arXiv:2303.02183v1 [math.MG] 3 Mar 2023

将 Wasserstein度量扩展到正测度

H. Leblanc*、T. Le Gouic†、J. Liandrat‡和 M. Tournus§¶

2023年3月7日

摘要

我们在有限正测度空间上定义了一种度量，将 2-瓦瑟斯坦度量扩展至该空间；也就是说，该度量在概率测度集合上的限制就是 2-瓦瑟斯坦度量。我们证明了其对偶表述和动态表述，并将 Wasserstein 空间的梯度流理论工具加以扩展。此外，我们还将该空间中的重心与归一化测度在 Wasserstein 空间中的重心联系起来。

1 引言

2-瓦瑟斯坦度量 W_2 是基于概率测度空间上最优传输的度量，定义于**实数集**^{d}上。由此得到的度量空间称为 Wasserstein 空间。其定义为

$$W_2^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \int |x-y|^2 \mathrm{d}\pi(x,y), \tag{1}$$

其中， $\Pi(\mu,\nu)$表示**实数集**$^d \times$ **实数集**^{d} 上的概率测度集合，其边缘分布为 μ 和 ν。这一形式由 Kantorovitch 提出 [Kan42; Kan48]，并以 Monge 的开创性工作为基础[Mon81]。Kantorovitch 还提出了著名的对偶形式

$$W_2^2(\mu,\nu) = \sup \int \phi \mathrm{d}\mu + \int \psi \mathrm{d}\nu, \tag{2}$$

其中，上确界取遍满足条件 $\phi(x)+\psi(y) \leq |x-y|^2$的函数对 $(\phi,\psi) \in \mathcal{C}_b$**(实数集**^{d}**)**^{2} 。这一形式是许多其他结果的核心，其中包括 Brenier定理：该定理指出，在正则情形下，原问题的最小值由一个支撑在函数图像上的耦合取得；这个函数称为 Brenier 或 Monge 映射 [Bre91]。

Benamou 和 Brenier [BB00]，提出了一种出于数值计算目的的动态表述，将 Wasserstein度量重新解释为动能泛函的最小化问题

$$W_2^2(\mu,\nu) = \inf \int_0^1 \int |v_t|^2 \mathrm{d}\mu_t \mathrm{d}t. \tag{3}$$

*hugo.leblanc@univ-amu.fr
†thibaut.le_gouic@math.cnrs.fr
‡jacques.liandrat@centrale-marseille.fr§magali.tournus@centrale-marseille.fr

¶马赛中央理工学院，I2M，UMR 7373，法国国家科学研究中心，艾克斯-马赛大学，马赛，13453，法国

这里，下确界取遍满足弱输运偏微分方程的概率测度值映射 $\mu_t : [0,1] \to \mathcal{P}_2$（**实数集**^{d}）和向量场值映射 $v_t : [0,1] \to L^2(\mu_t)$

$$\partial_t \mu_t + \operatorname{div}(v_t \mu_t) = 0, \quad \mu_0 = \mu,\ \mu_1 = \nu.$$

这一表述形式上赋予了 Wasserstein 空间一种类黎曼结构。因此，它具备一种伪切结构，从而使许多偏微分方程得以重新诠释为 Wasserstein 空间上的梯度流；这一趋势始于 Jordan、Kinderlehrer 和 Otto 的开创性工作 [JKO98]。Wasserstein 空间上由此产生的测地线结构，使其成为研究 Fr´echet 均值（即该空间上的重心）的自然对象。Wasserstein 重心由 [AC11] 引入，此后已应用于图像处理[Rab+12；Bon+15；SA20]和贝叶斯推断 [Bac+22]等领域，同时也引起了理论研究者的兴趣。

我们参阅专著 [Vil03；AGS05；San15] 以进一步了解该主题。 c.

度量计算的数值复杂度仍是一个非常活跃的研究课题，新的算法不断提出。在众多已提出的算法中，我们提到源自线性规划领域的单纯形法、Sinkhorn 算法 [Cut13；Thi+21；ANR17], ，以及较新的 Back-and-Forth 方法 [JL20] ，后者展现出令人瞩目的视觉效果。有关更多细节，读者可参阅 [PC19]。

W_2 的性质以及数值计算方面的进展，使 W_2 成为一种高效的多学科数据比较与插值工具。在计算机科学领域，图像处理、颜色迁移 [RFP14] 的许多应用都依赖于 Monge 映射的存在；风格迁移 [Mro19] 依赖于测地线的存在，以对艺术风格进行插值；在机器学习中， W_2 度量的对偶形式提供了一种更稳定的生成对抗网络训练方法 [LGS19]；在自然语言处理中，Wasserstein度量可用于学习词袋表示；在生物学中， W_2 通过测地线插值来研究细胞分化的轨迹 [Sch+19] ；在经济学中 [Pon+22]，通过比较关键绩效指标分布， W_2 度量能够更深入地分析不同商店的客户体验。

尽管两种测度质量相同是定义Wasserstein度量的核心要求，但这一要求往往构成很大的限制。例如，在点云之间进行点与形状匹配时 [She+21；BC19], ，或者在 [Sch+19] 中，细胞计数并非恒定，需要随时间进行匹配。实践中通常通过对测度进行归一化来应对这一问题，但这样会丢失总质量信息，也偏离了Wasserstein度量的理论框架。这一限制正是本文目标的出发点：寻找一种有意义的度量，将 W_2 扩展到更广泛的正有限测度集合。这也引出了一个简单的问题：在有限正测度上，是否存在一种非平凡度量，其在概率测度上的限制就是Wasserstein度量?

相关工作 迄今为止，人们以不同方式处理这一问题，并引入了其他基于最优传输的度量。为此，非平衡（或部分）最优传输（UOT）理论得以发展，该理论建立在经典最优传输的基础上。人们采用基于经典最优传输多重形式的不同方法，为更广泛的正有限测度空间定义有意义的 W_2 替代形式。这里我们介绍其中两种：熵传输和动态形式。

为了在广泛的框架下定义熵传输（ET）形式 [PR14；GGV21；LMS18；Chi+18b] ，我们首先回顾 Csiszar f-散度。设 μ, ν是 **实数集**^{d}上的正测度，且 f: **实数集**$_+ \to [0,+\infty]$ 是一个凸的下半连续函数，满足 $f(1) = 0$，并令 $f'_\infty(1) = \lim_{t\to+\infty} {}^{f_{(t)}}{}_t$，

则 f-散度为

$$D_f(\mu|\nu) = \int f\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\nu + f'_\infty(1)\nu^\perp(\mathbf{R}^d), \text{with the Lebesgue decomposition } \mu = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\nu + \nu^\perp.$$

对于 $c(x,y)$：$\mathbf{R}^d \times \mathbf{R}^d \to [0,+\infty]$，若其为下半连续成本函数，且 $c(x,x) = 0$ 对所有 $x \in \mathbf{R}^d$ 均成立，则熵-输运形式是 W_2（方程 (1)）原始表述的一种松弛：采用经典最优输运进行质量输运，同时通过 Csiszár f-散度惩罚质量的产生与消亡，从而放宽对输运测度边缘分布的约束。

$$\mathsf{ET}(\mu,\nu) = \inf_\pi \left\{ D_f(P_0\#\pi|\mu) + D_f(P_1\#\pi|\nu) + \int c(x,y)\mathrm{d}\pi(x,y) \right\}, \tag{4}$$

其中，下确界取自正测度，且 $P_0\#\pi$ 和 $P_1\#\pi$ 分别是 π的第一边缘和第二边缘。恰当选择 f 和代价函数 c，便可在正测度集合上得到不同的度量。

非平衡最优传输（UOT）的动态方法（Dyn）基于 W_2 (3) 的动态表述，具体如下。给定一个凸连续函数 $L(v,g)$：$\mathbf{R}^d \times \mathbf{R} \to \mathbf{R}$，其在 $L(0,0) = 0$ 处取最小值——称为拉格朗日函数，动态 UOT 形式通过最小化作用泛函给出

$$\mathsf{Dyn}(\mu,\nu) = \inf \int_0^1 \int L(v_t, g_t)\mathrm{d}\mu_t\mathrm{d}t, \tag{5}$$

其中，g_t 是用于表示质量产生与消亡的源项，下确界取遍满足含源输运方程的三元组 (v_t, g_t, μ_t)

$$\partial_t\mu_t + \mathrm{div}(v_t\mu_t) = g_t\mu_t, \quad \mu_0 = \mu,\ \mu_1 = \nu.$$

在某些假设下，这两种非平衡最优传输（UOT）表述的等价性已在 [LMS18；Chi+18b] 中得到证明。有关非平衡最优传输的更多细节，参见 [Chi+18a；LM15；Maa+15] 。

在基于非平衡最优传输的度量中，最受关注的是 Hellinger-Kantorovich 度量 HK（也称为 Wasserstein-Fisher-Rao 或 Hellinger-Fisher-Rao），它由 [Chi+18a；KMV15；LMS18] 同时提出。该度量可通过上述两种非平衡最优传输形式来定义

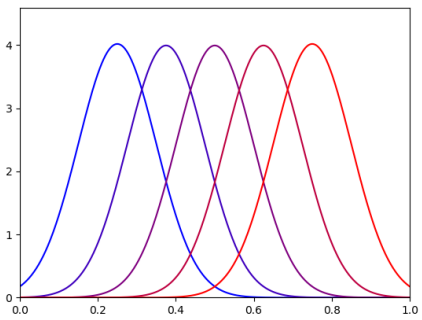
$$\mathsf{HK}^2(\mu,\nu) = \mathsf{ET}(\mu,\nu) = \mathsf{Dyn}(\mu,\nu),$$

其中，熵传输形式 (4) 中的 Csiszar f-散度定义为 $f(t) = t\log(t) + 1 - t$，输运代价 $c(x,y) = -2\log\left(\cos(|x-y| \wedge \frac{\pi}{2})\right)$。HK 的动态表述 (5) 中的拉格朗日函数 $L(v,g)$为 $L(v,g) = |v|^2 + \frac{1}{4}|g|^2$。

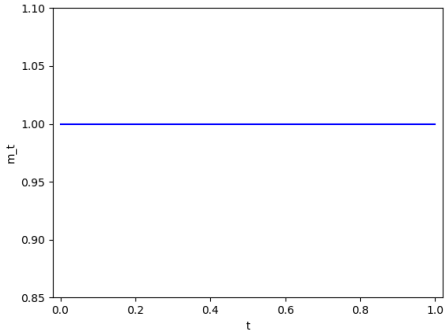
HK 度量常被用作类似 W_2的度量，以比较正测度，因为 HK 继承了 W_2 所具有的多种表述方式。然而，HK 并未将 W_2 扩展到正测度上，因为对于大多数概率测度 $\mu,\ \nu \in \mathcal{P}_2$(**实数集**^{d})，我们有 HK($\mu,\ \nu$) $< W_2(\mu,\ \nu)$；见图1。

本工作的一个核心观察是，在所有这些度量中，概率测度空间都不是测地凸的，即任意两个不同概率测度之间的测地线并不位于概率测度空间内。事实上，我们在第 5 节证明，借助非平衡最优传输框架，不可能将 W_2 扩展到**实数集**^{d} 上。这引出了以下问题：是否存在一种定义在所有正测度上的非平凡度量，其在概率测度上的限制为 Wasserstein距离?

