南華大學
UNIVERSITY OF SOUTH CHINA

毕业论文/设计外文翻译

题　　目	对深度学习的几何理解
学院名称	数理学院
指导老师	高有
职　　称	讲师
班　　级	信计 1802
学　　号	20184390213
学生姓名	朱柳承

2022 年 1 月 15 日

对深度学习的几何理解

摘要：这篇论文详细地给出了在深度学习中如何运用最优传输理论的相关知识的方法与技巧。这项工作引入了生成式对抗网络 (GANs) 的最优运输 (OT) 视图。自然数据集具有内在的模式，可以概括为流形分布原理：一类数据的分布接近于一个低维流形。GANs 主要完成两个任务：流形学习和概率分布转换。后者可以用经典的 OT 方法来进行。从 OT 的角度来看，生成器计算 OT 图，而鉴别器计算生成的数据分布和真实数据分布之间的瓦瑟斯坦距离；两者都可以简化为凸几何优化过程。此外，OT 理论还发现了发生器和鉴别器之间存在内在的协作关系，而不是竞争关系，以及模态崩溃的根本原因。我们还提出了一种新的生成模型，该模型使用自编码器 (AE) 进行流形学习，并使用 OT 映射进行概率分布变换。该 AE-OT 模型提高了理论的严谨性和透明度，以及计算的稳定性和效率；特别是，它消除了模式崩溃。实验结果验证了我们的假设，并证明了我们所提出的模型的优点。

关键词：生成性；对抗性；深度学习；最优传输；模式崩溃

1 介绍

生成对抗网络 (GANs) 已成为无条件图像生成的主要方法之一。当在多个数据集上进行训练时，GANs 能够产生真实和视觉上吸引人的样本。GAN 方法训练了一个无条件的生成器，从随机噪声中回归真实图像，以及一个鉴别器，测量生成的样本和真实图像之间的差异。GANs 已经得到了各种改进。生成对抗网络 (GANs) 已成为无条件图像生成的主要方法之一。当在多个数据集上进行训练时，GANs 能够产生真实和视觉上吸引人的样本。GAN 方法训练了一个无条件的生成器，从随机噪声中回归真实图像，以及一个鉴别器，测量生成的样本和真实图像之间的差异。GANs 已经得到了各种改进。

1.1 流形分布假说

差距的巨大成功可以解释为差距有效地发现真实数据集的内在结构，可以制定为流形分布假设：一个特定类的自然数据集中在低维流形嵌入在高维背景空间 [2]。

图 1 显示了 MNIST 数据库的流形结构。每个手写数字图像具有尺寸 28×28，并被视为图像空间中的一个点。MNIST 数据库集中在一个低维流形附近。通过使用 t-SNE 流形嵌入算法 [3]，MNIST 数据库被映射到一个平面域上，每个图像被映射到一个点上。代表相同数字的图像被映射到一个簇上，10 个簇被颜色编码。这表明 MNIST 数据库分布在嵌入单元立方体中的二维（2D）曲面附近。

1.2 GANs 的理论模型

图 2 为 GANs 的理论模型。真实的数据分布 v 集中在嵌入在环境空间 χ 中的流形 Σ 上。(Σ，v) 一起显示了真实数据集的内在结构。一个 GAN 模型计算一个从潜在空间 Z 到流形 Σ 的生成器映射 $g\theta$，其中 θ 表示一个深度神经网络 (DNN) 的参数。ζ 是潜在空间中的高斯分布，$g\theta$ 将 ζ 向前推到 $\mu\theta$。鉴别器计算真实数据分布 v 和生成的分布 $\mu\theta$ 之间的距离，如瓦瑟斯坦距离 Wc($\mu\theta$，v)，它相当于康

塔罗维奇的潜在 φξ

尽管 GANs 有优势，但它们也有严重的缺点。在理论上，对深度学习的基本原则的理解仍然是原始的。在实践中，GANs 的训练是棘手的，对超参数是敏感的；GANs 遭受模式崩溃。最近，Meschede 等人。[4] 研究了 9 种不同的 GAN 模型和变体，表明基于梯度下降的 GAN 优化并不总是局部收敛的。

尽管 GANs 有优势，但它们也有严重的缺点。在理论上，对深度学习的基本原则的理解仍然是原始的。在实践中，GANs 的训练是棘手的，对超参数是敏感的；GANs 遭受模式崩溃。最近，Meschede 等人。[4] 研究了 9 种不同的 GAN 模型和变体，表明基于梯度下降的 GAN 优化并不总是局部收敛的。

图 3 为生成器图 gθ=*h*□T 的分解，其中为从潜在空间到 h：Z□□ 环境空间中的数据流形 Σ，概率分布变换图 T：Z→Z。解码图 *h* 用于流形学习，图 T 用于测量传输。

1.3 最优传输的观点

OT 理论 [5] 研究了以最经济的方式将一种概率分布转化为另一种概率分布的问题。OT 提供了严格而强大的方法来计算最优映射，将一个概率分布转换为另一个分布，并确定它们之间的距离为 [6]。

