



HAL
open science

Sharp inequalities between Zolotarev and Wasserstein distances in $P_2(\mathbb{R}^d)$

Karol Bolbotowski, Guy Bouchitté

► **To cite this version:**

Karol Bolbotowski, Guy Bouchitté. Sharp inequalities between Zolotarev and Wasserstein distances in $P_2(\mathbb{R}^d)$. 2025. <hal-05346799>

HAL Id: hal-05346799

<https://hal.science/hal-05346799v1>

Preprint submitted on 4 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

SHARP INEQUALITIES BETWEEN ZOLOTAREV AND WASSERSTEIN DISTANCES IN $\mathcal{P}_2(\mathbb{R}^d)$

KAROL BOLBOTOWSKI AND GUY BOUCHITTÉ

ABSTRACT. Based on a new Kantorovich-Rubinstein duality principle for the Hessian that was recently established by the two authors, we extend the Rio inequality to any dimension $d \geq 1$ with an optimal constant. Similarly, we propose an optimal upper bound for the ratio of Zolotarev distance $Z_2(\mu, \nu)$ to Wasserstein distance $W_2(\mu, \nu)$ when $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ are centred probabilities with prescribed variances.

Keywords: Probability metrics, Optimal transport, Kantorovich-Rubinstein duality, Wasserstein distance, Zolotarev distance, Rio inequality.

2020 Mathematics Subject Classification: 26D10, 49K30, 60E05, 60E15.

1. INTRODUCTION

A considerable amount of literature is devoted to constructing metric structures on sets of measures. It begins with total variation (TVD) and then, based on the needs of statistics and information theory, moves on to the pioneering works of Kullback-Leibler divergence (KL divergence) and Hellinger-Bhattacharyya distance. More recently, Monge-Kantorovich optimal transport theory gave rise to transportation metrics, such as the Prokhorov distance (see for instance [9]) and the Wasserstein distances [21, 18]. In this vein, V.M. Zolotarev introduced in the 1970s the vast class of so-called ideal metrics [22]. This class includes TVD and the first Wasserstein metric.

This topic remains active today, providing a fundamental framework for several areas of research. Notably, it is used in mathematical physics to treat random measures as random variables in a Polish metric space. In statistics, it is used to evaluate rates of convergence in the Central Limit Theorem and to establish a large deviation principle for interacting particle models. It is also used in the promising field of machine learning.

In \mathbb{R}^d , or more generally in a complete metric space, the Wasserstein distance W_p is amongst the most popular choices for a metric between two elements of $\mathcal{P}_p(\mathbb{R}^d)$ (probabilities of finite p -th moment), where $p \in [1, +\infty]$. It is expressed through the Monge-Kantorovich optimal transport

problem: if $p < +\infty$, then

$$W_p(\mu, \nu) = \left(\inf \left\{ \iint |x - y|^p \gamma(dx dy) : \gamma \in \Gamma(\mu, \nu) \right\} \right)^{\frac{1}{p}},$$

where $\Gamma(\mu, \nu)$ is the set of admissible transport plans, i.e. elements of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ whose first and second marginals are μ and ν , respectively. Arguably, the most prominent cases of the Wasserstein distance are $p = 1, 2$. By virtue of the Kantorovich-Rubinstein duality, the Monge distance W_1 can be reset as the supremum with respect to 1-Lipschitz potentials,

$$W_1(\mu, \nu) = \sup \left\{ \int u d(\nu - \mu) : u \in C^{0,1}(\mathbb{R}^d), \text{lip}(u) \leq 1 \right\}. \quad (1.1)$$

With his seminal paper [7], Y. Brenier has established the fundamental importance of the quadratic case W_2 . He has proved that the optimal plans are supported on cyclically monotone subset of $(\mathbb{R}^d)^2$. Furthermore, under the condition of absolute continuity of one of the marginals, the unique optimal plan is induced by a transport map being the gradient of a convex potential.

Another family of metrics between probabilities on \mathbb{R}^d that we are interested in is the one proposed by V.M. Zolotarev [22]. For an integer¹ $r \geq 1$, the Zolotarev distance generalizes the Monge distance (1.1) to higher orders,

$$Z_r(\mu, \nu) = \sup \left\{ \int u d(\nu - \mu) : u \in C^{r-1,1}(\mathbb{R}^d), \|D^{r-1}u(x) - D^{r-1}u(y)\| \leq |x - y| \quad \forall x, y \in \mathbb{R}^d \right\},$$

where $D^k u$ is the k -th Fréchet derivative and $\|\cdot\|$ stands for the operator norm. Clearly, $Z_1 = W_1$. Note that the two-point inequality can be equivalently replaced by $\|D^r u\| \leq 1$ enforced a.e. in \mathbb{R}^d . The Zolotarev distance Z_r is an example of *ideal metric* of order r : it is scale- r -homogenous and it is subadditive under convolution. These properties render Z_r useful in obtaining Berry-Esseen-type bounds [10] in the Central Limit Theorem, see Chapters 14, 15 in [14] for more details.

In the short note [2] it was shown that the metrics Z_r and W_r induce the same topology on $\mathcal{P}_r(\mathbb{R}^d)$: the one of weak convergence together with convergence of r -th moments². Looking for quantitative results comparing the two families of distances is an active field of study. For instance, since Z_r and $(W_r)^r$ scale identically with respect to dilations, the inequalities of the kind $Z_r \geq C_r W_r^r$ can be pursued. Thus far, however, the results in this direction are mainly restricted to one dimension $d = 1$, where we can utilize the closed solutions for W_r that are expressed via the quantile functions of μ, ν . The first result on the real line can be traced back to the works of E. Rio, cf. [15]. The constants C_r were then improved in [16]. For certain orders r , an explicit formula for the constant was given. For instance, the inequality $Z_2 \geq C_2 W_2^2$ holds true with $C_2 = \frac{1}{4}$. Similar inequalities in 1D were established in [3] for a modified version of

¹The definition of Z_r can be extended to any real order $r > 0$: the right hand side of the two-point inequality must be then changed to $|x - y|^\alpha$ for $\alpha = r + 1 - [r]$.

²More accurately, the equivalence between the Z_r and W_r metrics holds true only on the relevant subspaces of $\mathcal{P}_r(\mathbb{R}^d)$ with fixed mixed moments of order less than r , cf. Remark 2.1 for the case $r = 2$.

Zolotarev distance where a uniform bound on the lower-order derivatives $u^{(k)}$ is included in the defining supremum.

In this paper we take advantage of the novel duality method, recently developed by the two authors in [4], to extend the validity of Rio's inequality for $r = 2$,

$$Z_2(\mu, \nu) \geq \frac{1}{4} W_2^2(\mu, \nu), \quad (1.2)$$

to any dimension $d \geq 1$, showing that the constant $\frac{1}{4}$ is optimal and that the inequality is strict unless $\mu = \nu$ (see Theorem 3.1).