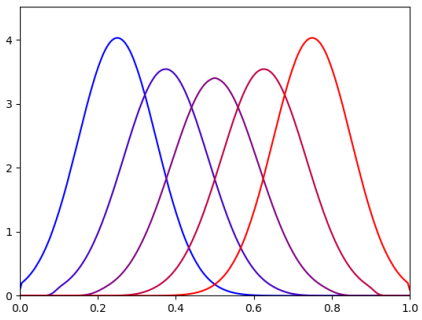
我们的回答是肯定的，并提出一种能够保留多种表述形式的度量。我们将阐明其测地线为何具有“意义”，且它继承了 W_2 度量的算法计算性质、插值和重心。



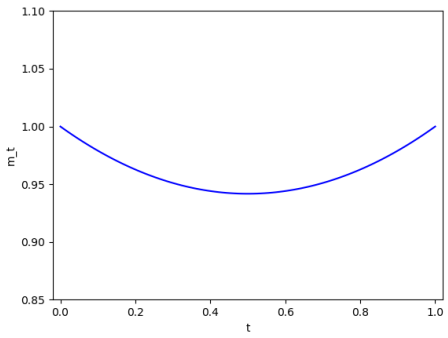
WOP测地线



WOP测地线的质量 m_t



HK测地线



HK测地线的质量 m_t

图1：WOP 度量与 HK 度量下高斯测度之间测地线的比较

概要 本文结构如下。第 2 节中，我们在正测度上定义 WOP度量，并证明 W_2。我们证明其对偶形式，并研究其拓扑和几何性质。第 3 节专门讨论 WOP 的动态表述及其后续梯度流。第 4 节中，我们描述并分析与 WOP 相关的重心；第 5 节则讨论进一步的结果和进展。

记号

- $\mathcal{P}_2$**(实数集**^{d}**)** 是具有二阶矩的正有限测度集合,
- $\mathcal{M}$**(实数集**^{d}**)** 是具有二阶矩的正有限测度集合,
- $\mathcal{M}_K$**(实数集**^{d}**)** 是二阶矩受控的正有限测度集合（见定义（6））,
- m_μ 是 μ的总质量,
- $\bar{\mu}$ 当 $m_\mu > 0$ 时为 μ/m_μ ， 否则为 δ_{x_0} ,
- $M_{x_0}(\mu) = \int d^2(x, x_0)\mathrm{d}\mu(x)$,
- $T_a(x) = a(x - x_0) + x_0$,
- $T\#\mu$是 μ在 T下的推前,

- $\Pi(\mu,\ \nu)$ 是 μ 和 ν之间的运输计划集合，
- $\Pi_o(\mu,\ \nu)$ 是针对成本函数 $|x-y|^2$，以 μ 和 ν 为对象的最优运输计划集合，
- $C_c^\infty(\Omega)$ 是 Ω 中具有紧支撑的无穷次可微函数集合。

2 WOP度量及其基本性质

如引言中所述，我们旨在为有限正测度空间$\mathcal{M}$(**实数集**^{d})定义一种度量，使其在概率测度集合 $\mathcal{P}_2$(**实数集**^{d})上与2-瓦瑟斯坦度量一致。鉴于有限正测度空间具有自然的锥结构，我们的度量受到实数集上的锥的启发（见图2）。

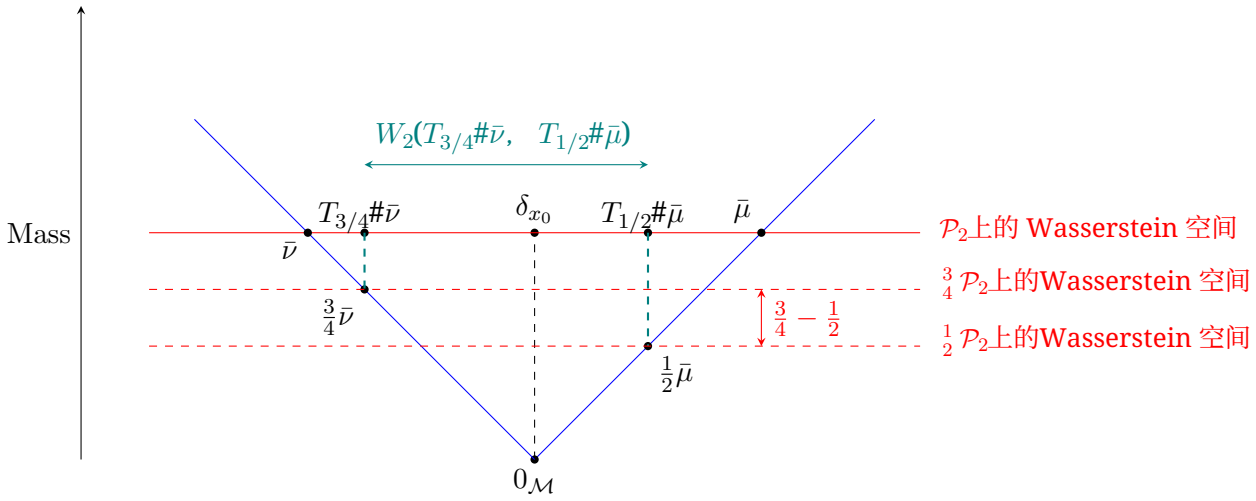


图2：类比于**实数集**^{d}，$\frac{3}{4}\bar{\nu}$与 $\frac{1}{2}\bar{\mu}$之间的平方距离可分解为 $T_{3/4}\#\bar{\nu}$ 与 $T_{1/2}\#\bar{\mu}$ 之间的平方距离（它们分别位于 δ_0 与 $\bar{\nu}$、$\bar{\mu}$之间的测地线上），再加上总质量为 $1/2$和总质量为 $3/4$ 的测度空间之间的平方距离——记为$(1/2-3/4)^2$。由此得到定义1中的WOP度量

定义1（正测度上的Wasserstein度量）。给定 $x_0 \in$ **实数集**^{d}，令（μ,ν）$\in \mathcal{M}$（**实数集**^{d}）^{2}。正测度上的*Wasserstein*（WOP）度量定义为

$$\mathrm{WOP}^2(\mu,\nu)=(m_\mu-m_\nu)^2+W_2^2(T_{m_\mu}\#\bar{\mu},T_{m_\nu}\#\bar{\nu})$$

其中 $T_a(x)=a(x-x_0)+x_0$，x_0 称为参考点。当 x_0 的取值与讨论相关时，我们将用记号 WOP_{x_0} 表示 WOP。

可以验证，WOP 是一个度量。

定理2（WOP 是一个度量。WOP 是空间 $\mathcal{M}$（**实数集**^{d}）上的一个度量。

证明。正性和对称性显然。对于正定性：

$$\mathrm{WOP}(\mu,\nu)=0 \iff \begin{cases} m_\mu=m_\nu \\ T_{m_\mu}\#\bar{\mu}=T_{m_\nu}\#\bar{\nu} \end{cases} \iff \mu=\nu$$

对于三角不等式，利用 $|.|$ 和 W_2 分别是**实数集** 和 $\mathcal{P}_2$**(实数集**^{d}**)**上的度量，以及闵可夫斯基不等式，我们证明如下。

$$\begin{aligned}\mathsf{WOP}(\mu,\nu) &= \sqrt{|m_\mu - m_\nu|^2 + W_2^2(T_{m_\mu}\#\bar{\mu}, T_{m_\nu}\#\bar{\nu})} \\ &\le \sqrt{\big[|m_\mu - m_\eta| + |m_\eta - m_\nu|\big]^2 + \big[W_2(T_{m_\mu}\#\bar{\mu}, T_{m_\eta}\#\bar{\eta}) + W_2(T_{m_\eta}\#\bar{\eta}, T_{m_\nu}\#\bar{\nu})\big]^2} \\ &\le \sqrt{|m_\mu - m_\eta|^2 + W_2^2(T_{m_\mu}\#\bar{\mu}, T_{m_\eta}\#\bar{\eta})} + \sqrt{|m_\eta - m_\nu|^2 + W_2^2(T_{m_\eta}\#\bar{\eta}, T_{m_\nu}\#\bar{\nu})} \\ &= \mathsf{WOP}(\mu,\eta) + \mathsf{WOP}(\eta,\nu).\end{aligned}$$

□

利用**实数集**^{d}上欧几里得距离的一项基本性质，我们可以用纯粹的度量方式重写 WOP度量。在**实数集**^{d}上，对所有 x， $y \in$ **实数集**^{d}和 a， $b \in$ **实数集**$_+$，

$$|a(x-x_0) - b(y-x_0)|^2 = a(a-b)|x-x_0|^2 + b(b-a)|y-x_0|^2 + ab|x-y|^2, \tag{6}$$

因此，WOP度量满足

$$\mathsf{WOP}^2(\mu,\nu) := (m_\mu - m_\nu)^2 + \inf_{\pi\in\Pi(\bar{\mu},\bar{\nu})} \int c^{x_0}_{m_\mu, m_\nu}(x,y)\mathrm{d}\pi(x,y)$$

其中

$$c^{x_0}_{a,b}(x,y) = a(a-b)\|(x-x_0\|^2 + b(b-a)\|y-x_0\|^2 + ab\|x-y\|^2.$$

注意，$c^{x_0}_{a,\ b}$ 的前两项与耦合无关。特别地，这表明最优耦合与归一化测度 $\bar{\mu}$ 和 $\bar{\nu}$ 之间的最优耦合相同，因此，测地线也与归一化测度之间的测地线相同（尽管速度不同，且总质量不同）。现在，令

$$M_{x_0}(\mu) := \int d^2(x,x_0)\mathrm{d}\mu(x),$$

我们可以将 WOP 重写为

$$\mathsf{WOP}^2(\mu,\nu) = (m_\mu - m_\nu)^2 + (m_\mu - m_\nu)(M_{x_0}(\mu) - M_{x_0}(\nu)) + m_\mu m_\nu W_2^2(\bar{\mu},\bar{\nu}). \tag{7}$$

注 3. 可以很容易验证，WOP 对参考点 x_0 的依赖仅体现在以 x_0为中心的矩上。事实上，对于任意两个参考点 x_0 和 y_0，以下等式成立

$$\mathsf{WOP}^2_{x_0}(\mu,\nu) = \mathsf{WOP}^2_{y_0}(\mu,\nu) + (m_\mu - m_\nu)[M_{x_0}(\mu) - M_{y_0}(\mu) + M_{y_0}(\nu) - M_{x_0}(\nu)].$$

WOP度量与 HK 度量之间的一个显著区别在于它们随测度总质量的缩放方式。HK 度量与质量的平方根成正比，而 WOP 度量与质量成正比。

性质4（1 次齐次性）。 WOP 度量满足以下 *1* 次齐次性，即，对于所有 $a \in$**实数集**$_+$，

$$\mathsf{WOP}(a\mu, a\nu) = a\mathsf{WOP}(\mu,\nu).$$

证明。 利用度量 **(7)** 的改写，并注意到 $m_{a\mu} = am_\mu$和 $M_{x_0}(a\mu) = aM_{x_0}(\mu)$，证明显而易见。

□

特别地，当 μ 是零测度 $0_{\mathcal{M}}$时，这一性质确保了一个测度与零测度之间的测地线位于原测度的缩放轨迹上。

现在我们来强调这个度量的主要性质：它将 Wasserstein度量扩展到了所有正有限测度。以下定理对此结果进行了表述。

定理5 （WOP 扩展 Wasserstein距离）。 *WOP* WOP 度量满足

- 对于所有 $\mu, \nu \in \mathcal{P}_2$（**实数集**^{d}），以及 $a \in$ **实数集**$_+$，

$$\mathsf{WOP}(a\mu, a\nu) = aW_2(\mu, \nu),$$

- 对于所有 $\mu \in \mathcal{P}_2$（**实数集**^{d}），以及 $a_0, a_1 \in$ **实数集**$_+$，

$$\mathsf{WOP}^2(a_0\mu, a_1\mu) = (a_0 - a_1)^2(1 + M_{x_0}(\bar{\mu})).$$

特别　　是，概率测度集合对于 WOP度量是测地凸的。

2.1 WOP的拓扑

在描述 WOP度量所诱导的拓扑之前，我们先从一个观察开始本节。给定一个定义在 $\mathcal{M}$（实数集^d）上的度量 D，且该度量扩展了 Wasserstein度量，可以预期，对于任意收敛到 0 的序列 $a_n \in$ 实数集$_+$ 以及任意概率测度序列 μ_n， $a_n\mu_n \to 0_{\mathcal{M}}$ 在 D的拓扑下成立。然而，如果 D 满足某种齐次性（即 D（$a\mu, a\nu$） $\approx a^p D$（μ, ν））， 那么取 $a_n \to 0$ 以及两个概率测度序列 μ_n 和 ν_n，

$$D(a_n\mu_n, a_n\nu_n) \approx a_n^p W_2(\mu_n, \nu_n).$$

因此，如果 W_2（μ_n, ν_n）趋于 $+\infty$ 的速度快于 a_n^p 趋于 0 的速度，那么测度 $a_n\mu_n$ 和 $a_n\nu_n$ 不可能同时收敛到 $0_{\mathcal{M}}$。为避免这一问题，我们引入正测度的一个子集，使得W_2（μ_n, ν_n）保持有界。

定义6 （二阶矩受控的正测度）。 对于 $K > 0$ 和该度量的一个参考点 x_0 WOP$_{x_0}$，我们将 $\mathcal{M}_K$（**实数集**^{d}） 定义为满足以下条件的正测度集合

$$M_{x_0}(\mu) \leq Km_\mu.$$

在该空间中，上述现象不会发生，并且 WOP 是完备度量。

性质7 （完备性）。 空间 （$\mathcal{M}_K$（**实数集**^{d}），WOP） 是完备的

证明。 设 μ_n 是 $\mathcal{M}_K$（**实数集**^{d}）中的柯西序列。由度量 $|.|$ 和 W_2 的完备性可知，m_{μ_n} 收敛于 $m \in$ **实数集**$_+$，且 $T_{m_{\mu_n}}\#\bar{\mu}_n$ 收敛于 $\lambda \in \mathcal{P}_2$（**实数集**^{d}）。由此可证明 μ_n 收敛。

1. 第一种情形：假设 $m > 0$，那么 $T_m(x)$ 与 $T_m^{-1}(x) = \frac{x-x_0}{m} + x_0$ 是双射，因此很容易满足 μ_n 收敛到 $\mu = m.(T_m)^{-1}\#\lambda$。剩下需要证明的是 $\mu \in$

