Multimarginal optimal transport with the repulsive harmonic cost

Alexander Zimin

alekszm@gmail.com

February 26, 2021

The multimarginal MK problem

We generalize the classical Monge-Kantorovich problem to the case of N probability measures $\mu_k \in \mathcal{P}(X_k)$ and a cost function $c\colon X_1 \times \cdots \times X_N \to \mathbb{R}$.

Problem

(Primal problem) *Find a transport plan π taken from the space of measures $\Pi(\mu_1, \ldots, \mu_N)$ with fixed marginals that minimizes the functional*

$$\int_{X_1 \times \cdots \times X_N} c(x_1, \ldots, x_N)\, \pi(dx_1, \ldots, dx_N) \to \inf .$$

(Dual problem) *Find the supremum of the functional*

$$\sum_{k=1}^{N} \int_{X_k} \varphi_k(x_k)\, \mu_k(dx_k) \to \sup$$

over all functions $(\varphi_k)_{k=1}^N$ satisfying the inequality $\sum_{k=1}^N \varphi_k(x_k) \le c(x_1, \ldots, x_N)$.

The multimarginal Wasserstein Barycenter

Problem (Wasserstein barycenter)

Given N marginals $\mu_k \in \mathcal{P}(\mathbb{R}^d)$, having finite second moment, and N real weights λ_i with $\lambda_1 + \cdots + \lambda_N = 1$. Solve

$$\inf \left\{ \sum_{i=1}^{N} \lambda_i W_2^2(\mu_k, \nu) \colon \nu \in \mathcal{P}(\mathbb{R}^d) \right\}, \qquad (\mathcal{CB})$$

where the measure ν is the barycenter and W_2 is the 2-Wasserstein distance, namely the optimal transport problem with the quadratic cost.

If $x_1, \ldots, x_N \in \mathbb{R}^d$, then the barycenter $\mathcal{B}(x) = \lambda_1 x_1 + \cdots + \lambda_N x_N$ minimizes the functional

$$x \to \sum_{i=1}^{N} \lambda_i |x_i - x|^2.$$

Example of OT barycenters

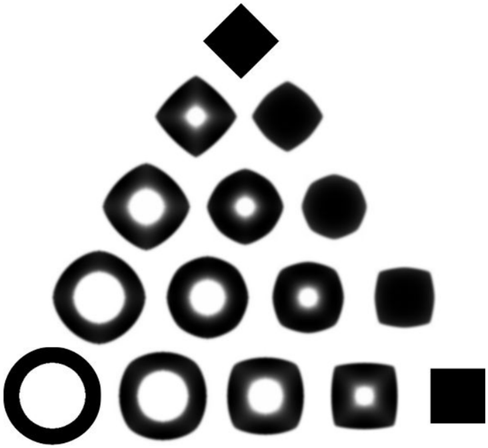


Figure: Example of OT barycenters with entropic regularization. This example was taken from the work of Luca Nenna [Nen16]

Multimarginal formulation of the Barycenter problem

For every $x := (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ define its Euclidean barycenter

$$\mathcal{B}(x) := \lambda_1 x_1 + \cdots + \lambda_N x_N.$$

Let us now introduce the multimarginal optimal transportation problem

$$\inf \left\{ \int \left(\sum_{i=1}^{N} \lambda_i |x_i - \mathcal{B}(x)|^2 \right) \gamma(dx_1, \dots, dx_N) \colon \gamma \in \Pi(\mu_1, \dots, \mu_N) \right\}. \quad (\mathcal{MB})$$

Theorem (Agueh & Carlier, [AC11])

Assume that μ_i vanishes on small sets for $i = 1, \dots, N$. Then optimal values in classical $(\mathcal{CB})$ and multimarginal $(\mathcal{MB})$ barycenter problems are the same and the solution of $(\mathcal{CB})$ is given by $\nu = \mathcal{B}_{\#}(\gamma)$, where γ is the solution of $(\mathcal{MB})$ and $\mathcal{B}(x) = \lambda_1 x_1 + \cdots + \lambda_N x_N$.

The repulsive harmonic cost

For the case of equal λ_i the cost function $\sum_{i=1}^{N} \lambda_i |x_i - \mathcal{B}(x)|^2$ is equivalent to the so-called Gangbo-Święch cost function $\sum_{i<j} |x_i - x_j|^2$, after the seminal work of Gangbo and Święch [GS98].

