arXiv:2602.08374v1 [stat.ML] 9 Feb 2026

Schr ödinger桥问题的经验风险最小化方法

Denis Belomestny* Alexey Naumov† Nikita Puchkin† Denis Suchkov †

摘要

我们研究端点分布仅通过样本获得时的 Schr¨odinger 桥问题。经典计算方法在经验测度上通过 Sinkhorn迭代估计 Schr¨odinger 势，随后通过对核平滑对偶解求导来构造时间非齐次漂移。与此不同，我们提出一种学习理论方法：将 Schr¨odinger 系统改写为关于单个正变换后势的形式，该势满足一个非线性不动点方程；然后在一个函数类上通过经验风险最小化来估计该势。我们在参考核和终端密度满足次高斯假设的条件下，证明经验风险与其总体对应量之间的一致集中性。我们将学习得到的势代入桥的随机控制表示式，以生成样本。我们通过数值实验展示所提方法的性能。

1 引言

薛定谔桥问题（SBP）提供了一种有原则的方法，用于在两个概率分布之间进行插值：在所有满足给定端点边缘分布的随机过程中，选取相对于参考动力学在相对熵意义下最接近的过程。形式化地，设 $(X_t)_{t\in[0,T]}$ 是定义在 $\mathbb{R}^d$ 上的马尔可夫过程，其在路径空间上的参考律为 Q，并且相对于勒贝格测度具有转移密度 $(q_t)_{t\in(0,T]}$，因此 $\mathsf{Q}\big(X_T \in [y, y+dy) \;\big|\; X_0 = x\big) = q_T(x,\ y)\ dy$。给定 $\mathbb{R}^d$ 上的两个概率密度 ρ_0 和 ρ_T，并考虑以下类

$$\mathcal{P}(\rho_0,\rho_T) \;:=\; \Big\{ P \ll \mathsf{Q} \;\Big|\; P\circ X_0^{-1} = \rho_0\,dx,\ P\circ X_T^{-1} = \rho_T\,dx \Big\}.$$

（动态）薛定谔桥问题旨在寻找概率测度 $P^\star \in \mathcal{P}(\rho_0,\ \rho_T)$，使其在相对熵意义下最接近 Q：

$$P^\star \;\in\; \arg\min_{P\in\mathcal{P}(\rho_0,\rho_T)} \mathcal{H}(P\,\|\,\mathsf{Q}), \qquad \mathcal{H}(P\,\|\,\mathsf{Q}) := \int \log\left(\frac{dP}{d\mathsf{Q}}\right) dP. \tag{1.1}$$

这是一个经典结果（例如参见Chen 等人 [2016]）：在温和条件下，极小化解 $P^\star$ 存在，并具有如下薛定¨格因子分解形式

$$\frac{dP^\star}{d\mathsf{Q}}(X) \;=\; \nu_0(X_0)\,\nu_T(X_T), \tag{1.2}$$

*杜伊斯堡-埃森大学，德国
†俄罗斯高等经济大学，俄罗斯联邦

对于某些非负可测函数（薛定¨格势） ν_0， ν_T： $\mathbb{R}^d \to (0, \infty)$。取时间–0和时间–$T$ 边缘分布于（1.2）可得如下方程组

$$\rho_0(x) = \nu_0(x) \int_{\mathbb{R}^d} q_T(x,z)\, \nu_T(z)\, dz, \tag{1.3}$$

$$\rho_T(y) = \nu_T(y) \int_{\mathbb{R}^d} q_T(x,y)\, \nu_0(x)\, dx. \tag{1.4}$$

相应的“静态”Schr¨odinger桥问题，是对 (X_0, X_T) 的耦合进行熵最小化：

$$\pi^\star \in \arg\min_{\pi \in \Pi(\rho_0,\rho_T)} \mathcal{H}\big(\pi \,\|\, \pi^{\mathrm{ref}}\big), \qquad \pi^{\mathrm{ref}}(dx,dy) = \rho_0(x)\, q_T(x,y)\, dx\, dy, \tag{1.5}$$

其中， $\Pi(\rho_0, \rho_T)$ 表示在 $\mathbb{R}^d \times \mathbb{R}^d$ 上边缘分布为 ρ_0 和 ρ_T 的概率测度集合。优化器 $\pi^\star$ 可写为

$$\pi^\star(dx,dy) \;=\; \nu_0(x)\, q_T(x,y)\, \nu_T(y)\, dx\, dy, \tag{1.6}$$

且势函数 (ν_0, ν_T) 满足 (1.3)–(1.4)。在 $q_T(x)$ 是随机微分方程 $dX_t = \sigma dW_t$ 的转移密度且 $\sigma > 0$时，(1.6) 与熵正则化最优传输（EOT）的解一致。从计算角度来看，EOT 通常通过在经验测度上执行 Sinkhorn迭代求解，从而得到 Schr¨odinger 势函数的离散近似。在这一视角基础上，近期研究将这些（离散）势函数代入桥的随机控制表示中，并在进一步平滑和求导后得到漂移场。尽管该流程在中等维度下较为有效，但从学习的角度来看，它带来了两个概念上的挑战。首先，势函数仅在经验样本的支撑集上进行估计，必须将其扩展到样本之外（例如通过核平滑），才能得到连续漂移。其次，总体误差混合了优化误差（有限次 Sinkhorn迭代）、统计误差（有限样本）以及离散化/平滑误差，这使得泛化分析变得复杂。

在本文中，我们将直接以函数形式估计 Schr¨odinger 势，并将其视为一个学习问题。我们将 Schr¨odinger 系统改写为关于单个正变换势 g 的形式，该势满足一个非线性不动点方程

$$g = \mathcal{C}[g], \tag{1.7}$$

对于一个非线性积分算子 $\mathcal{C}$，它依赖于 ρ_0， ρ_T 和 q_T。当只有样本可用时，

$$X_1, \dots, X_N \sim \rho_0, \qquad Y_1, \dots, Y_M \sim \rho_T,$$

^我们通过用经验平均替代期望，构造经验算子 $\mathcal{C}_{N,\,M}$。我们不通过迭代比例拟合（Sinkhorn）强制满足不动点条件，而是通过在正函数假设类 $\mathcal{G}$ （例如神经网络）上最小化经验残差损失来估计 g：

$$\widehat{g}_{N,M} \in \arg\min_{g \in \mathcal{G}} \frac{1}{M} \sum_{j=1}^{M} \ell\Big(g(Y_j),\, \widehat{\mathcal{C}}_{N,M}[g](Y_j)\Big), \tag{1.8}$$

其中 ℓ 在等号成立时取最小值（例如平方损失）。所得目标函数可以使用随机梯度方法进行优化。至关重要的是，与 Sinkhorn 类算法不同，后者仅在采样位置隐式生成势函数，并且需要通过插值或平滑来获得连续对象，而

学习得到的估计量在构造上就是连续的。这对于下游任务至关重要，例如：计算 Schr¨odinger 桥的漂移场、模拟受控扩散、进行敏感性分析以及基于梯度的控制。此外，函数学习视角的一个关键优势在于，它能够自然地适应 Schr¨odinger 势函数 g的稀疏表示。根据假设类的选择，可以获得：稀疏基展开（例如小波、傅里叶特征或核字典）、由瓶颈神经网络诱导的低秩表示，以及通过正则化实现的隐式稀疏性。这种稀疏性可以加快势函数及其梯度的评估。在平移不变场景中，稀疏性尤其有效，因为势函数通常呈现低频结构或局部特征，而少量活跃成分即可捕捉这些结构。

一旦学得势函数（或对数势函数），我们便采用 Schr¨odinger 桥的随机控制表示；参见 [Dai Pra, 1991]。令 $h(t,\ x)$ 表示随时间演化的 Schr¨odinger 势函数，其定义如下：

$$h(t,x) = \int \nu_T(y)\, q_{T-t}(x,y)\, dy.$$

则具有分布律 $P^\star$ 的过程 $(X_t^\star)^{t\in(0,T]}$ 的转移密度为

$$q^h(y,T|x,t) = \frac{q_{T-t}(x,y)h(T,y)}{h(x,t)}.$$

对于扩散过程 $dX_t = bdt + \sigma dW_t$，这对应于将基本过程 $(X_t)^{t\in(0,T]}$ 的漂移变换为 $a\nabla \log\ h$，其中 $a = \sigma\sigma^\top$。由此可得到一个实用的采样器：从 $x_0 \sim \rho_0$ 出发，利用学得的漂移模拟随机微分方程（例如采用欧拉–丸山方法），从而获得中间时刻及时间 T 的近似桥样本。

贡献 我们的主要贡献 可概括如下：

- 我们将 Schr¨odinger 系统重新表述为关于变换势 g 的单个非线性不动点方程，并提出一种经验风险最小化估计量，通过在某个变换势类 $\mathcal{G}$ 上最小化经验不动点残差来构造。由此为研究经验 Schr¨odinger 问题提供了一个灵活的框架。
- 当参考核 Q 为高斯时，我们证明，群体不动点 $g^\star$ 存在一个快速收敛的厄米函数展开。我们推导出厄米函数投影算子 n 阶近似的显式 L^2 误差界，并将其与经验风险的一致集中界结合，从而得到端到端风险保证；除多对数因子外，该保证对样本量呈近参数依赖。
- 我们通过数值实验展示了该方法在以下任务中的表现：(i) 二维瑞士卷到 S 形曲线的示例；(ii) 训练–测试分布偏移下的高斯混合输运；(iii) 单细胞群体插值。实验表明，该方法的表现与现有基线相当或优于现有基线。