$\mathcal{M}_K$(**实数集**^{d})。利用在 ($\mathcal{P}_2$(**实数集**^{d}) W_2) 中的收敛蕴含二阶矩收敛这一点（见 (8)），我们得到

$$\begin{aligned} M_{x_0}(\mu) &= m \int |T_m^{-1}(x) - x_0|^2 d\lambda \\ &= \frac{1}{m} \int |x - x_0|^2 d\lambda \\ &= \lim_n \frac{1}{m} \int |x - x_0|^2 dT_{m_{\mu_n}} \# \bar{\mu}_n \\ &= \lim_n \frac{m_{\mu_n}}{m} M_{x_0}(\mu_n) \\ &\leq \lim_n K \frac{m_{\mu_n}^2}{m} \\ &= Km. \end{aligned}$$

2. 第二种情况： $m = 0$ 那么我们要证明 μ_n 收敛到 $O_{\mathcal{M}}$。为此，我们证明 W_2 极限 $\lambda = \lim_n T_{m_n} \# \bar{\mu}_n$ 等于δ_{x_0}。利用三角不等式可得

$$W_2(\lambda, \delta_{x_0}) \leq W_2(\lambda, T_{m_{\mu_n}} \# \bar{\mu}_n) + W_2(T_{m_{\mu_n}} \# \bar{\mu}_n, \delta_{x_0}),$$

项 W_2(λ， $T_{m_{\mu_n}}$#$\bar{\mu}_n$)收敛到 0。对于第二项，我们利用 μ_n是一个 $\mathcal{M}_K$(**实数集**^{d})中的序列，因此可得上界

$$\begin{aligned} W_2(T_{m_{\mu_n}} \# \bar{\mu}_n, \delta_{x_0}) &= \int |x - x_0|^2 \mathrm{d} T_{m_{\mu_n}} \# \bar{\mu}_n(x) \\ &= \frac{1}{m_{\mu_n}} \int |T_{m_{\mu_n}}(x) - x_0|^2 \mathrm{d}\mu_n(x) \\ &= m_{\mu_n} \int |x - x_0|^2 \mathrm{d}\mu_n(x) \\ &\leq K m_{\mu_n}^2. \end{aligned}$$

因此 W_2($T_{m_{\mu_n}}$#$\bar{\mu}_n$， δ_{x_0})收敛到 0，这表明 W_2(λ， δ_{x_0}) $= 0$，或者等价地，$\lambda = \delta_{x_0}$，并且柯西序列 μ_n 收敛到 $0_{\mathcal{M}}$。

□

为了研究与 WOP度量相关的收敛性，我们回顾一下测度弱收敛的定义。

定义 8 （测度弱收敛）。 对于序列 $\mu_n \in \mathcal{M}$（**实数集**^{d}） $n \geq 1$，我们称μ_n 弱收敛到 μ，当且仅当对于所有 $\phi \in \mathcal{C}_b$（**实数集**^{d}），

$$\int \phi \, \mathrm{d}\mu_n \to \int \phi \, \mathrm{d}\mu.$$

我们用 $\mu_n \rightharpoonup \mu$ 表示这种收敛。

2-瓦瑟斯坦度量刻画了 $\mathcal{P}_2$(**实数集**^{d}) 中测度弱收敛以及二阶矩的收敛，也就是说，对于序列 (μ_n)$_{n\geq 1}$ in $\mathcal{P}_2$(**实数集**^{d}) 和 $\mu \in \mathcal{P}_2$(**实数集**^{d}),

$$W_2(\mu_n, \mu) \to 0 \iff \begin{cases} M_{x_0}(\mu_n) \to M_{x_0}(\mu) \\ \mu_n \rightharpoonup \mu \end{cases}. \tag{8}$$

WOP 也具有类似的性质。

性质 9 （WOP 将测度弱收敛度量化）。 设 μ_n 是 $\mathcal{M}_K$(**实数集**^{d})中的一个序列，且 $\mu \in \mathcal{M}_K$(**实数集**^{d})，则

$$\mathsf{WOP}(\mu_n, \mu) \to 0 \iff \begin{cases} M_{x_0}(\mu_n) \to M_{x_0}(\mu) \\ \mu_n \rightharpoonup \mu \end{cases}$$

证明。 设 μ_n 是 $\mathcal{M}_K$(**实数集**^{d}) 的一个序列，且设 $\mu \in \mathcal{M}_K$(**实数集**^{d})。那么根据该表述，我们有 ((7))

$$\mathsf{WOP}(\mu_n, \mu) \to 0 \iff \begin{cases} m_{\mu_n} \to m_\mu \\ (m_{\mu_n} - m_\mu)(M_{x_0}(\mu_n) - M_{x_0}(\mu)) + m_{\mu_n} m_\mu W_2^2(\bar{\mu}_n, \bar{\mu}) \to 0 \end{cases} . \tag{9}$$

我们将讨论分为两种情形:
在第一种情形中，我们考虑 $m_\mu > 0$。那么等价关系 (9) 变为

$$\begin{aligned} \mathsf{WOP}(\mu_n, \mu) \to 0 &\underset{(a)}{\iff} \begin{cases} m_{\mu_n} \to m_\mu \\ W_2(\bar{\mu}_n, \bar{\mu}) \to 0 \end{cases} \\ &\underset{(b)}{\iff} \begin{cases} m_{\mu_n} \to m_\mu \\ \bar{\mu}_n \rightharpoonup \bar{\mu} \\ M_{x_0}(\bar{\mu}_n) \to M_{x_0}(\bar{\mu}) \end{cases} \\ &\underset{(c)}{\iff} \begin{cases} \mu_n \rightharpoonup \mu \\ M_{x_0}(\mu_n) \to M_{x_0}(\mu) \end{cases} \end{aligned}$$

等价关系 (a) 成立，因为当 m_{μ_n} 收敛到 m_μ 时，$(m_{\mu_n} - m_\mu)(M_{x_0}(\mu_n) - M_{x_0}(\mu))$ 收敛到 0。事实上，

$$|(m_{\mu_n} - m_\mu)(M_{x_0}(\mu_n) - M_{x_0}(\mu))| \le |m_{\mu_n} - m_\mu|(M_{x_0}(\mu_n) + M_{x_0}(\mu)) \le |m_{\mu_n} - m_\mu|(K m_{\mu_n} + M_{x_0}(\mu)) \to 0$$

等价关系 (b) 成立，因为 W_2 度量了测度的弱收敛与 $+$ 矩收敛。

等价关系（c）直接由弱收敛的定义得出。
对于第二种情况，考虑 $m_\mu = 0$ 并回忆 $\mu_n \in \mathcal{M}_K$(**实数集**^{d})。于是，等价关系（9）变为

$$\begin{aligned} \mathsf{WOP}(\mu_n, 0_{\mathcal{M}}) \to 0 &\underset{(d)}{\iff} m_{\mu_n} \to 0 \\ &\underset{(e)}{\iff} \begin{cases} \mu_n \rightharpoonup 0_{\mathcal{M}} \\ M_{x_0}(\mu_n) \to 0 \end{cases} \end{aligned}$$

等价关系（d）成立，因为 $(m_{\mu_n} - m_\mu)(M_{x_0}(\mu_n) - M_{x_0}(\mu)) + m_{\mu_n} m_\mu W_2(\bar{\mu}_n,\ \bar{\mu})$ 在 m_{μ_n} 趋于 0 时也趋于 0。事实上，

$$|(m_{\mu_n} - m_\mu)(M_{x_0}(\mu_n) - M_{x_0}(\mu)) + m_{\mu_n} m_\mu W_2(\bar{\mu}_n, \bar{\mu})| = m_{\mu_n} M_{x_0}(\mu_n) \le K m_{\mu_n}^2 \to 0.$$

对于等价关系（e），蕴含关系（$\Leftarrow$）直接由弱收敛的定义得出。对于另一个蕴含关系（$\Rightarrow$），令 $\phi \in \mathcal{C}_b$(**实数集**^{d})，则有

$$\left| \int \phi d\mu_n \right| \le \|\phi\|_\infty m_{\mu_n} \to 0.$$

因此 $\mu_n \rightharpoonup 0_{\mathcal{M}}$。对于二阶矩，注意到

$$M_{x_0}(\mu_n) \le K m_{\mu_n} \to 0,$$

由此完成证明。 □

2.2 对偶形式

WOP 具有与式（2）中描述的 W_2 对偶形式度量相似的对偶形式。

定理 10 （对偶形式）。 WOP 度量的对偶形式为

$$\mathsf{WOP}^2(\mu,\nu)=\sup_{(\phi,\psi)\in\Gamma_{\mu,\nu}}\int\phi\mathrm{d}\mu+\int\psi d\nu,$$

其中 $\Gamma_{\mu,\ \nu}$ 是函数集

$$\Gamma_{\mu,\nu}=\left\{(\phi,\psi)\in\mathcal{C}_b(\mathbf{R}^d)^2\middle|m_\mu\phi(x)+m_\nu\psi(y)\leq(m_\mu-m_\nu)^2+|T_{m_\mu}(x)-T_{m_\nu}(y)|^2,\mu\otimes\nu\text{-}a.e.\right\}.$$

证明。 证明的思路是将 WOP 度量写成高维空间上的 2-瓦瑟斯坦度量，为测度的质量增加一个维度，并在该质量上取一个狄拉克测度。然后，利用对偶形式（公式（2））， 我们推导出 WOP 的一种类似对偶的形式。事实上，

$$\begin{aligned}\mathsf{WOP}^2(\mu,\nu)&=W_2^2(\delta_{m_\mu}\otimes T_{m_\mu}\#\bar{\mu},\delta_{m_\nu}\otimes T_{m_\nu}\#\bar{\nu})\\&=\sup_{(\phi,\psi)\in\Phi}\int\phi(\alpha,x)\mathrm{d}\delta_{m_\mu}(\alpha)\otimes T_{m_\mu}\#\bar{\mu}(x)+\int\phi(\beta,y)\mathrm{d}\delta_{m_\nu}(\beta)\otimes T_{m_\nu}\#\bar{\nu}(y)\\&=\sup_{(\phi,\psi)\in\Phi}\int\frac{\phi(m_\mu,T_{m_\mu}(x))}{m_\mu}\mathrm{d}\mu(x)+\int\frac{\phi(m_\nu,T_{m_\nu}(y))}{m_\nu}\mathrm{d}\nu(y).\end{aligned}$$

其中空间 Φ 是函数集 $(\phi,\ \psi)\in\mathcal{C}_b$(实数集$_+\times$实数集$^d)^2$，满足对于几乎处处由 $(\alpha,\ x,\ \beta,\ y)$ 定义的$\delta_{m_\mu}\otimes T_{m_\mu}\#\bar{\mu}\otimes\delta_{m_\nu}\otimes T_{m_\nu}\#\bar{\nu}$，都有不等式

$$\phi(\alpha,x)+\psi(\beta,y)\leq(\alpha-\beta)^2+|x-y|^2,$$

或等价地 等价地，对于几乎处处定义的所有 (x,y)，都有不等式

$$\phi(m_\mu,T_{m_\mu}(x))+\psi(m_\nu,T_{m_\nu}(y))\leq(m_\mu-m_\nu)^2+|T_{m_\mu}(x)-T_{m_\nu}(y)|^2.$$

令函数 $\tilde{\phi}(x)=\frac{\phi}{(m_\mu,T_{m_\mu}(x))}\ m_\mu$ 和 $\tilde{\psi}(y)=\frac{\phi}{(m_\nu,T_{m_\nu}(y))}\ m_\nu$，最终得到

$$\mathsf{WOP}^2(\mu,\nu)=\sup\int\tilde{\phi}(x)\mathrm{d}\mu(x)+\int\tilde{\psi}(y)d\nu(y),$$

其中函数 $(\tilde{\phi},\tilde{\psi})\in\mathcal{C}_b$(实数集$^d)^2$ 满足

$$m_\mu\tilde{\phi}(x)+m_\nu\tilde{\psi}(y)\leq(m_\mu-m_\nu)^2+|T_{m_\mu}(x)-T_{m_\nu}(y)|^2,$$

对于几乎处处定义的 (x,y)。 □

2.3 WOP 的测地线

回顾一下，若度量空间 $(D,\mathcal{X})$ 满足：对于 $\mathcal{X}^2$ 中任意一对点 (x,y)，都存在一条路径 σ_t：$[0,1]\longrightarrow\mathcal{X}$，称为（恒速）测地线，从 x 出发并以 y 为终点，使得

$$D(\sigma_t,\sigma_s)=|t-s|D(x,y),\quad\forall t,s\in[0,1].$$

在 $\left(W_2,\mathcal{P}_2\text{(实数集}^d\text{)}\right)$ 的情形下，测地线是众所周知的，可以用两个端点之间的 Monge 映射来表示。以下结果描述了 WOP 的测地线，并将其与归一化测度之间的 W_2 测地线联系起来。

性质 11（WOP的测地线）。给定 $\mu_0, \mu_1 \in \mathcal{M}$（**实数集**^{d}），测地线为

$$\mu_t = m_t.\hat{\mu}_{\lambda_t}, \quad t \in [0,1], \tag{10}$$

其中质量为 $m_t = (1-t)m_{\mu_0} + tm_{\mu_1}$，概率测度 $\hat{}\mu_t$ 是 $(\mathcal{P}_2(\mathbf{R}^d)\ W_2)$中连接 $\bar{\mu}_0$ 与 $\bar{\mu}_1$的测地线，参数化为 $\lambda_t = \frac{tm_{\mu_1}}{(1-t)m_{\mu_0}+tm_{\mu_1}}$。特别地，这些测地线~~与参考点~~ x_0无关。