Problem (Multimarginal OT with the repulsive harmonic cost)

The repulsive harmonic cost function is a function having the form $c(x_1, \dots, x_N) = -\sum_{i,j=1}^{N} |x_i - x_j|^2$. *In the multimarginal OT with the repulsive harmonic cost we need to optimize* $\int \sum_{i,j=1}^{N} -|x_i - x_j|^2 \, d\gamma \to \inf$, $\gamma \in \Pi(\mu_1, \dots, \mu_N)$.

Since the cost function is "repulsive", if Monge-type solutions exist, they should follow the rule "the further, the better!", which means that we want to move the mass as much as we can. In other words, in the present case, optimal transport plans tend to be as spread as possible.

The 2-marginals case

Consider the case of 2 marginals with $d = 1$. Then the cost function has the form $-(x-y)^2$, which is equivalent to the cost function $(x+y)^2$ up to one-dimensional functions. The following lemma shows that in this case the solution to the problem is a Monge-type solution and concentrated on the graph of non-increasing function.

Lemma

Let μ and ν be probability measures on the real line, and let $c(x,y) = h(x+y)$, where $h\colon \mathbb{R} \to [0,+\infty)$ is a convex function. Assume that μ has no atoms. Then there exists a ***non-increasing*** *function $T\colon \mathbb{R} \to \mathbb{R}$ such that $T_\#(\mu) = \nu$ and*

$$\int h(x + T(x))\, \mu(dx) = \min \left\{ \int h(x+y)\, \gamma(dx, dy) \colon \gamma \in \Pi(\mu, \nu) \right\}.$$

This case is similar to the case of the classical square-distance cost function.

Flat tuple of measures

The cost function $-\sum_{i,j=1}^{N} |x_i - x_j|^2$ is trivially equivalent to the cost function $|x_1 + \cdots + x_N|^2$. Since $|x_1 + \cdots + x_N|^2 \geq 0$ and the equality is attained on the hyperplane $x_1 + \cdots + x_N = 0$, every transport plan concentrated on this hyperplane is optimal. In [MGN17] authors prove the following generalization of this trivial observation:

Theorem (Di Marino & Gerolin & Nenna)

Let $\{\mu_k\}_{k=1}^{N}$ be probability measures on $\mathbb{R}^d$ and $h\colon \mathbb{R}^d \to \mathbb{R}$ be a strictly convex function and suppose $c\colon (\mathbb{R}^d)^N \to \mathbb{R}$ be a cost function of the form $c(x_1, \ldots, x_N) = h(x_1 + \cdots + x_N)$. Then if there exists a plan $\gamma \in \Pi(x_1, \ldots, x_N)$ concentrated on the hyperplane of the form $x_1 + \cdots + x_N = C$, this plan is optimal for the multimarginal problem with cost c.

Definition

In this case we will say that γ is a flat optimal plan and $\{\mu_k\}_{k=1}^{N}$ is a flat N-tuple of measures.

Economical applications

It would be useful to formulate the previous problem in economical terms. Assume that $X_1, \dots, X_N$ are random variables, corresponding to the random losses for given business lines or risk types, over a fixed time period. An aggregate loss variable S has the form

$$S = \sum_{i=1}^{N} X_i.$$

In practice, there exist efficient and accurate statistical techniques to estimate the respective marginal distributions of $X_1, \dots, X_N$. On the other hand, the joint dependence structure of $X = (X_1, \dots, X_N)$ is often much more difficult to capture.

At the same time, regulators and companies are usually more concerned about a risk measure $\rho(S)$ instead of the exact dependence structure of X itself. This scenario is referred to as **risk aggregation with dependence uncertainty**, and has been extensively studied in **quantitative risk management**.

Economical applications. Value-at-Risk (VaR)

One of the most famous risk aggregation measures is **the Value-at-Risk (VaR)**. VaR of the aggregate risk S, calculated at a probability level $\alpha \in (0,1)$, is the α-quantile of its distribution, defined as

$$\mathrm{VaR}_\alpha(S) = F_S^{-1}(\alpha) = \inf\{x \colon P(S \leq x) \geq \alpha\}.$$

Value-at-Risk is a key metric used as a risk management constraint within portfolio optimization. For α typically close to 1, $\mathrm{VaR}_\alpha(S)$ is a measure of extreme loss, i.e. $P(S > \mathrm{VaR}_\alpha(S)) \leq 1 - \alpha$ is typically small.

