From optimal transportation to optimal teleportation via the critical Sobolev embedding

G. Wolansky
Department of Mathematics, Technion, Haifa 32000, Israel

Abstract

The existence of a metric derivative of a measure valued path \( t \to \mu(t) \) in the Wasserstein space \( W_p, p > 1 \), implies the existence of \( \mathbb{L}_p(d\mu) \) valued vector-field transporting this path. In many cases such vector fields do not exist. This is generally the case where the support of \( \mu(t) \) is not connected, and the flow involves shifting positive mass over finite distance in zero time (teleportation).

In this paper we study such paths for which the distributional derivative \( \partial \mu/\partial t \) is a measure. In case of connected support of \( \mu(t) \) we find some estimates on the Holder exponent of such paths in \( W_p \). This exponent turns out to be dependent on some notion of dimension of \( \mu \). In particular we prove that \( \mu(t) \) is \( q \)-Holder in \( W_p \) where \( q = \min(1, 1/d + 1/p) \), \( d \) is the dimension of \( \mu \), provided \( 1/d + 1/p \neq 1 \).

The case \( 1/d + 1/p = 1 \) corresponds to the critical Sobolev embedding \( \mathbb{W}^{1,p'}(\mathbb{R}^d) \) where \( d = p'/p = 1/(p-1) \), and leads to a relaxation of the metric differentiability of \( \mu \) into a log-Lipschitz estimate.

The case where the support of \( \mu \) is not connected corresponds to \( d = \infty \), and the orbit is \( 1/p \)-Holder. In that case we provide the sharp Holder constant. This constant depends only on a finite number of parameters, related to a suitably defined discrete optimal transport on a finite metric graph.

1 Introduction

Consider the set \( \mathcal{P}(\Omega) \) of probability, Borel measures over a metric space \( \Omega \). In the following we assume for simplicity that \( \Omega \subset \mathbb{R}^k \) is open with a compact closure. The \( C^*(\Omega) \) topology over \( \mathcal{P}(\Omega) \) can be metrized by the Wasserstein \( p \)-distance function \( W_p \) where \( p \geq 1 \) (see, e.g. [V]). That is, given \( \mu_0, \mu_1 \in \mathcal{P}(\Omega) \)

\[
W_p(\mu_1, \mu_2) := \left( \inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_{\Omega} \int_{\Omega} |x - y|^p \pi(dx dy) \right)^{1/p}
\]

where

\[
\Pi(\mu_1, \mu_2) := \{ \pi \in \mathcal{P}(\Omega \times \Omega) : \pi_{\#1} = \mu_1; \pi_{\#2} = \mu_2 \}
\]

Here \( \pi_{\#} \) represents the first and second marginals of \( \pi \) on \( \Omega \), respectively.
The notion of metric differentiability of an orbit \( t \mapsto \mathcal{P}(\Omega) \) from a real interval (say, \( t \in [0, 1] \)) to \( W_p(\Omega) \) along with that of displacement interpolation involves two, seemingly unrelated but equivalent definitions. The first definition \([AGS]\) for metric differentiability at \( t_0 \) implies

\[
\limsup_{t \to t_0} \frac{W_p(\mu(t_0), \mu(t))}{|t - t_0|} < \infty. \tag{1}
\]

The second definition \([AGS, M]\) implies

\[
\inf_v \int_{\Omega} |v|^p d\mu(t_0) < \infty \tag{2}
\]

where \( v \) is a Borel vector field on \( \Omega \) which satisfies

\[
\frac{\partial \mu}{\partial t} + \nabla \cdot (v \mu) = 0 \tag{3}
\]

in \( \Omega \times (0, 1) \) in the sense of distributions.

One essential point to be noted is that the set of vector fields satisfying (3) can be empty. Consider for example \( \mu(t) = m(t)\delta_{x_0} + (1 - m(t))\delta_{x_1} \) where \( x_0 \neq x_1 \) and \( t \mapsto m(t) \in (0, 1) \) is a non-constant smooth function. The same holds even if \( \mu = \mu(t) \) have a smooth density in space and time, e.g. \( \mu(t) = \rho(x,t)dx \) where \( \rho \in C^\infty(\Omega \times [0, 1]) \) is supported on a disconnected set in \( \Omega \) for \( t_0 \in [0, 1] \) and \( \partial \rho/\partial t \) is not integrated to zero on some connected component of this support. In particular (2) (hence (1)) take infinite value along such orbit, and we get a non-differentiable (but still smooth(!)) orbit.