如前所述，GANs 完成了两个主要任务：流形学习和概率分布变换。后一种任务可以直接通过 OT 方法完全完成。具体地说，在图 3 中，可以用 OT 理论计算出概率分布变换图 T。该鉴别器计算生成的数据分布与真实数据分布之间的瓦瑟斯坦距离 Wc(μθ，v)，可以用 OT 方法直接计算。

从理论的角度来看，天然气的解释使黑盒子透明的一部分，概率分布变换减少凸优化过程使用 OT 理论，解决方案的存在性和唯一性理论保证，收敛速度和近似精度充分分析。

OT 的解释也解释了模态崩溃的根本原因。根据蒙格-安培方程的正则性理论，运输映射在某些奇异集合上是不连续的。然而，DNN 只能建模连续的函数/映射。因此，目标交通映射是在 GANs 可表示的功能空间之外的。这种内在的冲突

使模式的崩溃不可避免。

OT 解释也揭示了发生器和鉴别器之间更为复杂的关系。在当前的 GAN 模型中，生成器和鉴别器相互竞争，而不共享中间的计算结果。OT 理论表明，在 l2 代价函数下，生成器的最优解和鉴别器的最优解可以用封闭的形式相互表示。因此，应该用协作来代替发生器和鉴别器之间的竞争，并应该共享中间的计算结果，以提高效率。

1.4 自动编码器-最优运输模型

为了减少 GANs 的训练难度，特别是为了避免模式崩溃，我们提出了一个基于 OT 理论的更简单的生成模型：一个自动编码器 (AE)-OT 模型，如图 4 所示。

如前所述，生成模型的两个主要任务是流形学习和概率分布变换。AE 计算编码映射，$f\theta$：$Z\rightarrow\Sigma$，和解码映射 $g\xi$：$\Sigma\rightarrow Z$，用于流形学习。OT 图，T：$Z\rightarrow Z$，将白噪声 ζ 转换为由编码图推送转发的数据分布，$(f\theta)\#v$。

AE-OT 模型有很多优点。从理论的角度来看，OT 理论已经得到了很好的建立和充分的理解。通过对解码图和 OT 图进行解耦，可以提高生成模型的理论严谨性，使黑盒的一部分透明。在实际应用中，将 OT 映射简化为一个凸优化问题，保证了解的存在性和唯一性，使训练过程不会陷入局部最优状态。与 OT 映射相关的凸能量具有显式的黑森矩阵；因此，优化可以采用二阶收敛的牛顿方法，或采用超线性收敛的拟牛顿方法进行优化。相比之下，目前的生成模型基于线性收敛的梯度下降法；未知数等于训练样本，以避免过参数化问题；蒙特卡罗方法中的采样密度可以完全控制 OT 映射的误差界；具有自适应的层次算法进一步提高了效率；并行 OT 映射算法可以使用图形处理单元 (GPU) 实现。最重要的是，AE-OT 模型可以消除模式崩溃。

1.5 贡献

本工作利用 OT 理论解释了 GAN 模型。GANs 可以完成两个主要任务：流形学习和概率分布变换。后一种任务可以使用 OT 方法来完成。生成器计算一个 OT 图，而鉴别器计算生成的数据分布和真实数据分布之间的瓦瑟斯坦距离。利用布雷尼尔的理论，发电机和鉴别器之间的竞争可以被协作所取代；根据蒙格-安培方程的正则性理论，运输图的不连续导致了模态崩溃。我们进一步提出了利用 AE-OT 模型来解耦流形学习和概率分布变换，使部分黑箱透明，提高了训练效率，防止了模式崩溃。实验结果证明了该方法的有效性和有效性。

本文的组织结构如下：第 2 节简要回顾了 OT 和 GANs 中最相关的工作；第 3 节简要介绍了 OT 的基本理论和蒙格-安培方程的正则性理论；第 4 节介绍了计算 OT 的变分框架；第 5 节从 OT 的角度分析了发生器与鉴别器之间的协作（而不是竞争）关系，揭示了模式崩溃的内在原因；第 6 节报告了实验结果；论文在第 7 节中总结。

2 以前的工作

2.1 最优传输

OT 问题在各种领域中都起着重要的作用。有关详细的概述，我们可以参考读者的参考文献。[7] 和 [8]。

当输入域和输出域都是狄拉克质量时，OT 问题可以作为一个标准的线性规划 (LP) 任务来处理。为了将这个问题扩展到一个大的数据集，Ref 的作者。[9] 在原来的 LP 问题中添加了一个熵正则化器；因此，用辛角算法可以快速计算正则化问题。所罗门等人 [10] 随后通过引入快速卷积，提高了计算效率。