Economical applications. Value-at-Risk (VaR)

Since banks often have more precise information about the marginal distributions X_i, but less about the aggregate risk S, we are interested in estimation of lower and upper bound for $\mathrm{VaR}_\alpha(S)$ over all admissible risk S.
For given marginals $\mu_1, \dots, \mu_N$, define

$$\overline{\mathrm{VaR}}_\alpha(S) = \sup\{\mathrm{VaR}_\alpha(X_1 + \dots + X_N) \colon X_i \sim \mu_i\},$$
$$\underline{\mathrm{VaR}}_\alpha(S) = \inf\{\mathrm{VaR}_\alpha(X_1 + \dots + X_N) \colon X_i \sim \mu_i\},$$

The bounds $\overline{\mathrm{VaR}}_\alpha(S)$ and $\underline{\mathrm{VaR}}_\alpha(S)$ are the worst-possible and, respectively, the best-possible VaR for the aggregate risk S at the probability level α. Of course, the choice of words best versus worst is arbitrarily and depends on the specific application.

Economical applications. Value-at-Risk (VaR)

Theorem (Bernard & Jiang & Wang, [BJW14])

Assume that every marginal distribution μ_k has positive density on its support. For $1 \leq k \leq N$, denote by $\mu_{k,\alpha}$ the restriction of the marginal μ_k to the interval $\big[F_{\mu_k}^{-1}(\alpha), +\infty\big)$ (upper tail), and by μ_k^α the restriction of the marginal μ_k to the interval $\big(-\infty, F_{\mu_k}^{-1}(\alpha)\big]$ (lower tail).

1. *Assume that there is a flat transport plan γ_α concentrated on the hyperplane $\{x_1 + \cdots + x_N = C_\alpha\}$ with the marginals $(\mu_{1,\alpha}, \ldots, \mu_{N,\alpha})$. Then*
$$\overline{\mathrm{VaR}}_\alpha(S) = C_\alpha.$$
2. *Assume that there is a flat transport plan γ_α concentrated on the hyperplane $\{x_1 + \cdots + x_N = C^\alpha\}$ with the marginals $(\mu_1^\alpha, \ldots, \mu_N^\alpha)$. Then*
$$\underline{\mathrm{VaR}}_\alpha(S) = C^\alpha.$$

Normal distributions case

Example (Wang & Wang, [WW16])

Suppose that $\sigma_1, \sigma_2, \sigma_3 > 0$ satisfy the inequality

$$2 \max_{1 \le i \le 3} \sigma_i \le \sigma_1 + \sigma_2 + \sigma_3. \tag{1}$$

Consider the random variable $(X_1, X_2, X_3) \sim \mathcal{N}(0, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_3^2 - \sigma_1^2 - \sigma_2^2) & \sigma_2^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) \\ \frac{1}{2}(\sigma_2^2 - \sigma_1^2 - \sigma_3^2) & \frac{1}{2}(\sigma_1^2 - \sigma_2^2 - \sigma_3^2) & \sigma_3^2 \end{pmatrix}.$$

One can easily verify that Σ is positive semi-definite and that $\mathbb{E}[(X_1+X_2+X_3)^2] = 0$. Thus, the triplet of normal distributions $(\mathcal{N}(0, \sigma_1^2), \mathcal{N}(0, \sigma_2^2), \mathcal{N}(0, \sigma_3^2))$ is a flat tuple under the condition (1).

Uniform distributions

Theorem (Di Marino & Gerolin & Nenna, [MGN17])

Let $\mu_i = \mu = \mathcal{L}^d|_{[0,1]^d}$ be the uniform measure on the d-dimensional cube $[0,1]^d \subset \mathbb{R}^d$ and suppose that c is a cost function of the form $c(x_1, \ldots, x_N) = h(x_1 + \cdots + x_N)$ for some strictly convex function $h\colon \mathbb{R}^d \to \mathbb{R}$. Then there is a function $T\colon [0,1]^d \to [0,1]^d$ such that $T^N(x) = x$, $T_{\#}\mu = \mu$, and

$$\min_{\gamma \in \Pi(\mu_i)} \int c \, d\gamma = \int_{[0,1]^d} c(x, T(x), T^2(x), \ldots, T^{N-1}(x)) \, dx.$$

Moreover, T is not differentiable at any point and it is a fractal map.