In case where the left side \( \nu := \partial \mu/\partial t \) of (3) is a measure, definition (1) is related to the bound of

\[
\limsup_{\epsilon \searrow 0} \epsilon^{-1}W_p(\mu(t_0) + \epsilon \nu, \mu(t_0))
\]

which, in turn, is related to the bound of

\[
\limsup_{\epsilon \searrow 0} \epsilon^{-1}W_p(\mu(t_0) + \epsilon \nu_+, \mu(t_0) + \epsilon \nu_-)
\]

where \( \nu := \nu_+ - \nu_- \) is the factorization of \( \nu \) into positive and negative parts satisfying \( \int \nu = \int \nu_+ - \int \nu_- = 0 \).

Lemma 5.6 in \([W1]\) (see also Theorem 7.26 in \([V]\)) implies that for any such pair of positive measures \( \nu_+, \nu_- \)

\[
\lim_{\epsilon \searrow 0} \epsilon^{-1}W_p(\mu(t_0) + \epsilon \nu_+, \mu(t_0) + \epsilon \nu_-) \geq \sup_{\phi \in \mathcal{B}_p(\mu)} \int \phi(d\nu_+ - d\nu_-) \tag{4}
\]
where
\[ B_p(\mu) := \left\{ \phi \in C^1(\Omega); \int |\nabla \phi|^{p/(p-1)} d\mu \leq 1 \right\} \]  

while Lemma 5.7 implies the opposite inequality in (4) if \( \nu << \mu \) and both measures are regular enough.

In the current paper we attempt to deal with the case where the right side of (4) is infinite. In that case metric differentiability fails but, as it turns out, we can still give estimates on the Holder exponent of such orbits in \( W_p \). In case of disconnected support of \( \mu \) we show that the 1/p–Holder continuity is optimal and provide a sharp estimate on the Holder norm.

1.1 Main results

We start by posing some assumptions on a measure \( \mu \in \mathcal{P}(\Omega) \):

**Definition 1.1.** \( \mu \) is \( d \)–connected if for \( \mu \) a.e \( x_0, x_1, \exists L, K, \delta > 0 \) and \( \bar{x} : [0, 1] \to \Omega \) such that

i) \( \bar{x} \) is \( L \)–Lipshitz on \( [0, 1] \).

ii) \( \bar{x}(0) = x_0, \bar{x}(1) = x_1 \).

iii) For any \( s \in [0, 1] \), \( \mu(B_r(\bar{x}(s))) \geq Kr^d \) for any \( 0 < r < \delta \). Here \( B_r(x) \) is the ball of radius \( r \) centered at \( x \).

For technical reasons we need a stronger notion of \( d \)–connectedness of a measure as follows:

**Definition 1.2.** \( \mu \) is strongly \( d \)–connected if for \( \mu \) a.e \( x_0, x_1, \exists L, K > 0 \), a measure space \( (D, d\beta) \) and a mapping \( \Phi : J \equiv [0, 1] \times D \to \Omega \) such that

i') \( \Phi(\cdot, \beta) : [0, 1] \to \Omega \) is \( L \)–Lipshitz on \( [0, 1] \) for any \( \beta \in D \).

ii') \( \Phi \) is \( 1 – 1 \) on \( (0, 1) \times D \), \( \Phi(0, \beta) = x_0, \Phi(1, \beta) = x_1 \) for any \( \beta \in D \).

iii') There exists a density function \( h : J \to \mathbb{R}_+ \) such that
\[ h(s, \beta) \geq Ks^{d-1}(1 – s)^{d-1} \text{ and } \Phi^{-1}_#(d\mu) \geq h(s, \beta)dsd\beta \text{ on } J. \]

Note that condition (i’-iii’) above imply conditions (i-iii) in Definition 1.1. See Figure 1.

We also define:

**Definition 1.3.** A Borel measure \( \nu = \nu_+ – \nu_- \) is \( \mu \)–neutral if \( \text{supp}(\nu) \subset \text{supp}(\mu) \) and \( \int_{\Omega} d\nu = 0 \). Here \( \nu_\pm \) stands for the positive/negative parts of \( \nu \).
1.1.1 Connected support

**Theorem 1.1.** Suppose $\mu$ is strongly $d-$connected and $\nu$ is $\mu-$neutral. Then

$$\limsup_{\epsilon \downarrow 0} \epsilon^{-d} W_p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) < \infty$$

where $q = \min(1, 1/d + 1/p)$ provided $p \neq d/(d - 1)$.