此外，空间 $(\mathcal{M}(\mathbf{R}^d)$,WOP$)$ 是一个具有亚历山大罗夫意义下正曲率的测地空间。

证明。令 $\mu_0, \mu_1 \in \mathcal{M}(\mathbf{R}^d)$，利用 $(\mathbf{R}\ |.|)$ 和 $(\mathcal{P}_2(\mathbf{R}^d)\ W_2)$ 是测地空间这一事实，并根据 WOP 的定义，我们可推得 $(\mathcal{M}(\mathbf{R}^d)$,WOP$)$ 中连接 μ_0 与 μ_1 的测地线为

$$\mu_t = m_t.T_{m_t}^{-1}\#\eta_t,$$

其中 $m_t = (1-t)m_{\mu_0} + tm_{\mu_1}$，而 η_t 是 $(\mathcal{P}_2(\mathbf{R}^d)\ W_2)$ 中连接 $T_{m_{\mu_0}}\#\bar{\mu}_0$ 与 $T_{m_{\mu_1}}\#\bar{\mu}_1$ 的测地线。此外，在 $(\mathcal{P}_2(\mathbf{R}^d)\ W_2)$ 中，我们可以更详细地描述该测地线

$$\eta_t = P_t\#\pi,$$

其中 $P_t(x,\ y) = (1-t)x + ty$，并且 $\pi \in \Pi_o(T_{m_{\mu_0}}\#\bar{\mu}_0,\ T_{m_{\mu_1}}\#\bar{\mu}_1)$。因此，

$$\pi = (T_{m_{\mu_0}}, T_{m_{\mu_1}})\#\tilde{\pi},$$

其中 $\tilde{\pi} \in \Pi_o(\bar{\mu}_0,\ \bar{\mu}_1)$。因此，测地线可以改写如下

$$\begin{aligned}\mu_t &= m_t.T_{m_t}^{-1}\#P_t\#(T_{m_{\mu_0}}, T_{m_{\mu_1}})\#\tilde{\pi} \\ &= m_t.\big(T_{m_t}^{-1} \circ P_t \circ (T_{m_{\mu_0}}, T_{m_{\mu_1}})\big)\#\tilde{\pi}.\end{aligned}$$

令 $\lambda_t = \frac{tm_{\mu_1}}{(1-t)m_{\mu_0}+tm_{\mu_1}}$，我们可以简化 $\big(T_{m_t}^{-1} \circ P_t \circ (T_{m_{\mu_0}},\ T_{m_{\mu_1}})\big)$：

$$\begin{aligned}\big(T_{m_t}^{-1} \circ P_t \circ (T_{m_{\mu_0}}, T_{m_{\mu_1}})\big)(x,y) &= \frac{(1-t)\big[m_{\mu_0}(x-x_0)+x_0\big] + t\big[m_{\mu_1}(y-x_0)+x_0\big] - x_0}{(1-t)m_{\mu_0} + tm_{\mu_1}} + x_0 \\ &= \frac{(1-t)m_{\mu_0}x + tm_{\mu_1}y}{(1-t)m_{\mu_0} + tm_{\mu_1}} \\ &= (1-\lambda_t)x + \lambda_t y \\ &= P_{\lambda_t}(x,y)\end{aligned}$$

利用 $P_t\#\tilde{\pi}$ 是 $\bar{\mu}_0$ 与 $\bar{\mu}_1$ 之间的测地线，我们将其记为 $\hat{}\mu_t$，这意味着我们可以将该测地线解释为时间重参数化

$$\mu_t = m_t.P_{\lambda_t}\#\tilde{\pi} = m_t.\hat{\mu}_{\lambda_t}.$$

将测地线改写成这种形式的直接优势在于，它不依赖于参考点 x_0。最后，$(\mathcal{M}($**实数集**$^d)$,WOP$)$ 的正曲率直接源于 $($**实数集**$^d, |.|)$ 的 0 曲率和 $(\mathcal{P}_2($**实数集**$^d)\ W_2)$ 的正曲率。事实上，对于所有 $t \in [0,1]$，以及所有 $\mu_0, \mu_1 \in \mathcal{M}($**实数集**$^d)$，存在 μ_t，它是此前定义的连接 μ_0 和 μ_1 的测地线，使得对于所有 $\eta \in \mathcal{M}($**实数集**$^d)$，我们有：

$$
\begin{aligned}
\mathsf{WOP}^2(\mu_t,\ \eta) =& |m_t - m_\eta|^2 + W_2^2(T_{m_t}\#\bar{\mu}_t,\ T_{m_\eta}\#\bar{\eta}) \\
=& (1-t)|m_{\mu_0} - m_\eta|^2 + t|m_{\mu_1} - m_\eta|^2 - t(1-t)|m_{\mu_0} - m_{\mu_1}|^2 + W_2^2(T_{m_t}\#\bar{\mu}_t, T_{m_\eta}\#\bar{\eta}) \\
\geq& (1-t)\left[|m_{\mu_0} - m_\eta|^2 + W_2^2(T_{m_{\mu_0}}\#\bar{\mu}_0, T_{m_\eta}\#\bar{\eta})\right] + t\left[|m_{\mu_1} - m_\eta|^2 + W_2^2(T_{m_{\mu_1}}\#\bar{\mu}_1, T_{m_\eta}\#\bar{\eta})\right] \\
& - t(1-t)\left[|m_{\mu_0} - m_{\mu_1}|^2 + W_2^2(T_{m_{\mu_0}}\#\bar{\mu}_0, T_{m_{\mu_1}}\#\bar{\mu}_1)\right] \\
=& (1-t)\mathsf{WOP}^2(\mu_0,\eta) + t\mathsf{WOP}^2(\mu_1,\eta) - t(1-t)\mathsf{WOP}^2(\mu_0,\mu_1).
\end{aligned}
$$

□

3 WOP 上的一阶微积分

本节首先介绍 WOP 的动态解释，随后介绍其黎曼式解释：一大类定义在 $\mathcal{M}$(实数集^d)上的泛函的切空间及其梯度，以及梯度流方程。最后，我们给出一种直接方法，将定义在 $\mathcal{P}_2$(实数集^d)上的泛函扩展到 $\mathcal{M}$(实数集^d)上，同时保持质量守恒的梯度流。

3.1 动态表述

我们先推导动态表述的形式计算。回顾式 (10)， μ_0 与 μ_1 之间的测地线由下式给出

$$
\mu_t = m_t.\hat{\mu}_{\lambda_t},
$$

其中 $\lambda_t = \frac{tm_1}{\frac{1-t\ m_0+tm_1}{v=\nabla\phi-}\ \mu_t}$ W_2 $\bar{\mu}_0$ $\bar{\mu}_1$ $m_t = \ 1-t\ m_0+_{()}$，ˆ 是 与 之间的常速测地线，而 ( ) 。
tm_1 $\bar{\mu}_0$ $\bar{\mu}_1$
那么，对于 Id，作为 中从 指向 的切向量，

$$
\partial_t\mu_t = \partial_t m_t \cdot \hat{\mu}_{\lambda_t} + m_t \cdot \partial_t\lambda_t \cdot \partial_s\hat{\mu}_s|_{s=\lambda_t}.
$$

因此，

$$
\begin{aligned}
\partial_t\mu_t|_{t=0} &= (m_1 - m_0)\bar{\mu}_0 - m_0 \cdot \partial_{t|t=0}\lambda_t \cdot \mathrm{div}(v\bar{\mu}_0) \\
&= \left(\frac{m_1}{m_0} - 1\right)\mu_0 - \frac{m_1}{m_0}\mathrm{div}\,(v\mu_0) \\
&= \frac{m_0'}{m_0}\mu_0 - \mathrm{div}\left(v\left(1 + \frac{m_0'}{m_0}\right)\mu_0\right).
\end{aligned}
$$

在这个方程中，项 $m' := m_0' = m_1 - m_0$ 表示局部产生的质量， $u := v\left(1 + \frac{m'}{m_0}\right)$ 是一个使 $\bar{\mu}_0$ 发生位移的向量场。这两个量描述了从 μ_0 指向 μ_1 的位移方向，可以看作切平面中的元素。由于在 μ_0 处的切平面中， μ_1 与 μ_0 之间的距离和它们的实际距离相同，因此我们期望在 μ_0 处的切平面 $\|\cdot\|_{\mu_0}$ 中定义的距离满足

$$
\|(m', u)\|_{\mu_0}^2 = \mathsf{WOP}^2(\mu_0,\mu_1) = (m_1 - m_0)^2 + \int |y - x|^2 \mathrm{d}\pi(x,y),
$$

对于某个 $\pi \in \Pi_o$（$T_{m_0}\#\bar{\mu}_0$，$T_{m_1}\#\bar{\mu}_1$）。根据布雷尼尔定理，我们知道 $\pi = (T_{m_0}, T_{m_1})\#\big(\bar{\mu}_0 \otimes \delta_{\{y=\nabla\phi(x)\}}\big)$，因此

$$\begin{aligned}
\|(m', u)\|_{\mu_0}^2 &= m'^2 + \int |T_{m_1}\big(\nabla\phi(x)\big) - T_{m_0}(x)|^2 d\bar{\mu}_0(x) \\
&= m'^2 + \int |m_1\big(\nabla\phi(x) - x_0\big) - m_0(x - x_0)|^2 d\bar{\mu}_0(x) \\
&= m'^2 + \int |m_1 v + (m_1 - m_0)(x - x_0)|^2 d\bar{\mu}_0(x) \\
&= m'^2 + \int |m_0 u + m'(x - x_0)|^2 d\bar{\mu}_0(x).
\end{aligned}$$

换言之，切平面应由两个元素构成：一个实数 m'，用于表示质量的变化；以及一个向量场 u，用于表示归一化质量的输运；WOP度量则按照上述公式对这两个元素进行积分。我们将在第3.2节中更详细地讨论切空间。这些观察自然地导出了以下动态表述。

定理12（动态表述）**：** 对于所有 $\mu, \nu \in \mathcal{M}$（**实数集**^{d}），WOP 度量满足

$$\mathsf{WOP}^2(\mu, \nu) = \inf\left\{ \int |m_t'|^2 + |m_t u_t + m_t'(x - x_0)|^2 \mathrm{d}\bar{\mu}_t(x)\mathrm{d}t \right\}$$

其中，下确界是在满足以下条件的三元组 (u_t, m_t, μ_t) 中取的：$m_t \in \mathcal{C}^1([0,1])$，$u_t\colon [0,1] \to L^2(\mu_t)$，$\mu_t\colon [0,1] \to \mathcal{M}$（**实数集**^{d}），并且

$$\begin{cases}
\partial_t \mu_t + \operatorname{div}(u_t \mu_t) = \frac{m_t'}{m_t}\mu_t \\
m_0 = m_\mu, \quad m_1 = m_\nu, \\
\mu_0 = \mu, \quad \mu_1 = \nu.
\end{cases}$$

该方程应作弱意义下的理解，即需对具有紧支撑的测试函数集合 $\mathcal{C}^\infty$（**实数集**$^d \times [0,1]$）进行验证。

证明。 根据 WOP 定义中的 W_2 动态表述，我们得到

$$\begin{aligned}
\mathsf{WOP}^2(\mu_0, \mu_1) &= (m_0 - m_1)^2 + W_2^2(T_{m_0}\#\bar{\mu}_0, T_{m_1}\#\bar{\mu}_1) \\
&= \inf\left\{ \int |m_t'|^2 + \int |v_t|^2 \mathrm{d}\nu_t \mathrm{d}t :\ \nabla m_t' = 0, \partial_t \nu_t + \operatorname{div}(v_t \nu_t) = 0, \nu_i = T_{m_i}\#\bar{\mu}_i, i = 1, 2 \right\} \\
&= \inf\left\{ \int |m_t'|^2 + \int |v_t \circ T_{m_t}|^2 \mathrm{d}\bar{\mu}_t \mathrm{d}t :\ \nabla m_t' = 0, \partial_t \nu_t + \operatorname{div}(v_t \nu_t) = 0, \nu_i = T_{m_i}\#\mu_i, i = 1, 2 \right\}
\end{aligned}$$

对于每个 ν_t，我们令 $\mu_t = m_t \cdot T_{m_t}^{-1}\#\nu_t = m_t^{d+1} \cdot \nu_t \circ T_{m_t}$。因此，$\nabla\mu_t = m_t^{d+2}\nabla\nu_t \circ T_{m_t}$ 和

$$\begin{aligned}
\partial_t \mu_t &= (d+1) m_t^d m_t' \nu_t \circ T_{m_t} + m_t^{d+1} \partial_t(\nu_t \circ T_{m_t}) \\
&= \frac{(d+1)}{m_t} m_t' \mu_t + m_t^{d+1}\left[(\partial_t \nu_t) \circ T_{m_t} + \langle \partial_t T_{m_t}, \nabla\nu_t \circ T_{m_t} \rangle\right] \\
&= \frac{(d+1)}{m_t} m_t' \mu_t + m_t^{d+1}(\partial_t \nu_t) \circ T_{m_t} + m_t^{-1}\langle \partial_t T_{m_t}, \nabla\mu_t \rangle
\end{aligned}$$

