Calc. Var. (2019) 58:173
https://doi.org/10.1007/s00526-019-1610-4

Calculus of Variations

On multistochastic Monge–Kantorovich problem, bitwise operations, and fractals

Nikita A. Gladkov[1] · Alexander V. Kolesnikov[1] · Alexander P. Zimin[1]

Received: 11 March 2019 / Accepted: 31 July 2019 / Published online: 21 September 2019

Abstract
The multistochastic (n, k)-Monge–Kantorovich problem on a product space $\prod_{i=1}^{n} X_i$ is an extension of the classical Monge–Kantorovich problem. This problem is considered on the space of measures with fixed projections onto $X_{i_1} \times \cdots \times X_{i_k}$ for all k-tuples $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ for a given $1 \le k < n$. In our paper we study well-posedness of the primal and the corresponding dual problem. Our central result describes a solution π to the following important model case: $n = 3, k = 2, X_i = [0, 1]$, the cost function $c(x, y, z) = xyz$, and the corresponding two-dimensional projections are Lebesgue measures on $[0, 1]^2$. We prove, in particular, that the mapping $(x, y) \to x \oplus y$, where $\oplus$ is the bitwise addition (xor- or Nim-addition) on $[0, 1] \cong \mathbb{Z}_2^{\infty}$, is the corresponding optimal transportation. In particular, the support of π is the Sierpiński tetrahedron. In addition, we describe a solution to the corresponding dual problem.

1 Introduction

In this paper we consider a natural modification of the Monge–Kantorovich mass transportation problem which we call "multistochastic Monge–Kantorovich problem". To our best knowledge, this problem has never been studied before.

Assume we are given two probability measures μ, ν on measurable spaces X, Y and a function $c\colon X \times Y \to \mathbb{R}$. Let us remind the reader that the classical Kantorovich or transportation problem is a problem of minimization of the functional

Communicated by L. Ambrosio.

The second named author was supported by RFBR Project 17-01-00662 and DFG Project RO 1195/12-1. The article was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project '5-100'.

✉ Alexander V. Kolesnikov
Sascha77@mail.ru

Nikita A. Gladkov
gladkovna@gmail.com

Alexander P. Zimin
alekszm@gmail.com

[1] National Research University Higher School of Economics, Moscow, Russian Federation

$$\int_{X\times Y} c(x, y)d\pi, \tag{1.1}$$

on the set $\Pi(\mu, \nu)$ of probability measures on $X \times Y$ with fixed marginals μ, ν.

An important tool to attack this problem coming from the linear programming theory is the so-called dual transportation problem: maximize

$$\int f d\mu + \int g d\nu$$

on the set of couples of (integrable) functions (f, g) satisfying $f(x) + g(y) \leq c(x, y)$.

The most classical cost function c is given by the distance function, but the quadratic cost function $c(x, y) = |x - y|^2$ has gained incredible popularity because of impressive number of applications. For the quadratic cost function any standard solution π is concentrated on the graph of a mapping T, as proved by Brenier [8]. In this case T is a solution of the corresponding Monge problem which asks for a mapping minimizing the functional $\int c(x, T(x))d\mu$ in the class of measure preserving (i.e. pushing forward μ onto ν) mappings.

Since its revival at the end of eighties the Monge–Kantorovich theory attracts growing attention. The reader can find a lot of information on the classical mass transportation theory in many recent textbooks and survey papers [7,12,16,32,35,36].

Our research is motivated by a number of recent results appeared in several quickly developping branches of the mass transportation theory. Here is a short outline of the most important problems and ideas.

(1) Multimarginal transportation problem

The book of Rachev and Rüschendorf [32] contains rich material on the multimarginal transportation problem, in particular, a number of functional-analytical results on duality, probabilistic applications etc. However, until recently only the two-marginals case was important for the largest part of applications. The books of Villani [35,36] deal with the most important but specific two-marginals case.

The revival of interest to the multimarginal Monge–Kantorovich problem is partially motivated by economical applications (matching theory, multi-dimensional screening), see [12,16]. We refer to survey paper [31]. Many references on recent works on multimarginal duality theory for a wide class of cost functions can be found in [7].

(2) Doubly- and multistochastic measures

According to the classical Birkhoff–von Neumann theorem every bistochastic matrix is a convex combination of permutation matrices. More precisely, the permutation matrices are exactly the extreme points of the set of bistochastic matrices. The classical problem of Birkhoff asks for a generalization of this result for the set of bistochastic (doubly stochastic) measures $\Pi(\mu, \nu)$. This problem has been attacked by many researches (see [1,5,18,33]), let us in particularly mention the seminal paper [18], containing a characterization of supports of such measures. Using this characterization Ahmad, Kim, and McCann obtained in [1] interesting results on uniqueness of solution to the optimal transportation problem (see also [6, 29]). Exposition of relations between bistochastic measures, Markov operators, and Markov chains can be found in [34].

In this paper we deal with (n, k)-stochastic measures, which are probability measures on a product space

$$X = X_1 \times X_2 \times \cdots \times X_n$$

with fixed projections $\mu_I \in \mathbb{P}(X_I)$ for every $X_I = X_{i_1} \times \cdots \times X_{i_k}$, where $I = \{i_1, \ldots, i_k\}$ is a k-tuple of indices, $k < n$. The simplest (and most famous) example of such measures is given by the set of latin squares which is homeomorphic to $(3, 2)$-stochastic matrices. It is important to emphasize that for the set of (n, k)-stochastic matrices (measures) with $n > 2$ there is no analog of the Birkhoff–von Neumann theorem (see [21,27], see also [9] for description of extreme points for $k = 2, n = 3$ in the discrete case).

(3) Monge–Kantorovich problem with linear constraints and Monge–Kantorovich problem with symmetries

Apparenty the most famous example of a transportation problem with linear constraints is the optimal martingale transportation problem coming from financial mathematics. This problem is obtained from the classical one by adding an additional constraint: the measure π is assumed to be a martingale (to make the space of feasible measures non-empty μ should stochastically dominate ν). The dual martingale problem has a natural financial interpretation (see [3]). More information about martingale transportation the reader can find in [2,3,13,17]. Remarkably, a duality theorem for transportation problem with general linear constraints has been obtained only recently in [37]. This results covers, in particular, the case of martingale constraints. Another important class of linear constraints are various symmetric assumptions, in particular, invariance with respect to an action of some group of linear operators. This type of problem has been studied in [14,30,37]. Applications of symmetric problem to infinite-dimensional analysis and links with ergodic theory can be found in [22,23]. The Monge–Kantorovich problems with some convex constraints has been considered in [24–26].

In this paper we consider the problem of maximization/minimization of the functional

$$\int_X c(x_1, \ldots, x_n) d\pi$$

on a set of (n, k)-stochastic measures. We call it mutistochastic or (n, k)-stochastic Kantorovich problem. Clearly, for $k = 1$ one gets the multimarginal Kantorovich problem with n-marginals, and for $k = 1, n = 2$ one gets the standard Kantorovich problem.

The system of projections $\{\mu_I\}$ can not be arbitrary for $k > 1$, and in fact, it is a nontrivial question, when the set of (n, k)-stochastic measures is non-empty. This problem illustrates the main source of difficulties for the multistochastic problem: the constraints are highly non-independent, unlike the classical Monge–Kantorovich problem. We stress that existence of feasible measures is just one question among many others which have trivial solutions for the classical case, but not for the multistochastic one. On the other hand, the classical example of latin squares and its relation with discrete algebraic structures (groups and quasigroups) ensures that it is an interesting and non-artificial object.

We start with consideration of two basic questions of the mass transportation theory: duality and cyclical monotonicity. A natural guess that the dual problem should be the maximization problem

$$\sum_I \int f_I(x_{i_1}, \ldots, x_{i_k}) d\mu_I,$$

with the constraint $\sum_I f_I(x_{i_1}, \ldots, x_{i_k}) \le c$ is verified in Sect. 3 in a form analogous to the duality theorem considered in [35]. The proof is based on the minimax principle. In Sect. 4

we prove an analog of the cyclical monotonicity property. Unfortunately, applications of the cyclical nonotonicity are not that as straightforward as in the classical case. The main difficulty here is that the set of discrete competitors is essentially more complicated that the permutation cycles considered in the classical transportation theory. We don't know, whether any solution to a multistochastic problem (for a reasonable choice of the cost function c) is concentrated on the graph of a mapping (this is a standard corollary of the cyclical monotonicity property in the classical case). The uniqueness question is open as well. We were able, however, to deduce from the cyclical monotonicity property that any solution is singular to the Lebesgue measure under assumption that the projections have densities ($k = 2, n = 3, X_i = [0, 1]$).

In Sects. 5 and 6 we study our main example: $k = 2$, $n = 3$, $X_i = [0, 1]$, the two-dimensional projections are assumed to be Lebesque and

$$c(x, y, z) = xyz.$$

Let us consider the maximization problem $\int_{[0,1]^3} xyz \to \max$. We show that there exists a solution which is concentrated on the graph of the mapping

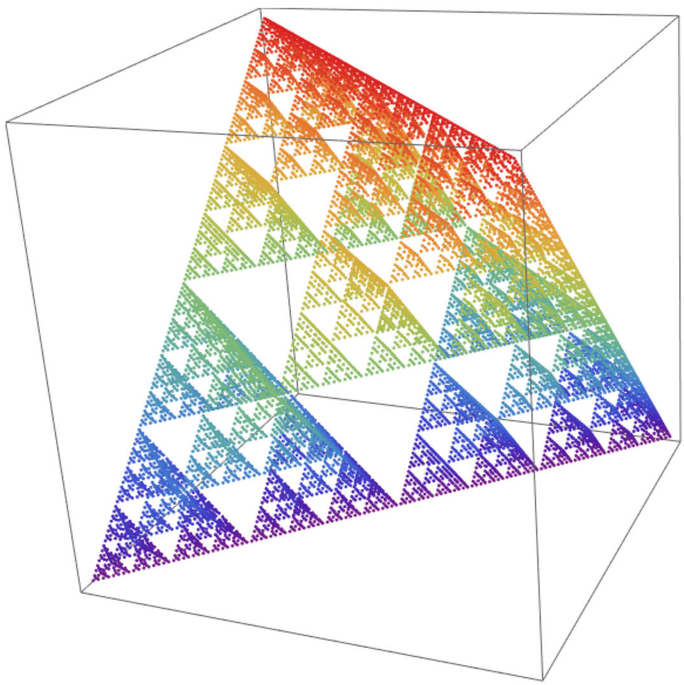
$$(x, y) \to 1 - x \oplus y,$$

where $\oplus$ is the bitwise addition, which is also called xor-addition or Nim-addition.

Similarly, for the minimization problem $\int_{[0,1]^3} xyz \to \min$ the solution π is concentrated on the graph of the mapping

$$(x, y) \mapsto x \oplus y.$$

It is known that the bitwise operations can be used to generate fractals (see [11,15]). In particular, the graph of this mapping $(x, y) \mapsto x \oplus y$ is the so-called Sierpiński tetrahedron.



This is a classical fractal self-similar set of dimension 2. In the book of Mandelbrot [28] it is briefly described under the name “fractal skewed web”: “Let us project it along a direction joining the midpoints of either couple of opposite sides. The initiator tetrahedron projects on a square, to be called initial. Each second-generation tetrahedron projects on a subsquare, namely 1/4-th initial square, etc. Thus, the web projects on the initial square. The subsquares' boundaries overlap.”