On the other hand, as we will show, the reverse inequality of the same form cannot hold true for any finite constant. However, the topologies induced by the Z_2 and W_2 distances are equivalent. By exploiting the same duality technique, we provide a quantitative confirmation of this topological equivalence by showing a modified reverse inequality under the assumption that the barycentres of μ and ν coincide,

$$Z_2(\mu, \nu) \leq \frac{1}{2}(\sigma_\mu + \sigma_\nu) W_2(\mu, \nu),$$

where σ_μ, σ_ν are the respective standard deviations of μ, ν (see Theorem 4.1). The factor $\frac{1}{2}(\sigma_\mu + \sigma_\nu)$ is optimal in the sense to be made precise. As a by-product, we deduce in Corollary 4.2 an alternative version of the upper bound inequality involving the variances and where the inequality is strict unless $\mu = \nu$,

$$Z_2(\mu, \nu) \leq \sqrt{\frac{\text{var } \mu + \text{var } \nu}{2}} W_2(\mu, \nu). \quad (1.3)$$

It is worth to note that similar reverse inequalities were proposed in [3] for the modified Zolotarev distance, again in one dimension only.

To conclude the introduction, let us briefly mention one of the possible applications of the new inequalities, which can be traced back to the paper [16]. Therein, Rio has derived his inequality to establish optimal convergence rates in the univariate Central Limit Theorem with respect to W_2 distance. More precisely, with μ_n being the law of the sum of iid random variables of zero mean and unit variance $n^{-1/2} \sum_{i=1}^n X_n$, and with ν being the standard normal distribution, he has shown that $W_2(\mu_n, \nu)$ decays as $n^{-1/2}$ provided that X_1 has a finite fourth moment.

Similarly, our multidimensional extension (1.2) could potentially pave the way towards finding new rates for the multivariate case of CLT in Wasserstein distance, see [5, 6] for the state of the art. Similarly as in [16], the strategy would be to work on the rates in Z_2 instead of directly in W_2 . One way of estimating the Zolotarev distance in CLT is by employing the Stein method [19], see also the survey [17]. For instance, the paper [11] has put forth the notion of higher-order Stein kernels which has allowed the author to provide $n^{-r/2}$ decay rates for Z_r under additional moment constraints and a Poincaré inequality condition. In [12] the Stein method was applied in the context of the smooth Wasserstein distance which is reminiscent of the Zolotarev distance.

The paper is organized as follows: in the preliminary Section 2, we present the duality scheme introduced in [4] for the Z_2 distance and its connection with convex order and the three marginal optimal transport formulation that involves martingale conditions. Then, we deduce an identity that will be crucial in establishing the upper bound inequality (see Lemma 2.5). In Section 3, we prove a sharp generalization of the Rio lower bound inequality to any dimension (see Theorem 3.1). In the last Section 4, we establish a new sharp upper bound inequality involving standard deviations (see Theorem 4.1) and its variant in Corollary 4.2.

Acknowledgments. This paper was conceived during the first author's visits to the University of Toulon and the second author's visits to the Lagrange Mathematics and Computing Research Center in Paris. We are grateful to the Lagrange Center for funding these visits. We would also like to thank G. Carlier, Q. Merigot, and F. Santambrogio for their valuable discussions, which greatly influenced the development of the probabilistic section of the paper [4], from which this work benefits. During the preparation of this paper, the first author was a postdoc at the Lagrange Center. The second author was partially supported by the French project ANR-23-CE40-0017.

Notations. Throughout the paper we will use the following notations.

- The Euclidean norm of $z \in \mathbb{R}^d$ is denoted by $|z|$, and $\langle \cdot, \cdot \rangle$ will be used to denote the canonical scalar product.
- By $C^{0,1}(\mathbb{R}^d)$ (resp. $C^{1,1}(\mathbb{R}^d)$) we understand the Banach space of these continuous functions $u \in C(\mathbb{R}^d)$ (resp. continuously differentiable functions $u \in C^1(\mathbb{R}^d)$) for which $\text{lip}(u) < +\infty$ (resp. $\text{lip}(\nabla u) < +\infty$).
- $\mathcal{P}(\mathbb{R}^d)$ is the set of probabilities on \mathbb{R}^d , i.e. positive Borel measures μ of unit total mass, $\mu(\mathbb{R}^d) = 1$. For any $p \geq 1$, $\mathcal{P}_p(\mathbb{R}^d)$ denotes the subset of whose elements of $\mathcal{P}(\mathbb{R}^d)$ which satisfy $\int |x|^p d\mu < +\infty$.
- δ_{x_0} is the Dirac delta measure at the point $x_0 \in \mathbb{R}^d$.
- For a probability μ and a μ -measurable map T , by $T^\#(\mu)$ we understand the push forward, i.e. $T^\#(\mu)(B) := \mu(T^{-1}(B))$ for every Borel set B .
- The topological support of $\mu \in \mathcal{P}(\mathbb{R}^d)$ is denoted by $\text{sp } \mu$.
- For a probability $\gamma \in \mathcal{P}(\mathbb{R}^d \times \dots \times \mathbb{R}^d)$ on the product of n ambient spaces, by γ_{k_1, \dots, k_m} we understand the marginal $\Pi_{k_1, \dots, k_m}^\#(\gamma)$ where, for $m \leq n$, Π_{k_1, \dots, k_m} is the projection onto the coordinates k_1, \dots, k_m .
- Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, we denote by $\Gamma(\mu, \nu)$ the convex subset of transport plans,

$$\Gamma(\mu, \nu) := \left\{ \gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \Pi_1^\# \gamma = \mu, \quad \Pi_2^\# \gamma = \nu \right\}.$$

- $[\mu]$ stands for the barycentre of a probability $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, whilst its variance $\text{var}(\mu) := \int |x - [\mu]|^2 \mu(dx)$ is finite if and only if $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. By $\sigma_\mu := \sqrt{\text{var}(\mu)}$ we will understand the standard deviation.

- $\nu \succeq_c \mu$ denotes the convex order between two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, whilst $\Gamma_M(\mu, \nu) \subset \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is the set of martingale transport plans from μ to ν .
- Assume $\mu \in \mathcal{P}(\mathbb{R}^n)$ and a map $x \mapsto \lambda^x \in \mathcal{P}(\mathbb{R}^m)$ that is μ -measurable in the sense that $x \mapsto \lambda^x(A)$ is μ -measurable for any Borel set $A \subset \mathbb{R}^m$. We will use the notation,

$$\gamma = \mu \otimes \lambda^x, \quad (1.4)$$

to define a probability $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^m)$ that satisfies,

$$\gamma(B) := \int \left(\int \chi_B(x, y) \lambda^x(dy) \right) \mu(dx),$$

for every Borel set $B \subset \mathbb{R}^n \times \mathbb{R}^m$, where χ_B is the characteristic function.