另一方面，

$$\begin{aligned}(\partial_t \nu_t)\circ T_{m_t} &= -\mathrm{div}(v_t\nu_t)\circ T_{m_t}\\ &= -\mathrm{div}(v_t)\circ T_{m_t}\cdot \nu_t\circ T_{m_t} - \langle v_t\circ T_{m_t}, \nabla\nu_t\circ T_{m_t}\rangle\\ &= -m_t^{-(d+1)}\mathrm{div}(v_t)\circ T_{m_t}\cdot\mu_t - m_t^{-(d+2)}\langle v_t\circ T_{m_t},\nabla\mu_t\rangle\\ &= -m_t^{-(d+2)}\mathrm{div}(v_t\circ T_{m_t})\cdot\mu_t - m_t^{-(d+2)}\langle v_t\circ T_{m_t},\nabla\mu_t\rangle\end{aligned}$$

因此，令 $m_t^{-1}v_t\circ T_{m_t}=\tilde{u}_t$

$$\begin{aligned}\partial_t\mu_t &= \frac{(d+1)}{m_t}m_t'\mu_t - m_t^{-1}\mathrm{div}(v_t\circ T_{m_t})\cdot\mu_t - m_t^{-1}\langle v_t\circ T_{m_t},\nabla\mu_t\rangle + m_t^{-1}\langle\partial_t T_{m_t},\nabla\mu_t\rangle\\ &= \frac{(d+1)}{m_t}m_t'\mu_t - \mathrm{div}(\tilde{u}_t)\cdot\mu_t - \langle\tilde{u}_t,\nabla\mu_t\rangle + m_t^{-1}\langle\partial_t T_{m_t},\nabla\mu_t\rangle\\ &= \frac{(d+1)}{m_t}m_t'\mu_t - \mathrm{div}(\tilde{u}_t\mu_t) + m_t^{-1}\langle\partial_t T_{m_t},\nabla\mu_t\rangle\\ &= \frac{m_t'}{m_t}\mu_t - \mathrm{div}(\tilde{u}_t\mu_t) + m_t^{-1}\mathrm{div}(\partial_t T_{m_t}\mu_t)\\ &= \frac{m_t'}{m_t}\mu_t - \mathrm{div}\left((\tilde{u}_t - m_t^{-1}\partial_t T_{m_t})\mu_t\right)\\ &= \frac{m_t'}{m_t}\mu_t - \mathrm{div}\left((\tilde{u}_t - \frac{m_t'}{m_t}(x-x_0)\mu_t\right)\end{aligned}$$

因此

$$\mathsf{WOP}^2(\mu_0,\mu_1) = \inf\left\{\int|m_t'|^2 + \int|m_tu_t|^2\mathrm{d}\bar{\mu}_t\mathrm{d}t:\ \nabla m_t'=0, \partial_t\mu_t + \mathrm{div}\left(\left(\tilde{u}_t - \frac{m_t'}{m_t}(x-x_0)\right)\mu_t\right) = \frac{m_t'}{m_t}\mu_t\right\}.$$

令 $u_t=\tilde{u}_t-\frac{m_t'}{m_t}(x-x_0)$，即可得到结果。 □

3.2 梯度流

如上一节所述，空间 $\left(\mathcal{P}_2(\mathbf{实数集}^d)\ W_2\right)$ 具有类黎曼结构，其中每个测度 μ 处的切空间 $\mathbf{T}_\mu\mathcal{P}_\mathbf{2}(\mathbf{实数集^d})$ 是 $L^2(\mu)$ 中向量场的一个子集。更准确地说，它定义为

$$\mathbf{T}_\mu\mathcal{P}_\mathbf{2}(\mathbf{R^d}) = \overline{\left\{\nabla\phi\middle|\phi\in\mathcal{C}_c^\infty(\mathbf{R}^d)\right\}}^{L^2(\mu)}.$$

更多细节请参阅 [AGS05，第 8 章。4]

受上一节的观察及动态表述的启发，空间$\left(\mathcal{M}(\mathbf{实数集}^d),\ \mathsf{WOP}\right)$ 也可以赋予一种类黎曼结构。

定义13 (切空间**)**。 我们将 $\mathcal{M}(\mathbf{实数集}^d)$ 在 μ 处的切空间定义为

$$\mathbf{T}_\mu\mathcal{M}(\mathbf{R^d}) := \mathbf{T}_{\bar{\mu}}\mathcal{P}_\mathbf{2}(\mathbf{R^d})\times\mathbf{R}.$$

对于切向量 $(u,m')\in\mathbf{T}_\mu\mathcal{M}(\mathbf{实数集^d})$，存在一条$(\mu_t)^{t\in[0,1]}$ 从$\mu_0=\mu$ 出发的测地线，使得

$$\partial_t\mu_t|_{t=0} = \frac{m'}{m_\mu}\mu - \mathrm{div}(u\mu). \tag{11}$$

我们在$\mathbf{T}_\mu\mathcal{M}$(**实数集**$^{\mathbf{d}}$) 上赋予如下定义的标量积:

$$\langle (u_1, m_1'), (u_2, m_2') \rangle_\mu = m_1' m_2' + \int \langle m_\mu u_1(x) + m_1'(x - x_0), m_\mu u_2(x) + m_2'(x - x_0) \rangle d\bar{\mu}(x). \quad (12)$$

利用切空间，我们可以仿照黎曼流形上的链式法则公式，在 WOP空间上定义梯度以及泛函的梯度流。

回顾一下，在黎曼流形 $(\mathcal{M},\ g)$ 上，光滑泛函 $F\colon \mathcal{X} \longrightarrow$**实数集** 的梯度流是曲线 $(x_t)^{t\geq 0}$，它在每一点都沿着 F 的负梯度方向。因此，它由以下方程刻画

$$\partial_t x_t = -\nabla F(x_t), \quad \forall t \geq 0.$$

定义14 (WOP 梯度)。 对于泛函 $F\colon \mathcal{M}($**实数集**$^d) \to$ **实数集**， F 在点 μ 处的WOP 梯度定义为满足以下条件的 $\nabla_{\mathsf{WOP}} F(\mu) \in \mathbf{T}_\mu\mathcal{M}($**实数集**$^{\mathbf{d}})$ ： 对于所有局部 ***Lipschitz*** 路径 $(\mu_t)^{t\in[0,1]}$，只要 $\mu_0 = \mu$ 且 $\partial_t \mu_t|_{t=0} = \frac{m'}{m_\mu}\mu - \mathrm{div}(\overline{u}\mu)$，

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\mu_t)|_{t=0} = \left\langle \nabla_{\mathsf{WOP}} F(\mu), (m', u) \right\rangle_\mu. \quad (13)$$

取值于 $\mathcal{M}($**实数集**$^d)$ 的连续路径$(\mu_t)^{t\geq 0}$是 F的梯度流，当且仅当 $\nabla_{\mathsf{WOP}}(F)(\mu_t) = (m_t', u_t)$ 且$(\mu_t)^{t\geq 0}$ 是**.....**的弱解

$$\partial_t \mu_t = \frac{m_t'}{m_{\mu_t}} \mu_t + \mathrm{div}(u_t \mu_t).$$

为了刻画 WOP空间上的梯度流，考虑泛函 F :$\mathcal{M}$**(实数集**^{d}**)** $\to$ **实数集**，使得对于 $\mu \in \mathcal{M}$**(实数集**^{d}**)**，存在一个函数 $\frac{\delta F}{\delta \mu}(\mu) \in \mathcal{C}^1$**(实数集**^{d}**)**，称为 F 在 μ 处的一阶变分，使得对于所有 $\nu \in \mathcal{M}$**(实数集**^{d}**)**

$$F(\nu) = F(\mu) + \int \frac{\delta F}{\delta \mu}(\mu) \mathrm{d}(\nu - \mu) + o\big(\mathsf{WOP}(\mu, \nu)\big). \quad (14)$$

定理 15 (WOP-梯度公式)。 设 F 是 $\mathcal{M}($**实数集**$^d)$ 上的泛函，满足 (14)，并以 $\frac{\delta F}{\delta \mu}(\mu)$ 表示其一阶变分。假设 $\|\nabla \frac{\delta F}{\delta \mu}(\mu)(x)\| \lesssim 1 + |x|$。那么 F的WOP 梯度，$\nabla_{\mathsf{WOP}} F(\mu) = (u_F(\mu)\ m_F'(\mu))$ 满足

$$m_F'(\mu) = \int \frac{\delta F}{\delta \mu}(\mu) \mathrm{d}\bar{\mu} - \int \langle \nabla \frac{\delta F}{\delta \mu}(\mu), (x - x_0) \rangle \mathrm{d}\bar{\mu}, \quad (15)$$

$$u_F(\mu) = \frac{1}{m_\mu} \left[\nabla \frac{\delta F}{\delta \mu}(\mu) - m_F'(\mu)(x - x_0) \right]. \quad (16)$$

特别地， F 的梯度流$(\nu_t)^{t\geq 0}$ 满足弱偏微分方程——即针对 $\mathcal{C}_c^\infty($**实数集**$^d \times$**实数集**$_+^\star)$中的测试函数，

$$\partial_t \nu_t = -\frac{m_F'(\nu_t)}{m_{\nu_t}} \nu_t + \mathrm{div}\big(u_F(\nu_t)\nu_t\big). \quad (17)$$

证明。 F 在 μ 处的梯度定义为满足链式法则的量 $\nabla_{\mathsf{WOP}} F(\mu) \in T_\mu\mathcal{P}_2($**实数集**$^d)$

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\mu_t)|_{t=0}, = \langle \nabla_{\mathsf{WOP}} F(\mu), (u, m') \rangle_\mu,$$

其中， μ_t 是时刻 t处的测地线，它从 μ 出发，沿由(u, m')表示的切向量前进，即满足（11）。事实上，得益于假设（14），只需对测地线验证（13）即可。再次利用关于一阶变分的假设（14），我们有

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}F(\mu_t) &= \lim_{t\to 0}\frac{F(\mu_t)-F(\mu)}{t}\\ &= \lim_{t\to 0}\frac{1}{t}\int\frac{\delta F}{\delta\mu}(\mu)\mathrm{d}(\mu_t-\mu))\\ &= \lim_{t\to 0}\frac{m_{\mu_t}-m_\mu}{t}\int\frac{\delta F}{\delta\mu}(\mu)\mathrm{d}\bar{\mu}+m_{\mu_t}\int\frac{\delta F}{\delta\mu}(\mu)\frac{1}{t}\mathrm{d}(\bar{\mu}_t-\bar{\mu}).\end{aligned}$$

第二个不等式利用了一阶变分的假设 (14)，以及对测地线有 $\mathrm{WOP}(\mu,\mu_t)/t\to 0$ as $t\to 0$ 这一事实。对于趋于 0 的 t，项 $\frac{m_{\mu_t}-m_\mu}{t}$ 收敛到 m'。对于第二项，由 WOP 的拓扑（性质 9）可知，m_{μ_t} 收敛到 m_μ。接着，注意到

$$\lim_{t\to 0}\int\frac{\delta F}{\delta\mu}(\mu)\frac{1}{t}\mathrm{d}(\bar{\mu}_t-\bar{\mu})=\frac{\mathrm{d}}{\mathrm{d}t}\int\frac{\delta F}{\delta\mu}(\mu)(x+tu)\mathrm{d}\bar{\mu}(x)\Big|_{t=0}$$

利用假设 $\nabla\frac{\delta F}{\delta\mu}(\mu)$ 具有线性增长 $|\nabla\frac{\delta F}{\delta\mu}(\mu)(x)|\lesssim 1+|x|$，我们可以在 $t=0$ 的邻域 $[0,\varepsilon[$ 内对以下导数进行估计，

$$\left|\frac{\partial}{\partial t}\left(\frac{\delta F}{\delta\mu}(\mu)(x+tu)\right)\right|=\left|\langle\nabla\frac{\delta F}{\delta\mu}(x+tu),u\rangle\right|\le 2\left|\nabla\frac{\delta F}{\delta\mu}(\mu)(x+tu)\right|^2+2|u|^2\lesssim 1+|x|^2+(1+\varepsilon)|u|^2$$

由于 u 是in $L^2(\bar{\mu})$，且 $\bar{\mu}$ 有二阶矩，莱布尼茨积分法则保证 s

$$\lim_{t\to 0}\int\frac{\delta F}{\delta\mu}(\mu)\frac{1}{t}\mathrm{d}(\bar{\mu}_t-\bar{\mu})=\int\langle\nabla\frac{\delta F}{\delta\mu}(x+tu),u\rangle\mathrm{d}\bar{\mu}\Big|_{t=0}=\int\langle\nabla\frac{\delta F}{\delta\mu},u\rangle\mathrm{d}\bar{\mu}.$$