The irregularity of this example is rather unexpected, since it is well-understood that the standard solutions to the classical Monge–Kantorovich problems are supported by regular surfaces. On the other hand the close relation of latin squares to groups makes the appearence of the xor-operation (equivalently, of the group $\mathbb{Z}_2^{\infty}$) natural. The appearence of the bitwise addition can be also illustrated by the baby (3, 2)-transportation problem on the cube $\{0, 1\}^3$ with $c = xyz$. All the competitors with uniform projections are convex combinations of two measures:

$$\{(0, 0, 0), (0, 1, 1), (1, 1, 0), (1, 0, 1)\},$$
$$\{(0, 1, 0), (0, 0, 1), (1, 0, 1), (1, 1, 1)\}$$

(we identify the point and the Dirac mass with weight 1/4 at the point), defined by equations

$$x + y + z = 0, \ x + y + z = 1, \ mod(2).$$

The first measure minimizes xyz and the second measure maximizes. The fact that essentially the same structure is preserved for the cube $[0, 1]^3$ is due to the symmetries of the corresponding continuous problem. We should stress that this is not the first example of this kind, see paper of Di Marino et al. [10], Theorem 4.6. They considered solution to the n-marginal Kantorovich problem with Lebesgue meaure projections and the cost function

$$c = h(x_1 + x_2 + \cdots + x_n).$$

Let us consider again the minimization problem $\int_{[0,1]^3} xyz \to \min$. We show in Sect. 7 that

$$F(x, y) = \int_0^x \int_0^y s \oplus t \, dsdt - \frac{1}{4} \int_0^x \int_0^x s \oplus t \, dsdt - \frac{1}{4} \int_0^y \int_0^y s \oplus t \, dsdt.$$

solves the corresponding dual problem

$$\int_{[0,1]^2} F(x, y)dxdy + \int_{[0,1]^2} F(x, z)dxdz + \int_{[0,1]^2} F(y, z)dydz \to \max,$$

$$F(x, y) + F(x, z) + F(y, z) \le xyz.$$

In particular, the corresponding optimal mapping T takes the form

$$(x, y) \mapsto \partial^2_{xy} F(x, y)$$

and the Sierpiński tetrahedron is the set of zeroes of the nonnegative function

$$xyz - F(x, y) - F(x, z) - F(y, z).$$

In addition, this function is almost everywhere differentiable and homogeneous with respect to factor 2. The first derivatives of this function are not differentiable, but have bounded variation.

It is an open question which particular properties of this solution are inherited by general solutions to the (3, 2)-problem. We discuss some related hypotheses in Sect. 8.

2 Multistochastic problem: basic properties

In this short section we define the main objects of our study and discuss their basic properties. We are given a finite number of spaces

$$X_1, \dots, X_n,$$

equipped with σ-algebras

$$\mathcal{B}_1, \dots, \mathcal{B}_n.$$

The product space

$$X = X_1 \times \cdots \times X_n,$$

is equipped with the standard product of σ-algebras

$$\mathcal{B} = \mathcal{B}_1 \times \cdots \times \mathcal{B}_n.$$

The projection

$$X \ni x \to (x_{i_1}, \dots, x_{i_k})$$

of X onto $X_{i_1} \times \cdots X_{i_k}$, $i_j \in \{1, \dots, n\}$ with $i_{j_1} \neq i_{j_2}$ for distinct j_1, j_2, will be denoted by

$$\mathrm{Pr}_{X_{i_1} \times \cdots \times X_{i_k}},\ \mathrm{Pr}_i,$$

where $i = (i_1, \dots, i_k)$.

Thoroughout the paper the following assumption holds:

Assumption I X_i are Polish spaces and $\mathcal{B}_i$ are the corresponding Borel σ-algebras.

Definition 2.1 (*Multistochastic Kantorovich problem*) For every fixed $1 \le k < n$ let $\mathcal{I}_k$ be the set of all ordered k-tuples of indices $i_j \in \{1, \dots, n\}$, $i_1 < i_2 < \cdots < i_{k-1} < i_k$. Assume that for every k-tuple

$$I = (i_1, \dots, i_k) \in \mathcal{I}_k$$

we are given a probability measure $\mu_I = \mu_{i_1,\dots,i_k}$ on $X_{i_1} \times \cdots \times X_{i_k}$. Denote by $\mathcal{P}_\mu$ the set of probability measures on X satisfying

$$\mathrm{Pr}_I \mu := \mathrm{Pr}_{X_{i_1} \times \cdots \times X_{i_k}} \mu = \mu_I.$$

Finally, assume that we are given a cost function

$$c \colon \prod_{i=1}^{n} X_i \to \mathbb{R}_+ \cup \{+\infty\}.$$

Then we say that $P \in \mathcal{P}_\mu$ is a solution to the (n, k)-Kantorovich minimization problem for c and $\{\mu_i\}$, $I \in \mathcal{I}_k$, if P gives minimum to the functional

$$P \to \int_X c\, dP$$

on $\mathcal{P}_\mu$.

We call the problem "(n, k)-Kantorovich maximization problem" if instead of minimum we are looking for maximum of $P \to \int_X c\, dP$.

Unlike the standard Kantorovich problem $\mathcal{P}_\mu$ can be empty. Let us briefly discuss some sufficient conditions assuring that $\mathcal{P}_\mu$ is not empty. For the sake of simplicity we restrict ourselves to the case $n = 3$, $k = 2$, $X_i = [0, 1]$.

A natural necessary assumption for $\mathcal{P}_\mu \neq \emptyset$ is the following Kolmogorov-type consistency condition.

Remark 2.2 If the set $\mathcal{P}_\mu$ is not empty, then

$$\mathrm{Pr}_1\mu_{1,2} = \mathrm{Pr}_1\mu_{1,3},\ \mathrm{Pr}_2\mu_{1,2} = \mathrm{Pr}_2\mu_{2,3},\ \mathrm{Pr}_3\mu_{2,3} = \mathrm{Pr}_3\mu_{1,3}. \tag{2.1}$$

Remark 2.3 One can naively think that (2.1) is a sufficient condition for $\mathcal{P}_\mu \neq \emptyset$. But this is not true. Consider the following example: $\mu_{1,2}$ is given by the (normalized) Lebesgue measure on the diagonal $\{x = y\}$, $x \in [0,\ 1]$, $y \in [0,\ 1]$ and $\mu_{1,3}$ is the two-dimensional Lebesgue measure on $[0,\ 1]^2$.

Since the projection of the set $\{x = y\} \times [0,\ 1]$ onto $0 \le x \le 1,\ 0 \le z \le 1$ along y is a one-to-one mapping, there exist the unique measure on $[0,\ 1]^3$ with projections $\mu_{1,2}$, $\mu_{1,3}$. It is the normalized Lebesgue measure on the diagonal $\{x = y = z\}$, $x \in [0,\ 1]$, $y \in [0,\ 1]$, $z \in [0,\ 1]$ Denote this measure by π.

In this construction $\mu_{2,3}$ was not used, so if $\mu_{2,3} \neq Pr_{2,3}(\pi)$, there exist no measure on I with projections $\mu_{1,2}$, $\mu_{1,3}$ and $\mu_{2,3}$. But one can easily find $\mu_{2,3}$ different from $Pr_{2,3}(\pi)$ such that (2.1) holds, for example $\mu_{2,3}$ equals Lebesgue measure on $[0,\ 1]^2$.

Some sufficient condition for $\mathcal{P}_\mu \neq \emptyset$ can be found in [19].

Remark 2.4 An important example of a non-empty set $\mathcal{P}_\mu$ is given by the following system of projections (for the sake of simplicity $k = 2$):

$$\mu_{i,j} = \mu_i \times \mu_j,$$

where every μ_i is a probability measure on X_i.

Assumption II It will be assumed throughout that $\mathcal{P}$ is non-empty.

The proof of the following result is omitted because it is a simple repetition of the proof of the corresponding fact for the standard Kantorovich problem (see [7,35]).

Theorem 2.5 *Assume that c is a lower semicontinuous function. Then there exists $P \in \mathcal{P}_\mu$ giving minimum to the functional $P \to \int c\,dP$ on $\mathcal{P}_\mu$.*

3 Duality

In this section we prove a duality theorem for the multistochastic problem. It can be deduced from the following general minimax result (see [35], Theorem 1.9) in the same way as the duality theorem for the standard Kantorovich problem. The arguments are essentially the same, we repeat the proof for the reader convenience. For the sake of simplicity we restrict ourselves to the case of compact spaces.

Theorem 3.1 *Let E be a normed vector space and E^* be the corresponding topologically dual space. Consider convex functionals Φ, Ψ on E with values in $\mathbb{R} \cup \{+\infty\}$. Let Φ^*, Ψ^* be their Legendre transforms. Assume that there exists a point $z \in E$ satisfying $\Phi(z) < +\infty$, $\Psi(z) < +\infty$ and Φ is continuous at z. Then*

$$\inf_E(\Phi + \Psi) = \max_{z \in E^*}(-\Phi^*(-z) - \Psi^*(z))$$

Theorem 3.2 *Let X_i be compact metric spaces and $c \geq 0$ be a continuous function on X. Then*

$$\min_{\pi \in \mathcal{P}} \int c d\pi = \sup \sum_{i \in \mathcal{I}_k} \int f_i(x_{i_1}, \dots, x_{i_k}) d\mu_i,$$

there the infimum is taken over the k-tuples $i = (i_1, \dots, i_k) \in \mathcal{I}_k$ and the functions $f_{i_1,\dots,i_k} \in L_1(X_{i_1} \times \cdots \times X_{i_k}, \mu_{(i_1,\dots,i_k)})$ satisfying

$$\sum_{i \in \mathcal{I}_k} f_i(x_{i_1}, \dots, x_{i_k}) \leq c.$$

Proof Let E be the space of continuous functions on X. By Radon's theorem E^* is the space of finite (signed) measures on X.

Set:

$$\Phi(u) = 0, \text{ if } u \geq -c$$

and $\Phi(u) = +\infty$ in the opposite case.

Let π_0 be a probability measure which belongs to $\mathcal{P}_\mu$. For every function u which has representation $u = \sum_{i \in \mathcal{I}_k} f_i$ we set

$$\Psi(u) = \int u d\pi_0 = \sum_{i \in \mathcal{I}_k} \int f_i d\mu_i, \text{ if } u = \sum_{i \in \mathcal{I}_k} f_i,$$

and $\Psi(u) = +\infty$ otherwise. It is easy to check that the functionals satisfy assumptions of Theorem 3.1.

Clearly

$$\inf_u \Big(\Phi(u) + \Psi(u)\Big) = - \sup_{\sum_{i \in \mathcal{I}_k} f_i \leq c} \sum_{i \in \mathcal{I}_k} \int f_i d\mu_i.$$

Let us find the Legendre transform of the functionals

$$\Phi^*(-\pi) = \sup_u \Big(-\int u d\pi - \Phi(u)\Big) = \sup_{u \geq -c} \Big(-\int u d\pi\Big) = \sup_{u \leq c} \int u d\pi.$$

It is easy to see that $\Phi^*(-\pi) = \int c d\pi$, if π is a non-negative measure. If not, then clearly $\Phi^*(-\pi) = +\infty$.

Let us compute $\Psi^*(\pi) = \sup_u \Big(\int u d\pi - \Psi(u)\Big)$. Clearly $\Psi^*(\pi) = 0$, if $\Pr_i \pi = \mu_i$ and $\Psi^*(\pi) = +\infty$ in the opposite case. This implies

$$\max_{z \in E^*} (-\Phi^*(-\pi) - \Psi^*(\pi)) = - \min_{Pr_i \pi = \mu_i} \int c d\pi.$$

The proof is complete. □

4 Cyclical monotonicity

Starting from this section we work with the following particular case:

$$n = 3, k = 2,$$
$$c(x, y, z) = xyz.$$

Here we consider the **maximization** problem

$$\int_{[0,1]^3} xyz d\pi \to \max, \ \pi \in \Pi(\mu_{12}, \mu_{23}, \mu_{13}).$$

This problem seems to be the simplest but important and illustrative particular case of the multistochastic Kantorovich problem. The choice of the cost function is natural in view of the examples which will be given below. In addition, it is analogous to the simplest quadratic Kantorovich problem with one-dimensional marginals. Indeed, the minimization of $\int |x-y|^2 d\pi$ on the set of measures $\Pi(\mu, \nu)$ with fixed marginals μ, ν is equivalent to maximization of $\int xyd\pi$ on the same set.

In the case of the standard Kantorovich problem with two marginals the well-known cyclical monotonicity property fully characterizes the solutions. In particular, if the marginals are one-dimensional, then the solution is concentrated on the graph of a monotone function. In this section we prove a weak analog of this property for our special multistochastic problem. Unlike the standard Kantorovich problem with one-dimensional marginals, the geometric structure of the sets which are cyclically monotone in our sense is essentially less clear.

To show cyclical monotonicity we follow an approach from [2] (Lemma 1.11).

Assume we are given three finite sets

$$X = \{x_1, \dots x_n\} \subset \mathbb{R}, \ Y = \{y_1, \dots y_n\} \subset \mathbb{R}, \ Z = \{z_1, \dots z_n\} \subset \mathbb{R}$$

of cardinality n.

In what follows we denote by

$$U(X, Y, Z)$$

the set of discrete probability measures on $X \times Y \times Z$ which have uniform projections onto $X \times Y$, $X \times Z$, $Y \times Z$. Among all measures in $U(X, Y, Z)$ let us consider a special important subclass of uniform distributions π_G on the sets of the type

$$G = (x, y, f(x, y)), \ x \in X, \ y \in Y, \ f : X \times Y \to Z,$$

where f admits the following property: fix any $z_i \in Z$, then for every $x_j \in X$ there exists exactly one $y_k \in Y$ and for every $y_l \in Y$ there exists exactly one $x_m \in X$ such that

$$z_i = f(x_i, y_k) = f(x_m, y_l).$$

Then the uniform measure π_G on G belongs to $U(X, Y, Z)$.