2. PRELIMINARY RESULTS

In this section, we provide an overview of the duality result in [4], which exposes the connection between the second-order Zolotarev distance and a new optimal transport formulation with martingale conditions. Let us recall that the definition of the second-order Zolotarev distance between two elements $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $[\mu] = [\nu]$,

$$Z_2(\mu, \nu) := \sup \left\{ \int u d(\nu - \mu) : u \in C^{1,1}(\mathbb{R}^d), \text{lip}(\nabla u) \leq 1 \right\}. \quad (2.1)$$

Remark 2.1. Note that the above supremum is infinite if the condition $[\mu] = [\nu]$ is violated. To avoid this, it is common to add constraints to u , namely that $u(x_0) = 0$ and $\nabla u(x_0) = 0$, where x_0 is usually the origin (see [2]). While this modified distance is well-defined on the whole set $\mathcal{P}_2(\mathbb{R}^d)$, it is no longer translation-invariant, which complicates the search for universal inequalities.

Given a pair $(\mu, \nu) \in (\mathcal{P}(\mathbb{R}^d))^2$ such that $[\mu] = [\nu]$, we are going to define a minimization problem in duality with the maximization problem (2.1), in which the unknown is a three marginal probability measure $\pi \in \mathcal{P}_2((\mathbb{R}^d)^3)$ (in short a *3-plan*), subject to linear constraints defining the set $\Sigma(\mu, \nu) \subset \mathcal{P}_2((\mathbb{R}^d)^3)$ as below .

2.1. The feasible set $\Sigma(\mu, \nu)$. It consists of all 3-plans π whose first and second marginal is μ and ν , respectively, and which satisfy the following equations,

$$\iiint \langle z - x, \Phi(x) \rangle \pi(dx dy dz) = \iiint \langle z - y, \Psi(y) \rangle \pi(dx dy dz) = 0, \quad (2.2)$$

for any smooth test functions $\Phi, \Psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ with compact support. Note that, by a density argument, the equality above immediately extends to Borel functions Φ, Ψ of linear growth, i.e. satisfying $|\Phi|, |\Psi| \leq C(1 + |x|)$. The relations (2.2) were introduced in [4] in the context of structural mechanics³ which might seem remote from our present concern. It turns out in fact that they also have a natural interpretation in probability theory through convex order

³They encode the moment equilibrium at junctions of beam systems.

and martingale transport. For the reader's convenience, let us recall some classical facts and definitions. For any coupling probability $\gamma \in \Gamma(\mu, \nu)$, there exists a unique disintegration of the form $\gamma = \mu \otimes \gamma^x$ (see the notation (1.4)), where $x \mapsto \gamma^x \in \mathcal{P}(\mathbb{R}^d)$ is a μ -measurable mapping, cf. for instance Section 2.5 in [1]. The class of martingale transport plans from μ to ν is then defined by,

$$\Gamma_M(\mu, \nu) := \left\{ \gamma \in \Gamma(\mu, \nu) : \gamma = \mu \otimes \gamma^x, [\gamma^x] = x \text{ for } \mu\text{-a.e. } x \right\}. \quad (2.3)$$

A consequence of the celebrated Strassen theorem [20] is that $\Gamma_M(\mu, \nu)$ is non-empty if and only if $\nu \succeq_c \mu$ in the sense of convex order, that is for every *convex* function $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ there holds the inequality,

$$\int \varphi d\nu \geq \int \varphi d\mu. \quad (2.4)$$

The stochastic interpretation of the set $\Sigma(\mu, \nu)$ appears in the following result in [4, Lemma 3.3] for which we refer for a proof.

Lemma 2.2. *Let $\pi \in \mathcal{P}_2((\mathbb{R}^d)^3)$ be a 3-plan with marginals (μ, ν, ρ) . Define the marginals $\pi_{1,3} := \Pi_{1,3}^\#(\pi)$ and $\pi_{2,3} := \Pi_{2,3}^\#(\pi)$, which are the push forwards of $\pi(dx dy dz)$ through the projection maps $(x, y, z) \mapsto (x, z)$ and $(x, y, z) \mapsto (y, z)$, respectively. Then,*

$$\pi \in \Sigma(\mu, \nu) \quad \Leftrightarrow \quad \begin{cases} \pi_{1,3} \in \Gamma_M(\mu, \rho), \\ \pi_{2,3} \in \Gamma_M(\nu, \rho). \end{cases} \quad (2.5)$$

Accordingly, an element $\rho \in \mathcal{P}_2(\mathbb{R}^d)$ is the third marginal of a 3-plan $\pi \in \Sigma(\mu, \nu)$ if and only if,

$$\rho \succeq_c \mu \quad \text{and} \quad \rho \succeq_c \nu. \quad (2.6)$$

2.2. The duality result. In [13] it was evidenced that a C^1 function u satisfies the constraint $\text{lip}(\nabla u) \leq 1$ if and only if the following 3-point inequality holds true,

$$(u(y) + \langle \nabla u(y), z - y \rangle) - (u(x) + \langle \nabla u(x), z - x \rangle) \leq \frac{1}{2} (|z - x|^2 + |z - y|^2) \quad \forall (x, y, z) \in \mathbb{R}^d, \quad (2.7)$$

see also Lemma 3.4 in [4] for an independent proof. Accordingly, for every u admissible in (2.1) and every 3-plan $\pi \in \Sigma(\mu, \nu)$, we find that,

$$\begin{aligned} \int u d(\nu - \mu) &= \iiint \left((u(y) + \langle \nabla u(y), z - y \rangle) - (u(x) + \langle \nabla u(x), z - x \rangle) \right) \pi(dx dy dz) \\ &\leq \iiint \frac{1}{2} (|z - x|^2 + |z - y|^2) \pi(dx dy dz). \end{aligned} \quad (2.8)$$

The equality above follows by testing (2.2) with $\Phi = \Psi = \nabla u$, and also by the fact that μ and ν are the first and, respectively, the second marginals of every feasible plan π . Taking the

supremum of the left hand side with respect to admissible potentials u , we find that each 3-plan furnishes an upper bound on the second Zolotarev distance,

$$Z_2(\mu, \nu) \leq \iiint \frac{1}{2} (|z - x|^2 + |z - y|^2) \pi(dx dy dz) \quad \forall \pi \in \Sigma(\mu, \nu).$$

Our new duality scheme [4] shows that for at least one feasible 3-plan $\pi = \bar{\pi}$ the gap vanishes. As in the case of the classical first-order Kantorovich-Rubinstein duality $Z_1(\mu, \nu) = W_1(\mu, \nu)$, the proof of this fact goes well beyond the standard use of the convex duality tools. It leads through the optimal convex dominance problem briefly recapitulated in Remark 2.4 below. One of the main techniques relies on the fact that convexification preserves the semi-concavity.

For the details on the second-order variant of the Kantorovich-Rubinstein duality we refer to [4]. Below, we state the result along with the optimality criteria.