Uniform distributions, construction

We express every $z \in [0,1]$ by its base-N system, $z = \sum_{k=1}^{+\infty} \frac{a_k}{N^k}$ with $a_k \in \{0,\dots,N-1\}$. Consider the map given by

$$S(z) = \sum_{k=1}^{+\infty} \frac{\mathcal{S}(a_k)}{N^k},$$

where $\mathcal{S}$ is the permutation of N symbols such that $\mathcal{S}(i) = i+1$ for $0 \le i \le N-2$ and $\mathcal{S}(N-1) = 0$. The uniform measure on the interval $(0,1)$ is invariant under this mapping. In addition,

$$S^0(z) + \dots + S^{N-1}(z) = \sum_{k=1}^{+\infty} \frac{\mathcal{S}^0(a_k) + \dots + \mathcal{S}^{N-1}(a_k)}{N^k} = \sum_{k=1}^{+\infty} \frac{N(N-1)}{2N^k} = \frac{N}{2}.$$

Uniform distributions, construction

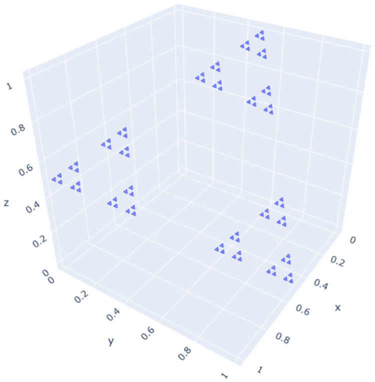


Figure: The graph of the mapping $x \to (x, T(x), T^2(x))$ for the case $N = 3$, $d = 1$.

Let $T\colon [0,1]^d \to [0,1]^d$ be the map defined by

$$T(x) = T(z_1, \ldots, z_d) = (S(z_1), \ldots, S(z_d)).$$

Since S maps the uniform distribution to itself, the measure $\mathcal{L}^d|_{[0,1]^d}$ is also invariant under the mapping T. Finally,

$$x + T(x) + \cdots + T^{N-1}(x) = \left(\frac{N}{2}, \ldots, \frac{N}{2}\right)$$

Necessary condition for flatness

Consider the case $d = 1$, and assume that the marginal μ_i is supported on an interval (l_i, r_i) for $i = 1, \ldots, N$. Denote by $\mathbb{E}(\mu_i)$ the first moment of the distribution μ_i: $\mathbb{E}(\mu_i) = \int_{l_i}^{r_i} x\, \mu_i(dx)$.

Proposition

Assume that the tuple $(\mu_1, \ldots, \mu_N)$ is flat and that γ is a flat optimal plan concentrated on the hyperplane $\{x_1 + \cdots + x_N = C\}$. Then $C = \mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N)$ and the inequality

$$l_1 + \cdots + l_N + (r_k - l_k) \leq \mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N) \leq r_1 + \cdots + r_N - (r_k - l_k)$$

holds for every $k = 1, \ldots, N$.

Necessary condition for flatness

Let's check the equation $C = \mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N)$:

$$\mathbb{E}(\mu_1) + \ldots \mathbb{E}(\mu_N) = \int_{l_1}^{r_1} x_1\, \mu_1(dx_1) + \cdots + \int_{l_N}^{r_N} x_N\, \mu_N(dx_N)$$
$$= \int_{\mathbb{R}^N} (x_1 + \cdots + x_N)\, \gamma(dx_1, \ldots, dx_N) = C.$$

For γ-a.a. points $(x_1, \ldots, x_N)$ we have $l_k \le x_k \le r_k$ and $x_1 + \ldots x_N = C = \mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N)$. Hence, $x_k = \mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N) - \sum_{i \ne k} x_i$, and therefore

$$\mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N) - \sum_{i \ne k} r_i \le x_k \le \mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N) - \sum_{i \ne k} l_i$$

for μ_k-a.a. points x_k. In particular,

$$\mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N) - \sum_{i \ne k} r_i \le l_k \quad \text{and} \quad r_k \le \mathbb{E}(\mu_1) + \cdots + \mathbb{E}(\mu_N) - \sum_{i \ne k} l_i.$$

Flatness for the case of marginals with non-increasing densities

Assume that the density function $d\mu_k/dx$ is non-increasing on the supporting interval (l_k, r_k). Then the necessary condition becomes a sufficient one.