求解 OT 问题的第二种方法是通过 OT 问题和凸几何之间的联系，通过最小化凸能量 [6] 来计算连续测度和点测度之间的 OT 映射。在参考文献中。[11]，作者然后通过勒让德对偶理论将凸几何观察的 OT 与坎托洛维奇对偶性联系起来。

该方法是对该方法在高维空间上的推广。如果输入和输出都是连续密度，则求解 OT 问题等价于求解著名的蒙格-安培方程，这是一个高度非线性的椭圆型偏微分方程 (PDE)。有了一个额外的虚拟时间维数，这个问题可以通过计算流体动力学 [12-14] 来放宽。

2.2 生成模型

在机器学习领域，能够生成复杂和高维数据的生成模型，最近正变得越来越重要和流行。具体来说，生成模型主要用于从给定的图像数据集中生成新的图像。一些方法，包括深度信念网络 [15] 和深度玻尔兹曼机器 [16] 已经在早期阶段被引入。然而，这些方法的训练通常是棘手的和低效的。后来，从变分 AEs(VAEs)[17] 的方案中取得了一个巨大的突破，其中解码器使用变分方法 [17,18] 从高斯分布中近似于真实的数据分布。最近遵循该方案的各种工作已经被提出，包括对抗性 AEs(AAEs)[19] 和瓦瑟斯坦 AEs(WAEs)[20]。虽然 vae 的训练相对简单，但它们生成的图像看起来很模糊。在某种程度上，这是因为显式表示的密度函数可能不能表示真实数据分布的复杂性，而不能学习高维数据分布 [21,22]。还提出了其他非危险性训练方法，包括 PixelCNN[23]、PixelRNN[24] 和 WaveNet[25]。然而，由于其自回归性质，新样本的生成无法并行。

2.3 对抗生成模型

为了解决上述模型的缺点，提出了 GANs[26]。尽管 GANs 是生成真实的样本的强大工具，但它们可能很难训练，并遭受模式崩溃。为了更好地进行 GAN 训练，人们提出了各种改进，包括改变损失函数 (例如，WGAN[1]) 和通过裁剪 [1]、梯度正则化 [4,27] 或光谱归一化 [28] 来正则化鉴别器为利普希茨。然而，gan 的训练仍然很棘手，需要仔细的超参数选择。

2.4 生成模型的评价

生成模型的评估仍然具有挑战性。早期的工作包括概率标准 [29]。然而，最近的生成模型 (特别是 GANs) 不适合接受这样的评估。传统上，GANs 的评估依赖于对少数例子或用户研究的视觉检查。最近，人们提出了若干定量评价标准。初始评分 (IS)[30] 同时衡量了多样性和图像质量。然而，它并不是一个距离度量标准。为了克服 IS 的缺点，在参考文献中引入了 Frechet 起始距离 (FID)。[31].FID 已被证明对图像损坏具有鲁棒性，并与视觉保真度很好地相关。在最近的一项工作 [32] 中，引入了分布的精度和查全率 (PRD) 来测量生成的数据分布和真实数据分布之间的精度和查全率。为了公平地比较 GANs，我们在参考文献中进行了大规模的比较。[33]，在统一的网络架构下比较了 7 个不同的 gan 和 vae，并建立了一个共同的评估基线。

2.5 非对抗性方法

最近也提出了各种非对抗性的方法。生成潜在优化 (GLO)[34] 采用了一种“无编码 AE”方法，其中生成模型用非对抗损失函数训练，获得比 VAE 更好的结果。隐式极大似然估计 (IMLE)[35] 提出了一种与迭代最近点 (ICP) 相关的生成模型训练方法。后来，Hoshen 和 Malik[36] 提出生成潜在的最近邻 (GLANN)，结合 GLO 和 GLANN 的优势，其中嵌入从图像空间潜在空间首先发现使用 GLO，然后任意分布和潜在之间的转换代码是使用 IMLE 计算。

其他方法利用具有可控雅可比矩阵 [37-39] 的 DNNs 直接逼近从噪声空间到图像空间的分布变换映射。最近，人们选择了基于能量的模型 [40-42]，通过用 [40-42] 表示能量函数，通过吉布斯分布来模拟图像的分布。这些方法利用现有的模型交替生成假样本，然后利用生成的假样本和真实样本对模型参数进行优化。

3 最优传输理论

在本节中，我们将介绍经典 OT 理论中的基本概念和定理，重点介绍布雷尼尔的方法及其推广到离散设置。细节可以在维拉尼的书《[5] 中找到。

3.1 Monge 问题

假设 $X \subset \mathbb{R}^d, Y \subset \mathbb{R}^d$ 是 d 维欧几里得空间 $\mathbb{R}^d$ 的两个子集，μ 和 ν 分别是定义在 X 和 Y 上的两个概率测度，密度函数如下：