Theorem 2.3. *For any pair $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $[\mu] = [\nu]$, the following statements hold true:*

(i) *one has the equality,*

$$Z_2(\mu, \nu) = \inf \left\{ \iiint \frac{1}{2} (|z - x|^2 + |z - y|^2) \pi(dx dy dz) : \pi \in \Sigma(\mu, \nu) \right\}, \quad (2.9)$$

where the infimum is achieved at least at one 3-plan $\bar{\pi}$;

(ii) *the maximization problem (2.1) has a solution \bar{u} ;*

(iii) *the 3-plan $\bar{\pi} \in \Sigma(\mu, \nu)$ and the function $\bar{u} \in C^{1,1}(\mathbb{R}^d)$ with $\text{lip}(\nabla \bar{u}) \leq 1$ solve the problems (2.9) and (2.1), respectively, if and only if the two following conditions are met:*

- *there exists a transport plan $\bar{\gamma} \in \Gamma(\mu, \nu)$ such that $\bar{\pi}$ satisfies*

$$\bar{\pi}(dx dy dz) = \bar{\gamma}(dx dy) \otimes \delta_{\bar{z}(x,y)}(dz), \quad (2.10)$$

where

$$z_{\bar{u}}(x, y) := \frac{x + y}{2} + \frac{\nabla \bar{u}(y) - \nabla \bar{u}(x)}{2}; \quad (2.11)$$

- *\bar{u} satisfies the two-point equality for $\bar{\gamma}$ -a.e. pair (x, y) :*

$$\bar{u}(y) - \bar{u}(x) = \frac{1}{2} \langle \nabla \bar{u}(x) + \nabla \bar{u}(y), y - x \rangle - \frac{1}{4} |\nabla \bar{u}(x) - \nabla \bar{u}(y)|^2 + \frac{1}{4} |x - y|^2. \quad (2.12)$$

Proof. The first two statements are a rephrasing of the assertion (i) of [4, Theorem 1.1]. Furthermore, the optimality conditions appearing in our assertion (iii) can be deduced directly from the assertion (ii) of the latter theorem and of the subsequent Corollary 1.2 in the same paper [4]. Nonetheless, with the equality (2.9) at hand, a short proof of the conditions (iii) can be devised without further reference to [4], which we will now do for the reader's convenience.

From (i) we see that the optimality of an admissible pair $(\bar{u}, \bar{\pi})$ is equivalent to the inequality in (2.8) being an equality. In turn, this is equivalent to the equality sign in the inequality (2.7) for every triple (x, y, z) in the support of $\bar{\pi}$. It is straightforward to check that, for fixed x, y , the point $z = z_{\bar{u}}(x, y)$ is the unique minimizer of the gap in the inequality (2.7). Therefore,

the disintegrated form of $\bar{\pi}$ in (2.10) is a necessary condition for the optimality. Once this form is enforced, we must make sure that the inequality (2.7), with z substituted by $z_{\bar{\pi}}(x, y)$, is an equality for all $(x, y) \in \text{sp } \bar{\gamma}$. After elementary rearrangements, this equality can be identified as (2.12). We have thus showed that the conditions (iii) are necessary, and similarly we deduce that they are also sufficient. \square

Remark 2.4. Thanks to the characterization (2.6) of the third marginals associated with admissible 3-plans, one infers that the optimal transport problem (2.9) can be reduced to a stochastic optimization problem under convex dominance constraints. Indeed, by developing the three-point cost and testing (2.2) with the identity maps, one deduces the following equality,

$$Z_2(\mu, \nu) = \mathcal{V}(\mu, \nu) - \frac{1}{2}(\text{var}(\mu) + \text{var}(\nu)), \quad \mathcal{V}(\mu, \nu) := \inf \left\{ \text{var}(\rho) : \rho \succeq_c \mu, \quad \rho \succeq_c \nu \right\}.$$

In turn, the problem $\mathcal{V}(\mu, \nu)$ above admits optimal solutions $\bar{\rho}$ from which optimal 3-plans $\bar{\pi}$ can be constructed by gluing *any* martingale transports in $\Gamma_M(\mu, \bar{\rho})$ and $\Gamma_M(\nu, \bar{\rho})$. For further details, we refer to [4]. Note that if $\nu \succeq_c \mu$, the unique minimizer for $\mathcal{V}(\mu, \nu)$ is trivially $\bar{\rho} = \nu$, while a solution to (2.1) is given by $\bar{u} = \frac{1}{2}|x|^2$. In particular, we have $Z_2(\mu, \nu) = \frac{1}{2}(\text{var}(\nu) - \text{var}(\mu))$ in this case.

2.3. An auxiliary identity. We will now derive an alternative expression for the Zolotarev-2 distance $Z_2(\mu, \nu)$ which was not presented in the paper [4]. This expression is crucial for establishing the optimal upper bound (1.3) announced in the introduction.

Lemma 2.5. *Let \bar{u} be a solution of the problem $Z_2(\mu, \nu)$, see (2.1). Then, the following formula holds true,*

$$Z_2(\mu, \nu) = \int \frac{1}{2} \langle x, \nabla \bar{u}(x) \rangle (\nu - \mu)(dx). \quad (2.13)$$

Proof. Assume a solution $\bar{\pi} \in \Sigma(\mu, \nu)$. Owing to Theorem 2.3, it is of the form $\bar{\pi}(dxdydz) = \bar{\gamma}(dxdy) \otimes \delta_{\bar{z}(x,y)}(dz)$, where $\bar{\gamma} \in \Gamma(\mu, \nu)$ and $\bar{z}(x, y) = \frac{x+y}{2} + \frac{\nabla \bar{u}(y) - \nabla \bar{u}(x)}{2}$. We get,

$$Z_2(\mu, \nu) = \iiint \frac{1}{2} (|z - x|^2 + |z - y|^2) \bar{\pi}(dxdydz) \quad (2.14)$$

$$= \iint \frac{1}{2} (|\bar{z}(x, y) - x|^2 + |\bar{z}(x, y) - y|^2) \bar{\gamma}(dxdy). \quad (2.15)$$

Thanks to the martingale conditions defining the set $\Sigma(\mu, \nu)$, for any Borel function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ of linear growth it holds that,

$$0 = \iiint \langle z - x, \Phi(x) \rangle \bar{\pi}(dxdydz) = \iint \langle \bar{z}(x, y) - x, \Phi(x) \rangle \bar{\gamma}(dxdy), \quad (2.16)$$

and similarly we get,

$$\iint \langle \bar{z}(x, y) - y, \Psi(y) \rangle \bar{\gamma}(dxdy) = 0. \quad (2.17)$$

We then deduce the following chain of equalities,

$$\begin{aligned}
Z_2(\mu, \nu) &= \iint \frac{1}{2} \left(\langle \bar{z}(x, y) - x, \bar{z}(x, y) - x \rangle + \langle \bar{z}(x, y) - y, \bar{z}(x, y) - y \rangle \right) \bar{\gamma}(dxdy) \\
&= \iint \frac{1}{2} \left(\langle \bar{z}(x, y) - x, \bar{z}(x, y) \rangle + \langle \bar{z}(x, y) - y, \bar{z}(x, y) \rangle \right) \bar{\gamma}(dxdy) \\
&= \iint \left\langle \bar{z}(x, y) - \frac{x+y}{2}, \bar{z}(x, y) \right\rangle \bar{\gamma}(dxdy) = \iint \left\langle \frac{\nabla u(y) - \nabla u(x)}{2}, \bar{z}(x, y) \right\rangle \bar{\gamma}(dxdy) \\
&= \frac{1}{2} \iint \langle \nabla u(y), \bar{z}(x, y) \rangle \bar{\gamma}(dxdy) - \frac{1}{2} \iint \langle \nabla u(x), \bar{z}(x, y) \rangle \bar{\gamma}(dxdy) \\
&= \frac{1}{2} \iint \langle \nabla u(y), x \rangle \bar{\gamma}(dxdy) - \frac{1}{2} \iint \langle \nabla u(x), y \rangle \bar{\gamma}(dxdy),
\end{aligned}$$

where:

- to pass from the second to the third line, we used (2.16), (2.17) with $\Phi(x) = x$, $\Psi(y) = y$;
- to pass to the last line, we used (2.16), (2.17) with $\Phi(x) = \nabla u(x)$ and $\Psi(y) = \nabla u(y)$.