Theorem (Wang & Wang, [WW16])

Assume that μ_k is concentrated on the interval (l_k, r_k) and that the density function $d\mu_k/dx$ is non-increasing on that interval. Then the tuple of marginals $(\mu_1, \dots, \mu_N)$ is flat if and only if the inequality

$$r_k - l_k \leq \mathbb{E}(\mu_1) + \dots + \mathbb{E}(\mu_N) - (l_1 + \dots + l_N)$$

is satisfied for all $k = 1, \dots, N$.

I proved this theorem in the preprint [Zim20], but recently I discovered that this theorem was already proven 5 years ago.

More examples

- Assume that $\mu_1 = \cdots = \mu_N = \mu$, where μ is concentrated on the interval (l, r) and $d\mu/dx$ is non-increasing on (l, r). Then $(\mu_1, \ldots, \mu_N)$ is flat if and only if $\mathbb{E}(\mu) \geq (1 - \frac{1}{N})l + \frac{1}{N}r$.
- If $\mu_k \sim U[l_k, r_k]$, then the tuple of measures $(\mu_1, \ldots, \mu_N)$ is flat if and only if the inequality

$$2(r_k - l_k) \leq (r_1 - l_1) + \cdots + (r_N - l_N)$$

holds for all k, i.e. the numbers $r_k - l_k$ are the lengths of the sides of some polygon with N vertices.

Applications to the product cost problem

Problem (Product cost problem)

Let $\mu_1, \ldots, \mu_N$ be probability measures on the real line.
(Primal problem) *Find a transport plan $\gamma \in \Pi(\mu_1, \ldots, \mu_N)$ minimizing the functional*

$$\int_{\mathbb{R}^N} x_1 x_2 \ldots x_N \, \gamma(dx_1, \ldots, dx_N) \to \min .$$

(Discrete "Monge" version) *Given samples $x_{i,k} \sim \mu_i$, $1 \le k \le M$. We need to find the permutations $\sigma_i \in S_M$ minimizing the total sum*

$$\frac{1}{M} \left(\sum_{k=1}^{M} x_{1,\sigma_1(k)} x_{2,\sigma_2(k)} \ldots x_{N,\sigma_N(k)} \right) \to \min .$$

Next, we will consider the case $N = 3$ and $\mu_1 = \mu_2 = \mu_3 = \mathcal{L}|_{[0,1]}$ (see [GZ20]).

C-monotonicity

Consider the discrete "Monge" version of the product cost problem. Let $(p_1, p_2, \ldots, p_N)$ and $(q_1, q_2, \ldots, q_N)$ be the elements of the optimal matching. If the replace these triples with $(q_1, p_2, p_3, \ldots, p_N)$ and $(p_1, q_2, q_3, \ldots, q_N)$ (we swap the first elements of tuples), we will also produce a matching. The total output will be changed by

$$\Delta = \frac{1}{M}\left(-p_1 p_2 \ldots p_N - q_1 q_2 \ldots q_N + q_1 p_2 p_3 \ldots p_N + p_1 q_2 q_3 \ldots q_N\right)$$
$$= -\frac{1}{M}(p_1 - q_1)(p_2 p_3 \ldots p_N - q_2 q_3 \ldots q_N)$$

Since the initial matching was with the minimal total output, we conclude that $\Delta \geq 0$. This means that if $p_1 < q_1$, then $p_2 p_3 \ldots p_N \geq q_2 q_3 \ldots q_N$ and vice versa.

Rearrangement algorithm

Using C-monotonicity, one can construct the following algorithm for construction the (sub)optimal solution. For simplicity, we consider only the case $N = 3$ and $\mu_1 = \mu_2 = \mu_3 = \mathcal{L}|_{[0,1]}$.