$$\mu(x) = f(x)dx$$
$$\nu(y) = g(y)dy$$

假设总测度相等，$\mu(X) = \nu(Y)$；即

$$\int_X f(x)dx = \int_Y g(y)dy \tag{1}$$

我们只考虑保留这些测度的映射。

定义 3.1 (保留测度映射) 如果映射 $T : X \to Y$ 对于任何可测量的集合 $B \subset Y$，集合 $T^{-1}(B)$ 是 μ 可测量的和 $[T^{-1}(B)] = \nu(B)$，即，

$$\int_{T^{-1}(B)} f(x)dx = \int_B g(y)dy \tag{2}$$

测度保持条件记为 $T_{\#\mu} = \nu$，其中 $T_{\#\mu}$ 是由 T 诱导的推向测度。

给定一个成本函数 $c(x, y) : X \times Y \to \mathbb{R}$，它表示将每个单位质量从源点移动到目标的成本，映射 $T : X \to Y$ 的总传输成本被定义为

$$C_t = \int_X c[x, T(x)]d\mu(x) \tag{3}$$

Monge 的 OT 问题来自于找到使总运输成本最小化的保测度映射。

问题 3.1 (蒙日【43】;MP) 给定一个传输成本函数 $c(x, y) : X \times Y \to \mathbb{R}_{\geq 0}$，找到

保测度映射 $X \to Y$，使总传输成本最小化：

$$(MP) \quad \min_{T_{\#\mu}=\nu} \int_X c[x, T(x)] d\mu(x) \tag{4}$$

定义 3.2 (OT 映射) 蒙日问题的解决方案被称为 *OT* 映射。一个 *OT* 映射的总运输成本被称为 μ 和 ν 之间的 *Wasserstein* 距离，记为 $W_c(\mu, \nu)$。

$$W_c(\mu, \nu) = \min_{T_{\#\mu}=\nu} \int_X c[x, T(x)] d\mu(x) \tag{5}$$

3.2 Kontarovich 的方法

根据成本函数和度量，在 $(X\ \mu)$ 和 $(Y\ \nu)$ 之间的 OT 映射可能不存在。Kontarovich 将传输映射放宽为传输方案，并定义了联合概率测测度 $\rho(x, y)$: $X \times Y \to \mathbb{R}_{\geq 0}$，使 ρ 的边缘概率分别等于 μ 和 ν。设投影映射形式化为 $\pi_x(x\ y) = x, \pi_y(x\ y) = y$，然后定义联合度量类如下：

$$\Pi(x, y) = \left\{\rho(x, y) : X \times Y \to \mathbb{R} : (\pi_x)_{\#}\rho = \mu, (\pi_y)_{\#}\rho = \nu\right\} \tag{6}$$

问题 3.2 (Kontarovich;KP) 给定一个传输成本函数 $c(x, y) : X \times Y \to \mathbb{R}_{\geq 0}$，找到保测度映射 $X \to Y$，使总传输成本最小化：

$$(KP) \quad W_c(\mu, \nu) = \min_{\rho \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\rho(x, y) \tag{7}$$

Kontarovich 问题 (KP) 可以用 LP 方法来求解。由于 LP 等式的二元性，因此 Eq.(7)（KP 方程）可以重新表述为对偶性问题 (DP)，如下：

问题 3.3 (对偶;DP) 给定一个传输成本函数 $c(x, y) : X \times Y \to \mathbb{R}_{\geq 0}$，找到使总运输成本最小化的联合概率测度 $\rho(x, y) : X \times Y \to \mathbb{R}_{\geq 0}$

$$(DP) \quad \max_{\varphi, \psi} \left[\int_X \varphi(x) d\mu + \int_Y \psi(y) d\nu : \varphi(x) + \psi(y) \geq c(x, y)\right] \tag{8}$$

等式 *(8)* 的最大值给出了 *Wasserstein* 距离。现有的 *WGAN* 模型大多是基于 L^1 代价函数下的对偶公式。

定义 3.3 (c-转换) $\varphi : X \to \mathbb{R}$ 的 c-变换被定义为 $\varphi^c : Y \to \mathbb{R}$:

$$\varphi^c(y) = \underset{x \in X}{inf}\left[c(x,y) - \varphi(x)\right] \tag{9}$$

然后，*DP* 可以重写如下:

$$(DP) \quad W_c(\mu,\nu) = \underset{\varphi}{max}\int_X \varphi(x)d\mu + \int_Y \varphi^c(y)d\nu \tag{10}$$

3.3 Brenier 方法

对于二次欧氏距离代价，用布雷内尔 [44] 证明了 OT 映射的存在性、唯一性和内在结构

定理 3.1 (Brenier【44】) 假设 X 和 Y 是欧几里得空间 $\mathbb{R}^d$ 的子集，运输代价是二次欧氏距离 $c(x,y) = \frac{1}{2}\left\|x-y\right\|^2$。此外，$\mu$ 是绝对连续的，μ 和 ν 有有限的二阶矩

$$\int_X \left\|x\right\|^2 d\mu(x) + \int_Y \left\|y\right\|^2 d\nu(x) < \infty \tag{11}$$