The asserted equality (2.13) now follows since $\bar{\gamma} \in \Gamma(\mu, \nu)$. \square

3. THE LOWER BOUND INEQUALITY

We now present a sharp extension of Rio inequality to \mathbb{R}^d . This inequality involves the ratio $\frac{Z_2(\mu, \nu)}{W_2^2(\mu, \nu)}$, which is dilation-invariant since the Wasserstein distance enters with a square.

Theorem 3.1. *Let μ, ν be two probability measures in $\mathcal{P}_2(\mathbb{R}^d)$. Then,*

$$\frac{1}{4} W_2^2(\mu, \nu) \leq Z_2(\mu, \nu), \quad (3.1)$$

and the constant $\frac{1}{4}$ cannot be improved. Moreover, the inequality is an equality if and only if $\mu = \nu$.

Let us remark that the inequality (3.1) is trivially true if $[\mu] \neq [\nu]$. Indeed, in this case the supremum in (2.1) is infinite and, by convention, $Z_2(\mu, \nu) = +\infty$.

Proof of the lower bound. It is enough to observe that for any triple $(x, y, z) \in (\mathbb{R}^d)^3$ we have,

$$\frac{1}{2} \left(|z - x|^2 + |z - y|^2 \right) \geq \frac{1}{4} |x - y|^2, \quad (3.2)$$

while the equality holds if and only if $z = \frac{x+y}{2}$. Let $\bar{\pi}$ be a solution of the three-point optimal transport formulation (2.9) and denote $\bar{\gamma} := \Pi_{1,2}^\# \bar{\pi}$. Then, by the assertion (i) of Theorem 2.3,

$$\begin{aligned}
Z_2(\mu, \nu) &= \iiint \frac{1}{2} \left(|z - x|^2 + |z - y|^2 \right) \bar{\pi}(dxdydz) \geq \frac{1}{4} \iiint |x - y|^2 \bar{\pi}(dxdydz) \\
&= \frac{1}{4} \iint |x - y|^2 \bar{\gamma}(dxdy) \\
&\geq \frac{1}{4} W_2^2(\mu, \nu).
\end{aligned} \quad (3.3)$$

To obtain the last inequality, we used the admissibility condition $\bar{\pi} \in \Sigma(\mu, \nu)$ which entails that $\bar{\gamma} \in \Gamma(\mu, \nu)$. The optimality of the constant $\frac{1}{4}$ follows from the Example 3.2 below.

Let us now pass to examining the criteria for (3.1) being an equality. Clearly the condition $\mu = \nu$ is sufficient. To show the necessity, we observe that the equality implies that the first inequality in (3.3) is an equality. In turn, this means that,

$$\bar{\pi}(dxdydz) = \bar{\gamma}(dxdy) \otimes \delta_{\frac{x+y}{2}}(dz). \quad (3.4)$$

The two martingale conditions (2.2) defining the set $\Sigma(\mu, \nu)$ furnish the relations,

$$\begin{aligned} 0 &= \iiint \langle \Phi(x), z - x \rangle \bar{\pi}(dxdydz) = \frac{1}{2} \iint \langle \Phi(x), y - x \rangle \bar{\gamma}(dxdy), \\ 0 &= \iiint \langle \Psi(y), z - y \rangle \bar{\pi}(dxdydz) = \frac{1}{2} \iint \langle \Psi(y), x - y \rangle \bar{\gamma}(dxdy), \end{aligned}$$

where $\Phi, \Psi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are any Borel maps of linear growth. By adding the two lines above, after taking $\Phi(x) = x$ and $\Psi(y) = y$, we infer that $\iint |y - x|^2 \bar{\gamma}(dxdy) = 0$. Therefore $\bar{\gamma}$ must be supported on the diagonal $\Delta = \{(x, x) : x \in \mathbb{R}^d\}$. It is thus straightforward that its marginals μ and ν must coincide. The proof is complete. \square

Example 3.2. For $b > a > 0$ consider the two centred probabilities on the real line:

$$\mu_{a,b} = \frac{b}{a+b} \delta_{-a} + \frac{a}{a+b} \delta_b, \quad \nu_{a,b} = \frac{a}{a+b} \delta_{-b} + \frac{b}{a+b} \delta_a.$$

The unique transport plan with a monotone support reads,

$$\gamma_{\text{mon}} = \frac{a}{a+b} \delta_{(-a,-b)} + \frac{b-a}{a+b} \delta_{(-a,a)} + \frac{a}{a+b} \delta_{(b,a)}.$$

It is straightforward to check that $\Pi_1^\# \gamma_{\text{mon}} = \mu_{a,b}$ and $\Pi_2^\# \gamma_{\text{mon}} = \nu_{a,b}$. Thus, the square of the Wasserstein distance equals,

$$\begin{aligned} W_2^2(\mu_{a,b}, \nu_{a,b}) &= \iint |x - y|^2 \gamma_{\text{mon}}(dxdy) \\ &= (b-a)^2 \frac{a}{a+b} + (2a)^2 \frac{b-a}{a+b} + (b-a)^2 \frac{a}{a+b} = 2a(b-a). \end{aligned}$$

In turn, let us propose an admissible 3-plan $\bar{\pi} \in \Sigma(\mu_{a,b}, \nu_{a,b})$,

$$\bar{\pi} = \frac{a}{a+b} \delta_{(-a,-b,-b)} + \frac{b-a}{a+b} \delta_{(-a,a,0)} + \frac{a}{a+b} \delta_{(b,a,b)}.$$

Since $\Pi_{1,2}^\# \bar{\pi} = \gamma_{\text{mon}}$, to show the admissibility of $\bar{\pi}$ it remains to check the martingale conditions.

