NON-MONOTONE CONVERGENCE IN THE QUADRATIC WASSERSTEIN DISTANCE

WALTER SCHACHERMAYER, UWE SCHMOCK, AND JOSEF TEICHMANN

ABSTRACT. We give an easy counter-example to Problem 7.20 from C. Villani's book on mass transport: in general, the quadratic Wasserstein distance between n-fold normalized convolutions of two given measures fails to decrease monotonically.

We use the terminology and notation from [5]. For Borel measures μ, ν on $\mathbb{R}^d$ we define the quadratic Wasserstein distance

$$\mathcal{T}(\mu,\nu) := \inf_{(X,Y)} \mathbb{E}\big[\|X - Y\|^2\big]$$

where $\|\cdot\|$ is the Euclidean distance on $\mathbb{R}^d$ and the pairs (X,Y) run through all random vectors defined on some common probabilistic space $(\Omega, \mathcal{F}, \mathbb{P})$, such that X has distribution μ and Y has distribution ν. By a slight abuse of notation we define $\mathcal{T}(U,V) := \mathcal{T}(\mu,\nu)$ for two random vectors U, V, such that U has distribution μ and V has distribution ν. The following theorem (see [5, Proposition 7.17]) is due to Tanaka [4].

Theorem 1. *For $a, b \in \mathbb{R}$ and square integrable random vectors X, Y, X', Y' such that X is independent of Y, and X' is independent of Y', and $\mathbb{E}[X] = \mathbb{E}[X']$ or $\mathbb{E}[Y] = \mathbb{E}[Y']$, we have*

$$\mathcal{T}(aX + bY, aX' + bY') \leq a^2 \mathcal{T}(X,X') + b^2 \mathcal{T}(Y,Y').$$

For a sequence of i.i.d. random vectors $(X_i)_{i\in\mathbb{N}}$ we define the normalized partial sums

$$S_m := \frac{1}{\sqrt{m}} \sum_{i=1}^{m} X_i, \qquad m \in \mathbb{N}.$$

If μ denotes the law of X_1, we write $\mu^{(m)}$ for the law of S_m. Clearly $\mu^{(m)}$ equals, up to the scaling factor $\sqrt{m}$, the m-fold convolution $\mu * \mu * \cdots * \mu$ of μ.

We shall always deal with measures μ, ν with vanishing barycenter. Given two measures μ and ν on $\mathbb{R}^d$ with finite second moments, we let $(X_i)_{i\in\mathbb{N}}$ and $(X_i')_{i\in\mathbb{N}}$ be i.i.d. sequences with law μ and ν, respectively, and denote by S_m and S_m' the corresponding normalized partial sums. From Theorem 1 we obtain

$$\mathcal{T}\big(\mu^{(2m)}, \nu^{(2m)}\big) \leq \mathcal{T}\big(\mu^{(m)}, \nu^{(m)}\big), \qquad m \in \mathbb{N},$$

from which one may quickly deduce a proof of the Central Limit Theorem (compare [5, Ch. 7.4] and the references given there).

However, we can *not* deduce from Theorem 1 that the inequality

$$(1) \qquad \mathcal{T}\big(\mu^{(m+1)}, \nu^{(m+1)}\big) \le \mathcal{T}\big(\mu^{(m)}, \nu^{(m)}\big)$$

holds true for all $m \in \mathbb{N}$. Specializing to the case $m = 2$, an estimate, which we can obtain from Tanaka's Theorem, is

$$\mathcal{T}\big(\mu^{(3)}, \nu^{(3)}\big) \le \frac{1}{3}\big[2\mathcal{T}\big(\mu^{(2)}, \nu^{(2)}\big) + \mathcal{T}(\mu, \nu)\big] \le \mathcal{T}(\mu, \nu).$$

This contains some valid information, but does not imply (1). It was posed as Problem 7.20 of [5], whether inequality (1) holds true for all probability measures μ, ν on $\mathbb{R}^d$ and all $m \in \mathbb{N}$.