In the critical case $p = d/(d - 1)$ (where $q = 1$)

$$\limsup_{\epsilon \downarrow 0} \frac{1}{\epsilon \ln^{1/p}(1/\epsilon)} W_p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) < \infty . \quad (6)$$

In particular, if $\mu$ is given by, say, the Lebesgue measure on $\Omega$ and $\Omega$ is a convex, bounded set in $\mathbb{R}^d$ with non-empty interior (hence strongly $d-$connected by definition), if $\nu_+ = \delta_{x_1}, \nu_- = \delta_{x_2}$ for some $x_1 \neq x_2 \in \Omega$ then (6) can be seemed as a critical case of the Sobolev embedding for $\mathcal{W}^{1,p'}(\Omega)$ where $p' := p/(p - 1) = d$ (cf. [4, 5]).

1.1.2 Disconnected support

**Assumption 1.1.**
1. $\mu$ is a Borel probability measure composed of a finite number $m$ of disjoint components $\mu = \sum_{j=1}^{m} \mu_j$ where $\text{supp}(\mu_i) \cap \text{supp}(\mu_j) = \emptyset$ for any $i \neq j$.

2. Each $\mu_i$ satisfies the condition of Theorem 1.1. In particular $A_i := \text{supp}(\mu_i)$ are the connected components of $\text{supp}(\mu)$.

3. $\nu$ is a $\mu$–neutral measure. In particular $\text{supp}(\nu) \subset \text{supp}(\mu)$.

**Definition 1.4.**

i) $\tilde{\nu}_j := \int_{A_j} \nu$. By Assumption 1.1–(3), $\sum_{j=1}^{m} \tilde{\nu}_j = 0$.

ii) $I := \{1, \ldots, m\}$, $I_+ := \{j \in I; \tilde{\nu}_j > 0\}$, $I_- := \{j \in I; \tilde{\nu}_j < 0\}$.

iii) For $i, j \in I$, $|E|_{i,j} := \text{dist}^p(A_i, A_j) \equiv \min_{x \in A_i, y \in A_j} |x - y|^p$.

iv) $G := (V, E)$ a complete graph whose vertices $V = I$ and the length $|E|_{i,j}$ of the edge $E_{i,j}$ connecting $i$ to $j$ is as above.

v) $\bar{G} := (\bar{V}, \bar{E}) \subset G$ is the complete graph such that $\bar{V} = I_+ \cup I_-$. For $\bar{E}_{i,j} \in \bar{E}$ connecting $i \in \bar{V}$ to $j \in \bar{V}$, $|\bar{E}|_{i,j}$ is the geodesic distance corresponding to $(V, E)$. That is:

$$|\bar{E}|_{i,j} := \min_{e_{i,j} \in \bar{E}_{i,j}} \sum_{t=1}^{|e_{i,j}| - 1} |E|_{k_t, k_{t+1}}$$

where $\bar{E}_{i,j}$ is the set of all orbits in $V$ connecting $i$ to $j$:

$$e_{i,j} \in \bar{E}_{i,j} \iff e_{i,j} = \{i = k_1, \ldots, k_n = j\} \quad |e_{i,j}| := n$$

See Figure 3 for an illustration.

vi) Let now $\tilde{\nu}_i > 0$ be the charge associated with the vertex $i \in I_+$, and $-\tilde{\nu}_j > 0$ the charge associated with $j \in I_-$. Let $\|\nu\|_\mu$ be the optimal cost of transportation of $\sum_{i \in I_+} \tilde{\nu}_i \delta_i$ to $\sum_{j \in I_-} (-\tilde{\nu}_j) \delta_j$ subject to the graph metric $d(i, j) = |\bar{E}|_{i,j}$. That is:

$$\|\nu\|_\mu := \min_{\Lambda \in \Lambda(\nu)} \sum_{i \in I_+} \sum_{j \in I_-} \lambda_{i,j} |\bar{E}|_{i,j} := \sum_{i \in I_+} \sum_{j \in I_-} \lambda_{i,j}^* |\bar{E}|_{i,j}$$

where $\Lambda(\nu)$ is the set of non-negative $|I_+| \times |I_-|$ matrices $\{\lambda_{i,j}\}$ which satisfy:

$$\sum_{j \in I_-} \lambda_{i,j} = \tilde{\nu}_i \quad \text{if} \quad i \in I_+$$

$$\sum_{i \in I_+} \lambda_{i,j} = -\tilde{\nu}_j \quad \text{if} \quad j \in I_-.$$
Theorem 1.2. If $\infty > p > 1$ and $\mu, \nu$ satisfy Assumption 1.1 then

$$\lim_{\epsilon \to 0} \epsilon^{-1/p} W_p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) = \|\nu\|_{\mu}^{1/p}$$