The set

$$L(X, Y, Z)$$

of such G can be identified with $n \times n$ **latin squares**.

Remark 4.1 Let us mention another important difference between the multistochastic and the standard Kantorovich problem. By the classical theorem of Birkhoff every bistochastic matrix is a convex combination of the permutation matrices. In the multistochastic case there is no analog of the Birkhoff theorem: not every multistochastic matrix is a convex combination of matrices with entries (a_{ij}) satisfying $a_{ij} = 0$ or $a_{ij} = 1$. An example is given by

$$\left[\begin{array}{ccc|ccc|ccc} 0.5 & 0 & 0.5 & 0 & 0.5 & 0.5 & 0.5 & 0.5 & 0 \\ 0 & 1 & 0 & 0.5 & 0 & 0.5 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0.5 & 0.5 & 0 & 0 & 0.5 & 0.5 \end{array}\right]$$

See [27] for explanations and [9] for descriptions of extremal points.

Definition 4.2 Let $\Gamma \subset \mathbb{R}^3$ be a finite set and π_Γ be the uniform measure on Γ. We say that Γ' is the competitor of Γ if π_Γ and $\pi_{\Gamma'}$ have the same projections onto the principal hyperplanes

$$Pr_{xy}\Gamma = Pr_{xy}\Gamma', \;\; Pr_{yz}\Gamma = Pr_{yz}\Gamma', \;\; Pr_{xz}\Gamma = Pr_{xz}\Gamma'.$$

Definition 4.3 The set $\Gamma \subset \mathbb{R}^3$ is called cyclically monotone if for every natural n and $G \subset \Gamma$ of cardinality n

$$\int xyz d\pi_{G'} \le \int xyz d\pi_G$$

for every competitor of G.

Theorem 4.4 *Let π be a solution to a* (3, 2)*-multistochastic Kantorovich maximization problem with $c(x, y, z) = xyz$. Then there exists a cyclically monotone set Γ with $\pi(\Gamma) = 1$.*

Proof For every n let

$$M_n = \Big\{ G \subset \mathbb{R}^3, \text{card}(G) = n, \text{ there exists a competitor } G' \text{ such that}$$
$$\int xyz d\pi_G < \int xyz d\pi_{G'} \Big\} \subset (\mathbb{R}^3)^n.$$

According to ([4], Proposition 2.1), derived from general duality result of Kellerer [20], one of the following two options holds:

(1) M_n is contained in a set of the type $\cup_{i=1}^n \mathbb{R}^3 \times \cdots \times M_n^i \times \cdots \times \mathbb{R}^3$ with $\pi(M_n^i) = 0$.
(2) There exists a measure γ on M_n such that $\gamma(M_n) > 0$ and $\Pr_i(\gamma) \le \pi$ for every i.

We will show that (2) is impossible. Thus (1) holds for every n and

$$\Gamma = \mathbb{R}^3 \setminus \cup_{n=1}^\infty \cup_{i=1}^n M_n^i$$

is the desired cyclically monotone set.

Assume that (2) holds for some n. Set $\gamma' = \frac{1}{n}\sum_{i=1}^n \Pr_i(\gamma)$. Clearly, for the uniform measure π_G on $G \in M_n$ there holds $\pi_G = \frac{1}{n}\sum_{i=1}^n \Pr_i(\delta(G))$ and

$$\gamma' = \int \pi_G \, d\gamma(G).$$

By definition of M_n for γ-a.e. G there exists a competitor G' such that and $\int xyz d\pi_G < \int xyz d\pi_{G'}$. Moreover, using linear programming algorithms one can make the correspondence $G \to G'$ measurable. Define

$$\tilde{\gamma} = \int \pi_{G'} d\gamma(G)$$

Clearly $\int xyz d\tilde{\gamma} > \int xyz d\gamma'$ and $\gamma', \tilde{\gamma}$ have the same projection onto the principal hyperplanes. Then we set $\pi' = \pi - \gamma' + \tilde{\gamma}$. The measures π, π' have the same projections onto the principal hyperplanes and π' has a larger total cost. We obtain a contradiction. □

Example 4.5 The simplest example of a set G which belongs to some $L(X, Y, Z)$ is given by the following four-points set with uniform projections on the products of some two-points sets

$$X_1 = (a_1, b_1, c_2), X_2 = (a_1, b_2, c_1), X_3 = (a_2, b_1, c_1), X_4 = (a_2, b_2, c_2).$$

The set G is cyclically monotone for $c = xyz$ if and only if

$$(a_1 - a_2)(b_1 - b_2)(c_1 - c_2) \le 0.$$

The well-known and by now classical result of Y. Brenier establishes existence of the so-called optimal transportation mapping in the classical setting. We don't know whether the multistochastic Kantorovich problem admits the same property. However, applying the cyclical monotonicity property proved in Proposition 4.4 we are able to show a weak version of the Brenier theorem saying that under natural assumptions π is a singular measure.

Let us denote by λ^n the standard n-dimensional Lebesgue measure.

Lemma 4.6 *Let $A \subset \mathbb{R}^3$ be a Borel set of positive Lebesgue measure. There exist numbers*

$$x_1 < x_2, y_1 < y_2, z_1 < z_2,$$

such that $\{x_1, x_2\} \times \{y_1, y_2\} \times \{z_1, z_2\} \subset A$.

Proof Without loss of generality let us consider bounded sets. By Fubini's theorem one gets that for every $\varepsilon > 0$ the set A_z of numbers $\tilde{z}$ satisfying

$$\lambda^2(A \cap \{z = \tilde{z}\}) > \varepsilon$$

has a non-zero Lebesgue measure. Hence there exists two points $z_1, z_2 \in A_z$, such that the projections $A \cap \{z = z_1\}$, $A \cap \{z = z_2\}$ onto the hyperplane xy have an intersection B of a positive measure. Hence $B \times \{z_1, z_2\} \subset A$. Next we apply the same arguments to the one-dimensional sections of B: $\{y = \tilde{y}\} \cap B$. This completes the proof. □

Corollary 4.7 *Every cyclically monotone set $\Gamma \subset \mathbb{R}^3$ satisfies $\lambda^3(\Gamma) = 0$.*

Proof Assume that $\lambda^3(\Gamma) > 0$. Then according to Lemma 4.6 there exist numbers $x_1 < x_2, y_1 < y_2, z_1 < z_2$, such that $\{x_1, x_2\} \times \{y_1, y_2\} \times \{z_1, z_2\} \subset \Gamma$. We get a contradiction with Example 4.5. □

5 Main example: primal problem

In this section we consider our main example: (3, 2)-Kantorovich problem on the unit three-dimensional cube $[0, 1]^3$, where the projections onto principal hyperplanes are equal to two-dimensional Lebesgue measure λ^2. The cost function is given by

$$c(x, y, z) = xyz.$$

The set of measures with such projections will be denoted by $\mathcal{P}_\lambda$. We are looking for

$$\max_{P \in \mathcal{P}_\lambda} \int xyz \, dP. \tag{5.1}$$

In this concerete example we are able to find an explicit solution. We emphasize that this is possible because the problem admits many symmetries. We don't know whether the problem has an explicit solution even after slight changes, for instance, in the case when the projections are equal to products $\mu_i \times \mu_j$, where $\{\mu_i\}$ are fixed one-dimensional distributions.

We denote by $\oplus$ the bitwise addition (xor). Given two couples of numbers $x, y \in [0, 1]$ we consider their diadic decompositions

$$x = \overline{0, x_1 x_2 x_3 \ldots}, \quad y = \overline{0, y_1 y_2 y_3 \ldots}, \quad x_i, y_i \in \{0, 1\}.$$

Then the xor operation is defined as follows:

$$x \oplus y = \overline{0, x_1 \oplus y_1 \; x_2 \oplus y_2 \; x_3 \oplus y_3 \ldots},$$

where $0 \oplus 0 = 1 \oplus 1 = 0, 0 \oplus 1 = 1 \oplus 0 = 1$.

Remark 5.1 The addition is not well-defined for dyadic rational numbers, because they can be written in two different ways. We agree that every dyadic rational number less then 1 has a finite numbers of units in its decomposition. The number $x = 1$ will be always decomposed in the following way:

$$1 = \overline{0, 11111 \ldots}.$$

Thus

$$x \oplus 1 = 1 - x.$$

This operation is continuous up to a countable set of dyadic numbers.

Theorem 5.2 *The image π of two-dimensional Lebesgue measure $\lambda(dx) \times \lambda(dy)$ under the mapping*

$$T : (x, y) \to (x, y, 1 - x \oplus y)$$

is a solution to problem (5.1).

If instead of maximizing the total cost function one asks for

$$\min_{P \in \mathcal{P}_\lambda} \int xyz \, dP,$$

then the corresponding mapping T is given by

$$T : (x, y) \to (x, y, x \oplus y).$$

Remark 5.3 We don't know whether this concrete problem and the problem in general setting (for an appropriate cost function) has unique solution. In this example there exists a corresponding optimal mapping, but we don't know whether the same is true for any $(3, 2)$-problem (under appropriate assumptions on the projections).

Proof Let us consider the following transformations of $[0, 1]^3$

$$T_{xy}(x, y, z) = (1 - x, 1 - y, z),$$
$$T_{xz}(x, y, z) = (1 - x, y, 1 - z),$$
$$T_{yz}(x, y, z) = (x, 1 - y, 1 - z).$$

All these transformations push forward arbitrary measure $\mu \in \mathcal{P}_\lambda$ onto a measure from $\mathcal{P}_\lambda$. We define

$$\mu^{xy} = \mu \circ T_{xy}^{-1}, \; \mu^{xz} = \mu \circ T_{xz}^{-1}, \; \mu^{yz} = \mu \circ T_{yz}^{-1}.$$

Next we note that every $\mu \in \mathcal{P}_\lambda$ satisfies

$$\int xyz d\mu^{xy} = \int (z - xz - yz + xyz) d\mu$$
$$= \int xyz d\mu + \int_0^1 z dz - \int_0^1 \int_0^1 xz dx dz - \int_0^1 \int_0^1 yz dy dz = \int xyz d\mu.$$

Thus the total cost $\int xyz d\mu$ is invariant with respect to T^{xy} (and with respect to T^{yz}, T^{xz}). Hence it follows that for every $\tilde{\pi}$ solving (5.1) the measures $\tilde{\pi}^{xy}, \tilde{\pi}^{yz}, \tilde{\pi}^{xz}$, and

$$\pi_1 = \frac{\tilde{\pi} + \tilde{\pi}^{xy} + \tilde{\pi}^{xz} + \tilde{\pi}^{yz}}{4}$$

are solutions to problem (5.1) as well. Note that π_1 is invariant with respect to T^{xy}, T^{yz}, T^{xz}. This follows from the relations

$$T^{xy} T^{xz} = T^{xz} T^{xy} = T^{yz}, \ T^{xy} T^{xy} = \mathrm{Id}.$$

Next we decompose $[0, 1]^3$ into sets I_1, I_2. Every $I_i, i \in \{1, 2\}$ is a union of four smaller cubes of volume $1/2^3$:

$$I_1 = \overline{[0, 1]^3 \backslash I_2}$$
$$I_2 = \left[0, \frac{1}{2}\right]^3 \bigcup \left(\left[\frac{1}{2}, 1\right]^2 \times \left[0, \frac{1}{2}\right]\right) \bigcup \left(\left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]^2\right)$$
$$\bigcup \left(\left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]\right).$$

Since every set I_1, I_2 is invariant under T_{xy}, T_{yz}, T_{xz}, the measures

$$\pi_{I_1} = (\pi_1)|_{I_1}, \ \pi_{I_2} = (\pi_1)|_{I_2}$$

are invariant as well. Hence the push-forward image

$$\pi_{I_2}^x = \pi_{I_2} \circ T_x^{-1}$$

of measure π_{I_2} with respect to $T_x \colon (x, y, z) \mapsto (1 - x, y, z)$ has the same hyperplane projections as π_{I_2}. Thus

$$\pi_{I_1} + \pi_{I_2}^x$$

belongs to $\mathcal{P}_\lambda$.