1.1 相关工作

薛定谔桥问题起源于薛定谔¨1932 年的工作 薛定谔¨ [1932] 关于在参考动力学下、给定两个边缘分布时，最可能随机演化过程的研究。现代研究强调它与互易过程、大偏差和最优传输之间的联系；背景介绍及更多参考文献可参见Leonard[2014] 。关于薛定谔桥问题（及相关互易扩散）的随机控制视角见 Dai Pra [1991], ；利用射影/希尔伯特度量结构的计算视角则见，例如Chen 等人 [2016]。

熵正则化最优传输与 Sinkhorn 算法 静态薛定谔桥问题与熵正则化最优传输相一致。后者因其稳定性和算法效率，已成为计算最优传输中的核心工具。熵正则化最优传输可以通过 Sinkhorn 算法（迭代比例拟合／矩阵缩放）求解其对偶问题；该算法由Cuturi [2013] 在机器学习领域推广，其根源可追溯至更早的矩阵缩放结果，例如 Franklin 和 Lorenz [1989]；全面综述另见Peyr´e 和 Cuturi [2019]。近期理论研究深化了我们对 Sinkhorn 算法在经典有界／紧集条件之外的压缩性和收敛性特征的理解，并给出了迭代结果及其梯度的非渐近界；例如参见 Conforti 等人 [2023], Greco 等人 [2023]。

从样本估计薛定谔桥¨ 在统计设定中，ρ_0 和 ρ_T 只能通过样本获取，一种标准方法是对经验测度求解熵正则化最优传输（通过 Sinkhorn），以获得薛定谔势的离散近似，然后为动态桥构造一个采样器。近期的一个代表性实例是 Pooladian 和 Niles-Weed 提出的 SinkhornBridge，该方法将（近似）对偶解代入随机控制表示，从而生成时间非齐次漂移。我们的方法在估计阶段有所不同：我们不再通过对经验测度进行 Sinkhorn迭代来计算离散势并随后对其进行扩展或平滑，而是通过在函数类（例如神经网络）上进行经验风险最小化，来估计一个连续势。这一视角专为学习理论分析（一致集中、逼近误差）而设计，并且从构造上保证所得势能够泛化到样本之外。

生成建模和数据到数据转换中的薛定谔桥问题 与受控扩散和基于分数的方法相结合，越来越多的研究将薛定谔桥问题与现代生成建模联系起来。例如，扩散薛定谔桥及其在生成建模中的应用 [De Bortoli 等人， 2021], 基于学习的薛定谔桥变体，例如神经拉格朗日薛定谔桥 [Koshizuka 和 Sato， 2022], 以及薛定谔桥匹配目标 [Shi 等人， 2023]。其他计算效率较高的方法包括 LightSB[Korotin 等人， 2024], LightSB-OU [Puchkin 等人， 2025]。在生物时间插值和动力学建模的应用中，薛定谔桥/最优传输思想也见于例如 TrajectoryNet [Tong 等人， 2020] 以及后续无需模拟或基于匹配的公式 [Tong 等人，2023年b], ，还有基于小批量最优传输的训练目标 [Tong 等人，2023年a]。连续归一化流基线通常采用稳定化正则化和架构约束进行训练；例如参见 Finlay 等人 [2020]。

记号 给定 $l, u : \mathbb{R}^d \to \mathbb{R}$ 括号 $[l, u]$ 定义为所有满足以下条件的函数的集合 $f : \mathbb{R}^d \to \mathbb{R}$对所有 $x \in \mathbb{R}^d$，均有 $l(x) \le f(x) \le u(x)$。对于函数类 $\mathcal{F}$，其函数为 $f : \mathbb{R}^d \to \mathbb{R}$，定义其括号数如下：$\mathcal{N}_{[]}(\mathcal{F}, \|\cdot\|_\infty, \delta) := \inf\left\{ m \in \mathbb{N} : \exists (l_j, u_j)_{j=1}^m \text{ s.t. } \|u_j - l_j\|_\infty \le \delta, \mathcal{G} \subseteq \bigcup_{j=1}^m [l_j, u_j] \right\}$。对于 $f : \mathbb{R}^d \to \mathbb{R}$，定义其傅里叶变换为 $\widehat{f}(\omega) := \int_{\mathbb{R}^d} f(x) e^{-i\omega \cdot x}\, dx$。对于 $K \subset \mathbb{R}^d$，我们使用记号 $\|f\|_{L^\infty(K)} := \sup_{x \in K} |f(x)|$。对于定义在 $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ 上的概率测度 μ，记作 $\|f\|_{L^p(\mu)} := \int_{\mathbb{R}^d} |f(x)| \mu(dx)$。我们还定义截断为 $\text{clip}(f, a, b) = f$若 $a \le f \le b, a$若 $f < a$，且 b若 $f > b$。

2 基于经验风险最小化的估计量

首先，我们将（1.3）、（1.4）改写为关于单个变换势的形式，并得到一个非线性不动点问题。定义

$$g(y) := \frac{\rho_T(y)}{\nu_T(y)}, \tag{2.1}$$

也就是说，$\nu_T(y) = \rho_T(y)/g(y)$。由（1.3）可得

$$\nu_0(x) = \frac{\rho_0(x)}{\int q_T(x,z)\,\nu_T(z)\,dz} = \frac{\rho_0(x)}{\int q_T(x,z)\,\dfrac{\rho_T(z)}{g(z)}\,dz}. \tag{2.2}$$

对于每个 $x \in \mathbb{R}^d$，定义

$$D_g(x) := \int_{\mathbb{R}^d} q_T(x,z)\,\frac{\rho_T(z)}{g(z)}\,dz. \tag{2.3}$$

将（2.2）和（2.1）代入（1.4）可得

$$\rho_T(y) = \frac{\rho_T(y)}{g(y)} \int_{\mathbb{R}^d} \frac{q_T(x,y)\rho_0(x)}{D_g(x)}\,dx.$$

约去两边的公因子 $\rho_T(y)$，再乘以 $g(y)$，可得非线性不动点方程

$$g(y) = \int_{\mathbb{R}^d} \frac{q_T(x,y)\rho_0(x)}{D_g(x)}\,dx =: \mathcal{C}[g](y) \tag{2.4}$$

因此，（静态）Schr¨odinger 桥问题可以等价地表述为：寻找一个正函数 g，使其满足非线性不动点方程 $g = \mathcal{C}[g]$。找到不动点 g $\mathcal{C}$ 后，Schr¨odinger 势可通过以下方式恢复：$\nu_T(y) = \rho_T(y)/g(y)$，并根据（2.2），通过按照（1.2）对参考过程 Q 进行倾斜，得到最优马尔可夫过程$P^\star$。

在许多应用中，边缘密度 ρ_0 和 ρ_T 无法显式获得。相反，我们观测到独立样本

$$X_1, \dots, X_N \sim \rho_0, \qquad Y_1, \dots, Y_M \sim \rho_T,$$

并寻求恢复满足 $g = \mathcal{C}[g]$ 的不动点 g。这便产生了 Schr¨odinger 桥问题的统计版本。定义经验测度

$$\widehat{\rho}_0^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_i}, \qquad \widehat{\rho}_T^M := \frac{1}{M}\sum_{j=1}^M \delta_{Y_j}.$$

将 ρ_T 替换为其在（2.3）中的经验对应项，我们得到

$$\widehat{D}(x) := \frac{1}{M}\sum_{k=1}^M q_T(x, Y_k)\,\frac{1}{g(Y_k)}\,.$$

在引理 B.1 中，我们证明，对于所有 $x \in \mathrm{supp}(\rho_0)$，$D \geq D_g(x) \geq D$ 成立。但我们无法为 $D(x)$证明相同的下界。我们引入一个形式为 $\mathcal{T}_{[a,b]}[f](x) = \mathrm{clip}(f, a, b)$的截断算子 $\mathcal{T}_{[a,b]}$。将ρ_0 和 ρ_T 替换为它们在（2.4）中的经验对应项，可得到经验非线性算子

$$\widehat{\mathcal{C}}_{N,M}[g](y) := \frac{1}{N}\sum_{i=1}^N \frac{q_T(X_i, y)}{\mathcal{T}_{[\underline{D},\overline{D}]}[\widehat{D}](X_i)}. \tag{2.5}$$

通过强制满足以下条件来近似不动点条件 $g = \mathcal{C}[g]$

$$g(Y_j) \approx \widehat{\mathcal{C}}_{N,M}[g](Y_j), \qquad j = 1, \dots, M.$$

为估计 g，我们将该系统重新表述为一个经验风险最小化（ERM）问题。设 $\ell:(0,\infty)\times(0,\infty)\to[0,\infty)$ 为一个在等式成立时取最小值的损失函数，例如 $\ell(u,v)=\frac{1}{2}(u-v)^2$。经验风险定义为

$$\widehat{\mathcal{R}}_{N,M}(g):=\frac{1}{M}\sum_{j=1}^{M}\ell\Big(g(Y_j),\,\widehat{\mathcal{C}}_{N,M}[g](Y_j)\Big)\,. \tag{2.6}$$

统计学上的 Schr¨odinger 桥估计量定义为

$$\widehat{g}_{N,M}\ \in\ \arg\min_{g\in\mathcal{G}}\widehat{\mathcal{R}}_{N,M}(g), \tag{2.7}$$

其中，$\mathcal{G}$ 是正函数或参数模型（例如神经网络）的一个可容许类。在下一^节中，我们将在一组现实的假设下研究 $\mathcal{R}_{N,M}(g)$ 的行为，这些假设使我们能够应用经验过程理论中的工具。

3 主要结果

我们陈述全文中使用的主要假设。首先介绍关于核 Q 和密度 ρ_0,ρ_T 的假设。

(Q) 存在常数 $c_-,c_+>0$ 和 $a_-,a_+>0$，使得

$$c_-\exp(-a_-\|x-y\|^2)\ \le\ q_T(x,y)\ \le\ c_+\exp(-a_+\|x-y\|^2),\qquad \forall x,y\in\mathbb{R}^d.$$

此外，q_T 是全局利普希茨的，常数为 L_q.