总而言之，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}F(\mu_t)|_{t=0} &= m'\int\frac{\delta F}{\delta\mu}\mathrm{d}\bar{\mu}+m_\mu\int\langle\nabla\frac{\delta F}{\delta\mu}(x),u\rangle\mathrm{d}\bar{\mu}\\ &= m'\left[\int\frac{\delta F}{\delta\mu}\mathrm{d}\bar{\mu}-\int\langle\nabla\frac{\delta F}{\delta\mu},(x-x_0)\rangle\mathrm{d}\bar{\mu}\right]+\int\langle m_\mu u+m'(x-x_0),\nabla\frac{\delta F}{\delta\mu}\rangle\mathrm{d}\bar{\mu}.\end{aligned}$$

利用标量积 (12)，我们将 WOP 梯度的质量变分识别为

$$m'_F(\mu)=\int\frac{\delta F}{\delta\mu}\mathrm{d}\bar{\mu}-\int\langle\nabla\frac{\delta F}{\delta\mu},(x-x_0)\rangle\mathrm{d}\bar{\mu},$$

向量场 $u_F(\mu)$ 定义为

$$\nabla\frac{\delta F}{\delta\mu}(\mu)(x)=m_\mu u_F(\mu)(x)+m'_F(\mu)(x-x_0)$$

i.e.

$$u_F(\mu)(x)=\frac{1}{m_\mu}\nabla\frac{\delta F}{\delta\mu}(\mu)(x)-\frac{m'_F(\mu)}{m_\mu}(x-x_0).$$

□

现在，我们沿用图2中对 WOP度量的相同直观解释，并描述一类泛函的直观特征：对于这些泛函，其梯度流满足守恒偏微分方程，即总质量 m_{ν_t} 保持不变。图3提示，泛函 F的梯度必须与半直线正交$\{aT_{\bar{1}}\#\bar{\mu}|a>0\}$，对每个测度$_a\mu$都成立。这等价于 F 满足质量不变方程

$$F(\mu)=F(aT_{\frac{1}{a}\#\mu}),\quad \forall a>0,\ \forall \mu\in\mathcal{M}(\mathbf{R}^d). \tag{18}$$

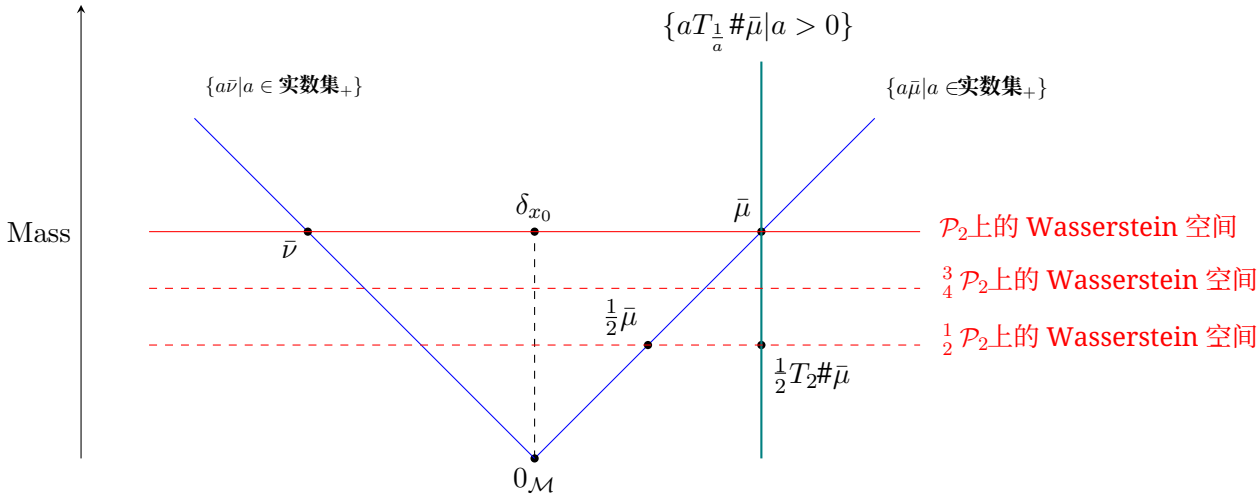


图3：借鉴图2的类比，我们着重说明恒质量平面（即与 $\mathcal{P}_2$ 成比例）和半直线 $\{aT_{\bar{1}}\#\bar{\mu}|a>0\}$.$_a$ 之间存在某种正交关系的直观理解。

事实上，以下结果成立。

定理16 (守恒泛函的刻画)。 设 F 是定义在 $\mathcal{M}$(**实数集**^{d}) 上且满足 (14) 的泛函。回顾一下，在 μ 处， F 的 WOP 梯度的总质量的变化由 $m'_F(\mu)=\int\frac{\delta F}{\delta\mu}\mathrm{d}\bar{\mu}-\int\langle\nabla\frac{\delta F}{\delta\mu},(x-x_0)\rangle\mathrm{d}\bar{\mu}$ 给出。设 ν_t 是 F 的一个梯度流（即满足方程 (17) ），且始于 ν_0，则以下断言等价。

1. $F(\mu)=F(aT_{\frac{1}{a}\#\mu}),\quad \forall\mu\in\mathcal{M}(\mathbf{R}^d)\setminus\{O_{\mathcal{M}}\},\ \forall a>0$
2. $m'_F(\mu)=0,\quad \forall\mu\in\mathcal{M}$(**实数集**^{d}) $\setminus\{O_{\mathcal{M}}\}$
3. 对所有 $t>0$ 和所有 ν_0， m_{ν_t} 为常数。

证明。 设 $\mu_a=aT_{\bar{1}}\#\mu$。我们首先证明断言 1. 和 2.等价。条目 1.$_a$ 等价于 $\mathrm{d}\,\mathrm{d}a^F(\mu_a)=0$，并利用一阶变分，对于$-\varepsilon>-a$,

$$\frac{F(\mu_{a+\varepsilon})-F(\mu_a)}{\varepsilon}=\underbrace{\frac{1}{\varepsilon}\int\frac{\delta F}{\delta\mu_a}\mathrm{d}(\mu_{a+\varepsilon}-\mu_a)}_{(A)}+o\left(\underbrace{\frac{1}{\varepsilon}\mathsf{WOP}(\mu_{a+\varepsilon},\mu_a)}_{(B)}\right)$$

我们首先计算这一项（B）

$$\frac{\mathsf{WOP}(\mu_{a+\varepsilon},\mu)}{\varepsilon}=\sqrt{\frac{((a+\varepsilon)-a)^2}{\varepsilon^2}+\frac{W_2^2(\bar{\mu},\bar{\mu})}{\varepsilon^2}}=1$$

现在对于这一项（A），采用与定理 15 的证明类似的论证，

$$\begin{aligned}
(A) =& \left(\frac{(a+\varepsilon)-a}{\varepsilon}\right)\int \frac{\delta F}{\delta \mu_a}\mathrm{d}T_{\frac{1}{a}}\#\bar{\mu} + (a+\varepsilon)\int \frac{\delta F}{\delta \mu_a}\frac{1}{\varepsilon}\mathrm{d}(T_{\frac{1}{a+\varepsilon}}\#\bar{\mu} - T_{\frac{1}{a}}\#\bar{\mu}) \\
=& \int \frac{\delta F}{\delta \mu_a}\mathrm{d}T_{\frac{1}{a}}\#\bar{\mu} + (a+\varepsilon)\int \frac{1}{\varepsilon}\left(\frac{\delta F}{\delta \mu_a}\big(T_{\frac{1}{a+\varepsilon}}\big) - \frac{\delta F}{\delta \mu_a}\big(T_{\frac{1}{a}}\big)\right)\mathrm{d}\bar{\mu} \\
\underset{\varepsilon\to 0}{\longrightarrow}& \int \frac{\delta F}{\delta \mu_a}\mathrm{d}T_{\frac{1}{a}}\#\bar{\mu} + \int \langle \big(\nabla \frac{\delta F}{\delta \mu_a}\big)\circ T_{\frac{1}{a}}, \partial_\varepsilon T_{\frac{1}{a+\varepsilon}}|_{\varepsilon=0}\rangle \mathrm{d}\bar{\mu} \\
=& \frac{1}{a}\int \frac{\delta F}{\delta \mu_a}\mathrm{d}\mu_a - \frac{1}{a^2}\int \langle \big(\nabla \frac{\delta F}{\delta \mu_a}\big)\circ T_{\frac{1}{a}}, (x-x_0)\rangle \mathrm{d}\bar{\mu} \\
=& \frac{1}{a}\left[\int \frac{\delta F}{\delta \mu_a}\mathrm{d}\mu_a - \int \langle \nabla \frac{\delta F}{\delta \mu_a}, (x-x_0)\rangle \mathrm{d}\mu_a\right] \\
=& \frac{m_{\mu_a}}{a} m'_F(\mu_a)
\end{aligned}$$

因此，对所有 $\mu \in \mathcal{M}$(实数集^d) 以及所有 $a>0$， $F(\mu_a) = F(\mu)$，等价于 $m'_F(\mu_a) = 0$，$\forall \mu \in \mathcal{M}$(实数集^d) $\setminus \{O_{\mathcal{M}}\}$，$\forall a>0$，而这又等价于 $m'_F(\mu) = 0$，$\forall \mu \in \mathcal{M}$(实数集^d) $\setminus \{O_{\mathcal{M}}\}$。

现在我们证明 2. 与 3.的等价性。给定 WOP梯度 $\nabla_{\mathsf{WOP}} F(\mu) = \big(m'_F(\mu)\ u_F(\mu)\big)$，梯度流方程由弱偏微分方程给出，并在集合 $\mathcal{C}_c^\infty$(实数集$^d\times]0,\ \infty[$) 上积分

$$\partial_t \nu_t = -\frac{m'_F(\nu_t)}{m_{\nu_t}}\nu_t + \mathrm{div}\big(u_F(\nu_t)\nu_t\big).$$

我们证明 m_{ν_t} 弱可微，且其弱导数为 $-m'_F(\nu_t)$。事实上，设 $(\phi_n(x))_n$ 是 $\mathcal{C}_c^\infty$(实数集^d) 中满足以下条件的函数序列

$$\begin{aligned}
&\phi_n(x) \underset{n\to+\infty}{\longrightarrow} 1 && \sup_{n\in\mathbb{N}} \|\phi_n\|_\infty < +\infty \\
&\nabla\phi_n(x) \underset{n\to+\infty}{\longrightarrow} 0 && \sup_{n\in\mathbb{N}} \|\nabla\phi_n\|_\infty < +\infty
\end{aligned}$$

令 $\psi(t) \in \mathcal{C}_c^\infty$(实数集$_+^\star$)，则 $\phi_n(x)\psi(t) \in \mathcal{C}_c^\infty$(实数集$^d\times$实数集$_+^\star$)，并将梯度流方程对这组函数积分，得到：

$$\int_{\mathbf{R}^d\times\mathbf{R}_+^\star} \phi_n\psi\partial_t\nu_t\mathrm{d}t = -\int_{\mathbf{R}^d\times\mathbf{R}_+^\star} \phi_n\psi\frac{m'_F(\nu_t)}{m_{\nu_t}}\mathrm{d}\nu_t\mathrm{d}t + \int_{\mathbf{R}^d\times\mathbf{R}^\star} \phi_n\psi\mathrm{div}\big(u_F(\nu_t)\nu_t\big)$$

或者等价地，

$$-\int_{\mathbf{R}^\star}\partial_t\psi\left[\int\phi_n\mathrm{d}\nu_t\right]\mathrm{d}t = -\int_{\mathbf{R}^\star}\psi m'_F(\nu_t)\left[\int\phi_n(x)\mathrm{d}\bar{\nu}_t\right]\mathrm{d}t - \int_{\mathbf{R}^\star}\psi\left[\int\nabla\phi_n u_F(\nu_t)\mathrm{d}\nu_t\right]\mathrm{d}t.$$

利用控制收敛定理，我们得到

$$-\int_{\mathbf{R}^\star}\partial_t\psi\ m_{\nu_t}\mathrm{d}t = -\int_{\mathbf{R}^\star}\psi m'_F(\nu_t)\mathrm{d}t.$$

利用 ν_t 的连续性，我们得出第三个等价关系。 □

这一结果使我们能够直接将定义在 $\mathcal{P}_2$(实数集^d)上的泛函扩展为定义在 $\mathcal{M}$(实数集^d)上的泛函，并保留 W_2 流的梯度。

推论17（泛函的扩展 $\mathcal{P}_2$）。设 F 是定义在 $\mathcal{P}_2$（**实数集**^{d}）上的泛函，我们在 $\mathcal{M}$（**实数集**^{d}）上定义如下泛函

$$\tilde{F}(\mu) = F(T_{m_\mu} \# \bar{\mu}).$$

那么，$\tilde{F}$ 与 F 在 $\mathcal{P}_2$(**实数集**^{d})上重合；$\tilde{F}$ 的WOP 梯度流具有恒定的质量，并且对于概率测度，它与 F 的 W_2 梯度流重合。