然后存在一个凸函数 $\mu : X \to \mathbb{R}$，即所谓的 *Brenier* 势，其梯度映射 $\bigtriangledown u$ 给出了 *Monge* 问题的解:

$$(\bigtriangledown u)_{\#}\mu = \nu \tag{12}$$

Brenier 的势是唯一不变的；因此，最佳的传输映射是唯一的。

假设 *Briener* 势是 C^2 光滑的，那么它就是以下 *Monge-Ampère* 方程的解:

$$det(\frac{\partial^2 u(x)}{\partial x_i \partial x_j}) = \frac{f(x)}{g \circ \bigtriangledown u(x)} \tag{13}$$

对于 $\mathbb{R}^d$ 中的 L^2 运输成本 $c(x,y) = \frac{1}{2}\left\|x-y\right\|^2$，*c*-变换和经典 *Legendre* 变换具有特殊的关系。

定义 3.4 (Legendre 变换) 给定一个函数 $\varphi : \mathbb{R}^n \to \mathbb{R}$，其 *Legendre* 变换定义如下:

$$\varphi^*(y) = \underset{x}{sup}\left[\langle x,y\rangle - \varphi(x)\right] \tag{14}$$

当 $c(x,y)=\frac{1}{2}\|x-y\|^2$ 可以证明，以下关系成立：

$$\frac{1}{2}\|y\|^2-\varphi^c(y)=\left[\frac{1}{2}\|x\|^2-\varphi(x)\right]^* \tag{15}$$

定理 3.2 (Briener 极性因子分解【44】) 假设 X 和 Y 是欧几里得空间 $\mathbb{R}^d$，μ 是绝对连续的 *Lebesgue* 测量，和映射 $\varphi:X\to Y$ 推动 μ 向前到 ν，$\varphi_{\#}\mu=\nu$，然后存在一个凸函数 $u:X\to\mathbb{R}$，这样 $\varphi=\bigtriangledown u(x)\circ s$ ，在 $s:X\to X$ 是保测度的，$s_{\#}\mu=\mu$。此外，这个因子分解是唯一的。

以下定理在 OT 理论中是众所周知的

定理 3.3 (Villani【44】) 给定紧凸区域 $\Omega\subset\mathbb{R}^d$ 的 μ 和 ν，对于代价 $c(x,y)=h(x-y)$ 存在一个 *OT* 方案 ρ，且 h 严格凸。它是唯一的，并且形式为 $(id\ T_{\#})\mu$ *(id*：标识映射*)*，前提是 μ 是绝对连续的，而 $\partial\Omega$ 可以忽略不计。此外，存在 *Kantorovich* 势 φ，T 可以表示为：

$$T(x)=x-(\bigtriangledown h)^{-1}\left[\bigtriangledown\varphi(x)\right]$$

当 $c(x,y)=\frac{1}{2}\|x-y\|^2$，我们有

$$T(x)=x-\bigtriangledown\varphi(x)=\bigtriangledown\left[\frac{1}{2}\|x\|^2-\varphi(x)\right]=\bigtriangledown u(x)$$

在这种情况下，*Brenier* 势在 μ 和 *Kantorovich* 势在 φ 有以下关系：

$$u(x)=\frac{1}{2}\|x\|^2-\varphi(x) \tag{16}$$

3.4 OT 映射的正则性

设 Ω 和 Λ 是 $\mathbb{R}^d$ 上的两个有界光滑开集，设 $\mu=fdx$ 和 $\nu=gdy$ 是该 $\mathbb{R}^d$ 上的两个概率测度 $f\mid_{\mathbb{R}^d\,\Omega}=0$ 和 $g\mid_{\mathbb{R}^d\,\Lambda}=0$。假设 f 和 g 在 Ω 和 Λ 上分别有界远离零和无穷大。

3.4.1 凸目标区域

定义 3.5 (Hölder 连续) 当存在非负实常数 C，$\alpha > 0$ 时，使得 $|f(x) - f(y)| \leq C\|x-y\|^{\alpha}$ 对于 f 域中的所有 x 和 y 在 d-维欧氏空间上的实值或复值函数 f 满足 *Hölder* 条件，或 *Hölder* 连续。

定义 3.6 (Hölder 空间) 霍尔德空间 $(C^{k,\alpha}(\Omega)$, 其中 Ω 是一些欧几里德空间和 $k \geq 0$ 的开放子集是一个整数，由 Ω 上的函数组成，这些函数具有 k 阶的连续导数，使得第 k 阶偏导数 *Hölder* 与指数 α 连续，其中 $0 < \alpha \geq 1$。$C^{k,\alpha}_{loc}(\Omega)$ 指上述条件适用于 Ω 的任何紧子集。

定理 3.4 (Caffarelli【45】) 如果 Λ 是凸的，那么 *Brenier* 势 u 是严格凸的；此外，

(1) 如果 $\lambda \geq f, g \geq \frac{1}{\lambda}$ 对于某些 $\lambda > 0$，则 $u \in C^{1,\alpha}_{loc}(\Omega)$ 。