(R0) 存在 $x_0\in\mathbb{R}^d$ 和两个正实数 $r_0,\ R_0$ 使得 $B(x_0,\ r_0)\ \subseteq\ \mathrm{supp}(\rho_0)\ \subseteq B(x_0,\ R_0)$。此外，存在常数$\rho_{0,\ -}>0$ 使得

$$\rho_0(x)\ge\rho_{0,-},\qquad \forall x\in B(x_0,r_0).$$

(RT) 存在常数 $0<c_T^-\le c_T^+<\infty$ 和 $b_T^-,\ b_T^+>0$ 使得

$$c_T^-\exp(-b_T^-\|y\|^2)\ \le\ \rho_T(y)\ \le\ c_T^+\exp(-b_T^+\|y\|^2),\qquad \forall y\in\mathbb{R}^d.$$

*(R0)*中的非退化条件可防止归一化因子在 $\mathrm{supp}(\rho_0)$ 上任意趋近于零，并用于得到 $D_g(x)$的一致界以及不动点 $g^\star$的双侧界。注意，我们并未假设 ρ_T 是紧支撑的，而是假设其具有双侧次高斯行为。现在，我们对损失函数施加额外假设。

(L) 损失函数 $\ell:(0,\ \infty)\times(0,\ \infty)\to\mathbb{R}$ 局部有界且联合利普希茨连续，即对于任意 $K>0$，存在常数 $B_\ell=B_\ell(K)\ L_\ell=L_\ell(K)\ >0$ 使得对所有$(u,\ v),\ (u',\ v')\ \in(0,\ K)^2$,

$$|\ell(u,v)|\le B_\ell,\qquad |\ell(u,v)-\ell(u',v')|\le L_\ell\big(|u-u'|+|v-v'|\big).$$

最后，我们对假设类施加假设。

(G) 假设类 $\mathcal{G}$ 满足以下假设：存在常数 $c_{\mathcal{G}}>0$ 和 $a_{\mathcal{G}}>0$，使得对所有 $g\in\mathcal{G}$,

$$c_{\mathcal{G}}e^{-a_{\mathcal{G}}\|y\|^2}\le g(y)\le c_{\mathcal{G}}^+,\qquad \forall y\in\mathbb{R}^d\,. \tag{3.1}$$

注意，*(G)* 要求所有候选 $g \in \mathcal{G}$ 一致上有界，并且至关重要的是，以高斯为下界。该下界可防止归一化因子 D_g 中的比值 $q_T(x,z)/g(z)$所引起的不稳定性，并为经验过程论证提供易于处理的包络函数。下文我们还假设$b_T^+ > 4a_G$，以确保这些包络函数在 ρ_T下平方可积。

定理 3.1。 假设满足(Q)、(R0)、(RT)、(L)、(G)，且 $b_T^+ > 4a_{\mathcal{G}}$ *and* $K = c_{\mathcal{G}}^+(1+(c_+/D)$。那么，对所有 N、$M \geq 1$，均有

$$\mathbb{E}[\mathcal{R}(\widehat{g}_{N,M})] \leq \inf_{g\in\mathcal{G}} \mathcal{R}(g) + 2\mathbb{E}\Big[\sup_{g\in\mathcal{G}} \big|\widehat{\mathcal{R}}_{N,M}(g) - \mathcal{R}(g)\big|\Big] \tag{3.2}$$

and

$$\mathbb{E}\Big[\sup_{g\in\mathcal{G}} \big|\widehat{\mathcal{R}}_{N,M}(g) - \mathcal{R}(g)\big|\Big] \leq C_1 \left(\sqrt{\frac{\log\log(N)}{N}} + \frac{1}{\sqrt{M}}\right) \int_0^{C_2} \sqrt{\log \mathcal{N}_{[]}\Big(\mathcal{G}, \|\cdot\|_\infty, \frac{\varepsilon}{C_3}\Big)}\, d\varepsilon\,,$$

其中 C_1、C_2 和 C_3 是正数常量，其取值取决于假设(Q)、(R0)、(RT)、(L)和(G)中的常数。

证明概要。 详细证明见B节。首先注意到，(3.2)由

$$\mathcal{R}(\widehat{g}_{N,M}) - \inf_{g\in\mathcal{G}} \mathcal{R}(g) = \mathcal{R}(\widehat{g}_{N,M}) - \widehat{\mathcal{R}}(\widehat{g}_{N,M}) + \inf_{g\in\mathcal{G}} \widehat{\mathcal{R}}(g) - \inf_{g\in\mathcal{G}} \mathcal{R}(g)\,.$$

为证明集中界，我们将其改写为

$$\widehat{\mathcal{R}}_{N,M}(g) - \mathcal{R}(g) = \big(\widehat{\mathcal{R}}_{N,M}(g) - \widetilde{\mathcal{R}}_M(g)\big) + \big(\widetilde{\mathcal{R}}_M(g) - \mathcal{R}(g)\big)\,,$$

其中 $\widetilde{\mathcal{R}}_M(g) := \frac{1}{M}\sum_{j=1}^M \ell\big(g(Y_j)\ \mathcal{C}[g](Y_j)\big)$ 是一个中间风险。

步骤 1 令 $\mathcal{F} := \{f_g :\ g \in \mathcal{G}\}$，其中 $f_g(y) := \ell(g(y)\ \mathcal{C}[g](y))$。引理 B.1 表明 $\mathcal{C}[g]$ 是利普希茨连续且有界的。更准确地说，存在一个常数$L_{\mathcal{C},\infty} > 0$ 使得

$$|\mathcal{C}[g](y) - \mathcal{C}[g'](y)| \leq L_{\mathcal{C},\infty}\|g - g'\|_{L^\infty(\mathbb{R}^d)}, \quad |\mathcal{C}[g](y)| \leq (c_+/\underline{D})\|g\|_{L^\infty(\mathbb{R}^d)}.$$

根据假设 *(L)* 以及任意 $g \in \mathcal{G}$的有界性，

$$|f_g(y) - f_{g'}(y)| \leq L_\ell \Big(|g(y) - g'(y)| + |\mathcal{C}[g](y) - \mathcal{C}[g'](y)|\Big)$$

与 $L_\ell = L_\ell(K)$. 结合 *(G)* 和 *(RT)*，我们得到 $\mathcal{F}$ 具有包络函数 F，其 $\|F\|_\infty \leq B_\ell = B_\ell(K)$。 Gin´e 和 Nickl [2016][定理 3.5。13]意味着

$$\mathbb{E}[\sup_{g\in\mathcal{G}} |\widetilde{\mathcal{R}}_M(g) - \mathcal{R}(g)|] \lesssim \frac{1}{\sqrt{M}} \int_0^{8B_\ell} \sqrt{\log(2N_{[]}(\mathcal{G}, L^2(\rho_T), \varepsilon/L_\ell))}\, d\varepsilon.$$

步骤 2 根据假设*(L)*,

$$\mathbb{E}[\sup_{g\in\mathcal{G}}|\widehat{\mathcal{R}}_{N,M}(g)-\widetilde{\mathcal{R}}_M(g)|]\le L_\ell\,\mathbb{E}\left[\sup_{g\in\mathcal{G}}\frac{1}{M}\sum_{j=1}^{M}\big|\widehat{\mathcal{C}}_{N,M}[g](Y_j)-\mathcal{C}[g](Y_j)\big|\right].$$

固定一个较大的紧集 $K\subset\mathbb{R}^d$（稍后定义），并考虑

$$\mathbb{E}\big[\sup_{g\in\mathcal{G}}\sup_{y\in K}\big|\widehat{\mathcal{C}}_{N,M}[g](y)-\mathcal{C}[g](y)\big|\big]\le\mathbb{E}[\sup_{g\in\mathcal{G}}\sup_{y\in K}|(\mathrm{A})|]+\mathbb{E}[\sup_{g\in\mathcal{G}}\sup_{y\in K}|(\mathrm{B})|],$$

其中 $(A)=\widehat{\mathcal{C}}_{N,M}[g](y)-\widetilde{\mathcal{C}}_N[g](y)$，$(B)=\widetilde{\mathcal{C}}_N[g](y)-\mathcal{C}[g](y)$ 且

$$\widetilde{\mathcal{C}}_N[g](y):=\frac{1}{N}\sum_{i=1}^{N}\frac{q_T(X_i,y)}{D_g(X_i)}.$$