Let us show that $\pi_{I_2} = 0$. To this end it is sufficient to show that

$$\int xyz d\mu < \int xyz d\hat{\mu},$$

where $\hat{\mu} = \mu \circ (T^x)^{-1}$, for every non-zero measure μ, which is invariant with respect to T^{xy}, T^{yz}, T^{xz}, and sastisfies supp($^-$) $\subset$ I$_2$. Indeed, if we show this, then we get

$$\int xyz d\pi_{I_1} + \int xyz d\pi_{I_2}^x > \int xyz d\pi_{I_1} + \int xyz d\pi_{I_2}.$$

The latter implies that measure $\pi_{I_1} + \pi_{I_2}^x$ gives better value to the total cost function.

Let ν be the projections of μ (hence, projections of $\hat{\mu}$) onto x-axis, and $\eta^x(dy, dz)$, $\hat{\eta}^x(dy, dz)$ are corresponding conditional measures

$$\mu = \nu(dx)\eta^x(dydz),$$
$$\hat{\mu} = \nu(dx)\hat{\eta}^x(dydz).$$

Note that η is invariant with respect to T^{yz} and

$$\hat{\eta}^x = \eta^x \circ T_y^{-1} = \eta^x \circ T_z^{-1} = \eta^{1-x} = \hat{\eta}^{1-x} \circ T_z^{-1}. \tag{5.2}$$

Hence

$$\begin{aligned}\int xyz(d\mu - d\hat{\mu}) &= \int \left(\int yz(d\eta^x - d\hat{\eta}^x)\right) x\nu(dx) \\ &= \int_0^{\frac{1}{2}} \left(\int yz(d\eta^x - d\hat{\eta}^x)\right) x\nu(dx) + \int_{\frac{1}{2}}^1 \left(\int yz(d\eta^x - d\hat{\eta}^x)\right) x\nu(dx) \\ &= \int_{\frac{1}{2}}^1 \left(\int yz(d\eta^x - d\hat{\eta}^x)\right) (2x-1)\nu(dx).\end{aligned}$$

Next, using T^{zy}-invariance of η and (5.2), one gets

$$\int yz(d\eta^x - d\hat{\eta}^x) = \frac{1}{2}\left(\int (yz + (1-z)(1-y))(d\eta^x - d\hat{\eta}^x)\right)$$
$$= \frac{1}{2}\int \big(yz + (1-z)(1-y) - (1-y)z - y(1-z)\big)d\eta^x = \frac{1}{2}\int (2y-1)(2z-1)d\eta^x.$$

Finally,

$$\int xyz(d\mu - d\hat{\mu}) = \frac{1}{2}\int_{\frac{1}{2}}^1 \left[\int (2y-1)(2z-1)d\eta^x(dzdy)\right](2x-1)\nu(dx).$$

Since the support of μ lies in I_2, one gets $\int xyz(d\mu - d\hat{\mu}) < 0$.

Thus we get that the support of π_1 belongs to the union of four disjoint cubes with volume $1/2^3$

$$J_1 = I_1 = C_1 \cup C_2 \cup C_3 \cup C_4.$$

Hence the restriction of π_1 onto every cube C_i is a solution of (2.1) for the same cost function with marginals which are restrictions of Lebesgue measure on projections of correspoding C_i. Hence the same arguments are applicable to every C_i and one gets a solution π_2 supported on a union of 16 cubes of volume $1/4^3$

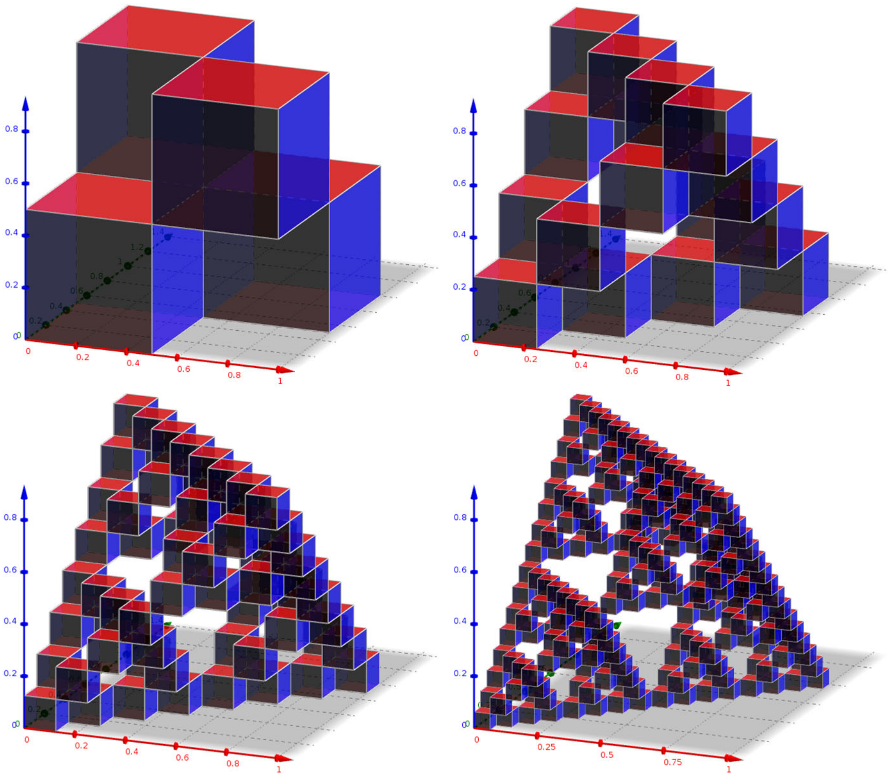
$$J_2 = \cup_{i=1}^4 \cup_{j=1}^4 C_{ij}.$$

Repeating this argument one gets a sequence of decreasing sets J_n such that each of them contains support of a measure π_n which solves (2.1). Clearly, the sequence $\{\pi_n\}$ admits a weak limit π supported on

$$J = \cap_{n=1}^{\infty} J_n.$$

We get immediately that π solves the desired problem, moreover J is a graph of $T(x, y)$ (up to a set which projection on xy has zero measure) and π is the unique measure supported on J with the desired projections. □

The following pictures represent the iteration procedure.



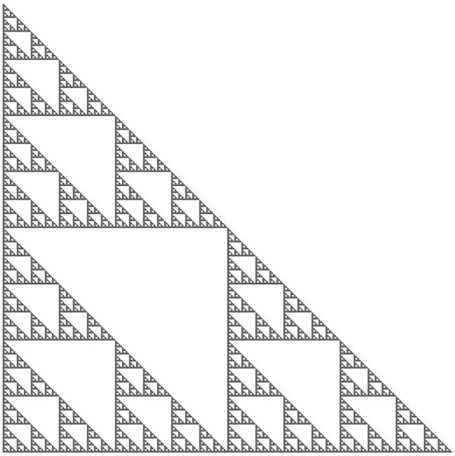
As we already mentioned, J is a self-similar fractal of Hausdorff dimension two, called "Sierpiński tetrahedron". This is a Kantor-type set which is a limit of iterations of unions of 4^n tetrahedrons. Remarkably, in our proof we get an alternative construction and obtain J as an intersection of collections of cubes.

Remark 5.4 The most trivial example of a fractal solution to the Monge–Kantorovich problem is apparently the (3, 2)-Kantorovich problem with Lebesgue measure projections and $c = 1 - x \oplus y$. Then the solution is again the Sierpiński tetrahedron. But this is due to a special choice of the cost function. Unlike this, our main example deals with the smooth cost function $c = xyz$ and the extremality of the presented solution is highly non-obvious. In addition, we will see in the subsequent sections that a solution to the corresponding dual problem provides a non-trivial representation of the Sierpiński tetrahedron as a set of zeroes of an a.e. differentiable function.

Less trivial example is given by measures supported on the set

$$T = \{x + y = x \oplus y,\ x \in [0, 1],\ y \in [0, 1]\}, \tag{5.3}$$

which is a variant of the Sierpiński triangle (see [11]).



Note that all $x \in [0,1]$, $y \in [0,1]$ satisfy

$$x + y \geq x \oplus y.$$

Let π be any probability measure on T with projections $Pr_x\pi = \mu$, $Pr_y\pi = \nu$. Consider the Monge–Kantorovich problem

$$\int_{[0,1]^2} x \oplus y \, dP \to \max, \ Pr_x P = \mu, \ Pr_y\pi = P. \tag{5.4}$$

By the Kantorovich duality principle the functions x, y solve the corresponding dual problem. Hence π is a solution to (5.4).

In particular, the self-similar measure π_0 on T solves problem (5.4) with marginals $\mu = \nu$, where μ can be described as the distribution of the series $\sum_{i=1}^{\infty} \frac{\xi_i}{2^i}$, where the sequence of i.i.d. Bernoulli random variables $\{\xi_i\}$ satisfies $\xi_i = 1$ with probability $1/3$ and $\xi_i = 0$ with probability $2/3$. Another example is the (normalized) Lebesgue measure on the main diagonal.

Another example of a fractal solution for smooth cost function can be found in the paper of Di Marino et al. [10], Theorem 4.6. The multimarginal Kantorovich prolem with Lebesgue projections and the cost function

$$h(x_1 + x_2 + \cdots + x_n),$$

where h is convex admits a fractal solution related to n-base decompositions.

6 Main example: dual problem

For the problem

$$\int xyz d\pi \to \min,$$

where π has Lebesgue projections onto principal hyperplane, let us consider the corresponding dual problem:

$$\int F(x,y)dxdy + \int G(y,z)dxdy + \int H(z,x)dxdz \to \max, \tag{6.1}$$

$$F(x,y) + G(y,z) + H(z,x) \le xyz. \tag{6.2}$$

It is clear that by symmetries of the problem one can reduce the general problem to the case

$$F = G = H,\ F(x,y) = F(y,x).$$

Let us remind to the reader that by the standard duality arguments any function F satisfying (6.2) and

$$F(x,y) + G(y,z) + H(z,x) = xyz,\ z = x \oplus y$$

(x, y)-almost everywere is a solution to (6.1).

Discretizing the problem and performing finite-dimensional linear programming algorithms we were able to guess reccurent relations for the restriction of F onto the set of dyadic rational numbers. Using these relations we prove the desired properties of our function. Finally, we will give an integral representation for the solution in the next section.

6.1 Definition and easy properties

Let $\mathbb{N}_0$ be the set of all non-negative integers.

Definition 6.1 Let $f : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{Z}$ be a function defined as follows. Set:

$$f(0,0) = 0,\ f(0,1) = f(1,0) = -1,\ f(1,1) = 2.$$

In all other points f is defined by the following recurrent relations:

$$f(a,b) = \begin{cases} 8f(\frac{a}{2}, \frac{b}{2}) \text{ if } a \equiv 0 \pmod 2 \text{ and } b \equiv 0 \pmod 2, \\ 4\left(f(\frac{a-1}{2}, \frac{b}{2}) + f(\frac{a+1}{2}, \frac{b}{2})\right) + 3 \text{ if } a \equiv 1 \pmod 2 \text{ and } b \equiv 0 \pmod 2, \\ 4\left(f(\frac{a}{2}, \frac{b-1}{2}) + f(\frac{a}{2}, \frac{b+1}{2})\right) + 3 \text{ if } a \equiv 0 \pmod 2 \text{ and } b \equiv 1 \pmod 2, \\ 2\left(f(\frac{a-1}{2}, \frac{b-1}{2}) + f(\frac{a-1}{2}, \frac{b+1}{2}) + f(\frac{a+1}{2}, \frac{b-1}{2}) + f(\frac{a+1}{2}, \frac{b+1}{2})\right) + 2, \\ \text{if } a \equiv 1 \pmod 2 \text{ and } b \equiv 1 \pmod 2. \end{cases} \tag{6.3}$$

The following properties can be immediately derived from the definition.

$$f(a,b) = f(b,a). \tag{6.4}$$

If a is odd and b is even, then

$$f(a,b) = \frac{1}{2}(f(a+1,b) + f(a-1,b)) + 3. \tag{6.5}$$

If a is odd and b is odd, then

$$f(a,b) = \frac{1}{2}(f(a+1,b) + f(a-1,b)) - 2. \tag{6.6}$$

$$f(a,b) \equiv a + b \pmod 2. \tag{6.7}$$

6.2 Continuity

Using the homogeneity relation

$$f(2a, 2b) = 8f(a, b)$$

with factor two one can define $f_C(x, y)$ for any non-negative binary-rational x and y. Namely, assume that $(x, y) = (\frac{a}{2^n}, \frac{b}{2^n})$, then one can set $f_C(x, y) = 8^{-n} f(a, b)$. It is easy to check that f_C is well-defined. In what follows we extend f_C to all pairs of non-negative real numbers by continuity. To this end we need some estimates of the increments of f.