Algorithm (Rearrangement algorithm [GZ20, EPR13])

1. Consider the matrix $X^{(0)}$ of the size $M \times 3$: $X_{ij}^{(0)} = \frac{i}{M}$.
2. (optional) Permute randomly the elements in each column of $X^{(0)}$.
3. Iteratively rearrange the jth column of the matrix $X^{(k)}$ so that it becomes oppositely ordered to the product of the other columns, for $1 \leq j \leq 3$. The matrix $X^{(k+1)}$ is found.
4. Repeat Step 3 until $\left|\sum_{i=1}^M X_{i1}^{(k)} X_{i2}^{(k)} X_{i3}^{(k)} - \sum_{i=1}^M X_{i1}^{(k+1)} X_{i2}^{(k+1)} X_{i3}^{(k+1)}\right| < \varepsilon$.
5. The sum of Dirac measures $\frac{1}{M}\sum_{i=1}^M \delta(X_{i1}, X_{i2}, X_{i3})$ approximates the optimal transport plan.

The support of the solution to the primal problem

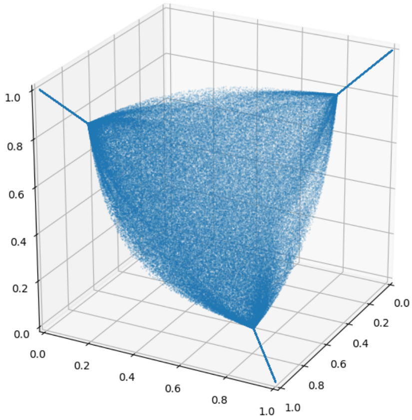


Figure: The support of the solution to the product cost problem for the case of uniform marginals

Let t be the solution of the transcendental equation

$$\ln(1-2t) - \ln(t) = 3 - 9t.$$

The support of the solution consists of 4 parts:

- 3 segments: the first connects the points $(1, 0, 0)$ and $(t, 1-2t, 1-2t)$, the second connects $(0, 1, 0)$ and $(1-2t, t, 1-2t)$, and the third connects $(0, 0, 1)$ and $(1-2t, 1-2t, t)$;
- the surface $x_1 x_2 x_3 = t(1-2t)^2$, where $t \le x_i \le (1-2t)$.

Connection with the repulsive harmonic cost

Let's make a change of variables $x_i = \exp(-y_i)$. Then

$$x_1 x_2 x_3 = \exp(-y_1 - y_2 - y_3) = h(y_1 + y_2 + y_3),$$

where $h(x) = \exp(-x)$ is a strictly convex function. Also,

$$d\widehat{\mu}_i = \exp(-y_i) \cdot \mathrm{Ind}(y_i \geq 0)\, dy_i,$$

so the density function is decreasing on $[0, +\infty)$. Unfortunately, the required inequality $\mathbb{E}(\widehat{\mu}) \geq \frac{2}{3} l + \frac{1}{3} r$ does not hold.

Connection with the repulsive harmonic cost

Remark

Assume that $\mu_i \sim U[a,b]$. Then

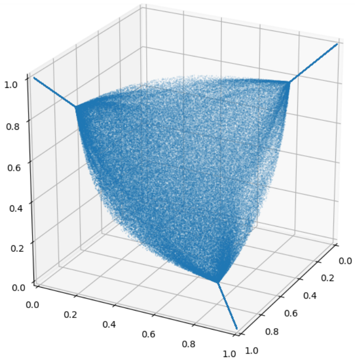
$$d\widehat{\mu}_i = \frac{\exp(-y_i)\cdot \mathrm{Ind}[-\ln(b) \le y_i \le -\ln(a)]}{b-a}\, dy_i.$$

In addition, $\mathbb{E}(\widehat{\mu}) = \frac{a\ln(a)-b\ln(b)}{b-a} + 1$. So, if the boundaries (a,b) satisfy the inequality

$$\frac{a\ln(a)-b\ln(b)}{b-a} + 1 \ge -\frac{2}{3}\ln(b) - \frac{1}{3}\ln(a) \Leftrightarrow 3(b-a) \ge (2a+b)(\ln(b)-\ln(a)),$$

then there is a measure π with the uniform projection supported on the surface $x_1x_2x_3 = C$, and this measure is optimal for any cost function $h(-\ln(x_1)-\ln(x_2)-\ln(x_3))$, where h is convex.

Idea of the construction



Idea of the construction for $\mu = \mathcal{L}|_{[0,1]}$.