对于 (B)，我们定义集合 $\Phi_K:=\{x\mapsto\phi_{g,y}(x):\ g\in\mathcal{G},y\in K\}$，其中

$$\phi_{g,y}(x):=q_T(x,y)/D_g(x).$$

于是可以验证

$$\log\mathcal{N}_{[]}\big(\Phi_K,L^2(\rho_0),\varepsilon\big)\ \le\ \log\mathcal{N}_{[]}\Big(\mathcal{G},\|\cdot\|_\infty,\varepsilon/2L_g\Big)\ +\ C_d\,\log\Big(L_y\,\mathrm{diam}(K)/\varepsilon\Big),$$

其中 $C_d\le d$ 是一个维数常数（来自对 K 在欧几里得范数下的覆盖），$L_g:=c_+\,L_{D,\infty}/D^2$以及 $L_y:=L_q/D$，其中$L_{D,\infty}$ 定义于引理 B.1 且 L_q 定义于 *(Q)*。随后由 Gin´e 与 Nickl [2016][定理 3.5.13],

$$\sup_{g\in\mathcal{G}}\sup_{y\in K}|(\mathrm{B})|\lesssim N^{-1/2}\int_0^{8F_\infty}\sqrt{\log\mathcal{N}_{[]}\Big(\mathcal{G},\|\cdot\|_\infty,\varepsilon/(2L_g)\Big)}\,d\varepsilon+N^{-1/2}\sqrt{\log\Big(L_y\,\mathrm{diam}(K)/F_\infty\Big)}.$$

对于 (A)，我们得到

$$\mathbb{E}\left[\sup_{g\in\mathcal{G}}\sup_{y\in K}|(\mathrm{A})|\right]\le c_+L_D\,\mathbb{E}\left[\sup_{g\in\mathcal{G}}\frac{1}{N}\sum_{i=1}^{N}\big|\widehat{D}_g(X_i)-D_g(X_i)\big|\right],$$

其中 $L_D:=D^2/D^2_\circ$。令 $\Psi:=\{(x,z)\mapsto\psi_g(x,z):\ x\in B(x_0,R_0)\ \ g\in\mathcal{G}\}$，其中

$$\psi_g(x,z):=q_T(x,z)/g(z).$$

注意，$\psi_g(x,z)\le F(z):=\frac{c_+}{c_\mathcal{G}}\,e^{a_\mathcal{G}\|z\|^2}$，并且

$$\|F\|_{L^2(\rho_T)}\le\frac{c_+}{c_\mathcal{G}}\sqrt{c_T^+}\left(\frac{\pi}{b_T^+-2a_\mathcal{G}}\right)^{d/4}.$$

作者： Gin´e 和 Nickl [2016][定理 3.5。13],

$$\mathbb{E}\Big[\sup_{x\in B(x_0,R_0)}\sup_{g\in\mathcal{G}}\big|\widehat{D}_g(x)-D_g(x)\big|\Big]\lesssim\frac{1}{\sqrt{M}}\int_0^{8\|F\|_{L^2(\rho_T)}}\sqrt{\log(2N_{[]}(\Psi,L^2(\rho_T),\epsilon))}\,d\varepsilon.$$

接下来需要估计前一个不等式右侧的括号数。注意，可以证明：

$$\log \mathcal{N}_{[]}\big(\Psi, L^2(\rho_T), \varepsilon\big) \;\le\; \log \mathcal{N}_{[]}\Big(\mathcal{G}, \|\cdot\|_\infty, \varepsilon/C_g\Big) \;+\; d\,\log\Big(C\,R_0/\varepsilon\Big),$$

其中 $C>0$ 是某个绝对常数，$C_g := c_+/(c_{\mathcal{G}}^2)\,M_\rho$，并且

$$M_\rho = \Big(\int e^{4a_{\mathcal{G}}\|z\|^2}\rho_T(z)\,dz\Big)^{1/2}.$$

步骤 3 最后，我们取 $K = B(0,\ R)$，其中 $R > R_0$。利用 $D_g(x) \ge \underline{D}$，以及 ρ_0 是概率密度，

$$|\mathcal{C}[g](y)| \le \frac{1}{\underline{D}}\int_{B(0,R_0)} \rho_0(x)\,q_T(x,y)\,dx \le \frac{c_+}{\underline{D}}\exp\big(-a_+(R-R_0)^2\big).$$

此外，利用 $\mathcal{T}_{[\underline{D},\overline{D}]}[D_g](X_i) \ge \underline{D}$ 和 $X_i \in B(0,\ R_0)$，

$$|\widehat{\mathcal{C}}_{N,M}[g](y)| \le \frac{1}{\underline{D}}\sup_{x\in B(0,R_0)} q_T(x,y) \le \frac{c_+}{\underline{D}}\exp\big(-a_+(R-R_0)^2\big).$$

因此，对于任意 $R > R_0$，

$$\mathbb{E}\left[\sup_{g\in\mathcal{G}}\sup_{y\notin B(0,R)}\big|\widehat{\mathcal{C}}_{N,M}[g](y) - \mathcal{C}[g](y)\big|\right] \le \frac{2c_+}{\underline{D}}\,\exp\big(-a_+(R-R_0)^2\big).$$

剩下的工作是选择阶为 $\sqrt{\log\ N}$ 的 R。 □

3.1 逼近误差

在本节中，我们证明，通过选择适当的函数类 $\mathcal{G}$，可以使$\mathbb{E}[\mathcal{R}(\widehat{g}_{N,M})]$取较小的值。为简单起见，我们假设转移核是高斯的。回想该函数具有如下形式

$$g^\star(y) = \int_{\mathbb{R}^d} w(x)\,q(y-x)\,dx$$

其中 $w(x) = \rho_0(x)/D_g(x)$，也就是说，它表示紧支撑函数与高斯核的卷积。在这种情况下，Hermite多项式类是函数类 G的自然候选。我们已将关于Hermite多项式的主要结果置于附录中，特别是 $g^\star$的Hermite系数估计。更确切地说，假设如下条件。

($\mathrm{Q}_{\mathrm{Gauss}}$) 参考核是高斯的：

$$q_T(z) = (2\pi T)^{-d/2}\exp\left(-\frac{\|z\|^2}{2T}\right), \qquad z\in\mathbb{R}^d.$$

下面的定理给出了 $\mathbb{E}[\mathcal{R}(\widehat{g}_{N,\ M})]$ 的一个界。

定理 3.2。 假设条件 (Q_{Gauss})、(R0)、(RT)，且满足 $b_T^+ > 4/T$ 和 (L)。那么存在一个类 $\mathcal{G}$，使得 (G)得到满足，并且

$$\mathbb{E}[\mathcal{R}(\widehat{g}_{N,M})] \lesssim \log^{d/2} \max(M,N)\Big(N^{-1/2} + M^{-1/2}\Big),$$

其中 $\lesssim$ 表示 不等式中与 M 和 d 无关的常数，以及双对数因子 *rs.*

证明。 利用 $\mathcal{R}(g^\star) = 0$ 和推论 C.2 可得

$$\inf_{g\in\mathcal{G}} \mathcal{R}(g) \le L_\ell\big(1 + L_{\mathcal{C},2}\big) \inf_{g\in\mathcal{G}} \|g - g^\star\|_{L^2(\rho_T)}. \tag{3.3}$$

设 $\{\psi_\alpha^{(\lambda)}\}_{\alpha\in\mathbb{N}_0^d}$ 为满足 $\lambda = T^{-1/2}$的缩放厄米基。设 Π_n 为到张成空间$\{\psi_\alpha^{(\lambda)}$上的 $L^2(\mathbb{R}^d)$–正交投影算子： $|\alpha| \le n\}$。定义

$$\mathcal{G} = \mathcal{G}_n(B) := \left\{ g = \sum_{|\alpha|\le n} c_\alpha \psi_\alpha^{(\lambda)} : \sum_{|\alpha|\le n} c_\alpha^2 \le B \right\}.$$

我们取 $B \asymp nd^d$。由命题 C.5，存在某个常数 $C > 0$ ，使得对于 $n \ge C\log\max(M,N)$, $\Pi_n g^\star \in \mathcal{G}$, 以及

$$\|g^\star - \Pi_n g^\star\|_{L^2(\mathbb{R}^d)} \lesssim \log^{d/2}\max(M,N)\Big(N^{-1/2} + M^{-1/2}\Big). \tag{3.4}$$

注意，由命题 A.1， $g^\star(y) \in \big[c_-^\star e^{-a_-^\star\|y\|^2}, c_+^\star\big] := I$ 对所有 y 根据 (A.2)。我们需要确保条件 (3.1)。定义截断算子 Π^{clip}： $\mathbb{R}\to\mathbb{R}$ 为

$$\Pi^{\text{clip}}(t) := \min\Big\{c_+^\star, \max\big\{t, c_-^\star e^{-a_-^\star\|y\|^2}\big\}\Big\}.$$