We compute,

$$\begin{aligned} \Pi_{1,3}^\# \bar{\pi} &= \frac{a}{a+b} \delta_{(-a,-b)} + \frac{b-a}{a+b} \delta_{(-a,0)} + \frac{a}{a+b} \delta_{(b,b)} \\ &= \frac{b}{a+b} \delta_{-a} \otimes \left(\frac{a}{b} \delta_{-b} + \frac{b-a}{b} \delta_0 \right) + \frac{a}{a+b} \delta_b \otimes \delta_b. \end{aligned}$$

Since the probability $\frac{a}{b} \delta_{-b} + \frac{b-a}{b} \delta_0$ has $-a$ as its barycentre, it is easy to see that the above plan is a martingale. Similarly, one shows that $\Pi_{2,3}^\# \bar{\pi}$ is a martingale as well. We can readily

compute an upper bound for the Zolotarev distance,

$$\begin{aligned} Z_2(\mu_{a,b}, \nu_{a,b}) &\leq \iiint \frac{1}{2} (|z-x|^2 + |z-y|^2) \bar{\pi}(dxdydz) \\ &= \frac{1}{2}(-b+a)^2 \frac{a}{a+b} + \frac{1}{2}(a^2+a^2) \frac{b-a}{a+b} + \frac{1}{2}(b-a)^2 \frac{a}{a+b} = \frac{ab}{a+b} (b-a). \end{aligned}$$

With $a > 0$ fixed, we look at the asymptotic quotient when b converges to a from above,

$$\limsup_{b \searrow a} \frac{Z_2(\mu_{a,b}, \nu_{a,b})}{W_2^2(\mu_{a,b}, \nu_{a,b})} \leq \lim_{b \searrow a} \frac{b}{2(a+b)} = \frac{1}{4}.$$

This establishes the optimality of the constant $\frac{1}{4}$ in the lower bound in Theorem 3.1 in dimension $d = 1$. Clearly this example can be embedded in \mathbb{R}^d by taking pairs of symmetric points on a straight line. Hence it holds for every $d \geq 1$ that,

$$\inf \left\{ \frac{Z_2(\mu, \nu)}{W_2^2(\mu, \nu)} : \mu, \nu \in \mathcal{P}_2(\mathbb{R}^d), \mu \neq \nu \right\} = \frac{1}{4}.$$

Remark 3.3. There does not exist a reverse inequality of the kind $Z_2(\mu, \nu) \leq C W_2^2(\mu, \nu)$ holding for pairs (μ, ν) such that $[\mu] = [\nu]$. This can be checked by showing that the ratio $\frac{Z_2(\mu, \nu_n)}{W_2^2(\mu, \nu_n)}$ blows up when $\mu := \frac{1}{2}(\delta_1 + \delta_{-1})$ and $\nu_n := \frac{1}{2}(\delta_{1+1/n} + \delta_{-1-1/n})$. Indeed, in that case, we have $W_2^2(\mu, \nu_n) = 1/n^2$, while, in view of Remark 2.4, it holds that $Z_2(\mu, \nu_n) = \frac{1}{2}(\text{var}(\nu_n) - \text{var}(\mu)) = \frac{1}{2}((1+1/n)^2 - 1)$, which goes to zero at the rate $1/n$. This shows that a reverse inequality cannot be true even if we confine ourselves to probability measures with supports contained in a fixed compact set $K \subset \mathbb{R}^d$. Another counter-example, but with non-compact supports, is given below for the Gaussian laws.

Example 3.4. For two positive numbers $\sigma_1 \leq \sigma_2$ consider two centred Gaussians on the real line with standard deviations equal to σ_1 and σ_2 , respectively, that is $\mathcal{N}(0, \sigma_1^2), \mathcal{N}(0, \sigma_2^2)$. The Wasserstein-2 distance between two Gaussian distributions admits a closed form in any dimension d , see e.g. [8]. In the simple case of two 1D centered Gaussians, the formula reduces to,

$$W_2\left(\mathcal{N}(0, \sigma_1^2), \mathcal{N}(0, \sigma_2^2)\right) = |\sigma_1 - \sigma_2| = \sigma_2 - \sigma_1. \quad (3.5)$$

It is also well known that every pair of centred Gaussians in 1D is in convex order. More precisely, in case when $\sigma_1 \leq \sigma_2$, we have,

$$\mathcal{N}(0, \sigma_1^2) \preceq_c \mathcal{N}(0, \sigma_2^2).$$

Meanwhile, for any measures in convex order $\mu \preceq_c \nu$, the Zolotarev distance equals $Z_2(\mu, \nu) = \frac{1}{2}(\text{var}(\nu) - \text{var}(\mu))$, cf. [4, Section 4.1]. Accordingly, for the Gaussian measures we have,

$$Z_2\left(\mathcal{N}(0, \sigma_1^2), \mathcal{N}(0, \sigma_2^2)\right) = \frac{\sigma_2^2 - \sigma_1^2}{2}. \quad (3.6)$$

By considering $\mu := \mathcal{N}(0, 1)$ and $\nu_n := \mathcal{N}(0, 1 + \frac{1}{n})$, we arrive at:

$$\frac{Z_2(\mu, \nu_n)}{W_2^2(\mu, \nu_n)} = \frac{2n+1}{2} \rightarrow +\infty.$$

4. UPPER BOUND INEQUALITIES

From now on, we consider pairs $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ such that $[\mu] = [\nu]$. Our goal is to derive an upper bound for the ratio $\frac{Z_2(\mu, \nu)}{W_2(\mu, \nu)}$ as a function of the variances of μ and ν . More precisely, we want to establish an explicit formula for the function $h(a, b) : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ given by,

$$h(a, b) := \sup \left\{ \frac{Z_2(\mu, \nu)}{W_2(\mu, \nu)} : \sigma_\mu \leq a, \sigma_\nu \leq b, \mu \neq \nu, [\mu] = [\nu] = 0 \right\}, \quad (4.1)$$

where $\sigma_\mu = \sqrt{\text{var}(\mu)}$, $\sigma_\nu = \sqrt{\text{var}(\nu)}$ are the standard deviations. By acknowledging how the two distances scale under dilations $T_\lambda(x) = \lambda x$, it is straightforward to check the homogeneity property $h(\lambda a, \lambda b) = \lambda h(a, b)$ holding for every $\lambda > 0$. On the other hand, if $a = 0$ (thus $\mu = \delta_0$), we get $Z_2(\mu, \nu) = \frac{1}{2}\sigma_\nu^2 = \frac{1}{2}b^2$, while $W_2(\mu, \nu) = \sigma_\nu = b$. Hence $h(0, b) = h(b, 0) = \frac{b}{2}$. We are going to show that, in fact, h is linear, i.e. $h(a, b) = \frac{1}{2}(a + b)$. This will give the optimal upper bound stated below.

Theorem 4.1. *Assume that the two probability distributions $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ share the barycentre, i.e. $[\mu] = [\nu]$. Then, we have,*

$$Z_2(\mu, \nu) \leq \frac{1}{2}(\sigma_\mu + \sigma_\nu) W_2(\mu, \nu), \quad (4.2)$$

and this upper bound is optimal in the sense that $h(a, b) = \frac{1}{2}(a + b)$ in (4.1).

Furthermore, the inequality becomes an equality if and only if the probability measures coincide up to a dilation centred at their common barycentre.