(2) 如果 $f \in C^{k,\alpha}_{loc}(\Omega)$ 和 $g \in C^{k,\alpha}_{loc}(\Lambda)$ 对于 $f, g > 0$，则 $u \in C^{k+2,\alpha}_{loc}(\Omega)$ 和 $(k \leq 0, \alpha \in (0,1))$ 。

3.4.2 非凸目标域

如果 Λ 不是凸的，并且存在光滑的 f 和 g 使得 $u \notin C^1(\Omega)$，那么 OT 映射 $\bigtriangledown u$ 在奇点上是不连续的。

定义 3.7 (次梯度) 给定一个开集 $\Omega \subset \mathbb{R}^d$ 和一个凸函数 $u : \Omega \to \mathbb{R}$，对于 $x \in \Omega$，u 在 **x** 处的次梯度（次微分）定义如下：

$$\partial u(x) = \{p \in \mathbb{R}^n : u(z) \leq u(x) + \langle p, z - x \rangle, \forall z \in \Omega\}$$

很明显，$u(x)$ 是一个闭凸集。几何上，如果 $\rho \in u(x)$，那么超平面 $l_{x,p}(z) = u(x) + \langle p, z - x \rangle$ 在 x 从下面接触 u；也就是说，在 Ω 和 $l_{x,p}(x) = u(x)$ 上的 $l_{x,p} \leq u$，其中 $l_{x,p}$ 是 u 在 **x** 上的支撑平面。

如果 *Brenier* 势 u 的次梯度 $\partial u(x)$ 是一个单例的，那么他在 **x** 上是可微的。我们根据点的子梯度的维数对点进行分类，并定义集合 $\sum_k(u) = \{x \in \mathbb{R}^d \mid \dim[\partial u(x)] = k\}, k = 0, 1, 2, \cdots, d$。

很明显，$\sum_0(u)$ 是正则点的集合，而 $\sum_k(u)$，其中 $k>0$，是奇异点的集合。我们还定义了在 **x** 处的可达的子梯度如下：

$$\partial_* u(x)=\left\{\lim_{x\to\infty}\bigtriangledown u(x_k)\mid x_k\in\sum_0, x_k\to x\right\}$$

众所周知，该子梯度等于可达子梯度的凸包：

$$\partial u(x)=Convex \quad hull \quad \bigtriangledown_* u(x)$$

定理 3.5 (规律性) 设 Ω，Λ 是两个有界开集，设在 Ω 和 Λ 之外的两个为零的概率密度 $f,g:\mathbb{R}^d\to\mathbb{R}^+$，并且在 Ω 和 Λ 上分别从零和无穷远处有界。用 $T=\bigtriangledown u:\Omega\to\Lambda$ 表示由定理 *(3.1)* 提供的 *OT* 映射。然后存在两个相对封闭的集 $\sum_\Omega\subset\Omega$ 和 $\sum_\Lambda\Lambda$ 与 $\sum_\Omega=\sum_\Lambda=0$，这样 $T:\Omega\quad\sum_\Omega\to\Lambda\quad\sum_\Lambda$ 是某些 $\alpha>0$ 的 $C^{0,\alpha}_{loc}$ 的同态类。

我们称 $\sum_\Omega$ 为 *OT* 映射的奇异集 $\bigtriangledown u:\Omega\to\Lambda$。图 *5* 说明了奇异集的结构，使用基于定理 *4.2* 的算法计算。我们得到的计算结果如下：

$$\sum_0=\frac{\Omega}{\{\sum_1\cup\sum_2\}},\sum_1=\bigcup_{k=0}^{3}\gamma_k,\sum_2=\{x_0,x_1\}$$

x_0 的次梯度，$\partial u(x_0)$，是 Λ 的整个内孔，而 $\partial u(x_1)$ 是阴影三角形。对于 $\gamma_k(t)$ 上的每个点，$\partial u\left[\gamma_k(t)\right]$ 是 Λ 外的一条线段。x_1 是 γ_1,γ_2 和 γ_3 的分岔点。$\sum_1$ 和 $\sum_2$ 上的 *Brenier* 势是不可微的，它们上的 *OT* 映射 $\bigtriangledown u$ 是不连续的。

4 计算算法

Brenier 定理可以直接推广到离散的情况下。在 GAN 模型中，源测度 μ 被定义为紧凸域 Ω 上的均匀（或高斯）分布；目标测度 ν 表示为经验测度，它是狄拉克测度的和：

$$\nu=\sum_{i=1}^{n}\nu_i\delta(y-y_i) \tag{17}$$

其中，$Y=\{y_1,y_2,\cdots,y_n\}$ 为训练样本，具有权重：$\sum_{i=1}^{n}\nu_i=\mu(\Omega)$；$\delta$ 为特征函数。