那么，对于任意 $f\in\mathcal{G}$， $\Pi^{\text{clip}}(f)(y)$ 均满足（3.1）。对于每个固定的 y，映射 $t\mapsto\Pi(t)$ 是到区间 I上的度量投影，因此它是1–利普希茨的。由此可得，对于任意 $y\in\mathbb{R}^d$，逐点收缩性成立。

$$|g^\star(y) - \Pi^{\text{clip}}(\Pi_n g^\star(y))| = \big|\Pi^{\text{clip}}(g^\star(y)) - \Pi^{\text{clip}}(\Pi_n g^\star(y))\big| \le |g^\star(y) - \Pi_n g^\star(y)|,$$

特别地，截断只能减小 任意 $L^2(\mathbb{R}^d)$ 误差:

$$\|g^\star(y) - \Pi^{\text{clip}}(\Pi_n g^\star(y))\|_{L^2(\mathbb{R}^d)} \le \|g^\star(y) - \Pi_n g^\star(y)\|_{L^2(\mathbb{R}^d)}.$$

由最后一个不等式、（3.3）和（3.4）可得

$$\inf_{g\in\mathcal{G}} \mathcal{R}(g) \lesssim \log^{d/2}\max(M,N)\Big(N^{-1/2} + M^{-1/2}\Big).$$

注意，对于选定的 n，引理C.7 和命题C.3意味着存在某个常数 $C' > 0$ 使得

$$\log \mathcal{N}_{[]}\Big(\mathcal{G}, \|\cdot\|_\infty, \varepsilon\Big) \lesssim \log^{d/2}\max(M,N)\log(C'/\varepsilon).$$

因此，我们可以应用定理 3.1 得出该结论。 □

4 数值实验

本节将对所提出的算法进行实验分析，并将其与 SinkhornBridge [Pooladian 和 Niles-Weed, 2024]进行比较。选择该算法进行比较，是因为其训练过程与我们的方法最为接近。然而，由于我们的算法学习的是连续的 Schr¨odinger 对数势函数，而非最优传输问题的离散解，因此我们能够证明，该算法在多项生成建模和数据到数据的转换任务中具有显著优势。下文中，为简洁起见，我们将所提出的算法称为 ERM-Bridge。该算法的正式描述、超参数值以及其他实验细节见附录 D.6。

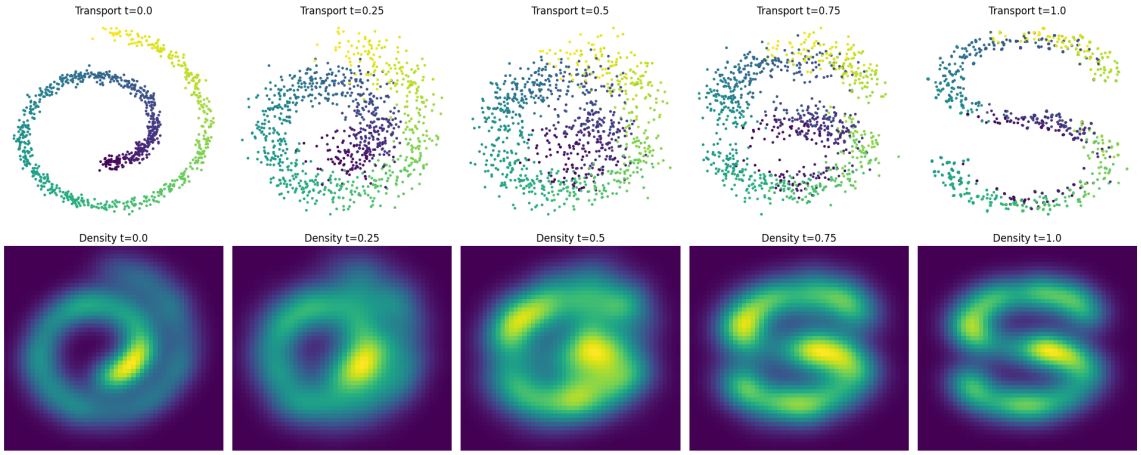


图 1：从瑞士卷到 S 曲线的样本迁移，以及时间 $t \in [0, 0.25, 0.5, 0.75, 1]$. 时 ERM-Bridge 的密度图

瑞士卷到 S 曲线实验 我们在经典的二维示例上评估了该算法的有效性。在此实验中，我们考察了瑞士卷分布向 S 曲线的迁移，并展示了时间 $t \in [0, 0.25, 0.5, 0.75, 1]$ 时的分布，详见图 1。图中清楚地表明，该算法能够可靠地学习分布和正确的密度图（密度是通过在二维曲面上使用核密度估计近似得到的）。还值得注意的是，该算法在此问题上展现出最佳的训练和采样时间。结果和超参数见附录D.6。

高斯混合的评估 为进一步说明网络近似对数势的重要性，我们在高斯混合问题上评估了该算法。在此实验中，我们使用了由均匀网格上的 25 个高斯分布组成的混合分布，其协方差矩阵相同且为标准协方差矩阵。我们使用 Wasserstein 距离的切片近似来衡量传输质量 $\mathbb{W}_1$ 作为度量，度量定义见附录D.2。

在这项实验中，两种算法均使用来自定义在$[-10, 10]$上的截断正态分布的 *3000* 个样本进行训练，并以来自定义在 $[-1, 1]$上的截断正态分布的起始点作为采样点 $[-2, 2]$。 $[-5, 5]$.这项实验展示了训练集与采样集中的起始分布可能不同时，各算法的表现情况，见图2。这种情况在数据到数据的

传输中相当常见，因为起始分布和目标分布都只能通过样本获得，而且无法确保训练集与测试集完全一致。显然，与离散最优传输 SinkhornBridge 相比，我们的算法通过神经网络学习到的连续势函数表现更好。基线在 $\mathbb{W}_1 = 1.3115$ 情形下达到 $[-1, 1]$，而我们的算法达到 $\mathbb{W}_1 = 0.3818$. 。超参数值见附录D.6。

图 2：ERM-Bridge 与 SinkhornBridge 的切片 Wasserstein 距离随训练分布与采样分布之间 KL散度 变化的曲线图。

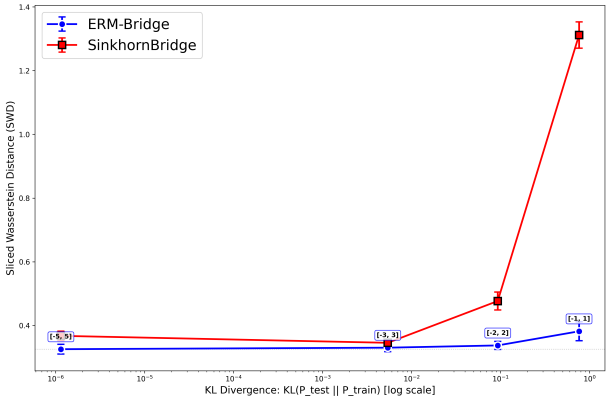


表 1：包括 ERM-Bridge 和 SinkhornBridge 在内的多种算法在单细胞数据上的中间分布恢复质量。

求解器	$\mathbb{W}_1$ 指标
OT-CFM [Tong 等人, 2023年a]	0.790 ± 0.068
[SF]^{2}M-Exact [Tong 等人, 2023年b]	0.793 ± 0.066
ERM-Bridge	0.809 ± 0.030
LightSB-OU [Puchkin 等人, 2025]	0.815 ± 0.016
SinkhornBridge [Pooladian 和 Niles-Weed, 2024]	0.818 ± 0.028
LightSB [Korotin 等人, 2024]	0.823 ± 0.017
Reg. CNF [Finlay 等人, 2020]	$0.825 \pm$ 不适用[a]
T. Net [Tong 等人, 2020]	$0.848 \pm$ 不适用[a]
DSB [De Bortoli 等人, 2021]	0.862 ± 0.023
I-CFM [Tong 等人, 2023a]	0.872 ± 0.087
[SF]^{2}M-Geo [Tong 等人, 2023b]	0.879 ± 0.148
NLSB [Koshizuka 和 Sato, 2022]	$0.970 \pm$ 不适用[a]
[SF]^{2}M-Sink [Tong 等人, 2023b]	1.198 ± 0.342
SB-CFM [Tong 等人, 2023a]	1.221 ± 0.380
DSBM [Shi 等人, 2023]	1.775 ± 0.429

[a]作者未报告标准差。

单细胞数据评估 为了与 SinkhornBridge 进行更全面的比较，我们在生物数据上进行了实验 [Tong 等人, 2020]。按照原论文，我们将问题表述为将时间 t_{i-1} 的细胞分布输送至时间 t_{i+1}，持续 $i \in \{1, 2, 3\}$。我们采用了其他方法的结果，出处为 [Tong 等人, 2023b],，其作者是最早在 [Tong 等人, 2020]中考虑这一设置的。随后，我们预测中间时间 t_i 的细胞分布，并计算预测分布与真实值之间的 $\mathbb{W}_1$ Wasserstein 距离。对所有三种设置的结果取平均（$i = 1, 2, 3$），见表 1。

为确保统计稳健性，我们重复进行了五次实验。基线的表现为 $\mathbb{W}_1$ =0.818，而我们的算法的表现为 $\mathbb{W}_1 = 0.809$。超参数值见附录D.6。