2 Proof of Theorem 1.1

We first show that it is enough to prove the Theorem for $\nu_+ = \delta_{x_1}$ and $\nu_- = \delta_{x_2}$ for any $x_1, x_2 \in \text{supp}(\mu)$. Recall that (see, e.g. [R])

$$W_p^p(\mu_1, \mu_2) = \sup_{(\phi, \psi) \in C_p(\Omega)} \int_\Omega \phi d\mu_1 - \int_\Omega \psi d\mu_2$$

where

$$C_p(\Omega) := \{(\phi, \psi) \in C(\Omega) \times C(\Omega); \phi(x) - \psi(y) \leq |x - y|^p \ \forall (x, y) \in \Omega \times \Omega\}$$

Given $\delta > 0$ let $(\bar{\phi}, \bar{\psi}) \in C_p(\Omega)$ for which

$$W_p^p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) \leq \int_\Omega \bar{\phi} d(\mu + \epsilon \nu_+) - \int_\Omega \bar{\psi} d(\mu + \epsilon \nu_-) + \delta$$

$$\leq \int_\Omega \bar{\phi} d(\mu + \epsilon |\nu| \delta_{x_2}) - \int_\Omega \bar{\psi} d(\mu + \epsilon |\nu| \delta_{x_2}) + \delta \leq W_p^p(\mu + \epsilon |\nu| \delta_{x_1}, \mu + \epsilon |\nu| \delta_{x_2}) + \delta$$

where $|\nu| := \int \nu_+ = \int \nu_-$, $\delta_x$ is the Dirac delta function and $x_{1,2}$ is the point of maximum (minimum) of $\bar{\phi}$ (res. $\bar{\psi}$) in $\Omega$. Since $\delta$ is arbitrary small it follows that $\nu_{\pm} = \delta_{x_{1,2}}$ can be chosen for $\nu_{\pm}$ in the proof of Theorem 1.1 up to a non-relevant multiplicative constant.

2.1 Simplifying assumption

To illustrate the proof we start by assuming that $\Omega$ is one dimensional, e.g $\Omega = [0, 1]$ and $x_1 = 0, x_2 = 1$. For $\mu_1, \mu_2 \in \mathcal{P}[0, 1]$, let $M_i(s) := \mu_i[0, s]$ be the accumulation functions of $\mu_i$ for $i = 1, 2$ respectively. Let $S^{(i)}$ be the generalized inverses of $M_i$. Then (Theorem 2.18 in [V] for the case $p = 2$ and Remark 2.19 there for the general case)

$$W_p^p(\mu_1, \mu_2) = \int_0^1 |S^{(1)}(m) - S^{(2)}(m)|^p ds .$$

In our case $M_1$ is the cumulative function of $\mu + \epsilon \delta_0$ while $M_2$ the cumulative function of $\mu + \epsilon \delta_1$. Setting $M = M(s)$ the cumulative function of $\mu$ and $S = S(m)$ its generalized inverse, then $M_1(s) = M(s) + \epsilon$ on $(0, 1]$ and $M_2(s) = M(s)$ on $s \in [0, 1)$, $M_2(1) = 1 + \epsilon$. The corresponding inverses are
i) $S^{(1)}(m) = 0$ for $s \in [0,\epsilon]$, $S^{(1)}(m) = S(m - \epsilon)$ for $\epsilon \leq m \leq 1 + \epsilon$.

ii) $S^{(2)}(m) = S(m)$ for $m \in [0,1]$ and $S^{(2)}(m) = 1$ for $s \in [1,1 + \epsilon]$.