Consider a family of integer segments I_n: $I_n = [0, 2^{n+1}]$, $n \geq 0$. Note that for any $a \in I_n$ with $n \geq 1$ the numbers $\frac{a}{2}$ for even a, and $\frac{a+1}{2}$ and $\frac{a-1}{2}$ for odd a, belong to the segment I_{n-1}.

Set:

$$N_{n,m} = \max(|f(a+1, b) - f(a, b)| : a, (a+1) \in I_n, b \in I_m).$$

Lemma 6.2 *There exists universal constant C, such that $N_{n,m} \leq C(4^n + 4^m)$.*

Proof It will be convenient to prove more general inequality $N_{n,m} \leq C_1(4^n + 4^m) + C_2$ applying induction method. At the end we obtain that C_2 can take negative values.

Base of induction for $n = m = 0$ can be checked directly: $N_{0,0} = 15 \leq 2C_1 + C_2$.

To prove the step of induction let us estimate $|f(a+2, b) - f(a, b)|$, where $b \in I_m$, $a, (a+2) \in I_n$ and a is even.

Let b be even. Then $|f(a+2, b) - f(a, b)| = 8|f(\frac{a}{2}+1, \frac{b}{2}) - f(\frac{a}{2}, \frac{b}{2})|$. If n and m are both strictly positive, we obtain by induction hypothesis

$$8\left|f\left(\frac{a}{2}+1, \frac{b}{2}\right) - f\left(\frac{a}{2}, \frac{b}{2}\right)\right| \leq N_{m-1,n-1}.$$

If only one number (say, m) is positive, then

$$8\left|f\left(\frac{a}{2}+1, \frac{b}{2}\right) - f\left(\frac{a}{2}, \frac{b}{2}\right)\right| \leq N_{m-1,0}.$$

In any case one gets

$$\begin{aligned} 8\left|f\left(\frac{a}{2}+1, \frac{b}{2}\right) - f\left(\frac{a}{2}, \frac{b}{2}\right)\right| &\leq 8\left(C_1\left(4^{n-1} + 4^{m-1} + \frac{3}{4}\right) + C_2\right) \\ &= 2C_1(4^n + 4^m) + (6C_1 + 8C_2). \end{aligned} \tag{6.8}$$

Here we used inequality $4^{\max(n-1,0)} + 4^{\max(m-1,0)} \leq 4^{n-1} + 4^{m-1} + \frac{3}{4}$, which holds provided one of the numbers n, m is positive.

Using that $a+1$ is odd and applying the recurrent relations (6.5) one gets

$$\begin{aligned} f(a+1, b) &= \frac{1}{2}(f(a, b) + f(a+2, b)) + 3, \\ f(a+2, b) - f(a+1, b) &= \frac{1}{2}(f(a+2, b) - f(a, b)) - 3, \\ f(a+1, b) - f(a, b) &= \frac{1}{2}(f(a+2, b) - f(a, b)) + 3. \end{aligned}$$

These estimates imply that $|f(a+1,b)-f(a,b)|$ and $|f(a+2,b)-f(a+1,b)|$ can be estimated from above by

$$\begin{aligned}&\frac{1}{2}|f(a+2,b)-f(a,b)|+3\leq C_1(4^n+4^m)+(3C_1+4C_2+3)\\&\leq C_1(4^n+4^m)+C_2,\end{aligned}\tag{6.9}$$

provided $3C_1+4C_2+3\leq C_2$.

Hence we obtain that for any even $b\in I_m$ and for any even a, $(a+1)\in I_n$ the following inequality holds: $|f(a+1,b)-f(a,b)|\leq C_1(4^n+4^m)+C_2$.

Let now b be odd. We estimate $|f(a+2,b)-f(a,b)|$ for any even a satisfying a, $(a+2)\in I_n$ in a similar manner. Using recurrent relations (6.3) we obtain:

$$\begin{aligned}f(a+2,b)-f(a,b)&=4\left[f\left(\frac{a}{2}+1,\frac{b+1}{2}\right)-f\left(\frac{a}{2},\frac{b+1}{2}\right)\right]\\&\quad+4\left[f\left(\frac{a}{2}+1,\frac{b-1}{2}\right)-f\left(\frac{a}{2},\frac{b-1}{2}\right)\right]+6\\&\leq 8\left[C_1\left(4^{n-1}+4^{m-1}+\frac{3}{4}\right)+C_2\right]+6=2C_1(4^n+4^m)\\&\quad+(6C_1+8C_2+6).\end{aligned}$$

Next we estimate $|f(a+1,b)-f(a,b)|$ and $|f(a+2,b)-f(a+1,b)|$. Since $a+1$ and b are odd, one gets applying (6.6)

$$\begin{aligned}f(a+1,b)&=\frac{1}{2}(f(a,b)+f(a+2,b))-2,\\f(a+2,b)-f(a+1,b)&=\frac{1}{2}(f(a+2,b)-f(a,b))+2,\\f(a+1,b)-f(a,b)&=\frac{1}{2}(f(a+2,b)-f(a,b))-2.\end{aligned}$$

Finally,

$$\begin{aligned}&|f(a+1,b)-f(a,b)|,|f(a+2,b)-f(a+1,b)|\\&\leq\frac{1}{2}|f(a+2,b)-f(a,b)|+2\leq C_1(4^n+4^m)+3C_1+4C_2+5\\&\leq C_1(4^n+4^m)+C_2,\end{aligned}$$

provided that $3C_1+4C_2+5\leq C_2$.

Now we get that for all odd $b\in I_m$ and for all a, $(a+1)\in I_n$ one has

$$|f(a+1,b)-f(a,b)|\leq C_1(4^n+4^m)+C_2.$$

This implies $N_{n,m}\leq C_1(4^n+4^m)+C_2$, which completes the induction step.

To conclude it is sufficient to find solutions C_1 and C_2 to the following system of inequalities

$$\begin{cases}2C_1+C_2\geq 15,\\3C_1+4C_2+3\leq C_2,\\3C_1+4C_2+5\leq C_2.\end{cases}\tag{6.10}$$

Set: $C_1=17$, $C_2=-19$. This completes the proof. □

In what folows we consider the square

$$I = [0, 2^{N+1}] \times [0, 2^{N+1}].$$

Assume that dyadic rational numbers $x, \Delta x, y, \Delta y$ satisfy $(x, y), (x + \Delta x, x + \Delta y) \in I$.

Lemma 6.3 $|f_C(x + \Delta x, x + \Delta y) - f_C(x, y)| \le 2^{2N+1} C(|\Delta x| + |\Delta y|)$.

Proof There exist an integer number M, such that $2^M x, 2^M y, 2^M \Delta x, 2^M \Delta y$ are non-negative integers. Then the desired result follows from the line of inequalities

$$\begin{aligned}
&|f_C(x + \Delta x, x + \Delta y) - f_C(x, y)| \\
&= \frac{1}{8^M} |f(2^M(x + \Delta x), 2^M(y + \Delta y)) - f(2^M x, 2^M y)| \\
&\le \frac{1}{8^M} 2^M (|\Delta x| + |\Delta y|) N_{N+M,N+M} \\
&\le \frac{1}{4^M} C(4^{N+M} + 4^{N+M})(|\Delta x| + |\Delta y|) = 2^{2N+1} C(|\Delta x| + |\Delta y|).
\end{aligned}$$

□

This statement immediately implies that for every Cauchy sequence (x_i, y_i) the sequence $f_C(x_i, y_i)$ is a Cauchy sequence as well. Thus f_C can be extended to a continuous function on the set of non-negative real numbers. In what follows f_C denotes this extension.

From the properties of f and continuity of f_C we infer the important homogeneity property:

Proposition 6.4

$$f_C(2x, 2y) = 8 f_C(x, y).$$

6.3 Solution to the dual problem

In this section we prove our main duality result. Namely, let us set

$$F(a, b, c) = f(a, b) + f(b, c) + f(c, a)$$

and

$$F_C(x, y, z) = f_C(x, y) + f_C(y, z) + f_C(z, x).$$

We show that function $\frac{1}{8} F_C$ solves the dual problem. Note that Theorem 3.2 does not establish existence of a solution to the dual problem. In this concrete example we construct it explicitly.

The following theorem is the main result of this section.

Theorem 6.5 *Function F_C satisfies*

$$F_C(x, y, z) \le 8xyz.$$

The case of equality $F_C(x, y, z) = 8xyz$ holds if and only if (x, y, z) belongs to the closure of the set

$$x \oplus y \oplus z = 0.$$

In particular, the triple $\frac{1}{8} f_C(x, y)$, $\frac{1}{8} f_C(x, z)$, $\frac{1}{8} f_C(y, z)$ solves problem (6.1).

Proof See Corollary 6.7 and Proposition 6.11. □

Proposition 6.6 *Function $F(a, b, c)$ satisfies inequality*

$$F(a, b, c) \leq 8abc.$$

The equality case

$$F(a, b, c) = 8abc \tag{6.11}$$

can hold only if $a + b + c \equiv 0 \pmod 2$.

In particular, continuity of f_C implies

Corollary 6.7

$$F_C(x, y, z) \leq 8xyz.$$

Proof Let us prove the claim by induction. Base of induction is easy to check. Note that $F(a, b, c) = f(a, b) + f(b, c) + f(c, a) \equiv (a + b) + (b + c) + (c + a) \equiv 0 \pmod 2$ because of (6.7). The latter implies $F(a, b, c) \leq 8abc - 2$ provided $F(a, b, c) < 8abc$.

To prove the induction step we consider several cases.

- All of a, b, c are even. From (6.3) we infer $F(a, b, c) = 8F(\frac{a}{2}, \frac{b}{2}, \frac{c}{2})$. By induction hypothesis $F(a, b, c) = 8F(\frac{a}{2}, \frac{b}{2}, \frac{c}{2}) \leq 8 \cdot 8 \cdot \frac{a}{2} \cdot \frac{b}{2} \cdot \frac{c}{2} = 8abc$.
- Assume that one of the numbers a, b, c (say, a) is odd and the other are even. We need to check

$$F(a, b, c) \leq 8abc - 2, \tag{6.12}$$

because $a + b + c \equiv 1 \pmod 2$. Applying (6.3) one gets

$$\begin{aligned} F(a, b, c) &= f(a, b) + f(a, c) + f(b, c) \\ &= \left[4\left(f\left(\frac{a-1}{2}, \frac{b}{2}\right) + f\left(\frac{a+1}{2}, \frac{b}{2}\right)\right) + 3\right] \\ &\quad + \left[4\left(f\left(\frac{a-1}{2}, \frac{c}{2}\right) + f\left(\frac{a+1}{2}, \frac{c}{2}\right)\right) + 3\right] + 8f\left(\frac{b}{2}, \frac{c}{2}\right) \\ &= 4\left(F\left(\frac{a-1}{2}, \frac{b}{2}, \frac{c}{2}\right) + F\left(\frac{a+1}{2}, \frac{b}{2}, \frac{c}{2}\right)\right) + 6. \end{aligned}$$

One of the triples $\left(\frac{a-1}{2}, \frac{b}{2}, \frac{c}{2}\right), \left(\frac{a+1}{2}, \frac{b}{2}, \frac{c}{2}\right)$ admits even sum of elements, hence satisfies (6.12).