Before giving the proof which is postponed to the end of this section, we present a straightforward variant of the inequality (4.2) which, in contrast, is strict whenever $\mu \neq \nu$.

Corollary 4.2. *Under the assumptions of Theorem 4.1, we have the following upper bound,*

$$Z_2(\mu, \nu) \leq \sqrt{\frac{\text{var} \mu + \text{var} \nu}{2}} W_2(\mu, \nu), \quad (4.3)$$

where the inequality is an equality if and only if $\mu = \nu$.

Proof. The inequality (4.3) follows from (4.2) and from the inequality $\frac{a+b}{2} \leq \sqrt{\frac{a^2+b^2}{2}}$. Noticing that the latter inequality is strict if $a \neq b$, we infer that an equality in (4.3) implies the equality in (4.2) along with $\sigma_\mu = \sigma_\nu$. Owing to the last statement of Theorem 4.1, we conclude that $\mu = \nu$ since the only possible dilation sending μ to ν is the identity. \square

Remark 4.3. From Theorem 3.1 and Theorem 4.1, we infer that on the subset of centered probabilities with finite second order moment,

$$\mathcal{P}_2^{(0)}(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}_2(\mathbb{R}^d) : [\mu] = 0 \right\},$$

the topologies induced by the W_2 and the Z_2 distances are equivalent. We thus recover the result of [2]. Clearly, by the invariance with respect to translations, the same statement holds true if we enforce any fixed barycentre $a \in \mathbb{R}^d$, not necessarily the origin.

Proof of Theorem 4.1. Without any loss of generality, we can assume that the two probabilities are centered, i.e. $[\mu] = [\nu] = 0$. Let \bar{u} be a solution of the problem (2.1). As $\mu - \nu$ vanishes on affine functions, it is not restrictive to assume that $\bar{u}(0) = 0$ and $\nabla\bar{u}(0) = 0$. Since $\text{lip}(\nabla\bar{u}) \leq 1$, this guarantees that $|\nabla\bar{u}(x)| \leq |x|$ for any $x \in \mathbb{R}^d$.

Step 1. [*proving the upper bound and the sufficient condition for equality*] Let us take an optimal plan $\hat{\gamma} \in \Gamma(\mu, \nu)$ realizing the distance $W_2(\mu, \nu)$. Thanks to Lemma 2.5, we obtain,

$$2Z_2(\mu, \nu) = \iint \left(\langle y, \nabla\bar{u}(y) \rangle - \langle x, \nabla\bar{u}(x) \rangle \right) \hat{\gamma}(dxdy) \quad (4.4)$$

$$\begin{aligned} &= \iint \left(\frac{1}{2} \langle y - x, \nabla\bar{u}(x) + \nabla\bar{u}(y) \rangle + \frac{1}{2} \langle x + y, \nabla\bar{u}(y) - \nabla\bar{u}(x) \rangle \right) \hat{\gamma}(dxdy) \\ &= \frac{1}{2} \iint \langle \nabla\bar{u}(x), y - x \rangle \hat{\gamma}(dxdy) + \frac{1}{2} \iint \langle \nabla\bar{u}(x), y - x \rangle \hat{\gamma}(dxdy) \\ &\quad + \frac{1}{2} \iint \langle x, \nabla\bar{u}(y) - \nabla\bar{u}(x) \rangle \hat{\gamma}(dxdy) + \frac{1}{2} \iint \langle y, \nabla\bar{u}(y) - \nabla\bar{u}(x) \rangle \hat{\gamma}(dxdy) \\ &\leq \frac{1}{2} \left(\iint |\nabla\bar{u}(x)|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \left(\iint |y - x|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \quad (4.5) \end{aligned}$$

$$\begin{aligned} &\quad + \frac{1}{2} \left(\iint |\nabla\bar{u}(y)|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \left(\iint |y - x|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \\ &\quad + \frac{1}{2} \left(\iint |x|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \left(\iint |\nabla\bar{u}(y) - \nabla\bar{u}(x)|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \\ &\quad + \frac{1}{2} \left(\iint |y|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \left(\iint |\nabla\bar{u}(y) - \nabla\bar{u}(x)|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \\ &\leq \left(\int |x|^2 \mu(dx) \right)^{\frac{1}{2}} \left(\iint |y - x|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \quad (4.6) \\ &\quad + \left(\int |y|^2 \nu(dy) \right)^{\frac{1}{2}} \left(\iint |y - x|^2 \hat{\gamma}(dxdy) \right)^{\frac{1}{2}} \\ &= \sigma_\mu W_2(\mu, \nu) + \sigma_\nu W_2(\mu, \nu). \end{aligned}$$

Above, the first inequality (4.5) uses Cauchy-Schwarz inequality in $L^2_{\hat{\gamma}}(\mathbb{R}^d \times \mathbb{R}^d; \mathbb{R}^d)$ four times. To obtain the second inequality (4.6) we used the fact that $|\nabla\bar{u}(x)| \leq |x|$ and $|\nabla\bar{u}(y) - \nabla\bar{u}(x)| \leq |y - x|$. Eventually we pass to the last line by using the optimality of $\hat{\gamma}$ and the fact that μ, ν are assumed to be centred. This proves the upper bound inequality (4.2) and the fact that h given by (4.1) satisfies $h(a, b) \leq \frac{1}{2}(a + b)$.

Next, in order to show that (4.2) is optimal in the sense that $h(a, b) = \frac{1}{2}(a + b)$, we check that the equality is saturated by all centred pairs (μ, ν) such that μ and ν coincide up to a dilation. Without any loss of generality, we can assume that $\nu = T_\lambda^\# \mu = (\lambda x)^\# \mu$ with $\lambda \geq 1$. Then, it is easy to see that $\mu \preceq_c \nu$ ⁴. Accordingly, as pointed out in Remark 2.4, we have $Z_2(\mu, \nu) = \frac{1}{2}(\sigma_\nu^2 -$

⁴by Strassen theorem, the probability kernel $\gamma^x = \frac{\lambda-1}{\lambda} \nu + \frac{1}{\lambda} \delta_{\lambda x}$ renders $\gamma = \mu \otimes \gamma^x$ an element of $\Gamma_M(\mu, \nu)$

$\sigma_\mu^2) = \frac{1}{2}(\lambda^2 - 1)\sigma_\mu^2$. On the other hand, the transport map $T_\lambda = \nabla(\frac{\lambda}{2}|\cdot|^2)$ is a gradient of a convex function. Therefore, by Brenier theorem [7], $W_2(\mu, \nu) = (\int |T_\lambda(x) - x|^2 \mu(dx))^{1/2} = (\lambda - 1)\sigma_\mu$. It follows that $Z_2(\mu, \nu) = \frac{1}{2}(\lambda + 1)\sigma_\mu W_2(\mu, \nu) = \frac{1}{2}(\sigma_\mu + \sigma_\nu)W_2(\mu, \nu)$.