Then (11) implies $W_p^p(\mu + \epsilon \delta_0, \mu + \epsilon \delta_1) =$

$$
\int_0^\epsilon |S(m)|^p dm + \int_\epsilon^{1+\epsilon} |S(m) - S(m - \epsilon)|^p dm + \int_1^{1+\epsilon} |S(m - \epsilon) - 1|^p dm
$$

(12)

Since $S$ is monotone non decreasing:

$$
\int_0^\epsilon |S(m)|^p dm + \int_1^{1+\epsilon} |S(m) - 1|^p dm \leq \epsilon [S^p(\epsilon) + |1 - S(1 - \epsilon)|^p]
$$

(13)

while

$$
\int_\epsilon^1 |S(m) - S(m - \epsilon)|^p dm = \epsilon^p \int_\epsilon^{1-\epsilon} \left| \frac{dS}{dm} \right|^p dm (1 + o(1)).
$$

(14)

Now we use the one-dimensional version of Assumption 1.2-(iii'), namely that $\mu(ds) \geq h(s) ds$ where $h(s) \geq K s^{d-1} (1 - s)^{d-1}$ to obtain the estimates $M(s) \geq \tilde{K} s^d$ and $dM/ds \geq \tilde{K} \min(s^{d-1}, (1 - s)^{d-1})$ hence

$$
S(m) \leq \tilde{K} m^{1/d}, \quad \frac{dS}{dm} \leq \tilde{K} m^{1/d - 1} (1 - m)^{1/d - 1}; \quad m \in (0,1)
$$

for some $\tilde{K} > 0$. It follows from (11-14) that

i) If $p < d/(d - 1)$ then $W_p^p(\mu + \epsilon \delta_0, \mu + \epsilon \delta_1) \leq \tilde{K} \epsilon^p$.

ii) if $p = d/(d - 1)$ then $W_p^p(\mu + \epsilon \delta_0, \mu + \epsilon \delta_1) \leq \tilde{K} \epsilon^p \ln(1/\epsilon)$.

iii) if $p > d/(d - 1)$ then $W_p^p(\mu + \epsilon \delta_0, \mu + \epsilon \delta_1) \leq \tilde{K} \epsilon^{p/d + 1}$.

### 2.2 General case

Let now consider $J, \Phi$ as in Definition 1.2. We may replace $\mu$ by the measure

$\hat{\mu} := \Phi_#(h_1 d\sigma \beta)$ since, by assumption, $\hat{\mu} \leq \mu$ and the inequality

$$
W_p^p(\mu + \epsilon \delta x_1, \mu + \epsilon \delta x_2) \leq W_p^p(\hat{\mu} + \epsilon \delta x_1, \hat{\mu} + \epsilon \delta x_2)
$$

(15)

is evident.

Let $(X_\epsilon, \sigma)$ be a reference measure space such that $\int_{X_\epsilon} d\sigma = \int_{\Omega} d\hat{\mu} + \epsilon$. If $T^{(i)} : X_\epsilon \to \Omega, i = 1,2$, is a pair of Borel mappings such that $T_#^\epsilon \sigma = \hat{\mu} + \epsilon \delta x_1$, then

$$
W_p^p(\hat{\mu} + \epsilon \delta x_1, \hat{\mu} + \epsilon \delta x_2) \leq \int_{X_\epsilon} \left| T^{(1)}(x) - T^{(2)}(x) \right|^p \sigma(dx).
$$

(16)
We now construct \((X_\epsilon, \sigma)\) as follows:

Let

\[ M(s, \beta) := \int_0^s h(w, \beta) dw \quad \text{on} \quad J. \tag{17} \]

Set \(\bar{M}(\beta) := M(1, \beta)\) on \(D\). Then

\[ X_\epsilon := \{(m, \beta) ; \quad \beta \in D, \quad 0 \leq m \leq \bar{M}(\beta) + \epsilon\} \]

and \(\sigma \propto dmd\beta\) is a multiple of the Lebesgue measure on \(X_\epsilon\), normalized according to

\[ \int_{X_\epsilon} d\sigma = \int_\Omega d\hat{\mu} + \epsilon. \]

Let \(S(m, \beta)\) the generalized inverse of \(M\) \(\tag{17}\) with respect to \(s\). In analogy with one-dimensional case above, set

i) \(S^{(1)}(m, \beta) = 0\) for \(s \in [0, \epsilon]\), \(S^{(1)}(m, \beta) = S(m - \epsilon, \beta)\) for \(\epsilon \leq m \leq \bar{M}(\beta) + \epsilon\).

ii) \(S^{(2)}(m, \beta) = S(m, \beta)\) for \(m \in [0, \bar{M}(\beta)]\) and \(S^{(2)}(m, \beta) = \bar{M}(\beta)\) for \(m \in [\bar{M}(\beta), \bar{M}(\beta) + \epsilon]\).