The subsequent easy example shows that the answer is no, even for $d = 1$ and symmetric measures. We can choose $\mu = \mu_n$ and $\nu = \nu_n$ for sufficiently large $n \ge 2$, as the proposition (see also Remark 1) shows.

Proposition 1. *Denote by μ_n the distribution of $\sum_{i=1}^{2n-1} Z_i$, and by ν_n the distribution of $\sum_{i=1}^{2n} Z_i$ with $(Z_i)_{i \in \mathbb{N}}$ i.i.d. and $\mathbb{P}(Z_1 = 1) = \mathbb{P}(Z_1 = -1) = \frac{1}{2}$. Then*

$$(2) \qquad \lim_{n \to \infty} \sqrt{n}\, \mathcal{T}(\mu_n * \mu_n, \nu_n * \nu_n) = \frac{2}{\sqrt{2\pi}},$$

while $\mathcal{T}(\mu_n * \mu_n * \mu_n, \nu_n * \nu_n * \nu_n) \ge 1$ for all $n \in \mathbb{N}$.

Remark 1. *If one only wants to find a counter-example to Problem 7.20 of [5], one does not really need the full strength of Proposition 1, i.e. the estimate that $\mathcal{T}(\mu_n * \mu_n, \nu_n * \nu_n) = \mathcal{O}(1/\sqrt{n})$. In fact, it is sufficient to consider the case $n = 2$ in order to contradict the monotonicity of inequality (1). Indeed, a direct calculation reveals that*

$$\mathcal{T}(\mu_2 * \mu_2, \nu_2 * \nu_2) = 0.625 < \frac{2}{3} \le \left(\frac{\sqrt{2}}{\sqrt{3}}\right)^2 \mathcal{T}(\mu_2 * \mu_2 * \mu_2, \nu_2 * \nu_2 * \nu_2).$$

Proof of Proposition 1. We start with the final assertion, which is easy to show. The 3-fold convolutions of the measures μ_n and ν_n, respectively, are supported on odd and even numbers, respectively. Hence they have disjoint supports with distance 1 and so the quadratic transportation costs are bounded from below by 1.

For the proof of (2), fix $n \in \mathbb{N}$, define $\sigma_n = \mu_n * \mu_n$ and $\tau_n = \nu_n * \nu_n$, and note that σ_n and τ_n are supported by the even numbers. For $k = -(2n-1), \ldots, (2n-1)$ we denote by $p_{n,k}$ the probability of the point $2k$ under σ_n, i.e.

$$p_{n,k} = \binom{4n-2}{k+2n-1} \frac{1}{2^{4n-2}}.$$

We define $p_{n,k} = 0$ for $|k| \ge 2n$. We have $\tau_n = \sigma_n * \rho$, where ρ is the distribution giving probability $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ to $-2, 0, 2$, respectively. We deduce that for $0 \le k \le 2n-2$,

$$
\begin{aligned}
\tau_n(2k+2) &= \frac{1}{4}p_{n,k} + \frac{1}{4}p_{n,k+2} + \frac{1}{2}p_{n,k+1} \\
(3) \qquad &= \frac{1}{4}(p_{n,k} - p_{n,k+1}) + \frac{1}{4}(p_{n,k+2} - p_{n,k+1}) + \sigma_n(2k+2) \\
&= \frac{1}{4}p_{n,k}\Big(1 - \frac{p_{n,k+1}}{p_{n,k}}\Big) + \frac{1}{4}p_{n,k+1}\Big(\frac{p_{n,k+2}}{p_{n,k+1}} - 1\Big) + \sigma_n(2k+2).
\end{aligned}
$$

Notice that $p_{n,k} \geq p_{n,k+1}$ for $0 \leq k \leq 2n - 1$. The term in the first parentheses is therefore non-negative. It can easily be calculated and estimated via

$$0 \leq 1 - \frac{p_{n,k+1}}{p_{n,k}} = 1 - \frac{\binom{4n-2}{k+2n}}{\binom{4n-2}{k+2n-1}} = 1 - \frac{2n-k-1}{k+2n} = \frac{2k+1}{2n+k} \leq \frac{2k+1}{2n},$$

for $0 \leq k \leq 2n - 1$.