每个训练样本 yi 对应于 Brenier 势的一个支撑平面，表示如下：

$$\pi_{h,i}(x)=\langle x,y_i\rangle+h_i \tag{18}$$

其中，高度 h_i 是一个未知的变量。我们将所有的高度变量表示为 $h=(h_1,h_2,\cdots,h_n)$。

欧几里得空间中一个超平面族的包络是一个超曲面，它在某一点上与族的每个成员相切，这些切点一起构成了整个包络。如图 6 所示，布雷尼尔势 $u_h:\Omega\rightarrow\mathbb{R}$ 是一个由 h 决定的分段线性凸函数，它是其所有支撑平面的上包络线：

$$u_h(x)=\max_{i=1}^{n}\left[\pi_{h,i}(x)\right]=\max_{i=1}^{n}\left[\langle x,y_i\rangle+h_i\right] \tag{19}$$

布雷尼尔势的图是一个凸的多面体。每个支撑平面 $\pi_{h,i}$，i 对应于多面体的一个面。多面体的投影诱导了 Ω 的胞腔分解，其中每个支撑平面 $\pi_i(x)$ 投射到一个胞胞 $W_i(h),p$ 上是 $\mathbb{R}^d$ 中的任意点：

$$\Omega=\bigcup_{i=1}^{n}w_i(h)\cap\Omega,W_i(h)=\{p\in\mathbb{R}^d\mid\triangledown u_h(p)=y_i\} \tag{20}$$

胞腔分解是一个幂次图。$W_i\cap\Omega$ 的 μ-测度记为 $w_i(h)$：

$$w_i(h)=\mu\left[W_i(h)\cap\Omega\right]=\int_{W_i(h)\cap\Omega}\mathrm{d}\mu \tag{21}$$

梯度映射：$\triangledown u_h:\Omega\rightarrow Y$ 将每个单元格 $W_i(h)$ 映射到一个单点 y_i：

$$\triangledown u_h:W_i(h)\rightarrow y_i,i=1,2,\cdots,n \tag{22}$$

给定等式 (17) 中的目标度量 ν，在等式 (19) 中存在一个离散的布雷尼尔势，其每个面 $w_i(h)$ 的投影 μ 体积等于给定的目标度量 ν_i。Alexandrov[46] 在凸几何中证明了这一点。

定理 4.1 (Alexandrov【46】) 假设 Ω 是一个紧凑的凸多面体，在 $\mathbb{R}^d$ 中内部为非空，$n_1,\cdots,n_k\subset\mathbb{R}^{n+1}$ 是不同的 k 个单位向量，$(n+1)$ 坐标为负，$v_1,\cdots,v_k>0$，

因此 $\sum_{i=1}^{k}\nu_i = vol(\Omega)$ 。然后存在一个有精确 k 余维-*1* 面 $F_1,\cdots,F_k$ 的凸多面体 $P\subset\mathbb{R}^{n+1}$，其中 n_i 是 F_i 的法向量，并且在 Ω 和 F_i 的投影之间有体积 v_i。此外，这种 P 在垂直平移之前是唯一的。

Alexandrov 的解的存在性证明是基于代数拓扑，这不是建设性的。最近，*Gu* 等人 *[6]* 提供了一个基于变分方法的建设性证明。

定理 4.2 (参考文献【6】) 设 μ 是定义在 $\mathbb{R}^d$ 紧凸域 Ω 上的概率测度，设 $Y=\{y_1,y_2,\cdots,y_n\}$ 是 $\mathbb{R}^d$ 中的一个不同点。然后对于 $\sum_{i=1}^{n}v_i = \mu(\Omega)$ 的任何 $v_1,v_2,\cdots,v_n>0$，对于所有 i，存在唯一的添加一个常数 $(c,c,\cdots,c)$ 使得 $w_i(h)=v_i$ 的 $h=(h_1,h_2,\cdots,h_n)\in\mathbb{R}^n$。向量 $\mathbf{h}$ 是以下凸能量的唯一最小值参数：

$$E(h)=\int_0^h\sum_{i=1}^{n}w_i(\eta)\mathrm{d}\eta_i-\sum_{i=1}^{n}h_iv_i \tag{23}$$

定义在一个开的凸集上

$$H=\{h\in\mathbb{R}^d: w_i(h)>0, i=1,2,\cdots,n\} \tag{24}$$

此外，$\bigtriangledown u_k$ 使二次代价最小化

$$\frac{1}{2}\int_\Omega\|x-T(x)\|^2\,\mathrm{d}\mu(x) \tag{25}$$

在所有的传输映射中，$T_{\#\mu}=\nu$。

上述凸能量在等式 *(23)* 中的梯度值如下：

$$\bigtriangledown E(h)=[w_1(h)-v_1,w_2(h)-v_2,\cdots,w_n(h)-v_n]^T \tag{26}$$

能量的第 i 行和第 j 列元素如下：

$$\frac{\partial w_i}{\partial h_j}=-\frac{\mu(W_i\cap W_j\cap\Omega)}{\|y_i-y_j\|},\frac{\partial w_i}{\partial h_i}=\sum_{j\neq i}\frac{\partial w_i}{\partial h_j} \tag{27}$$