1. The measure $\mu \sim U[0,1]$ is splitted onto 2 components (tails and a center part):

$$\mu = (\mathcal{L}|_{[0,l]} + \mathcal{L}|_{[r,1]}) + \mathcal{L}|_{[l,r]}.$$

2. Next, we find a measure μ concentrated on the surface $x_1 x_2 x_3 = C$ such that $\mathrm{pr}_i(\pi) = \mathcal{L}|_{[l,r]}$. The existence of such measure is guaranteed by the inequality $3(r-l) \geq (2l+r)(\ln(r)-\ln(l))$.
3. Find a measure π_1 on a first segment such that $\mathrm{pr}_1(\pi_1) = \mathcal{L}|_{[0,l]}$ and $\mathrm{pr}_{2,3}(\pi_1) = \mathcal{L}|_{[r,1]}$. The measures π_2 and π_3 are symmetric. Then $2\mu([0,l]) = \mu([r,1])$, or $r = 1-2l$.

Independence on the function h

Theorem ([GZ20])

Assume that $\mu_1, \mu_2, \mu_3 = \mathcal{L}|_{[0,1]}$ *and* $c(x_1, x_2, x_3) = h(-\ln(x_1) - \ln(x_2) - \ln(x_3))$ *for some strictly convex function* h. *Then the measure* $\pi \in \Pi(\mu_1, \mu_2, \mu_3)$ *constructed above minimizes the functional*

$$\int_{[0,1]^3} c(x_1, x_2, x_3)\, \pi(dx_1, dx_2, dx_3) \to \min .$$

In particular, this measure is independent on h.

Closed form solution to the dual problem

One can prove the optimality of the constructed measure with the help of the dual problem.

Theorem ([GZ20])

Consider the cost function $c(x_1, x_2, x_3) = x_1 x_2 x_3$. *Then the triple of functions*

$$\varphi_k(x) = \begin{cases} c\ln(l) - \frac{1}{3}(c\ln(c) - c) + \frac{1}{6}((2x-1)^3 - (2l-1)^3), & \text{if } 0 \le x \le l, \\ c\ln(x) - \frac{1}{3}(c\ln(c) - c), & \text{if } l \le x \le r, \\ c\ln(r) - \frac{1}{3}(c\ln(c) - c) + \frac{1}{4}(x^2 - r^2) - \frac{1}{6}(x^3 - r^3), & \text{if } r \le x \le 1 \end{cases}$$

is a solution to the dual problem, where l *is the solution of the equation*

$$3(1 - 3l) = \ln(1 - 2l) - \ln(l),$$

$r = 1 - 2l$, *and* $c = lr^2$.

The multistochastic case

Problem (The particular case of the multistochastic MK problem)

Let $X_1 = X_2 = X_3 = [0,1]$*, and let* μ_{ij} *be the restriction of the Lebesgue measure to the unit square* $[0,1]^2$*. We need to maximize*

$$\int_{[0,1]^3} x_1 x_2 x_3 \, \pi(dx_1, dx_2, dx_3)$$

over all probability measures π *on* $X_1 \times X_2 \times X_3$ *with the property that its projection on* $X_i \times X_j$ *is* μ_{ij} *for all* $1 \le i < j \le 3$*.*

Unlike the multimarginal product cost problem, we fix not only marginals of the transport plan, but also the projections on 2-dimensional faces. In the general case of a multistochastic problem, we fix the projections on all k-dimensional faces.

Fractal example for the multistochastic case

Given a couple of numbers $x, y \in [0,1]$, we consider their binary representations

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

Then the xor operation is defined as follows:

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

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

Theorem ([GKZ19])

Denote by π the image of the Lebesgue measure $\mathcal{L}^2|_{[0,1]^2}$ under the mapping $(x_1, x_2) \mapsto (x_1, x_2, x_1 \oplus x_2)$. Then $\mathrm{pr}_{ij}\pi = \mu_{ij}$ for all $1 \le i < j \le 3$, and π is the solution to the problem considered above.