Therefore we can write:

$$\begin{aligned} &4\left(F\left(\frac{a-1}{2}, \frac{b}{2}, \frac{c}{2}\right) + F\left(\frac{a+1}{2}, \frac{b}{2}, \frac{c}{2}\right)\right) + 6 \\ &\leq 4((a-1)bc + (a+1)bc - 2) + 6 = 8abc - 2. \end{aligned}$$

- Assume that there are exactly two odd numbers among a, b, c. Without loss of generality they are a and b. Check that $F(a, b, c) \leq 8abc$, because $a+b+c \equiv 0 \pmod 2$. Applying

(6.3) one gets

$$
\begin{aligned}
F(a,b,c) &= f(a,b) + f(b,c) + f(c,a) \\
&= \left[2 \sum_{\Delta a, \Delta b \in \{-1,1\}} f\left(\frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}\right) + 2 \right] \\
&\quad + \left[4 \sum_{\Delta b \in \{-1,1\}} f\left(\frac{b+\Delta b}{2}, \frac{c}{2}\right) + 3 \right] \\
&\quad + \left[4 \sum_{\Delta a \in \{-1,1\}} f\left(\frac{a+\Delta a}{2}, \frac{c}{2}\right) + 3 \right] \\
&= 2 \sum_{\Delta a, \Delta b \in \{-1,1\}} F\left(\frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}, \frac{c}{2}\right) + 8.
\end{aligned}
$$

Note that triples of the type $(\frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}, \frac{c}{2})$ there are exactly two with even sum of elements, so by induction hypothesis for at most two triples (6.11) holds.
Hence

$$
2 \sum_{\Delta a, \Delta b \in \{-1,1\}} F\left(\frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}, \frac{c}{2}\right) + 8 \leq 2((2a)(2b)c - 2 \cdot 2) + 8 = 8abc.
$$

- Finally let us assume that all a, b, c are odd. Thus $a + b + c \equiv 1 \pmod 2$, so we need to check $F(a, b, c) \leq 8abc - 2$. Again, (6.3) implies

$$
\begin{aligned}
F(a,b,c) &= f(a,b) + f(b,c) + f(c,a) \\
&= \left[2 \sum_{\Delta a, \Delta b \in \{-1,1\}} f\left(\frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}\right) + 2 \right] \\
&\quad + \left[2 \sum_{\Delta b, \Delta c \in \{-1,1\}} f\left(\frac{b+\Delta b}{2}, \frac{c+\Delta c}{2}\right) + 2 \right] \\
&\quad + \left[2 \sum_{\Delta c, \Delta a \in \{-1,1\}} f\left(\frac{c+\Delta c}{2}, \frac{a+\Delta a}{2}\right) + 2 \right] \\
&= \sum_{\Delta a, \Delta b, \Delta c \in \{-1,1\}} F\left(\frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}, \frac{c+\Delta c}{2}\right) + 6.
\end{aligned}
$$

Counting the equality cases and repeating the arguments from above one gets

$$
\begin{aligned}
&\sum_{\Delta a, \Delta b, \Delta c \in \{-1,1\}} F\left(\frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}, \frac{c+\Delta c}{2}\right) + 6 \\
&\qquad \leq (2a)(2b)(2c) - 2 \cdot 4 + 6 = 8abc - 2.
\end{aligned}
$$

Step of induction is verified in all possible cases. □

6.4 Some nice identities

Here we prove some other useful identities for $f(a,b)$ and their continious analogues for $f_C(x,y)$.

Proposition 6.8 *Let* $0 \le a, b \le 2^n$. *Then*

$$f(2^n+a, 2^n+b) = 2\cdot 8^n + 6\cdot 4^n(a+b) + f(a,b).$$

Proof Apply induction by n. The case $n=0$ is easy to check.

Let a and b be even. Then

$$\begin{aligned} f(2^n+a, 2^n+b) &= 8f\left(2^{n-1}+\frac{a}{2}, 2^{n-1}+\frac{b}{2}\right) \\ &= 8\left(2\cdot 8^{n-1}+6\cdot 4^{n-1}\frac{a+b}{2}+f\left(\frac{a}{2},\frac{b}{2}\right)\right) \\ &= 2\cdot 8^n + 6\cdot 4^n(a+b) + f(a,b). \end{aligned}$$

Let exactly one of the numbers a or b be odd. Without loss of generality assume that a is odd. Then

$$\begin{aligned} & f(2^n+a, 2^n+b) \\ &= 4\left(f\left(2^{n-1}+\frac{a+1}{2}, 2^{n-1}+\frac{b}{2}\right)+f\left(2^{n-1}+2^{n-1}+\frac{a-1}{2},\frac{b}{2}\right)\right)+3 \\ &= 4\left[2\cdot 8^{n-1}+6\cdot 4^{n-1}\frac{a+b+1}{2}+f\left(\frac{a+1}{2},\frac{b}{2}\right)\right] \\ &\quad +4\left[2\cdot 8^{n-1}+6\cdot 4^{n-1}\frac{a+b-1}{2}+f\left(\frac{a-1}{2},\frac{b}{2}\right)\right] \\ &\quad +4\left[f\left(\frac{a+1}{2},\frac{b}{2}\right)+f\left(\frac{a-1}{2},\frac{b}{2}\right)\right]+3 \\ &= 2\cdot 8^n + 6\cdot 4^n(a+b) + f(a,b). \end{aligned}$$

Let both of a and b be odd. Similarly by the induction hypothesis:

$$\begin{aligned} f(2^n+a, 2^n+b) &= 2\sum_{\Delta a,\Delta b\in\{-1,1\}} f\left(2^{n-1}+\frac{a+\Delta a}{2}, 2^{n-1}+\frac{b+\Delta b}{2}\right)+2 \\ &= 2\sum_{\Delta a,\Delta b\in\{-1,1\}}\left[2\cdot 8^{n-1}+6\cdot 4^{n-1}\frac{a+b+\Delta a+\Delta b}{2}+f\left(\frac{a+\Delta a}{2},\frac{b+\Delta b}{2}\right)\right]+2 \\ &= 2\cdot 8^n+6\cdot 4^n(a+b)+2\sum_{\Delta a,\Delta b\in\{-1,1\}} f\left(\frac{a+\Delta a}{2},\frac{b+\Delta b}{2}\right)+2 \\ &= 2\cdot 8^n + 6\cdot 4^n(a+b) + f(a,b). \end{aligned}$$

□

Proposition 6.9 *For* $0 \le a, b \le 2^n$ *one has*

$$f(2^n+a, b) = -8^n - 6\cdot 4^n a + 4^{n+1}b + 8\cdot 2^n ab + f(a,b).$$

Proof The proof is similar to the proof of Proposition 6.8. Apply the induction by n. The base for $n = 0$ can be checked by an easy computation.

Consider $n \geq 1$. Set $g_1(n, a, b) = -8^n$, $g_2(n, a, b) = 6 \cdot 4^n a$, $g_3(n, a, b) = 4^{n+1} b$, $g_4(n, a, b) = 8 \cdot 2^n ab$. For induction step it is sufficient to check that the following identities hold:

$$g_i(n, a, b) = 8g_i\left(n-1, \frac{a}{2}, \frac{b}{2}\right),$$
$$g_i(n, a, b) = 4g_i\left(n-1, \frac{a-1}{2}, \frac{b}{2}\right) + 4g_i\left(n-1, \frac{a+1}{2}, \frac{b}{2}\right),$$
$$g_i(n, a, b) = 4g_i\left(n-1, \frac{a}{2}, \frac{b-1}{2}\right) + 4g_i\left(n-1, \frac{a}{2}, \frac{b+1}{2}\right),$$
$$g_i(n, a, b) = 2\sum_{\Delta a, \Delta b \in \{-1,1\}} g_i\left(n-1, \frac{a+\Delta a}{2}, \frac{b+\Delta b}{2}\right).$$

Next we prove the desired identity by considering four different cases: a is odd(even), b is odd(even) and applying an appropriate identity for all summands in the right hand side. For any of 8^n, $4^n a$, $4^n b$ and $2^n ab$ these properties are obviously true. □

Clearly, the continious analogues of these identities look as follows.

Proposition 6.10 *Let* $0 \leq x, y \leq \frac{1}{2}$. *Then:*

$$f_C\left(\frac{1}{2}+x, \frac{1}{2}+y\right) = \frac{1}{4} + \frac{3}{2}(x+y) + f_C(x, y),$$
$$f_C\left(\frac{1}{2}+x, y\right) = -\frac{1}{8} - \frac{3}{2}x + y + 4xy + f_C(x, y).$$

6.5 Case of equality

Proposition 6.11 *The relation* $x \oplus y \oplus z = 0$ *implies* $F_C(x, y, z) = 8xyz$.

Proof Assume the opposite and consider the maximum of $8xyz - F_C(x, y, z)$ on the closure S of the set of points (x, y, z) satisfying $x \oplus y \oplus z = 0$. This maximum C exists since the set is compact and $8xyz - F_C(x, y, z)$ is continuous. It is sufficient to show that C is not strictly positive. Find a point (x_0, y_0, z_0) with $z_0 = x_0 \oplus y_0$ such that $8x_0y_0z_0 - F_C(x_0, y_0, z_0) > C/2$.

The first numbers in the binary representations of x_0, y_0, z_0 contains either all zeroes of exactly two units, because $x_0 \oplus y_0 \oplus z_0 = 0$. If they all are zeroes, then $2x_0 \oplus 2y_0 \oplus 2z_0 = 0$. Thus $8(2x_0)(2y_0)(2z_0) - F_C(2x_0, 2y_0, 2z_0) > 4C > C$, this contradicts to the choice of C. If the numbers contain two units, without loss of generality assume $x_0 = y_0 = 1$. Set $x_0 = \frac{1}{2} + x_1$, $y_0 = \frac{1}{2} + y_1$. The identities (6.10) imply

$$\begin{aligned}
8x_0y_0z_0 - F_C(x_0, y_0, z_0) &= 8\left(x_1+\frac{1}{2}\right)\left(y_1+\frac{1}{2}\right)z_0 - F_C\left(x_1+\frac{1}{2}, y_1+\frac{1}{2}, z_0\right)\\
&= 8x_1y_1z_0 + 4x_1z_0 + 4y_1z_0 + 2z_0\\
&\quad - f_C\left(x_1+\frac{1}{2}, y_1+\frac{1}{2}\right) - f_C\left(x_1+\frac{1}{2}, z_0\right) - f_C\left(y_1+\frac{1}{2}, z_0\right)\\
&= 8x_1y_1z_0 + 4x_1z_0 + 4y_1z_0 + 2z_0 - \left[\frac{1}{4} + \frac{3}{2}(x_1+y_1) + f_C(x_1, y_1)\right]
\end{aligned}$$

$$
\begin{aligned}
&-\left[-\frac{1}{8}-\frac{3}{2}x_1+z_0+4x_1z_0+f_C(x_1,z_0)\right]\\
&-\left[-\frac{1}{8}-\frac{3}{2}y_1+z_0+4y_1z_0+f_C(y_1,z_0)\right]\\
&=8x_1y_1z_0-F_C(x_1,y_1,z_0).
\end{aligned}
$$

Note that $x_1 \oplus y_1 \oplus z_0 = 0$, moreover, the function $8xyz - F_C(x, y, z)$ takes at the point (x_1, y_1, z_0) the same value $C/2$. Note that $x_1, y_1, z_0 \leq \frac{1}{2}$, but we have already shown that this is impossible. We got a contradiction. □

7 Integral representation of $f_C(a, b)$

The solution to the dual problem in our main example has a simple relation to the (cumulative) distribution function

$$I(a,b)=\int_0^a\int_0^b x\oplus y\; dydx,\;\; a,b\in\mathbb{R}_+.$$

of the measure $x \oplus y\; dxdy$. This function admits the following properties:

Property 7.1 *Symmetry:* $I(a, b) = I(b, a)$.

Property 7.2 *Homogeneity with respect to factor* 2*:* $I(2a, 2b) = 8I(a, b)$.

Proof Note that for almost all x, y and integer number n one has $2^n x \oplus 2^n y = 2^n(x \oplus y)$. This yields

$$
\begin{aligned}
I(2a,2b)&=\int_0^{2a}\int_0^{2b} x\oplus y\; dydx=\begin{bmatrix} x=2u\\ y=2v\end{bmatrix}\\
&=4\int_0^a\int_0^b 2u\oplus 2v\; dvdu=8I(a,b).
\end{aligned}
$$

□

Property 7.3 *For all* $0 \leq a \leq 1$

$$I(a,1)=\frac{a}{2}.$$

Proof To this end we need the following lemma:

Lemma 7.4 *For every couple* $0 \leq x, y \leq 1$*, where neither* x *nor* y *is binary rational, the following relation holds:* $x \oplus y + x \oplus (1 - y) = 1$.