例如，定义在 $\mathcal{P}_2$**(实数集**^{d}**)**上的玻尔兹曼熵泛函为

$$E(\mu) = \begin{cases} \int \log(\mu) \mathrm{d}\mu & \text{if } \mu \ll dx \\ +\infty & \text{otherwise} \end{cases} \quad \forall \mu \in \mathcal{P}_2(\mathbf{R}^d),$$

E 关于 W_2 度量的梯度流是热方程 [JKO98]

$$\partial_t \nu_t = \Delta \nu_t. \tag{19}$$

WOP度量下玻尔兹曼熵的扩展为

$$\tilde{E}(\mu) = \begin{cases} \int \log(\bar{\mu}) \mathrm{d}\bar{\mu} - d\log(m_\mu) & \text{if } \mu \ll dx \\ +\infty & \text{otherwise} \end{cases} \quad \forall \mu \in \mathcal{M}(\mathbf{R}^d),$$

˜ E 的WOP梯度流满足以下方程

$$\partial_t \nu_t = \frac{1}{m_{\nu_t}^2} \Delta \nu_t, \tag{20}$$

当梯度流 μ_0 的初始测度是概率测度时，该方程与热方程（19）重合。

以下推论汇总了泛函及其 WOP梯度的其他示例。

推论18。将 F 在 μ 处的WOP 梯度记为

$$\nabla_{\mathsf{WOP}} F(\mu) = (u_F, m'_F)(\mu)$$

对于 $F(\mu) = m_\mu$，梯度为

$$m'_F = 1 \quad and \quad u_F(x) = -\frac{x - x_0}{m_\mu}.$$

对于满足以下条件的 F：对所有 μ，有 $F(\mu) = F(\bar{\mu})$，其梯度满足

$$m'_F = -\frac{1}{m_\mu} \int \langle \nabla \frac{\delta F(\bar{\mu})}{\delta \bar{\mu}}, (x - x_0) \rangle \mathrm{d}\bar{\mu} \quad and \quad u_F(x) = \frac{1}{m_\mu} \left[\nabla \frac{\delta F(\mu)}{\delta \mu} - m'(x - x_0) \right].$$

特别地，如果 $F(\mu) = M_{x_0}(\bar{\mu})/2$，则

$$m'_F = -\frac{M_{x_0}(\bar{\mu})}{m_\mu} \quad and \quad u_F(x) = \frac{1}{m_\mu^2} \left[(1 + M_{x_0}(\bar{\mu}))(x - x_0) \right].$$

对于 $F(\mu) = m_\mu^2 (1 + M_{x_0}(\bar{\mu}))/2$，利用 $\nabla(fg) = f\nabla g + g\nabla f$，我们有

$$m'_F = m_\mu \quad and \quad u_F(x) = 0.$$

同样，$F(\mu) = m_\mu^2 M_{x_0}(\bar{\mu})/2$ 的梯度为

$$m'_F = 0 \quad and \quad u_F(x) = x - x_0.$$

4 重心

现在我们转而讨论空间 ($\mathcal{M}_K$(**实数集**^{d})，WOP) 中的重心。回顾一下，对于定义在 ($\mathcal{M}$(**实数集**^{d})，WOP) 上的概率测度$\mathbb{P}=\sum_{i=1}^n \lambda_i\delta_{\mu_i}$，其重心 $\mu_B\in\mathcal{M}$(**实数集**^{d}) 定义为以下函数的极小化元:

$$\mu\mapsto V(\mu):=\sum_i \lambda_i \mathsf{WOP}^2_{x_0}(\mu,\mu_i)=\int \mathsf{WOP}^2(\mu,\tilde{\mu})\mathrm{d}\mathbb{P}(\tilde{\mu}).$$

由于 WOP 扩展了 Wasserstein距离，因此可以预期，Wasserstein距离与 WOP 的重心之间存在简单的关系。以下结果阐述了这一点。

定理19 (重心**)**。 设 $\mathbb{P}=\sum_i \lambda_i\delta_{\mu_i}\in\mathcal{P}(\mathcal{M}$(**实数集**^{d})$)$。那么，它在 μ_B *on* ($\mathcal{M}$(**实数集**^{d}),WOP_{x_0}) 上的重心质量为 $m_{\mu_B}=\sum_i \lambda_i m_{\mu_i}$，且 $\bar{\mu}_B$ 是 $\sum_i \lambda_i m_{\mu_i}\delta_{\bar{\mu}_i}$ 的 ***Wasserstein*重心**。特别地，尽管 $\mathsf{WOP}(\mu_0,\mu_1)$ 的值取决于 x_0，但 μ_0 与 μ_1 之间的测地线并不取决于 x_0。

该空间中的重心不依赖于 x_0，且形式简单。

定理20 (重心**)**。 设 $\mathbb{P}=\sum_i \lambda_i\delta_{\mu_i}\in\mathcal{P}(\mathcal{M}$(**实数集**^{d})$)$。其重心 $\mu_B\in\mathcal{M}$(**实数集**^{d}) 定义为下式的极小化元

$$\mu\mapsto V(\mu):=\sum_i \lambda_i \mathsf{WOP}^2(\mu,\mu_i)=\int \mathsf{WOP}^2(\mu,\tilde{\mu})\mathrm{d}\mathbb{P}(\tilde{\mu}).$$

那么，μ_B 不依赖于 x_0。此外，$m_{\mu_B}=\sum_i \lambda_i m_{\mu_i}$ 和 $\bar{\mu}_B$ 是 $\sum_i \lambda_i m_{\mu_i}\delta_{\bar{\mu}_i}$的 ***Wasserstein*重心**。

证明。 利用（7），重心 μ_B 因此使下式最小化

$$\mu\mapsto \sum_i \lambda_i(m_\mu-m_{\mu_i})^2+\sum_i \lambda_i(m_\mu-m_{\mu_i})\big(m_\mu M_{x_0}(\bar{\mu})-m_{\mu_i}M_{x_0}(\bar{\mu}_i)\big)+m_\mu\sum_i \lambda_i m_{\mu_i}W_2^2(\bar{\mu},\bar{\mu}_i).$$

令 $\tilde{\lambda}_i=\lambda_i m_{\mu_i}$表示 $i=1$、.....、n、 $\tilde{\lambda}_0=m_\mu-\sum_{i=1}^n\lambda_i$和

$$\bar{x}=\frac{\sum_{i=0}^n \tilde{\lambda}_i x_i}{\sum_{i=0}^n \tilde{\lambda}_i}=\sum_{i=0}^n \frac{\tilde{\lambda}_i x_i}{m_\mu}$$

将通过拼接 $\bar{\mu}$与$\bar{\mu}_i$之间的最优耦合得到的 $\bar{\mu}$、 $\bar{\mu}_1$、.....、 $\bar{\mu}_n$的耦合记为 γ。那么，

$$\begin{aligned}\sum_i \lambda_i\big(m_{\mu_i}W_2^2(\bar{\mu},\bar{\mu}_i)+(m_\mu-m_{\mu_i})M_{x_0}(\bar{\mu})\big)&=\int\sum_{i=1}^n \tilde{\lambda}_i|x_i-x|^2+\tilde{\lambda}_0|x_0-x|^2\mathrm{d}\gamma(x,x_1,\ldots,x_n)\\&\geq\int\sum_{i=1}^n \tilde{\lambda}_i|x_i-\bar{x}|^2+\tilde{\lambda}_0|x_0-\bar{x}|^2\mathrm{d}\gamma(x,x_1,\ldots,x_n)\\&=\sum_i \lambda_i\big(m_{\mu_i}W_2^2(\bar{\mu},\bar{\mu}_i)+(m_\mu-m_{\mu_i})M_{x_0}(\bar{\mu})\big)\end{aligned}$$

现在，利用 $\sum_{i=0}^n \frac{\lambda_i}{m_\mu}|x_i-\bar{x}|^2=\sum_{i=1}^n \frac{\lambda_i}{m_\mu}|x_i-x_0|^2-|\bar{x}-x_0|^2$，我们得到

$$\sum_i \lambda_i m_{\mu_i}W_2^2(\bar{\mu},\bar{\mu}_i)+(m_\mu-\sum_i \lambda_i m_{\mu_i})M_{x_0}(\bar{\mu})=\sum_i \lambda_i m_{\mu_i}M_{x_0}(\bar{\mu}_i)-m_\mu M_{x_0}(\bar{\mu})$$

最后，

$$m_\mu^2 M_{x_0}(\bar{\mu}) = m_\mu^2 \int |\bar{x} - x_0|^2 \mathrm{d}\gamma(x, x_1, \ldots, x_n)$$
$$= \int |\sum_{i=1}^n \tilde{\lambda}_i (x_i - x_0)|^2 \mathrm{d}\gamma(x, x_1, \ldots, x_n)$$

因此，对于 $\bar{\mu}$ 固定时，使 $V(\mu)$ 最小的 m_μ 满足 $m_\mu = \sum_i \lambda_i m_{\mu_i}$。事实上

$$\mathrm{d},$$

$$V(\mu) = \sum_i \lambda_i (m_\mu - m_{\mu_i})^2 - \sum_i \lambda_i (m_\mu - m_{\mu_i}) m_{\mu_i} M_{x_0}(\bar{\mu}_i) + m\big(\sum_i \lambda_i m_{\mu_i} M_{x_0}(\bar{\mu}_i) - m_\mu M_{x_0}(\bar{\mu})\big)$$
$$= \sum_i \lambda_i (m_\mu - m_{\mu_i})^2 - m_\mu^2 M_{x_0}(\bar{\mu}) + \sum_i \lambda_i m_{\mu_i}^2 M_{x_0}(\bar{\mu}_i)$$
$$= \sum_i \lambda_i (m_\mu - m_{\mu_i})^2 \underbrace{- \int |\sum_{i=1}^n \tilde{\lambda}_i (x_i - x_0)|^2 \mathrm{d}\gamma(x, x_1, \ldots, x_n) + \sum_i \lambda_i m_{\mu_i}^2 M_{x_0}(\bar{\mu}_i)}_{\text{does not depend on } m_\mu}.$$

对于证明的最后一个断言，注意到 $V(\mu)$ 仅通过 $\int |\sum_{i=1}^n \tilde{\lambda}_i (x_i - x_0)|^2 \mathrm{d}\gamma(x, x_1, \ldots, x_n)$ 依赖于 $\bar{\mu}$，其最优耦合不依赖于 x_0。 □

5 进一步的发展

本节研究扩展 2-瓦瑟斯坦度量时的一些局限性，以及熵-输运形式不可能扩展 2-瓦瑟斯坦度量的问题。我们还简要介绍一种将 WOP度量扩展到任意指数 $p \geq 1$ 的尝试。以下性质是 $\mathcal{M}$(**实数集**^{d}) 上任何概括 Wasserstein度量的度量都存在的局限性。

不可能性结果 我们从观察零测度与概率测度空间之间的距离开始。

任意测度 $\mathcal{P}_2$(**实数集**^{d}) 到零测度 $0_{\mathcal{M}}$ 的 WOP 距离并非常数。事实上，对于狄拉克测度 δ_x，

$$\mathsf{WOP}(\delta_x, 0_{\mathcal{M}}) = \sqrt{1 + |x - x_0|^2}.$$

性质 21 ($\mathcal{P}$(**实数集**^{d}) 与零测度之间的距离)。 在 $\mathcal{M}(\mathcal{X})$ 上不存在满足以下性质的度量 $D(\mu, \nu)$，其中 $\mathcal{X} \subset$**实数集**^{d} 同时满足以下性质。

1. D 与 $\mathcal{P}_2(\mathcal{X})$ 上的 2-瓦瑟斯坦度量一致，即对所有$(\mu, \nu) \in \mathcal{P}_2(\mathcal{X})^2$,, $D(\mu, \nu) = W_2(\mu, \nu)$
2. $D(0, \mu)$ 有界，即存在 $\alpha > 0$ 使得 $D(0, \mu) \leq \alpha$, $\forall \mu \in \mathcal{P}(\mathcal{X})$,
3. $\mathcal{X}$ 无界

因此，若要推广 Wasserstein度量，例如应用于高斯测度，就必须放弃 $\mathcal{P}^2$(**实数集**^{d})中测度到零测度的等距性。相比之下，HK 度量不满足第 1 点，但满足第 2 点和第 3 点。

证明。令 $\mathcal{X} \subset$ **实数集**^{d}，并假设 $(\mathcal{M}(\mathcal{X}), D)$ 是满足泛化条件 (1) 的度量空间。那么，对于任意测度 $(\mu, \nu) \in \mathcal{P}_2(\mathcal{X})^2$，利用三角不等式可得：

$$W_2(\mu,\nu) = D(\mu,\nu) \leq D(\mu, 0_{\mathcal{M}}) + D(0_{\mathcal{M}}, \nu) \leq 2\alpha.$$