Following [5] we know that the quadratic Wasserstein distance $\mathcal{T}$ can be given by a cyclically monotone transport plan $\pi = \pi_n$. We define the transport plan π via an intuitive transport map T. It is sufficient to define T for $0 \leq k \leq 2n - 1$, since it acts symmetrically on the negative side. T moves mass $\frac{1}{4}p_{n,k}\frac{2k+1}{2n+k}$ from the point $2k$ to $2k + 2$ for $k \geq 1$. At $k = 0$ the transport T moves $\frac{1}{8n}p_{n,0}$ to every side, which is possible, since there is enough mass concentrated at 0.

By equation (3) we see that the transport T moves σ_n to τ_n, since, for $1 \leq k \leq 2n - 2$, the first terms corresponds to the mass, which arrives from the left and is added to σ_n, and the second term to the mass, which is transported away: summing up one obtains τ_n. For $k = 2n - 1$, mass only arrives from the left. At $k = 0$ mass is only transported away. By the symmetry of the problem around 0 and by the quadratic nature of the cost function (the distance of the transport is 2, hence cost 2^2), we finally have

$$\mathcal{T}(\sigma_n, \tau_n) \leq 2 \sum_{k=0}^{2n-1} \frac{2^2}{4} p_{n,k} \frac{2k+1}{2n+k} \leq \sum_{k=0}^{2n-1} p_{n,k} \frac{2k+1}{n}.$$

By the Central Limit Theorem and uniform integrability of the function $x \mapsto x_+ := \max(0, x)$ with respect to the binomial approximations, we obtain

$$\lim_{n\to\infty} \frac{1}{2\sqrt{n}} \sum_{k=0}^{2n-1} (2k)p_{n,k} = \int_0^\infty \frac{x}{\sqrt{2\pi}} e^{-x^2/2} \, dx.$$

Hence

$$\limsup_{n\to\infty} \sqrt{n}\, \mathcal{T}(\sigma_n, \tau_n) \leq \frac{2}{\sqrt{2\pi}} \approx 0.79788.$$

In order to obtain equality we start from the local monotonicity of the respective transport maps on non-positive and non-negative numbers. It easily follows that the given transport plan is cyclically monotone and hence optimal (see [5, Ch. 2]). The subsequent equality allows also to consider estimates from below. Rewriting (3) yields

$$\tau_n(2k+2) = \frac{1}{4}p_{n,k+1}\left(\frac{p_{n,k}}{p_{n,k+1}} - 1\right) + \frac{1}{4}p_{n,k+2}\left(1 - \frac{p_{n,k+1}}{p_{n,k+2}}\right) + \sigma_n(2k+2)$$

for $0 \leq k \leq 2n - 3$, and

$$\tau_n(2k+2) = \frac{1}{4}p_{n,k+1}\left(\frac{p_{n,k}}{p_{n,k+1}} - 1\right) + \sigma_n(2k+2)$$

for $k = 2n - 2$. Furthermore,

$$\frac{p_{n,k}}{p_{n,k+1}} - 1 = \frac{\binom{4n-2}{k+2n-1}}{\binom{4n-2}{k+2n}} - 1 = \frac{k+2n}{2n-k-1} - 1 = \frac{2k+1}{2n-k-1} \geq \frac{2k+1}{2n}$$

for $0 \leq k \leq 2n - 2$. This yields by a reasoning similar to the above that

$$\mathcal{T}(\sigma_n, \tau_n) \geq \sum_{k=0}^{2n-2} p_{n,k+1} \frac{2k+1}{n},$$

hence

$$\liminf_{n\to\infty} \sqrt{n}\, \mathcal{T}(\sigma_n, \tau_n) \geq \frac{2}{\sqrt{2\pi}}.$$