Proof Note that for $a = x \oplus y$ and $b = x \oplus (1 - y)$ the i-th digits satisfy $a_i = x_i \oplus y_i$, $b_i = x_i \oplus \overline{y_i}$. Clearly, $a_i \oplus b_i = 0$. □

This can be used for computation of $I(a, 1)$:

$$
\begin{aligned}
I(a,1)&=\int_0^a\int_0^1 x\oplus y\; dydx\\
&=\frac{1}{2}\int_0^a\int_0^1 (x\oplus y+x\oplus(1-y))\; dydx=\frac{1}{2}\int_0^a\int_0^1 1\; dydx=\frac{a}{2}.
\end{aligned}
$$

□

Applying homogeneity property one immediately gets

Corollary 7.5 *For every* $0 \le a \le \frac{1}{2^n}$

$$I\left(a, \frac{1}{2^n}\right) = \frac{a}{2^{2n+1}}.$$

In the following proposition we establish a recurent relation for f_C:

Proposition 7.6 *For all* $0 \le a, b \le \frac{1}{2}$ *the following identity holds:*

$$I\left(\frac{1}{2} + a, b\right) = \frac{1}{2}ab + \frac{1}{8}b + I(a, b).$$

Proof Represent the integral as a sum of two parts

$$\begin{aligned} I\left(\frac{1}{2} + a, b\right) &= \int_0^{\frac{1}{2}+a} \int_0^b x \oplus y \, dydx \\ &= \int_{\frac{1}{2}}^{\frac{1}{2}+a} \int_0^b x \oplus y \, dydx + \int_0^{\frac{1}{2}} \int_0^b x \oplus y \, dydx. \end{aligned}$$

Making the change of variable $x = \frac{1}{2} + t$ one gets

$$\begin{aligned} \int_{\frac{1}{2}}^{\frac{1}{2}+a} \int_0^b x \oplus y \, dydx &= \int_0^a \int_0^b \left(\frac{1}{2} + t\right) \oplus y \, dydt \\ = \int_0^a \int_0^b \left(t \oplus y + \frac{1}{2}\right) dydt &= \frac{1}{2}ab + I(a, b). \end{aligned}$$

Hence

$$I(a, b) = \frac{1}{2}ab + I(a, b) + I\left(\frac{1}{2}, b\right) = \frac{1}{2}ab + \frac{1}{8}b + I(a, b).$$

□

Let us prove another similar relation

Proposition 7.7 *For every* $0 \le a, b \le \frac{1}{2}$ *one has*

$$I\left(\frac{1}{2} + a, \frac{1}{2} + b\right) = \frac{1}{16} + \frac{3}{8}a + \frac{3}{8}b + I(a, b).$$

Proof Similarly to the arguments of the previous proposition one obtains

$$\begin{aligned} I\left(\frac{1}{2} + a, \frac{1}{2} + b\right) &= \int_0^{\frac{1}{2}+a} \int_0^{\frac{1}{2}+b} x \oplus y \, dydx \\ = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} x \oplus y \, dydx &+ \int_{\frac{1}{2}}^{\frac{1}{2}+a} \int_0^{\frac{1}{2}} x \oplus y \, dydx \\ + \int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^{\frac{1}{2}+b} x \oplus y \, dydx &+ \int_{\frac{1}{2}}^{\frac{1}{2}+a} \int_{\frac{1}{2}}^{\frac{1}{2}+b} x \oplus y \, dydx. \end{aligned}$$

Clearly

$$\int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} x \oplus y \, dydx = I\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{16}.$$

To compute the second integral let us make the variables change $x = \frac{1}{2} + t$:

$$\int_{\frac{1}{2}}^{\frac{1}{2}+a} \int_0^{\frac{1}{2}} x \oplus y \, dydx = \int_0^a \int_0^{\frac{1}{2}} \left(\frac{1}{2} + t\right) \oplus y \, dydt$$
$$= \int_0^a \int_0^{\frac{1}{2}} \left(\frac{1}{2} + t \oplus y\right) \, dydt = \frac{1}{4}a + I\left(\frac{1}{2}, a\right)$$
$$= \frac{1}{4}a + \frac{1}{8}a = \frac{3}{8}a.$$

In the same way one gets the following formula for the third integral:

$$\int_0^{\frac{1}{2}} \int_{\frac{1}{2}}^{\frac{1}{2}+b} x \oplus y \, dydx = \frac{3}{8}b.$$

To compute the last integral, let us set $x = \frac{1}{2} + t$, $y = \frac{1}{2} + u$:

$$\int_{\frac{1}{2}}^{\frac{1}{2}+a} \int_{\frac{1}{2}}^{\frac{1}{2}+b} x \oplus y \, dydx = \int_0^a \int_0^b \left(t + \frac{1}{2}\right) \oplus \left(u + \frac{1}{2}\right) \, dudt$$
$$= \int_0^a \int_0^b t \oplus u \, dudt = I(a, b).$$

Finally,

$$I\left(\frac{1}{2} + a, \frac{1}{2} + b\right) = \frac{1}{16} + \frac{3}{8}(a + b) + I(a, b).$$

□

It remains to relate f_C and I.

Theorem 7.8 *For all non-negative $x, y \in \mathbb{R}_+$ the following relation holds:*

$$f_C(x, y) = 8I(x, y) - 2I(x, x) - 2I(y, y).$$

Proof By homogeneity $f_C(x, y)$ and $I(x, y)$ it is sufficent to prove this relation on $[0, 1]^2$.

Set $f_1(x, y) = 8I(x, y) - 2I(x, x) - 2I(y, y)$. We prove that f_1 satifies the same relation as f_C (see Proposition 6.10). Indeed, for all, $0 \le x, y \le \frac{1}{2}$:

$$f_1\left(\frac{1}{2} + x, y\right) = 8I\left(\frac{1}{2} + x, y\right) - 2I\left(\frac{1}{2} + x, \frac{1}{2} + x\right) - 2I(y, y)$$
$$= 4xy + y + 8I(x, y) - 2\left(\frac{1}{16} + \frac{3}{8}(x + x) + I(x, x)\right) - 2I(y, y)$$
$$= -\frac{1}{8} - \frac{3}{2}x + y + 4xy + f_1(x, y),$$
$$f_1\left(\frac{1}{2} + x, \frac{1}{2} + y\right) = 8I\left(\frac{1}{2}+x, \frac{1}{2}+y\right) - 2I\left(\frac{1}{2}+x, \frac{1}{2}+x\right) - 2I\left(\frac{1}{2} + y, \frac{1}{2} + y\right)$$
$$= \frac{1}{2} + 3x + 3y + 8I(x, y) - 2\left(\frac{1}{16} + \frac{3}{8}(x + x) + I(x, x)\right)$$
$$-2\left(\frac{1}{16} + \frac{3}{8}(y + y) + I(y, y)\right) = \frac{1}{4} + \frac{3}{2}(x + y) + f_1(x, y).$$

It remains to show that $M = \sup_{0\le x\le 1, 0\le y\le 1} |f - f_1| = 0$. Note that the supremum is attained on $\left[0, \frac{1}{2}\right]^2$, because $f - f_1$ is invariant with respect to the shifts $x \to x + \frac{1}{2}$, $y \to y + \frac{1}{2}$. If M is larger than zero and attained at some point (x_0, y_0), where $0 \le x_0, y_0 \le \frac{1}{2}$, then the value of $|f - f_1|$ at $(2x_0, 2y_0)$ equals $8M$. We obtained a contradiction. □

Applying the above result we obtain the following integral representation theorem for our solution to the dual problem.

Theorem 7.9 *The function*

$$F(x, y) = \int_0^x \int_0^y s \oplus t \, dsdt - \frac{1}{4}\int_0^x \int_0^x s \oplus t \, dsdt - \frac{1}{4}\int_0^y \int_0^y s \oplus t \, dsdt$$

solves the dual problem

$$\int_{[0,1]^2} F(x, y)dxdy + \int_{[0,1]^2} F(x, z)dxdz + \int_{[0,1]^2} F(y, z)dydz \to \max,$$
$$F(x, y) + F(x, z) + F(y, z) \le xyz$$

to the primal $(3, 2)$*-Kantorovich problem*

$$\int xyzd\pi \to \min, \ (x, y, z) \in [0, 1]^3,$$

considered on the space of measure which projections onto principal hyperplanes are Lebesgue measures on $[0, 1]^2$.

8 Concluding remarks

Numerical experiments visually reveal fractal structure of the solutions to (3,2)-Kantorovich problem for other cost functions and projections. This happens even under absence of symmetry, which, in turn, means that the solutions do not posess dyadic structure. Which properties of our main example are preserved in general case? Here we discuss several natural hypotheses.

Question 8.1 *Consider the* $(3, 2)$*-Kantorovich problem on the set* $X \times Y \times Z$*, where*

$$X = \{x_0 < x_1 \cdots < x_{2^n-1}\},$$
$$Y = \{y_0 < y_1 \cdots < y_{2^n-1}\},$$
$$Z = \{z_0 < z_1 \cdots < z_{2^n-1}\}.$$

As usual, $c = xyz$ *and the projections are supposed to be uniform. We want to maximize* $\int xyzd\pi$.

Is it true that unifrom measure concentrated on the points (x_i, y_j, z_k) *with* $i \oplus j \oplus k = 0$ *is optimal?*

Question 8.2 *Consider the dual* $(3, 2)$*-Kantorovich problem on the set* $[0, 1]^3$.

$$\int F(x, y)d\mu_{xy} + \int G(x, z)d\mu_{xz} + \int H(y, z)d\mu_{yz} \to \max,$$
$$F(x, y) + G(x, z) + H(y, z) \le xyz$$

for some triple of measures $\mu_{xy}, \mu_{xz}, \mu_{yz}$.

Is it true that F satisfies inequality

$$F(x+\Delta x, y+\Delta y) + F(x,y) - F(x+\Delta x, y) - F(x, y+\Delta y) \geq 0$$

for every $x, y, \Delta x \geq 0, \Delta y \geq 0$? Equivalently, F has the representation

$$F(x,y) = m([0,\ x]\times[0,\ y]) + f(x) + g(y)$$

for some nonnegative measure m and some functions f, g?

Numerical computations demonstrate that Question 8.2 has a negative answer. The answer to Question 8.1 is negative in general, but remarkably the answer is affirmative for $n = 2$.

Example 8.3 Consider the discrete cube $8\times 8\times 8$,

$$X = Y = Z = \{0, \varepsilon, 2\varepsilon, 1-4\varepsilon, 1-3\varepsilon, 1-2\varepsilon, 1-\varepsilon, 1\}.$$

For sufficiently small ε, the uniform measure M', concentrated on the points (x_i, y_j, z_k) with $i\oplus j\oplus k = 0, i, j, k \in \{0, 1, \dots, 2^3-1\}$, is not optimal. Let us say that numbers 0, 1, 2 are **small**. Other numbers are **large**. Consider the following competitor: measure M'' assigns to a point (x_i, y_j, z_k) the following value :

$$\begin{cases} \frac{1}{3}, & \text{if all three indexes } i, j \text{ and } k \text{ are small;} \\ 0, & \text{if two indexes are small and one is large;} \\ \frac{1}{5}, & \text{if one index is small and two are large;} \\ \frac{2}{25}, & \text{if all three indexes } i, j \text{ and } k \text{ are large.} \end{cases}$$

Integrals $\int xyz dM'$ and $\int xyz dM''$ are the polynomials in ε. Their free terms are equal to 12 and $125\times\frac{2}{25} = 10$ respectively. Thus $\int xyz dM' > \int xyz dM''$ for sufficiently small epsilon.

Let $I = [0,1]^3$ be the unit cube and μ be arbitrary measure on I. We denote by F_μ the distribution function of μ

$$F_\mu(a,b,c) = \mu([0,a]\times[0,b]\times[0,c]).$$

Lemma 8.4 *Let μ be a measure on I. Then the following identity holds:*

$$\int_I (1-x)(1-y)(1-z)d\mu = \int_I F_\mu(x,y,z)dxdydz.$$

Proof Let I' be the unit cube endowed with the uniform Lebesgue measure ω. One can consider the product $I\times I'$ with the product measure $d\mu\otimes d\omega$. Set:

$$D = \{(p,q)\in I\times I' \mid p \text{ is not larger than } q \text{coordinatewise}\}.$$

Let us find $(\mu\otimes\omega)(D)$. We apply to this end the Fubini theorem

$$\int_D d\mu\otimes d\omega = \int_{(x,y,z)\in I} \int_{\substack{(x_1,y_1,z_1)\in I', \\ (x_1,y_1,z_1)\geq(x,y,z)}} d\omega d\mu = \int_{(x,y,z)\in I} (1-x)(1-y)(1-z)d\mu.$$

On the other hand,

$$\int\limits_{(x,y,z)\in I} = \int\limits_{(x,y,z)\in I'} \int\limits_{\substack{(x_1,y_1,z_1)\in I,\\ (x_1,y_1,z_1)\le(x,y,z)}} d\mu d\omega = \int\limits_{(x,y,z)\in I} F_\mu(x,y,z)d\omega.$$

□

Let $\mu_{xy}, \mu_{yz}, \mu_{zx}$ be projections of μ onto the corresponding principal hyperplanes. On can rewrite the integral as follows:

$$\begin{aligned}\int\limits_I (1-x)(1-y)(1-z)d\mu &= 1 - \int\limits_{I_{xy}} xy\, d\mu_{xy} - \int\limits_{I_{yz}} yz\, d\mu_{yz} - \int\limits_{I_{zx}} zx\, d\mu_{zx}\\ &+ \int\limits_{I_{xy}} x\, d\mu_{xy} + \int\limits_{I_{yz}} y\, d\mu_{yz} + \int\limits_{I_{zx}} z\, d\mu_{zx} - \int\limits_I xyz\, d\mu\\ &= C(\mu_{xy}, \mu_{yz}, \mu_{zx}) - \int\limits_I xyz\, d\mu,\end{aligned}$$

where $C(\mu_{xy}, \mu_{yz}, \mu_{zx})$ only depends on the projections of μ onto the principal hyperplanes.

