(\mu - \nu) \to \sup, \ \|u\|_{\mathrm{Lip}} \leq 1.$$

Corollary: the distance W_1 is a **norm** on the space of probability measures. It is invariant under addition of a mass
$W_1(\mu, \nu) = W_1(\mu + \sigma, \nu + \sigma)$.

References

Amb-Pra AMBROSIO L., PRATELLI A., “Existence and stability results in the L^1 theory of optimal transportation”, Optimal transportation and applications (Martina Franca, 2001), p. 123–160, Lecture Notes in Math., V. 1813, Springer, Berlin, 2003.

BK BOGACHEV V.I., KOLESNIKOV A.V. The Monge-Kantorovich problem: achievements, connections, and perspectives. 2012 Russ. Math. Surv. 67(5), 785–890.

BGMS BEIGLBÖCK M., GOLDSTERN M., MARESCH G., AND SCHACHERMAYER W. Optimal and better transport plans. J. Funct. Anal., 256(6) : 1907–1927, 2009.

Beig BEIGLBÖCK M., "Cyclical monotonicity and the ergodic theory".

Ferg FERGUSON T.S. "Linear programming: a concise introduction". http://www.math.ucla.edu/ tom/LP.pdf

G-Mc GANGBO W., MCCANN R.J. "The geometry of optimal transportation". Acta Math., 177 (1996), 113–161.

References

GG GHOUSOOUB N., GALISHON A., Variational representations for N-cyclically monotone vector felds. arXiv: 1207.2408.

Kn-Sm KNOTT M, SMITH C.S., “On the optimal mapping of distributions”, J. Optim. Theory Appl., 43 (1984), 39–49.

KKN KOLESNIKOV A.V., KUDRYAVTSEVA O.V., NAGAPETYAN T., Remarks on the Afriat’s theorem and the Monge–Kantorovich problem. arXiv:1302.4695.

Olivier OLIVIER Y., "A survey of Ricci curvature for metric spaces and Markov chains", Probabilistic approach to geometry, Adv. Stud. Pure Math. 57, Math. Soc. Japan (2010), 343–381.

M-St MACLAGAN D., STURMFELS B. "Introduction to Tropical Geometry", 2009.

Ruesch RÜSCHENDORF L., "On *c*-optimalrandom variables", Statist. Probab. Lett. 27:3 (1996), 267–270.

RR RACHEV S.T., RÜSCHENDORF L., Mass transportation problems, V. I, II, Springer, New York, 1998.

References

Vill VILLANI C., Topics in optimal transportation, Amer. Math. Soc. Providence, Rhode Island, 2003.

Vill2 VILLANI2 Villani C., Optimal transport, old and new, Springer, New York, 2009.