By construction, \(S^{(i)} : X_\epsilon \to J\) satisfy

\[ S^{(1)}_# \sigma = h(ds\beta) + \epsilon \delta_{(0,D)} ; \quad S^{(2)}_# \sigma = h(ds\beta) + \epsilon \delta_{(1,D)}, \]

where \(\delta_{(0,D)}, \delta_{(1,D)}\) stand for the Dirac function concentrated on \((0, D) \subset J,\)

\((1, D) \subset J\) respectively. From Definition \ref{def:1.2} it follows that \(T^{(j)} := \Phi \circ S^{(j)}\)

satisfy \(T^{(j)}_# \sigma = \bar{\mu} + \epsilon \delta_{X_j}\). Then Definition \ref{def:1.2}-(i') yields

\[ \int_{X_\epsilon} \left| T^{(1)}(m, \beta) - T^{(2)}(m, \beta) \right|^p dmd\beta \leq L^p \int_{X_\epsilon} \left| S^{(1)}(m, \beta) - S^{(2)}(m, \beta) \right|^p dmd\beta. \]

We now proceed as in the one-dimensional case (section 2.1) to obtain the proof from \((15, 16)\) via \((12,14)\), in the general case.

3 Proof of Theorem \ref{thm:1.2}

We start by proving the inequality

\[ \lim_{\epsilon \to 0} \epsilon^{-1/p} W_p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) \geq ||\nu||_1^{1/p}. \tag{18} \]

Recall the dual formulation:

\[ W^p_p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) \geq \int \psi d(\mu + \epsilon \nu_+) - \int \phi d(\mu + \epsilon \nu_-) \]
for any \((\phi, \psi) \in \mathcal{C}_p(\Omega)^9\). In fact, it is enough to restrict to \((\phi, \psi) \in \mathcal{C}_p(\text{supp}(\mu)) \equiv \mathcal{C}_p(\bigcup A_i)\). In the special case \(\psi(x) = \phi(x) := z_i\) is a constant over \(A_i\) we get

\[
W_p^p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) \geq \epsilon \sum_{i \in \bar{V}} z_i \bar{\nu}_i \tag{19}
\]

provided \(z_i - z_j \leq \min_{x \in A_i, y \in A_j} |x - y|^p\) for any \(i, j\) and \(x \in A_i, y \in A_j\).

In particular, if \(z_i - z_j \leq |\bar{E}|_{i,j}\) (see definition \ref{def:metric}(iii, vi)). From \ref{eq:19} and Definition \ref{def:metric}(ii) we get that

\[
W_p^p(\mu + \epsilon \nu_+, \mu + \epsilon \nu_-) \geq \epsilon \sup_{\{z\}} \sum_{i \in \bar{V}} z_i \bar{\nu}_i \tag{20}
\]

where the supremum is on all possible values of \(\{z_1, \ldots, z_{\#\bar{V}}\}\) which satisfy \(z_i - z_j \leq |\bar{E}|_{i,j}\) for any \(i, j \in \bar{V}\). Since \(|\bar{E}|\) is a metric on the graph \((\bar{V}, \bar{E})\) via Definition \ref{def:metric}(vi) we recall the dual formulation of the metric Monge problem, or the so called Kantorovich Rubinstein Theorem (Theorem 1.14 in \cite{V} or \cite{R1}) in discrete version:

\[
\|\nu\|_{\mu} = \sup_{\{z\}} \sum_{i \in \bar{V}} z_i \bar{\nu}_i ; \quad z_i - z_j \leq |\bar{E}|_{i,j} \tag{21}
\]

(see also Definition \ref{def:metric}(vii)). Then \ref{eq:18} follows from \ref{eq:20}\ref{eq:21}.

To prove the opposite inequality we need some additional definitions:

**Definition 3.1.**

1. Denote \(Z_{i,j} \in A_i\) to be the closest point in \(A_i\) to \(A_j\). (see Definition \ref{def:metric}(iii)).

2. For any \(i \in I_+, j \in I_-\), \(\bar{e}_{i,j} \in \mathcal{E}_{i,j}\) is an orbit which realizes \ref{eq:7} in Definition \ref{def:metric}(vi).