We want to find a measure π which minimizes $\int xyz d\pi$ on the set of all (3, 2)-stochastic measures on $X \times Y \times Z$.

Finally, consider

$$\begin{aligned} X &= \{x_0 < x_1 \cdots < x_{2^n-1}\},\\ Y &= \{y_0 < y_1 \cdots < y_{2^n-1}\},\\ Z &= \{z_0 < z_1 \cdots < z_{2^n-1}\}.\end{aligned}$$

Without loss of generality assume that $X \times Y \times Z \subset I$. Let $\mu_\oplus$ be a measure on I which is supported on $X \times Y \times Z$ and defined by

$$\mu_\oplus(x_i, y_j, z_k) = \frac{1}{4^n},$$

if $i \oplus j \oplus k = 0$, and

$$\mu_\oplus(x_i, y_j, z_k) = 0$$

in the opposite case.

Theorem 8.5 *Assume that $|X| = |Y| = |Z| = 4$. Let μ be arbitrary measure $X \times Y \times Z$ with uniform projections on $X \times Y$, $X \times Z$, $Y \times Z$. Then*

$$\int xyz\, d\mu \ge \int xyz\, d\mu_\oplus.$$

Moreover,

$$F_{\mu_\oplus} \ge F_\mu$$

at every point.

Proof Since the projections of μ and $\mu_\oplus$ onto the hyperplanes are equal, one has the following equivalence relation

$$\int\limits_I xyz\, d\mu \geq \int\limits_I xyz\, d\mu_\oplus \Leftrightarrow \int\limits_I F_\mu(x,y,z)\, dxdydz \leq \int\limits_I F_{\mu_\oplus}(x,y,z)\, dxdydz.$$

Let us prove that $F_{\mu_\oplus} \geq F_\mu$. Since the measures are discrete, it is suficient to check the desired inequality at the points $(x_i, y_j, z_k) \in X \times Y \times Z$. Without loss of generality let $i \leq j \leq k$.

If $k = 3$, the distribution function satisfies $F_{\mu_\oplus}(x_i, y_j, z_3) = F_\mu(x_i, y_j, z_3) = \frac{(i+1)(j+1)}{16}$. This follows from the fact that μ and $\mu_\oplus$ have uniform projections onto $X \times Y$.

Let $i = 0$. Then $F_{\mu_\oplus}(x_0, y_j, z_k) = \frac{1}{16}\min(j+1, k+1) = \frac{j+1}{16}$. Indeed, for $i = 0$ measure $\mu_\oplus$ is concentrated at the points $(x_0, y_t, z_t), t \in \{0,1,2,3\}$. Hence $F_{\mu_\oplus}(x_0, y_j, z_k) = \frac{1}{16}\#(t \mid 0 \leq t \leq j, 0 \leq t \leq k)$. On the other hand $F_\mu(x_0, y_j, z_k) \leq F_\mu(x_0, y_j, z_3) = \frac{j+1}{16}$.

It remains to consider the cases when every i, j, k equals 1 or 2.

Let $k = 2$. Compute $F_{\mu_\oplus}(x_i, y_j, z_2)$. To this end we count all triples (a, b, c) satisfying $0 \leq a \leq i, 0 \leq b \leq j, 0 \leq c \leq 2$ and $a \oplus b \oplus c = 0$. For every couple (a, b) there exists the unique c having this property except for the case $a \oplus b = 3$. This happens if and only if $\{a, b\} = \{1, 2\}$. It is easy to check that amount of couples with this property is exactly the number of indices i, j which takes value 2, i.e. $i + j - 2$. Thus the total amount of such triples (a, b, c) equals $(i+1)(j+1) - i - j + 2 = ij + 3$. Hence $F_{\mu_\oplus}(x_i, y_j, z_2) = \frac{ij+3}{16}$.

Represent the number $F_\mu(x_i, y_j, z_2)$ as follows:

$$\begin{aligned} F_\mu(x_i, y_j, z_2) &= \sum_{\substack{x\in[0,x_i]\\ y\in[0,y_j]\\ z\in[0,z_2]}} \mu(x,y,z) = \sum_{\substack{x\in[0,x_i]\\ y\in[0,y_j]\\ z\in[0,z_3]}} \mu(x,y,z) - \sum_{\substack{x\in[0,x_i]\\ y\in[0,y_j]}} \mu(x,y,z_3) \\ &= F_\mu(x_i, y_j, z_3) - \sum_{\substack{x\in[0,x_i]\\ y\in[0,y_3]}} \mu(x,y,z_3) + \sum_{\substack{x\in[0,x_i]\\ y\in[y_{j+1},y_3]}} \mu(x,y,z_3), \end{aligned}$$

where the sum is taken over the atoms of μ.

We know that $F_\mu(x_i, y_j, z_3) = \frac{(i+1)(j+1)}{16}$, because the projection of μ onto $X \times Y$ is uniform. Analogously, the same facts about projections onto $X \times Z$ and $Y \times Z$ imply

$$\sum_{\substack{x\in[0,x_i]\\ y\in[0,y_3]}} \mu(x,y,z_3) = \frac{i+1}{16}.$$

$$\sum_{\substack{x\in[0,x_i]\\ y\in[y_{j+1},y_3]}} \mu(x,y,z_3) \leq \sum_{\substack{x\in[0,x_3]\\ y\in[y_{j+1},y_3]}} \mu(x,y,z_3) = \frac{3-j}{16},$$

Hence

$$F_\mu(x_i, y_j, z_2) \leq \frac{(i+1)(j+1)}{16} - \frac{i+1}{16} + \frac{3-j}{16} = \frac{ij+3}{16}.$$

It remains to consider the case $i = j = k = 1$. One gets immediately $F_{\mu_\oplus}(x_i, y_j, z_k) = \frac{4}{16}$, and $F_\mu(x_i, y_j, z_k) \leq F_\mu(x_i, y_j, z_3) = \frac{(i+1)(j+1)}{16} = \frac{4}{16}$. □

References

1. Ahmad, N., Kim, H.-K., McCann, R.J.: Extremal doubly stochastic measures and optimal transportation. Bull. Math. Sci. **1**(1), 13–32 (2011)

2. Beiglböck, M., Juillet, N.: On a problem of optimal transport under marginal martingale constraints. Ann. Probab. **44**(1), 42–106 (2016)
3. Beiglböck, M., Henry-Labordère, P., Penkner, F.: Model independent bounds for option prices: a mass transport approach. Finance Stoch. **17**(3), 477–501 (2013)
4. Beiglböck, M., Goldstern, M., Maresch, G., Schachermayer, W.: Optimal and better transport plans. J. Funct. Anal. **256**(6), 1907–1927 (2009)
5. Beneš, V., Štěpán, J.: The support of extremal probability measures with given marginals. In: Puri, M.L., Revesz, P., Wertz, W. (eds.) Mathematical Statistics and Probability Theory, vol. A, pp. 33–41. Springer, Dordrecht (1987)
6. Bianchini, S., Caravenna, L.: On the extremality, uniqueness and optimality of transference plans. Bull. Inst. Math. Acad. Sin. (New Series) **4**, 353–454 (2009)
7. Bogachev, V.I., Kolesnikov, A.V.: The Monge–Kantorovich problem: achievements, connections, and perspectives. Russ. Math. Surv. **67**(5), 785–890 (2012)
8. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Commun. Pure Appl. Math. **44**(4), 375–417 (1991)
9. Cui, L.-B., Li, W., Ng, M.K.: Birkhoff–von Neumann theorem for multistochastic tensors. SIAM J. Matrix Anal. Appl. **35**(3), 956–973 (2014)
10. Di Marino, S., Gerolin, A., Nenna, L.: Optimal transportation theory with repulsive costs. In: Topological Optimization and Optimal Transport, Vol. 17 of Radon Series on Computational and Applied Mathematics, pp. 204–256. De Gruyter, Berlin (2017)
11. Fraenkel A., Kontorovich A.: The Sierpiński sieve of nim-varieties and binomial coefficients. In: Combinatorial Number Theory: Proceedings of the 'Integers Conference 2005' in Celebration of the 70th Birthday of Ronald Graham, Carrollton, Georgia, USA, October 27–30 (2005)
12. Galichon, A.: Optimal Transport Methods in Economics. Princeton University Press, Princeton (2016)
13. Galishon, A., Henry-Labordère, M., Touzi, N.: A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. Ann. Appl. Probab. **24**(1), 312–336 (2014)
14. Ghoussoub, N., Moameni, A.: Symmetric Monge-Kantorovich problems and polar decompositions of vector fields. Geom. Funct. Anal. **24**(4), 1129–1166 (2014). https://doi.org/10.1007/s00039-014-0287-2
15. Gibbons K.: The Geometry of Nim. https://arxiv.org/pdf/1109.6712.pdf
16. Guillen N., McCann R.J.: Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the Seminaire de Mathematiques Superieure (SMS) Montreal 2011. G. Dafni et al. (eds.), pp. 145–180. American Mathematical Society, Providence (2013)
17. Henry-Labordère, P.: Model-Free Hedging: A Martingale Optimal Transport Viewpoint. Chapman and Hall/CRC Financial Mathematics Series. Chapman and Hall/CRC, New York (2017)
18. Hestir, K., Williams, S.C.: Supports of doubly stochastic measures. Bernoulli **1**, 217–243 (1995)
19. Kellerer, H.G.: Verteilungsfunktionen mit gegebenen Marginalverteilungen. Z. Wahrsch. Verw. Gebiete **3**(91964), 247–270 (1964)
20. Kellerer, H.G.: Duality theorems for marginal problems. Z. Wahrsch. Verw. Gebiete **67**, 399–432 (1984). https://doi.org/10.1007/BF00532047
21. Kiloran N.: Supports of extremal doubly and triply stochastic measures—master's project. (2007). http://www.math.toronto.edu/mccann/papers/Killoran.pdf
22. Kolesnikov, A.V., Zaev, D.: Exchangeable optimal transportation and log-concavity. Theory Stoch. Process. **20**(2), 54–62 (2015)
23. Kolesnikov, A.V., Zaev, D.A.: Optimal transportation of processes with infinite Kantorovich distance. Independence and symmetry. Kyoto J. Math. **57**(2), 293–324 (2017)
24. McCann, R.J., Korman, J.: Optimal transportation with capacity constraints. Trans. Am. Math. Soc. **367**, 1501–1521 (2015)
25. McCann, R.J., Korman, J., Seis, C.: An elementary approach to linear programming duality with application to capacity constrained transport. J. Convex Anal. **22**, 797–808 (2015)
26. Levin, V.L.: The problem of mass transfer in a topological space and probability measures with given marginal measures on the product of two spaces. Dokl. Akad. Nauk SSSR **276**(5), 1059–1064 (1984)
27. Linial, N., Luria, Z.: On the vertices of the d-dimensional Birkhoff polytope. Discrete Comput. Geom. **51**(1), 161–170 (2014)
28. Mandelbrot, B.: The Fractal Geometry of Nature. W. H. Freeman and Co., San Francisco (1982)
29. Moameni, A.: Supports of extremal doubly stochastic measures. Can. Math. Bull. **59**(2), 1–10 (2015)
30. Moameni, A.: Invariance properties of the Monge–Kantorovich mass transport problem. Discrete Contin. Dyn. Syst. A **36**(5), 2653–2671 (2016)
31. Pass, B.: Multi-marginal optimal transport: theory and applications. ESAIM: M2AN **49**, 1771–1790 (2015)
32. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems. V. I, II. Springer, New York (1998)

33. Sudakov, V.N.: Geometric problems in the theory of infinite dimensional probability distributions. Trudy Mat. Inst. Steklov **141**, 3–191 (1976) (**in Russian**)
34. Vershik, A.M.: What does a generic Markov operator look like? St. Petersb. Math. J. **17**(5), 763–772 (2006)
35. Villani, C.: Topics in Optimal Transportation. American Mathematical Society, Providence (2003)
36. Villani, C.: Optimal Transport, Old and New. Springer, New York (2009)
37. Zaev, D.A.: On the Monge–Kantorovich problem with additional linear constraints. Math. Notes **98**(5), 725–741 (2015)