$\square$

Remark 2. *Let $p \geq 2$ be an integer. By slight modifications of the proof of Proposition 1 we can construct sequences of measures $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$, such that the quadratic Wasserstein distances of k-fold convolutions are bounded from below by 1 for all k which are not multiples of p, while*

$$\lim_{n\to\infty} \mathcal{T}(\mu_n^{(p)}, \nu_n^{(p)}) = 0.$$

Remark 3. *Assume the notations of [5]. In the previous considerations we can replace the quadratic cost function by any other lower semi-continuous cost function $c : \mathbb{R}^2 \to [0, +\infty]$, which is bounded on parallels to the diagonal and vanishes on the diagonal. For example, if we choose $c(x,y) = |x - y|^r$ for $0 < r < \infty$, then we obtain the same asymptotics as in Proposition 1 (with a different constant).*

Remark 4. *We have used in the above proof that τ_n is obtained from σ_n by convolving with the measure ρ. In fact, this theme goes back (at least) as far as L. Bachelier's famous thesis from 1900 on option pricing [2, p. 45]. Strictly speaking, L. Bachelier deals with the measure assigning mass $\frac{1}{2}$ to -1, 1 and considers consecutive convolutions, instead of the above ρ. Hence convolutions with ρ correspond to Bachelier's result after two time steps. Bachelier makes the crucial observation that this convolution leads to a* radiation *of probabilities: Each stock price x radiates during a time unit to its neighboring price a quantity of probability proportional to the difference of their probabilities. This was essentially the argument which allowed us to prove (1). Let us mention that Bachelier uses this argument to derive the fundamental relation between Brownian motion (which he was the first to define and analyse in his thesis) and the heat equation (compare e.g. [3] for more on this topic).*

Remark 5. *Having established the above counterexample, it becomes clear how to modify Problem 7.20 from [5] to give it a chance to hold true. This possible modification was also pointed out to us by C. Villani.*

Problem 1. *Let μ be a probability measure on $\mathbb{R}^d$ with finite second moment and vanishing barycenter, and γ the Gaussian measure with same first and second moments. Does $(\mathcal{T}(\mu^{(n)}, \gamma))_{n\geq 1}$ decrease monotonically to zero?*

When entropy is considered instead of the quadratic Wasserstein distance the corresponding question on monotonicity was answered affirmatively in the recent paper [1].

One may also formulate a variant of Problem 7.20 as given in (1) by replacing the measure ν through a log-concave probability distribution. This would again generalize problem 1.

References

[1] S. Artstein, K. M. Ball, F. Barthe and A. Naor, *Solution of Shannon's Problem on the Monotonicity of Entropy*, Journal of the AMS **17**(4), 2004, pp. 975–982.

[2] L. Bachelier, *Theorie de la Speculation*, Paris, 1900, see also: http://www.numdam.org/en/.

[3] W. Schachermayer, *Introduction to the Mathematics of Financial Markets*, LNM 1816 - Lectures on Probability Theory and Statistics, Saint-Flour summer school 2000 (Pierre Bernard, editor), Springer Verlag, Heidelberg (2003), pp. 111–177.

[4] H. Tanaka, *An inequality for a functional of probability distributions and its applications to Kac's one-dimensional model of a Maxwell gas*, Zeitschrift fr Wahrscheinlichkeitstheorie und verwandte Gebiete **27**, 47–52, 1973

[5] C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Mathematics **58**, American Mathematical Society, Providence Rhode Island, 2003.

Financial and Actuarial Mathematics, Technical University Vienna, Wiedner Hauptstrasse 8–10, A-1040 Vienna, Austria.