如图 *6* 所示，黑森矩阵有一个显式的几何解释。图 *6(a)* 显示了离散的布雷尼尔势 u_h，而图 *6(b)* 显示了其使用定义 *(3.3)* 的勒让德变换。勒让德变换可以用几何方法构造：对于每个支撑平面 $\pi_{h,i}$，我们构造对偶点 $\pi^*_{h,i}$；对偶点的凸包

$\{\pi^*_{h,1}, \pi^*_{h,2}, \cdots, \pi^*_{h,n}\}$ 是勒让德变换的图 u^*_h。

u^*_h 的投影得到了一个的三角剖分 $Y = \{y_1, y_2, \cdots, y_n\}$，即加权的德劳内三角剖分。如图 7 所示，等式 *(20)* 中的能量图和加权的德劳内三角剖分是庞加莱对偶的：如果在幂图中，$W_i(h)$ 和 $W_j(h)$ 在一个 $(d-1)$-维胞腔上相交，那么在加权的德劳内三角剖分中，y_i 与 y_j 连接。在等式 *(27)* 中的黑森矩阵的元素是能量图中 $(d-1)$ 单元的 μ 体积与加权德劳内三角剖分中双边的长度之间的比值。

传统的幂次图可以与上述定理密切相关。

定义 4.1 (power 距离) 给定一个功率权重为 ψ_i 的点 $y_i \in \mathbb{R}^d$，*power* 距离如下：

$$pow(x, y_i) = \|x - y_i\|^2 - \psi_i \tag{28}$$

定义 4.2 (power 图) 给定加权点 $(y_1, \psi_1), \cdots, (y_k, \psi_k)$，幂图为单元分解的 $\mathbb{R}^d$：

$$\mathbb{R}^d = \bigcup_{i=1}^{k} W_i(\psi) \tag{29}$$

其中，每个单元格都是一个凸面多面体：

$$W_i(\psi) = \left\{x \in \mathbb{R}^d \mid pow(x, y_j) \le pow(x, y_j)\right\} \tag{30}$$

加权德劳内三角剖分，记为 $T(\psi)$，是幂图的庞加莱对偶；如果 $W_i(\psi) \cap W_j(\psi) \neq \phi$，那么在加权德劳内三角剖分中有一条边连接 y_i 和 y_j。注意，$pow(x, y_i) \le pow(x, y_j)$ 等价于

$$\langle x, y_i\rangle + \frac{1}{2}(\psi_i - \|y_i\|^2) \ge \langle x, y_j\rangle + \frac{1}{2}(\psi_j - \|y_j\|^2) \tag{31}$$

让 $h_i = \frac{1}{2}(\psi_i - \|y_i\|^2)$；然后我们将其定义重写如下：

$$W_i(\psi) = \left\{x \in \mathbb{R}^d \mid \langle x, y_i\rangle + h_i \ge \langle x, y_j\rangle + h_j, \forall j\right\} \tag{32}$$

在实践中，我们的目标是计算离散的布雷尼尔的潜在等式 (19) 通过优化凸能量等式 (23). 对于低维情况，我们可以通过计算牛顿方法直接计算梯度等式 (26) 和黑森矩阵等式 (27)。对于深度学习应用，直接计算黑森矩阵是不可行的；

相反，我们可以使用梯度下降法或具有超线性收敛性的拟牛顿法。梯度的关键是估计 μ 体积 $w_i(h)$。这可以用蒙特卡罗方法来实现：我们从分布 中抽取 n 个随机样本，并计算在 $W_i(h)$ 范围内的样本数量，这是收敛于 μ 体积的比值。这种方法是纯并行的，可以使用 GPU 来实现。此外，我们还可以使用分层的方法来进一步提高效率：首先，我们将目标样本分类为聚类，并计算到聚类质量中心的 OT 映射；其次，对于每个聚类，我们计算从相应的细胞到聚类内原始目标样本的 OT 映射。

为了避免模态坍塌，我们需要在 Ω 中找到奇异点集。如图 8 所示，目标狄拉克测度有两个簇；源是单位平面圆盘上的均匀分布。布雷尼尔势函数的图是一个中间有一个脊的凸多面体。脊在圆盘上的投影是奇异集 $\sum_1(u)$，最优映射在 $\sum_1$ 上是不连续的。在一般情况下，如果两个单元格 $W_i(h)$ 和 $W_j(h)$ 相邻，则我们计算法线与相应支撑平面之间的夹角：

$$\theta_{ij} = \frac{\langle x_i, y_j \rangle}{\|y_i\| \|y_j\|}$$

如果 θ_{ij} 大于一个阈值，则公共面 $W_i(h) \cap W_j(h)$ 在不连续奇异集合中。

5 GANs 和最优传输