5 结论

我们研究了从有限样本中估计 Schr¨odinger 桥势函数的统计问题。我们将 Schr¨odinger 系统改写为关于变换势的单个非线性不动点方程 $g = C[g]$，并提出了一种经验风险最小化估计量：通过在假设类上最小化经验不动点残差得到该估计量。由此构造出的估计量必然是连续势函数估计量，并且自然地与 Schr¨odinger 桥的随机控制表示相结合。未来仍有若干研究方向，包括将结果推广到次高斯尾部假设之外，以及推导 $g_{M,\,N}$ 与 $g^\star$ 之间误差的界。

后者需要研究 $\mathcal{C}[g]$ 在希尔伯特度量下 $g^\star$ 附近的导数局部行为（谱隙条件）。

参考文献

T. Akiba, S. Sano, T. Yanase, T. Ohta 和 M. Koyama。Optuna：新一代超参数优化框架。载于第 25 届 ACM SIGKDD知识发现与数据挖掘国际会议，页码 2623–2631，2019。Y. Chen, T. T. Georgiou 和 M. Pavon。熵插值与位移插值：一种基于希尔伯特度量的计算方法。《SIAM应用数学杂志》，76(6):2375–2396，2016。doi：10.1137/ 16M1061382。URL https://doi.org/10.1137/16M1061382。G. Conforti, A. Durmus 和 G. Greco。Sinkhorn 算法的定量收缩率：超越有界代价与紧边缘分布。arXiv 电子预印本，2023 年 4 月。doi：10.48550/arXiv.2304.04451。URL https://doi.org/10.48550/arXiv.2304.04451。M. Cuturi。Sinkhorn距离：以闪电般的速度计算最优传输。载于《神经信息处理系统进展》，第 26 卷，页码 2292–2300，2013。URL https://papers.nips.cc/paper/4927-sinkhorn-distances-lightspeed-computation-of-optimal-transport。P. Dai Pra。一种研究互易扩散过程的随机控制方法。《应用数学与优化》，23(1):313–329，1991。doi：10.1007/BF01445134。V. De Bortoli, J. Thornton, J. Heng 和 A. Doucet。扩散 Schr¨odinger 桥及其在基于分数的生成建模中的应用。载于《神经信息处理系统进展》，第 34 卷，页码 17695–17709，2021。C. Finlay, J.-H. Jacobsen, L. Nurbekyan 和 A. Oberman。如何训练神经常微分方程：雅可比正则化与动能正则化的世界。载于第 37 届国际机器学习会议论文集，《机器学习研究论文集》第 119 卷，页码 3154–3164。PMLR，2020。URL https://proceedings.mlr.press/v119/finlay20a.html。J. Franklin 和 J. Lorenz。多维矩阵的缩放。《线性代数及其应用》，114–115:717–735，1989。doi：10.1016/0024-3795(89)90490-4。URL https://doi.org/10. 1016/0024-3795(89)90490-4。E. Gin´e 和 R. Nickl。《无限维统计模型的数学基础》，剑桥统计与概率数学丛书第 40 卷。剑桥大学出版社，2016。ISBN 9781107043169。doi：10.1017/CBO9781107337312。G. Greco, M. Noble, G. Conforti 和 A. Durmus。Sinkhorn迭代序列及其梯度的非渐近收敛界：一种耦合方法。载于第 36 届学习理论会议论文集，《机器学习研究论文集》第 195 卷，页码 716–746。PMLR，2023。URL https: // proceedings.mlr.press/v195/greco23a.html。A. Korotin, N. Gushchin 和 E. Burnaev。轻量级 Schr¨odinger 桥。载于国际学习表征会议，2024。doi：10.48550/arXiv.2310.01174。URL https://doi.org/10.48550/ arXiv.2310.01174。T. Koshizuka 和 I. Sato。神经拉格朗日 Schr¨odinger 桥：用于种群动

力学建模的扩散建模。arXiv 预印本 arXiv:2204.04853，2022。doi：10.48550/arXiv:2204.04853。URL https: //doi.org/10.48550/arXiv:2204.04853。C. Leonard。Schr¨odinger 问题及其与最优传输若干联系的综述。《离散与连续动力系统》，34(4):1533–1574，2014。G. Peyr´e 和 M. Cuturi。计算最优传输。《机器学习基础与趋势》，11(5–6):355–607，2019 年 2 月。doi：10.1561/2200000073。URL https://doi.org/10.1561/ 2200000073。A.-A. Pooladian 和 J. Niles-Weed。Schr¨odinger 桥的插件估计。arXiv 预印本 arXiv:2408.11686，2024。doi：10.48550/arXiv:2408.11686。URL https://doi.org/10.48550/arXiv.2408.11686。N. Puchkin, D. Suchkov, A. Naumov 和 D. Belomestny。非配对数据转换中 Schr¨odinger 势函数估计的紧界，2025。URL https://arxiv.org/abs/2508.07392。¨E. Schr¨odinger。自然规律的逆转。普鲁士科学院会议录，物理数学部，页码 144–153，1932。Y. Shi, V. De Bortoli, A. Campbell 和 A. Doucet。扩散 Schr¨odinger 桥匹配。arXiv 预印本 arXiv:2303.16852，2023。doi：10.48550/arXiv:2303.16852。URL https://doi.org/10.48550/arXiv.2303.16852。A. Tong, J. Huang, G. Wolf, D. van Dijk 和 S. Krishnaswamy。TrajectoryNet：一种用于细胞动力学建模的动态最优传输网络。载于第 37 届国际机器学习会议论文集，《机器学习研究论文集》第 119 卷，页码 9526–9536。PMLR，2020。URL https://proceedings.mlr.press/v119/tong20a.html。A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf 和 Y. Bengio。通过小批量最优传输改进并泛化基于流的生成模型。arXiv 预印本 arXiv:2302.00482，2023a。doi：10.48550/arXiv:2302.00482。URL https://doi.org/10.48550/ arXiv.2302.00482。2023a。

A. Tong、N. Malkin、K. Fatras、L. Atanackovic、Y. Zhang、G. Huguet、G. Wolf 和 Y. Bengio。通过分数匹配和流匹配实现无需模拟的 Schr¨ödinger 桥。 *arXiv* 预印本 *arXiv:2307.03672*，2023b。doi：10.48550/arXiv.2307.03672。URLhttps://doi.org/10.48550/arXiv.2307.03672。2023b。

次高斯假设下经验算子的性质

命题 A.1 （不动点的双侧高斯界）。 假设 (Q),(R0), (RT)。设 $g^\star$： $\mathbb{R}^d \to (0,\infty)$ 可测且满足

$$g^\star(y) = \mathcal{C}[g^\star](y), \qquad y \in \mathbb{R}^d, \qquad \textit{and} \qquad \int_{\mathbb{R}^d} \frac{1}{g^\star(z)}\, \rho_T(z)\, dz = 1. \tag{A.1}$$

令 $a_-^\star := 2a_-$, $a_+^\star := \frac{a_+}{2}$则存在常数 $c_-^\star, c_+^\star > 0$， 它们仅依赖于 q_T, ρ_0, ρ_T（而不依赖于$g^\star$），使得

$$c_-^\star \exp(-a_-^\star \|y\|^2) \;\le\; g^\star(y) \;\le\; c_+^\star \exp(-a_+^\star \|y\|^2), \qquad \forall y \in \mathbb{R}^d. \tag{A.2}$$

注释 A.2. 需要注意的是，函数 $g^\star$ 的行为仅由对转移核 q和密度 ρ_0, ρ_T的条件决定。

证明。 对每个 $x \in \mathbb{R}^d$，定义

$$D_{g^\star}(x) := \int_{\mathbb{R}^d} q_T(x,z)\,\frac{\rho_T(z)}{g^\star(z)}\,dz.$$

于是，不动点方程可写为

$$g^\star(y) = \int_{\mathbb{R}^d} \frac{\rho_0(x)}{D_{g^\star}(x)}\,q_T(x,y)\,dx.$$

我们声称存在常数 $0 < \underline{D}^\star \le D^\star < \infty$，它们仅取决于 q_T、ρ_0、ρ_T，使得

$$\underline{D}^\star \;\le\; D_{g^\star}(x) \;\le\; \overline{D}^\star, \qquad \forall x \in \operatorname{supp}(\rho_0). \tag{A.3}$$

由 *(Q)*,

$$q_T(x,z) \le c_+ \exp(-a_+\|x-z\|^2) \le c_+,$$

因此，利用 (A.1) 中的归一化,

$$D_{g^\star}(x) = \int q_T(x,z)\,\frac{\rho_T(z)}{g^\star(z)}\,dz \le c_+ \int \frac{\rho_T(z)}{g^\star(z)}\,dz = c_+. \tag{A.4}$$

因此，我们可以取

$$\overline{D}^\star := c_+. \tag{A.5}$$

定义

$$m(x) := \int_{\mathbb{R}^d} q_T(x,z)\,\rho_T(z)\,dz.$$

利用*(Q)*、*(RT)* 以及 $B(x_0, R_0)$ 的紧性，可验证

$$m_- \;\le\; m(x) \;\le\; m_+, \qquad \forall x \in \operatorname{supp}(\rho_0)$$

其中 $m_+ := c_+$,并且

$$m_- := c_- c_T^- \exp\big(-2a_- R_\star^2\big)\Big(\frac{\pi}{2a_- + b_T^-}\Big)^{d/2}$$