3. For any \(k = k_l \in \bar{e}_{i,j}\), \(l < |\bar{e}_{i,j}|\) define \(\bar{e}_{i,j}^+(k) = k_{l+1}\). Likewise, \(\bar{e}_{i,j}^-(k) = k_{l-1}\) if \(l > 1\). We also refer to an edge \(E \in \bar{e}_{i,j}\) if \(E = E_{k_l, k_{l+1}}\) for some vertex \(k_l\) along the orbit \(\bar{e}_{i,j}\).

4. For any edge \(E\) in \(G\)

\[
\lambda^*_E := \sum_{\{i,j : E \in \bar{e}_{i,j}\}} \lambda^*_{i,j},
\]

see \ref{eq:8} for \(\lambda^*_{i,j}\) and definition-3 above.
5. For \( k \in V \)

\[
\hat{\nu}^+_k := \sum_{\{i,j : k \in \bar{e}_{i,j}-\{j\}\}} \lambda^*_i,j \delta_{Z_{k,\bar{e}_{i,j}^+(k)}}.
\]

Here \( \delta_x \) is the Dirac delta function at \( x \).

6. For \( i,j \in V \), let \( B_r(Z^j_i) \) be the ball of radius \( r \) centered at \( Z^j_i \in A_j \). Given \( \epsilon > 0 \) let \( r^j_i,\epsilon > 0 \) be the radius of the ball such that \( \mu(A_j \cap B_r(Z^j_i)) = \epsilon \).

Let \( \hat{\mu}_{i,j}^\epsilon \) be the restriction of the measure \( \mu \) to the set \( A_j \cap B_r(Z^j_i) \) defined above. See Figure 2.

7. Let

\[
\hat{\nu}^-_k(\epsilon) := \sum_{l \in V} \sum_{E=(l,k)} \hat{\mu}_{l,k}^* \lambda^+_l, (22)
\]

cf. definitions 3, 4, 6 above.

8. \( \hat{\nu}^+_\epsilon := \sum_{k \in V} \hat{\nu}^+_k; \hat{\nu}^-_\epsilon(\epsilon) := \sum_{k \in V} \hat{\nu}^-_k(\epsilon); \hat{\nu}(\epsilon) := \epsilon \hat{\nu}^+ - \hat{\nu}^-_\epsilon(\epsilon). \)

From the metric property of \( W_p \) and the triangle inequality

\[
W_p(\mu + \epsilon \nu^+, \mu + \epsilon \nu^-) \leq W_p(\mu + \epsilon \nu^+, \mu + \hat{\nu}^+_\epsilon) + W_p(\mu + \epsilon \hat{\nu}^-_\epsilon, \mu + \epsilon \nu^-) + W_p(\mu + \epsilon \hat{\nu}^+_\epsilon, \mu + \hat{\nu}^-_\epsilon) \quad (23)
\]
From the second condition of the Theorem and Theorem 1.1 we obtain that the first two terms on the right of (23) is controlled by $o(\epsilon)$. To complete the proof we need to show Proposition 3.1.

Proposition 3.1.

$$W_p^p(\mu + \epsilon \hat{\nu}_+, \mu + \nu_-(\epsilon)) \leq \epsilon \|\nu\|_\mu + o(\epsilon).$$

In fact, it is enough to show, via Definition 3.1 that

$$W_p^p(\mu, \mu + \hat{\nu}(\epsilon)) \leq \epsilon \|\nu\|_\mu + o(\epsilon),$$

since

$$W_p^p(\mu, \mu + \hat{\nu}(\epsilon)) \geq W_p^p(\mu + \epsilon \hat{\nu}_+, \mu + \nu_-(\epsilon)).$$

For the proof we construct a transport plan $\pi$ from $\mu$ to $\mu + \hat{\nu}(\epsilon)$. To illustrate this construction by a particular example see the directed tree in Figure 3. A detailed description of the plan is given below.