Step 2. [*necessary condition for equality*] Now we start with the assumption that (4.2) is an equality for some pair (μ, ν) . We may assume without any loss of generality that $\sigma_\nu \geq \sigma_\mu$. Furthermore, we can exclude the cases when either $\mu = \delta_{x_0}$ or $\mu = \nu$ since both trivially satisfy the dilation condition. Accordingly, we can assume that both σ_μ and σ_ν are strictly positive, and we want to show that $\nu = T_\lambda^\# \mu$ for a suitable $\lambda \in (1, +\infty)$.

Going back to the chain of relations starting from (4.4), we see that the equality in (4.2) implies that the inequalities in (4.5) and in (4.6) are actually equalities. From the equality in (4.6), we infer that, for $\hat{\gamma}$ -a.e. (x, y) , we have,

$$|\nabla \bar{u}(x)| = |x|, \quad |\nabla \bar{u}(y)| = |y|, \quad |\nabla \bar{u}(y) - \nabla \bar{u}(x)| = |y - x|, \quad (4.7)$$

By the continuity of $\nabla \bar{u}$, the equalities above hold true for every $(x, y) \in \text{sp } \hat{\gamma}$. Since $\mu \neq \nu$, $x - y$ is not zero as an element of $L_\gamma^2(\mathbb{R}^d \times \mathbb{R}^d; \mathbb{R}^d)$, and the same can be deduced for the function $\nabla \bar{u}(y) - \nabla \bar{u}(x)$ which shares the same norm by virtue of the third equality in (4.7). With that information at our disposal, we may identify the conditions under which the inequality (4.5) is an equality. The four Cauchy-Schwarz inequalities in the vector valued space $L_\gamma^2(\mathbb{R}^d \times \mathbb{R}^d; \mathbb{R}^d)$ must be equalities, which means that for any $(x, y) \in \text{sp } \hat{\gamma}$,

$$\nabla \bar{u}(x) = a(y - x), \quad \nabla \bar{u}(y) = b(y - x), \quad (4.8)$$

$$x = c(\nabla \bar{u}(y) - \nabla \bar{u}(x)), \quad y = e(\nabla \bar{u}(y) - \nabla \bar{u}(x)), \quad (4.9)$$

for suitable constants $a, b, c, e \geq 0$ which *do not depend* on x, y . The condition $\sigma_\mu > 0$ implies that the projection of $\text{sp } \hat{\gamma}$ on the first component does not reduce to $\{0\}$. Hence, from the first equality in (4.9), we infer that c must be strictly positive. The equalities in the sequel of the proof will hold true for each pair $(x, y) \in \text{sp } \hat{\gamma}$. From (4.8), we get $\nabla \bar{u}(y) - \nabla \bar{u}(x) = (b - a)(y - x)$. Going back to (4.7), this means that $\varepsilon := b - a \in \{-1, 1\}$. In turn, from (4.9), we get $x = c\varepsilon(y - x)$, which we can rewrite as,

$$y = \lambda x, \quad \text{where } \lambda := 1 + \frac{1}{c\varepsilon}. \quad (4.10)$$

Therefore, for λ as above, it holds that $\hat{\gamma} = (\text{id}, T_\lambda)^\# \mu$, hence $\nu = T_\lambda^\# \mu$. Moreover, our assumptions $\sigma_\nu \geq \sigma_\mu$, $\mu \neq \nu$ imply that $\lambda > 1$, which gives $\varepsilon = 1$ and $\mu \preceq_c \nu$ (so that an optimal potential \bar{u} is given by $\bar{u} = \frac{1}{2}|x|^2$). This ends the proof. \square

REFERENCES

- [1] L. Ambrosio, N. Fusco, D. Pallara: *Functions of bounded variation and free discontinuity problems*. Clarendon Press, Oxford, 2000.
- [2] N. Belili, H. Heinich: Distances de Wasserstein et de Zolotarev. *C. R. Acad. Sci. Paris*, t. 330, Srie I, p. 811–814, 2000.

- [3] S.G. Bobkov: Zolotarev-type distances. Preprint, 2023.
- [4] K. Bolbotowski, G. Bouchitté: Kantorovich-Rubinstein duality theory for the Hessian. *arXiv preprint arXiv:2412.00516*, 2024 (to appear in *Duke Math. J.* <https://projecteuclid.org/journals/dmj/duke-mathematical-journal/acceptedpapers>).
- [5] T. Bonis: Stein’s method for normal approximation in Wasserstein distances with application to the multivariate Central Limit Theorem. *Probab. Theory Relat. Fields*, 178:827–860, 2020.
- [6] T. Bonis: Improved rates of convergence for the multivariate Central Limit Theorem in Wasserstein distance. *Electron. J. Probab.*, 29:1–18, 2024.
- [7] Y. Brenier: Polar factorization and monotone rearrangement of vector-valued functions. *Commun. Pure Appl. Math.*, 44:375–417, 1991.
- [8] D.C. Dowson, B.V. Landau: The Fréchet distance between multivariate normal distributions. *J. Multivar. Anal.*, 12:450–455, 1982.
- [9] R.M. Dudley: *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [10] C.G. Esseen: On mean central limit theorems. *Kungl. Tekn. Högsk. Handl. Stockholm*, 121, 1958.
- [11] M. Fathi: Higher-order stein kernels for Gaussian approximation. *Stud. Math.*, 256:241–258, 2021.
- [12] R.E. Gaunt: Stein’s method for functions of multivariate normal random variables. *Ann. inst. Henri Poincaré (B) Probab. Stat.*, 56:1484–1513, 2020.
- [13] E. Le Gruyer. Minimal Lipschitz extensions to differentiable functions defined on a Hilbert space. *Geom. Funct. Anal.*, 19:1101–1118, 2009.
- [14] S.T. Rachev: *Probability metrics and the stability of stochastic models*. Wiley, 1991.
- [15] E. Rio: Distances minimales et distances idéales. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(9):1127–1130, 1998.
- [16] E. Rio: Upper bounds for minimal distances in the central limit theorem. In *Ann. inst. Henri Poincaré (B) Probab. Stat.*, 45:802–817, 2009.
- [17] N. Ross: Fundamentals of Stein’s method. *Probab. Surv.* 8:210–293, 2011.
- [18] F. Santambrogio. *Optimal transport for applied mathematicians*. Birkhäuser, 2015.
- [19] C. Stein: Approximate computation of expectations. In: *Institute of Mathematical Statistics Lecture Notes—Monograph Series*, Vol. 7. Institute of Mathematical Statistics, 1986.
- [20] V. Strassen: The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [21] C. Villani: *Topics in optimal transportation*. American Mathematical Society, 2003.
- [22] V.M. Zolotarev: Ideal metrics in the problem of approximating distributions of sums of independent random variables. *Theory Probab. Appl.*, 22:433–449, 1978.

Karol Bolbotowski:

Faculty of Mathematics, Informatics and Mechanics, University of Warsaw

2 Banacha Street, 02-097 Warsaw - POLAND

`k.bolbotowski@mimuw.edu.pl`

Guy Bouchitté:

Laboratoire IMATH, Université de Toulon

BP 20132, 83957 La Garde Cedex - FRANCE

`bouchitte@univ-tln.fr`