其中 $R_\star := \|x_0\| + R_0$. 。这里用到了这样一个事实：对任意 $\gamma > 0$,

$$\int_{\mathbb{R}^d} \exp(-\gamma\|z\|^2)\,dz = \left(\frac{\pi}{\gamma}\right)^{d/2} < \infty, \tag{A.6}$$

and

$$\|x-z\|^2 \le 2\|x\|^2 + 2\|z\|^2\,. \tag{A.7}$$

对于每个这样的 x，定义一个概率测度 ν_x，如下所示

$$\nu_x(dz) := \frac{q_T(x,z)\,\rho_T(z)}{m(x)}\,dz.$$

Then

$$D_{g^\star}(x) = m(x)\,\mathbb{E}_{\nu_x}\Big[\frac{1}{g^\star(Z)}\Big].$$

现在比较 ν_x 和 ρ_T。拉东–尼科迪姆导数为

$$\frac{d\nu_x}{d\rho_T}(z) = \frac{q_T(x,z)}{m(x)}.$$

由*(Q)*中的双侧高斯界以及对 m 的界，存在常数 $0 < k_- \le k_+ < \infty$，使得

$$k_- \;\le\; \frac{d\nu_x}{d\rho_T}(z) \;\le\; k_+, \qquad \forall x \in \mathrm{supp}(\rho_0),\; z \in \mathbb{R}^d.$$

因此，对于任意非负可测的 h，有

$$k_- \int h(z)\,\rho_T(z)\,dz \;\le\; \int h(z)\,\frac{d\nu_x}{d\rho_T}(z)\,\rho_T(z)\,dz \;\le\; k_+ \int h(z)\,\rho_T(z)\,dz,$$

也就是说，

$$k_-\,\mathbb{E}_{\rho_T}[h(Z)] \;\le\; \mathbb{E}_{\nu_x}[h(Z)] \;\le\; k_+\,\mathbb{E}_{\rho_T}[h(Z)].$$

将 $h(z) = {g^\star}^{-1}(z) \ge 0$代入。利用 (A.1) 中的归一化，我们得到

$$k_- \le \mathbb{E}_{\nu_x}\Big[\frac{1}{g^\star(Z)}\Big] \le k_+.$$

因此，

$$D_{g^\star}(x) = m(x)\,\mathbb{E}_{\nu_x}\Big[\frac{1}{g^\star(Z)}\Big] \ge m_-\,k_- =: \underline{D}^\star > 0 \tag{A.8}$$

对所有 $x \in \mathrm{supp}(\rho_0)$成立。结合上界，这就证明了 (A.3)。利用不动点方程和 (A.3)，

$$g^\star(y) = \int_{\mathrm{supp}(\rho_0)} \frac{\rho_0(x)}{D_{g^\star}(x)}\,q_T(x,y)\,dx \;\le\; \frac{1}{\underline{D}} \int_{B(x_0,R_0)} \rho_0(x)\,q_T(x,y)\,dx.$$

根据*(Q)*中的高斯上界，

$$q_T(x,y) \le c_+ \exp\big(-a_+\|x-y\|^2\big).$$

利用

$$\|x-y\|^2 \ge \frac{1}{2}\|y\|^2 - \|x\|^2 \ge \frac{1}{2}\|y\|^2 - R_\star^2,$$

我们得到

$$q_T(x,y) \le c_+ \exp(a_+ R_\star^2)\,\exp\Big(-\frac{a_+}{2}\|y\|^2\Big), \qquad x \in B(x_0,R_0).$$

因此

$$g^\star(y) \le \frac{c_+ \exp(a_+R_\star^2)}{\underline{D}^\star} \int_{\mathrm{supp}(\rho_0)} \rho_0(x)\,dx\,\exp\Big(-\frac{a_+}{2}\|y\|^2\Big).$$

由于 ρ_0 是概率密度，$\int_{\mathrm{supp}(\rho_0)} \rho_0(x)\,dx = 1$，我们可以令

$$c_+^\star := \frac{c_+ \exp(a_+R_\star^2)}{\underline{D}^\star}, \qquad a_+^\star := \frac{a_+}{2},$$

以获得

$$g^\star(y) \;\le\; c_+^\star \exp(-a_+^\star\|y\|^2), \qquad \forall y \in \mathbb{R}^d.$$

再次利用不动点方程和（A.3），

$$g^\star(y) = \int_{\mathrm{supp}(\rho_0)} \frac{\rho_0(x)}{D_{g^\star}(x)}\, q_T(x,y)\, dx \;\geq\; \frac{1}{\overline{D}^\star} \int_{\mathrm{supp}(\rho_0)} \rho_0(x)\, q_T(x,y)\, dx.$$

根据*(R0)*,

$$g^\star(y) \geq \frac{\rho_{0,-}}{\overline{D}^\star} \int_{B(x_0,r_0)} q_T(x,y)\, dx.$$

利用*(Q)*中的高斯下界,

$$q_T(x,y) \geq c_- \exp\bigl(-a_- \|x-y\|^2\bigr),$$

并且 $\|x\| \leq \|x_0\| + r_0$ 对于 $x \in B(x_0, r_0)$，我们有

$$\|x-y\| \leq \|x\| + \|y\| \leq \|x_0\| + r_0 + \|y\|$$

因此

$$\|x-y\|^2 \leq (\|x_0\| + r_0 + \|y\|)^2 \leq 2r_\star^2 + 2\|y\|^2,$$

其中 $r_\star = \|x_0\| + r_0$。于是,

$$q_T(x,y) \geq c_- \exp\bigl(-a_-\,(2r_\star^2 + 2\|y\|^2)\bigr), \qquad x \in B(x_0, r_0).$$

由此可得

$$\int_{B(x_0,r_0)} q_T(x,y)\, dx \geq c_- \exp(-2a_- r_\star^2)\, \exp(-2a_-\|y\|^2)\, \lambda^d(B(x_0,r_0)),$$

其中 $\lambda^d(B(x_0,\ r_0))$ 是 $B(x_0,\ r_0)$ 的勒贝格测度。因此

$$g^\star(y) \geq \frac{\rho_{0,-}\, c_-\, \lambda^d(B(x_0,r_0))\, \exp(-2a_- r_\star^2)}{\overline{D}^\star} \exp(-2a_-\|y\|^2).$$

令

$$c_-^\star := \frac{\rho_{0,-}\, c_-\, \lambda^d(B(x_0,r_0))\, \exp(-2a_- r_\star^2)}{\overline{D}^\star}, \qquad a_-^\star := 2a_-,$$

我们得到

$$g^\star(y) \;\geq\; c_-^\star \exp(-a_-^\star \|y\|^2), \qquad \forall y \in \mathbb{R}^d.$$

结合上界和下界，即完成了 (A.2) 的证明。 □

B 经验风险的一致集中

引理 B.1。 假设(Q)、(R0)、(RT) 和(G)成立。此外，假设 $b_T^+ > 2a_{\mathcal{G}}$。那么存在常数 $\underline{D}$、 $\overline{D}$、$L_{D,\,\infty}$、$L_{\mathcal{C},\,\infty} > 0$，使得

$$\underline{D} \;\leq\; D_g(x) \;\leq\; \overline{D}, \qquad x \in B(x_0, R_0),\ g \in \mathcal{G}; \tag{B.1}$$

$$\|D_g - D_h\|_{L^\infty(B(x_0,R_0))} \leq L_{D,\infty} \|g-h\|_\infty; \tag{B.2}$$

$$\|\mathcal{C}[g] - \mathcal{C}[h]\|_\infty \leq L_{\mathcal{C},\infty}\, \|g-h\|_{L^\infty(\mathbb{R}^d)}; \tag{B.3}$$

$$\|\mathcal{C}[g]\|_\infty \leq (c_+/\underline{D})\, \|g\|_{L^\infty(\mathbb{R}^d)}\,. \tag{B.4}$$

注意， $\underline{D}, \overline{D}$ 不依赖于类 $\mathcal{G}$。

证明。 (B.1)的证明重复了命题 A.1的证明。我们做出必要的修改。首先，我们将(A.4)替换为

$$D_g(x) = \int q_T(x,z)\,\frac{\rho_T(z)}{g(z)}\,dz \le c_+ c_T^+ c_{\mathcal{G}}^{-1} \int \exp\Big(-\Big(b_T^+ - a_{\mathcal{G}}\Big)\|z\|^2\Big)\,dz$$
$$= c_+ c_T^+ c_{\mathcal{G}}^{-1} \left(\frac{\pi}{b_T^+ - a_{\mathcal{G}}}\right)^{d/2} =: \overline{D}\,.$$

回顾（A.8）的证明。注意到

$$\mathbb{E}_{\rho_T}[1/g] \ge (c_{\mathcal{G}}^+)^{-1}\,,$$

并且我们可以取 $\underline{D} := m_-\, k_-(c_{\mathcal{G}}^+)^{-1}$。我们证明（B.2）。根据定义，

$$D_g(x) - D_h(x) = \int_{\mathbb{R}^d} q_T(x,z)\,\rho_T(z)\left(\frac{1}{g(z)} - \frac{1}{h(z)}\right)dz.$$

利用

$$\left|\frac{1}{g(z)} - \frac{1}{h(z)}\right| = \frac{|g(z)-h(z)|}{g(z)h(z)} \le \|g-h\|_\infty \cdot \frac{1}{g(z)h(z)}$$