For any positive measure $\sigma \in M(\Omega)$ and $x \in \Omega$ define $\Delta(x, \sigma) \in M(\Omega \times \Omega)$ by its action on test functions $\phi \in C(\Omega \times \Omega)$

$$<\Delta(x, \sigma), \phi>_{\Omega \times \Omega} = \int_\Omega \phi(x, y) d\sigma(y). \quad (24)$$

Note that $\Delta(x, \sigma) = 0 \in M(\Omega \times \Omega)$ for $\sigma = 0 \in M(\Omega)$. More generally,

$$<\Delta, x, \sigma, 1>_{\Omega \times \Omega} =<\sigma, 1>_{\Omega} \quad .$$

Let

$$\mu^- := \mu - \hat{\nu}_-(\epsilon) \quad (25)$$

and $\pi_{\mu^-}$ be the transport plan which preserve the measure $\mu^-$, that is:

$$\pi_{\mu^-} := (Id)_\# \mu^- \quad (26)$$

where $Id : \Omega \to \Omega \times \Omega$ is defined as $Id(x) = (x, x)$. Let now

$$\pi^\epsilon := \pi^- + \sum_{l \in V} \sum_{E = E_{l,k}} \Delta(Z_{l,k}, \mu^\epsilon_{l,k}). \quad (27)$$

First, observe that $\pi^\epsilon \in \Pi(\mu, \nu + \hat{\nu}(\epsilon))$. In fact, from (24) and (27)

$$\int_\Omega \pi^\epsilon(\cdot, dx) = \int_\Omega \pi^-_\#(\cdot, dx) + \sum_{l \in V} \sum_{E = E_{l,k}} \hat{\mu}_{l,k}^\epsilon(dx) \quad (28)$$
Figure 3: Transfer plan via a directed tree: $|\overline{E}|_{13} = |E|_{13}$ ; $|\overline{E}|_{15} = |E|_{13} + |E|_{35}$ ; $|\overline{E}|_{16} = |E|_{13} + |E|_{35} + |E|_{56}$ ; $|\overline{E}|_{17} = |E|_{13} + |E|_{34} + |E|_{47}$ ; $|\overline{E}|_{25} = |E|_{23} + |E|_{35}$ ; $|\overline{E}|_{26} = |E|_{23} + |E|_{35} + |E|_{56}$ ; $|\overline{E}|_{27} = |E|_{23} + |E|_{34} + |E|_{47}$ ; $|\overline{E}|_{43} = |E|_{43}$ ; $|\overline{E}|_{45} = |E|_{43} + |E|_{35}$ ; $|\overline{E}|_{47} = |E|_{47}$ ; $|\overline{E}|_{46} = |E|_{47} + |E|_{76}$ ; $|\overline{E}|_{23} = |E|_{23}$ ;
From (22, 25, 26, 28) we obtain

\[ \int_\Omega \pi^\epsilon(\cdot, dx) = \mu \; , \]

while (25, 24, 27) and Definition 3.1-(5,8) imply

\[ \int_\Omega \pi^\epsilon(dx, \cdot) = \mu + \hat{\nu}(\epsilon) \; . \]

Hence \( \pi^\epsilon \in \Pi(\mu, \mu + \hat{\nu}(\epsilon)) \) as claimed.

It then follows from (24) that

\[ W^p_p(\mu, \mu + \hat{\nu}(\epsilon)) \leq \sum_{l \in V} \sum_{E = E_{l,k}} \int_\Omega \int_\Omega |Z_{l,k} - y|^p \mu^\epsilon_{l,k}(dy) \tag{29} \]

From Definition 3.1(1,6,7) we obtain

\[ \int_\Omega \int_\Omega |Z_{l,k} - y|^p \epsilon_\mu^\epsilon_{l,k}(dy) = \epsilon \sum_{(i,j), E \in \bar{E}_{i,j}} \lambda^*_i |E|_{i,k} + o(\epsilon) \]

This, Definition 3.1(5) and (7, 29) imply

\[ W^p_p(\mu, \mu + \hat{\nu}(\epsilon)) \leq \epsilon \sum_{i,j \in \bar{V} \times \bar{V}} \lambda^*_i |\bar{E}|_{i,j} + o(\epsilon) \; . \]

\[ \square \]
References

[AGS] Ambrosio, L, Gigli, N and Savare, G.: Gradient Flows in metric spaces and in the space of probability measures, Lec. in Mathematics, Birkhauser, 2005

[V] Villani, C  Topics in Optimal Transportation, Graduate Studies in Mathematics, Vol. 58, AMS

[M] McCann, R, J.: A convexity principle for interacting gases, Adv. Math. 128 (1997), no. 1, 153-179.

[R] Rachev, S.T andþuschendorf, L.R: Mass Transportation Problems, Vol 1, Springer, 1998

[W1] Wolansky, G: Limit theorems for optimal mass transportation, Calc. Var. Partial Diferential Equations 42 (2011), no. 3-4, 487-516