The Sierpinski tetrahedron

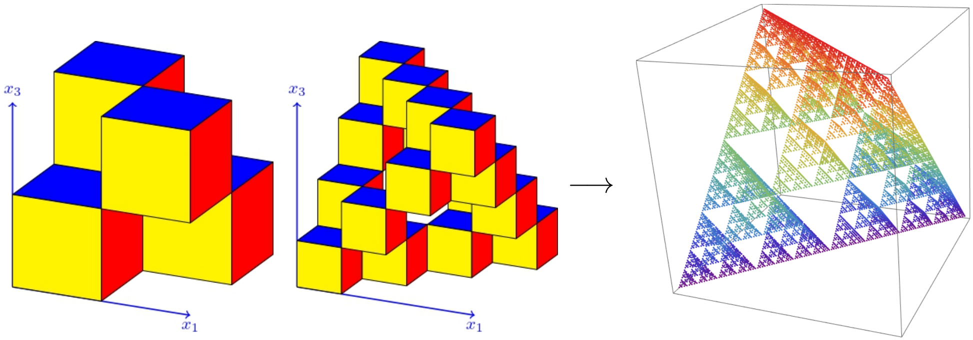


Figure: The support of the optimal measure π. The measure π is concentrated on the set $\{(x_1, x_2, x_3) \colon x_1 \oplus x_2 \oplus x_3 = 0\}$, which is called *the Sierpinski tetrahedron*. One can find the support by the following process. Start with the unit cube $[0, 1]^3$. On each step, split every cube on 8 equal subcubes, paint all of them in a checkerboard pattern, and remove all the subcubes colored in the second color.

Dual to the multistochastic MK problem

Problem (Dual to the multistochastic MK problem)

Find a triple of functions $(\varphi_{12}(x_1, x_2), \varphi_{13}(x_1, x_3), \varphi_{23}(x_2, x_3))$ *satisfying the inequality*

$$\varphi_{12}(x_1, x_2) + \varphi_{13}(x_1, x_3) + \varphi_{23}(x_2, x_3) \geq x_1 x_2 x_3$$

for all $0 \leq x_1, x_2, x_3 \leq 1$ *and minimizing the sum*

$$\int_0^1 \int_0^1 \varphi_{12}(x_1, x_2)\, dx_1 dx_2 + \int_0^1 \int_0^1 \varphi_{13}(x_1, x_3)\, dx_1 dx_3 + \int_0^1 \int_0^1 \varphi_{23}(x_2, x_3)\, dx_2 dx_3$$

Theorem (Kantorovich duality)

The supremum in the primal problem and the infimum in the dual problem are the same. Moreover, both of them are achieved.

Closed form solution to the dual problem

Consider the function

$$I(a, b) = \int_0^a \int_0^b x \oplus y \, dxdy.$$

Theorem (Solution to the dual problem)

The triple of functions

$$\varphi_{12}(a, b) = \varphi_{13}(a, b) = \varphi_{23}(a, b) = I(a, b) - \frac{1}{4} I(a, a) - \frac{1}{4} I(b, b)$$

solves the dual problem.

Thank you for your attention!

References I

[AC11] Martial Agueh and Guillaume Carlier, *Barycenters in the wasserstein space*, SIAM J. Math. Analysis **43** (2011), 904–924.

[BJW14] Carole Bernard, Xiao Jiang, and Ruodu Wang, *Risk aggregation with dependence uncertainty*, Insurance: Mathematics and Economics **54** (2014), 93–108.

[EPR13] Paul Embrechts, Giovanni Puccetti, and Ludger Rüschendorf, *Model uncertainty and var aggregation*, Journal of Banking & Finance **37** (2013), no. 8, 2750–2764.

[GKZ19] Nikita A. Gladkov, Alexander V. Kolesnikov, and Alexander P. Zimin, *On multistochastic monge–kantorovich problem, bitwise operations, and fractals*, Calculus of Variations and Partial Differential Equations **58** (2019), no. 5, 173.

References II

[GS98] Wilfrid Gangbo and Andrzej Święch, *Optimal maps for the multidimensional monge-kantorovich problem*, Communications on Pure and Applied Mathematics **51** (1998), no. 1, 23–45.

[GZ20] Nikita A. Gladkov and Alexander P. Zimin, *An explicit solution for a multimarginal mass transportation problem*, SIAM J. Math. Analysis **52** (2020), no. 4, 3666–3696.

[MGN17] Simone Di Marino, Augusto Gerolin, and Luca Nenna, *Optimal transportation theory for repulsive costs*, Topological Optimization and Optimal Transport In the Applied Sciences (2017).

[Nen16] Luca Nenna, *Numerical methods for multi-marginal optimal transportation*, Ph.D. thesis, 12 2016.

References III

[WW16] Bin Wang and Ruodu Wang, *Joint mixability*, Mathematics of Operations Research **41** (2016), 808–826.

[Zim20] Alexander P. Zimin, *On existence of measure with given marginals supported on a hyperplane*, 2020.