这个关于 W_2 的有界条件等价于 $\mathrm{diam}(\mathcal{X}) \leq 2\alpha$。事实上，对于蕴含关系 ($\Rightarrow$)，我们可以取 $\mu = \delta_x$ 和 $\nu = \delta_y$，使得 $(x, y) \in \mathcal{X}^2$，于是有：

$$W_2(\delta_x, \delta_y) = |x - y| \leq 2\alpha$$

蕴含关系 ($\Leftarrow$) 可证明如下：

$$W_2(\mu,\nu) = \sqrt{\inf_{\pi\in\Pi(\mu,\nu)} \int |x-y|^2 \mathrm{d}\pi} \leq \sqrt{\inf_{\pi\in\Pi(\mu,\nu)} \int \mathrm{diam}(\mathcal{X})^2 \mathrm{d}\pi} = \mathrm{diam}(\mathcal{X}) \leq 2\alpha.$$

因此，$\mathcal{X}$ 有界，这与我们的假设矛盾。 □

这里，我们在熵-输运表述这一宽泛框架之外，论证我们对度量的推广。

我们先回顾非常一般情形下 Csiszar f-散度和 ET 表述的定义。设 f: **实数集**$_+ \longrightarrow [0, +\infty]$ 是凸的下半连续函数，且 $f(1) = 0$，定义 $f'_\infty(1) = \lim_{t\to+\infty} {}^{f(t)}\, t$。设 μ，$\nu \in \mathcal{M}$(**实数集**^{d})，μ 关于 ν 的勒贝格分解为 $\mu = \frac{d\mu}{d\nu}\nu + \nu^\perp$。则 f-散度定义为

$$D_f(\mu|\nu) = \int f(\frac{d\mu}{d\nu})\mathrm{d}\nu + f'_\infty(1).m_{\nu^\perp},$$

约定 $+\infty.0 = 0$。现在，为了在一般情形下定义 ET，设 f：**实数集**$_+ \to [0, +\infty]$，使得 f 是凸、下半连续的，且满足 $f(1) = 0$；设 $c(x, y)$：**实数集**$^d \times$ **实数集**$^d \to [0, +\infty]$ 为下半连续函数，且满足 $c(x, x) = 0$。那么，熵传输定义为

$$\mathsf{ET}(\mu,\nu) = \inf_{\gamma\in\mathcal{M}(\mathbf{R}^d\times\mathbf{R}^d)} \left\{ D_f(P_0\#\gamma|\mu) + D_f(P_1\#\gamma|\nu) + \int c(x,y)\mathrm{d}\gamma(x,y) \right\},$$

其中，$P_0\#\gamma$ 是 γ 的第一边缘，$P_1\#\gamma$ 是 γ的第二边缘。

定理 22（ET度量无法扩展 W_2）。如果 $\min\left(f(0)\ f'_\infty(1)\right) < +\infty$，那么 ET不可能成为一个将 W_2 ^{d} $(f\ 0\ \ f'_\infty\ 1\ ) =$扩展到 **实** $+\infty$ $\mu\ 0_{\mathcal{M}} = +\infty$ $\mu \in \mathcal{M}\ \ ^d \setminus \{0_{\mathcal{M}}\}$ **数集**上正测度的度量。反之，如果 $\min()()$，那么对于所有 (**实数集**)，ET(,)。

证明. 对于 $\mu \in \mathcal{P}_2$(**实数集**^{d})，我们有如下界：

$$\mathsf{ET}(\mu, 0_{\mathcal{M}}) \leq \min\left(f(0), f'_\infty(1)\right). \tag{21}$$

事实上，取次优运输计划 $\gamma = 0_{\mathcal{M}}$，可得

$$\mathsf{ET}(\mu, 0_{\mathcal{M}}) \leq D_f(0_{\mathcal{M}}, \mu) + D_f(0_{\mathcal{M}}, 0_{\mathcal{M}}) + \int c(x,y)\mathrm{d}0_{\mathcal{M}} = f(0).m_\mu = f(0)$$

而取次优输运计划 $\gamma = (Id,\ Id)\#\mu$，可得

$$\mathsf{ET}(\mu, 0_{\mathcal{M}}) \leq D_f(\mu,\mu) + D_f(\mu, 0_{\mathcal{M}}) + \int c(x,x)\mathrm{d}\mu = f(1).m_\mu + f(0).0 + f'_\infty(1).m_\mu = f'_\infty(1).$$

结合定理 21 使用我们的界（21），我们可以得出结论：基于熵-输运形式的度量都无法将 W_2 推广到**实数集**^{d}上的测度。 □

推广 W_p 本节最后讨论如何将这一思想推广到 W_p 度量。我们可以如下定义 $p \geq 1$ 的 WOP$_p$ 度量，

$$\mathsf{WOP}_p^p(\mu,\nu) = |m_\mu - m_\nu|^p + W_p^p(T_{m_\mu}\#\bar{\mu}, T_{m_\nu}\#\bar{\nu})$$

并且，空间 $\mathcal{M}_{K,p}$(**实数集**^{d}) 是由正测度构成的集合，其中在 x_0 处的 p 阶矩，记作 $M_{x_0,p}(\mu)$，满足以下不等式

$$M_{x_0,p}(\mu) \leq K m_\mu$$

那么 WOP$_p$ 满足 WOP 所满足的几乎所有性质：

- WOP$_p$ 是 $\mathcal{M}$(**实数集**^{d}) 上的一个度量
- WOP$_p$ 推广了 p-Wasserstein
- ($\mathcal{M}_{K,p}$(**实数集**^{d}),WOP$_p$) 对所有 $K > 0$ 都是完备的
- WOP$_p$ 将 $\mathcal{M}_{K,p}$(**实数集**^{d}) 中的弱收敛 $+$ 阶矩的收敛度量化
- WOP$_p$ 满足 1 次齐次性
- WOP$_p$ 是一种测地度量，两个测度 μ_0 和 μ_1 之间的测地线为：

$$\mu_t = m_t.\eta_t$$

其中，$m_t = (1-t)m_{\mu_0} + t m_{\mu_1}$ 和 η_t 是测度$T_{m_{\mu_0}}\#\bar{\mu}_0$ 和 $T_{m_{\mu_1}}\#\bar{\mu}_1$之间在 ($\mathcal{P}_2$(**实数集**^{d}) W_p) 中的测地线

请注意，在 $p \neq 2$ 的情况下，测地线和重心取决于参考点 x_0 的选择。

参考文献

[AC11] M. Agueh 和 G. Carlier。“Wasserstein 空间中的重心”。载于：*SIAM* 数学分析杂志 43.2（2011），第 904–924 页。[AGS05] L. Ambrosio、N. Gigli 和 G. Savar´e。梯度流：度量空间和概率测度空间中。Birkh¨auser Basel，2005 年。[ANR17] J. Altschuler、J. Niles-Weed 和 P. Rigollet。“通过 Sinkhorn 迭代实现最优传输的近线性时间近似算- 法”。载于：神经信息处理系统进展 30（2017）。[Bac+22] J. Backhoff-Veraguas、J. Fontbona、G. Rios 和 F. Tobar。“基于 Wasserstein 重心的贝叶斯学习”。载于：*ESAIM*：概率与统计 26（2022），第 436–472 页。[BB00] J.-D. Benamou 和 Y. Brenier。“Monge-Kantorovich 质量转移问题的计算流体力学解法”。载于：《数值数学》 84.3（2000），第 375–393 页。[BC19] N. Bonneel 和 D. Coeurjolly。“Spot：切片部分最优传输”。载于：*ACM* 图形学汇刊（*TOG*） 38.4（2019），第 1–13 页。

[Bon+15] N. Bonneel、J. Rabin、G. Peyr´e 和 H. Pfister。“测度的切片和 Radon Wasserstein 重- 心”。载于：《数学成像与视觉杂志》51（2015），第 22–45 页。[Bre91] Y. Brenier。“向量值函- 数的极分解和单调重排”。载于：《纯粹与应用数学通讯》44.4（1991），第 375–417 页。[Chi+18a] L. Chizat、G. Peyr´e、B. Schmitzer 和 F.-X. Vialard。“最优传输与 Fisher–Rao 度量之- 间的插值距离”。载于：《计算数学基础》18.1（2018），第 1–44 页。[Chi +18b] L. Chizat、G. Peyr´e、B. Schmitzer 和 F.-X. Vialard。“非平衡最优传输：动态和 Kantorovich 表述”。载于：《泛函分析杂志》274.11（2018），第 3090–3123 页。[Cut13] M. Cuturi。“Sinkhorn 距离：最优传输的光速计算”。载于：《神经信息处理系统进展》26（2013）。[GGV21] T. Gallou¨et、R. Ghezzi 和 F.-X. Vialard。“非平衡最优传输的正则性理论和几何”。载于：arXiv 预印本 arXiv:2112.11056（2021）。[JKO98] R. Jordan、D. Kinderlehrer 和 F. Otto。“Fokker–Planck 方程的变分表述”。载于：《SIAM 数学分析杂志》29.1（1998），第 1–17 页。[JL20] M. Jacobs 和 F. L´eger。“最优传输的快速方法：Back-and-Forth 方法”。载于：《数值数学》146.3（2020），第 513–544 页。[Kan42] L. V. Kantorovich。“关于质量迁移”。载于：《苏联科学院通报》（NS）。第 37 卷。1942 年，第 199–201 页。[Kan48] L. V. Kantorovich。“关于 Monge 的一个问题”。载于：《苏联科学院通报》（NS）。第 3 卷。1948 年，第 225–226 页。[KMV15] S. Kondratyev、L. Monsaingeon 和 D. Vorotnikov。有限 Radon 测度空间上的一种新最优传输距离。2015 年。[LGS19] H. Liu、X. Gu 和 D. Samaras。“具有二次传输代价的 Wasserstein 生成对抗网络”。载于：IEEE/CVF 国际计算机视觉会议论文集。2019 年，第 4832–4841 页。[LM15] D. Lombardi 和 E. Maitre。“用于非平衡最优传输的欧拉模型和算法”。载于：《ESAIM：数学建模与数值分析》49.6（2015），第 1717–1744 页。[LMS18] M. Liero、A. Mielke 和 G. Savar´e。“最优熵传输问题以及正测度之间一种新的 Hellinger–Kantorovich 距离”。载于：《数学发明》211.3（2018），第 969–1117 页。[Maa+15] J. Maas、M. Rumpf、C. Sch¨onlieb 和 S. Simon。“包含耗散和密度调制的图像最优传输广义模型”。载于：《ESAIM：数- 学建模与数值分析》49.6（2015），第 1745–1769 页。[Mon81] G. Monge。“关于土方与填方理´论的回忆录”。载于：《皇家科学院数学与物理回忆录》（1781），第 666–704 页。[Mro19] Y. Mroueh。“Wasserstein 风格迁移”。载于：CoRR abs/1905.12828（2019）。arXiv：1905.12828。

[PC19] G. Peyr´e 和 M. Cuturi. “计算最优传输”。载于：机器学习基础与趋势 11.5-6 (2019), 第 355–607 页。[Pon+22] A. Ponti、I. Giordani、M. Mistri、A. Candelieri 和 F. Archetti. “Wasserstein距离在分析门店网络关键绩效指标中的「非凡」效用”。载于：大数据与认知计算 6.4 (2022), 第 138 页。[PR14] B. Piccoli 和 F. Rossi. “广义Wasserstein距离及其在带源项输运方程中的应用”。载于：《理性力学与分析档案》211.1 (2014), 第 335–358 页。[Rab+12] J. Rabin、G. Peyr´e、J. Delon 和 M. Bernot. “Wasserstein重心及其在纹理混合中的应用”。载于：《计算机视觉中的尺度空间与变分方法：第三届国际会议，*SSVM 2011*，以色列 *Ein-Gedi*，*2011* 年 *5* 月 *29* 日至 *6* 月 *2* 日，修订精选论文 *3* 》. Springer. 2012, 第 435–446 页。[RFP14] J. Rabin 、 S. Ferradans 和 N. Papadakis. “基于松弛最优传输的自适应颜色迁移”。载于：*2014 IEEE*国际图像处理会议（*ICIP*）. IEEE. 2014, 第 4852–4856 页。[SA20] D. Simon 和 A. Aberdam. “自然图像的重心：用于图像变形的约束Wasserstein重心”。载于：*IEEE/CVF*计算机视觉与模式识别会议论文集. 2020, 第 7910–7919 页。[San15] F. Santambrogio. “应用数学家的最优传输”。载于：$Birk¨auser,\ NY$55.58-63 (2015), 第 94 页。[Sch+19] G. Schiebinger 等. “单细胞基因表达的最优传输分析揭示了重编程中的发育轨迹”。载于：《细胞》176.4 (2019), 第 928–943 页。[She+21] Z. Shen、J. Feydy、P. Liu、A. H. Curiale、R. San Jose Estepar、R. San Jose Estepar 和 M. Niethammer. “基于鲁棒最优传输的精确点云配准”。载于：神经信息处理系统进展34 (2021), 第 5373–5389 页。[Thi+21] A. Thibault、L. Chizat、C. Dossal 和 N. Papadakis. “用于正则化最优传输的超松弛Sinkhorn–Knopp算法”。载于：算法14.5 (2021), 第 143 页。[Vil03] C. Villani. 最优传输专题. 第 58 卷。美国数学学会，2003。