以及 *(G)* 中的下界（分别应用于 g 和 h），可得对所有 z,

$$\frac{1}{g(z)h(z)} \le c_{\mathcal{G}}^{-2} \exp(2a_{\mathcal{G}}\|z\|^2).$$

因此

$$|D_g(x) - D_h(x)| \le \|g-h\|_\infty\,(c_{\mathcal{G}})^{-2} \int_{\mathbb{R}^d} q_T(x,z)\,\rho_T(z)\,\exp(2a_{\mathcal{G}}\|z\|^2)dz. \tag{B.5}$$

现在我们对积分进行关于 $x \in B(x_0, R_0)$ 的一致界定。根据 *(Q)* 以及不等式

$$\|x-z\|^2 \ge \frac{1}{2}\|z\|^2 - \|x\|^2,$$

对于所有 $x \in B(x_0,\ R_0)$,

$$q_T(x,z) \le c_+ \exp(-a_+\|x-z\|^2) \le c_+ \exp(a_+ R_\star^2)\exp\Big(-\frac{a_+}{2}\|z\|^2\Big)$$

其中 $R_\star := \|x_0\| + R_0$. 将其与 *(RT)* 结合可得

$$q_T(x,z)\,\rho_T(z)\,\exp(2a_{\mathcal{G}}\|z\|^2) \le c_+ c_T^+ \exp(a_+R_\star^2)\,\exp\Big(-\Big(b_T^+ - 2a_{\mathcal{G}} + \frac{a_+}{2}\Big)\|z\|^2\Big)\,.$$

根据（A.6），对所有 $x \in B$（$x_0,\ R_0$），

$$|D_g(x) - D_h(x)| \le L_{D,\infty}\|g-h\|_\infty\,,$$

其中

$$L_{D,\infty} := c_{\mathcal{G}}^{-2} c_+ c_T^+ \exp(a_+R_\star^2)\left(\frac{\pi}{b_T^+ - 2a_{\mathcal{G}} + a_+/2}\right)^{d/2}. \tag{B.6}$$

这就证明了（B.2）。对于每个 $y \in \mathbb{R}^d$,

$$\mathcal{C}[g](y) - \mathcal{C}[h](y) = \int_{\mathrm{supp}(\rho_0)} \rho_0(x)\, q_T(x,y) \left(\frac{1}{D_g(x)} - \frac{1}{D_h(x)} \right) dx.$$

由（B.1）和（B.2）可得

$$\begin{aligned} |\mathcal{C}[g](y) - \mathcal{C}[h](y)| &\le \frac{L_{D,\infty}\|g-h\|_\infty}{\underline{D}^2} \int_{\mathrm{supp}(\rho_0)} \rho_0(x)\, q_T(x,y)\, dx \\ &\le L_{\mathcal{C},\infty}\|g-h\|_\infty, \end{aligned}$$

其中

$$L_{\mathcal{C},\infty} = L_{D,\infty} c_+ \,. \tag{B.7}$$

这证明了（B.3）。

□

定理 *3.1* 的证明。 定义一个在 Y_j 中使用经验平均、但使用总体算子 $\mathcal{C}[g]$ 的中间风险：

$$\widetilde{\mathcal{R}}_M(g) := \frac{1}{M} \sum_{j=1}^{M} \ell\big(g(Y_j), \mathcal{C}[g](Y_j)\big).$$

Then

$$\widehat{\mathcal{R}}_{N,M}(g) - \mathcal{R}(g) = \underbrace{\big(\widehat{\mathcal{R}}_{N,M}(g) - \widetilde{\mathcal{R}}_M(g)\big)}_{\text{(I)}} + \underbrace{\big(\widetilde{\mathcal{R}}_M(g) - \mathcal{R}(g)\big)}_{\text{(II)}}.$$

对 $g \in \mathcal{G}$ 取上确界并取期望，可得

$$\mathbb{E} \sup_{g\in\mathcal{G}} \big|\widehat{\mathcal{R}}_{N,M}(g) - \mathcal{R}(g)\big| \;\le\; T_1 + T_2,$$

其中

$$T_1 := \mathbb{E} \sup_{g\in\mathcal{G}} \big|\widehat{\mathcal{R}}_{N,M}(g) - \widetilde{\mathcal{R}}_M(g)\big|, \qquad T_2 := \mathbb{E} \sup_{g\in\mathcal{G}} \big|\widetilde{\mathcal{R}}_M(g) - \mathcal{R}(g)\big|.$$

~对于固定的 g， $\mathcal{R}_M(g)$ 是独立同分布变量的经验平均

$$f_g(Y_j) := \ell\big(g(Y_j), \mathcal{C}[g](Y_j)\big), \qquad j = 1, \dots, M.$$

令 $\mathcal{F} := \{f_g : \; g \in \mathcal{G}\}$.根据引理 B.1,

$$|\mathcal{C}[g](y) - \mathcal{C}[g'](y)| \le L_{\mathcal{C},\infty}\|g-g'\|_{L^\infty(\mathbb{R}^d)}, \quad |\mathcal{C}[g](y)| \le (c_+/\underline{D})\|g\|_{L^\infty(\mathbb{R}^d)}.$$

根据*(L)* 以及任意 $g \in \mathcal{G}$的有界性,

$$|f_g(y) - f_{g'}(y)| \le L_\ell \Big(|g(y) - g'(y)| + |\mathcal{C}[g](y) - \mathcal{C}[g'](y)| \Big)$$

结合 $L_\ell = L_\ell\big(c_{\mathcal{G}}^+ \,(1 + (c_+/\underline{D}))\big)$. 再结合 *(G)* 和 *(RT)*，我们得到 $\mathcal{F}$ 具有包络函数 F，其中 $\|F\|_\infty \le B_\ell = B_\ell((1 + (c_+/\underline{D}))\, c_{\mathcal{G}}^+)$。GinGin´e 和 Nickl [2016] 蕴含

$$
\begin{aligned}
T_2 &\lesssim \frac{1}{\sqrt{M}} \int_0^{8\|F\|_\infty} \sqrt{\log(2N_{[]}(\mathcal{F}, L^2(\rho_T), \varepsilon))}\, d\varepsilon \\
&\lesssim \frac{1}{\sqrt{M}} \int_0^{8B_\ell} \sqrt{\log(2N_{[]}(\mathcal{G}, L^2(\rho_T), \varepsilon/L_\ell))}\, d\varepsilon.
\end{aligned} \tag{B.8}
$$

现在我们给出上界

$$
T_1 = \mathbb{E} \sup_{g\in\mathcal{G}} \big|\widehat{\mathcal{R}}_{N,M}(g) - \widetilde{\mathcal{R}}_M(g)\big|.
$$

根据 Lipschitz 性质*(L)*，对于每个 g,

$$
\begin{aligned}
\big|\widehat{\mathcal{R}}_{N,M}(g) - \widetilde{\mathcal{R}}_M(g)\big| &= \left| \frac{1}{M} \sum_{j=1}^M \Big(\ell\big(g(Y_j), \widehat{\mathcal{C}}_{N,M}[g](Y_j)\big) - \ell\big(g(Y_j), \mathcal{C}[g](Y_j)\big) \Big) \right| \\
&\le \frac{L_\ell}{M} \sum_{j=1}^M \big|\widehat{\mathcal{C}}_{N,M}[g](Y_j) - \mathcal{C}[g](Y_j)\big|.
\end{aligned}
$$

对 $g \in \mathcal{G}$ 取上确界并求期望,

$$
T_1 \le L_\ell\, \mathbb{E} \left[\sup_{g\in\mathcal{G}} \frac{1}{M} \sum_{j=1}^M \big|\widehat{\mathcal{C}}_{N,M}[g](Y_j) - \mathcal{C}[g](Y_j)\big| \right].
$$

固定一个较大的紧集 $K \subset \mathbb{R}^d$（稍后定义），并考虑

$$
\mathbb{E} \left[\sup_{g\in\mathcal{G}} \sup_{y\in K} \big|\widehat{\mathcal{C}}_{N,M}[g](y) - \mathcal{C}[g](y)\big| \right].
$$

根据引理B.1,

$$
0 < \underline{D} \;\le\; D_g(x) \;\le\; \overline{D} < \infty, \quad x \in B(x_0, R_0), \quad g \in \mathcal{G}. \tag{B.9}
$$

定义

$$
\widetilde{\mathcal{C}}_N[g](y) := \frac{1}{N} \sum_{i=1}^N \frac{q_T(X_i, y)}{D_g(X_i)}.
$$

然后，对于每个 $(g,\ y)$,

$$
\widehat{\mathcal{C}}_{N,M}[g](y) - \mathcal{C}[g](y) = \underbrace{\big(\widehat{\mathcal{C}}_{N,M}[g](y) - \widetilde{\mathcal{C}}_N[g](y)\big)}_{\text{(A)}} + \underbrace{\big(\widetilde{\mathcal{C}}_N[g](y) - \mathcal{C}[g](y)\big)}_{\text{(B)}}.
$$

因此,

$$
\sup_{g\in\mathcal{G}} \sup_{y\in\mathbb{R}^d} \big|\widehat{\mathcal{C}}_{N,M}[g](y) - \mathcal{C}[g](y)\big| \le \sup_{g\in\mathcal{G}} \sup_{y\in K} |(\text{A})| + \sup_{g\in\mathcal{G}} \sup_{y\in K} |(\text{B})|.
$$

对于固定的 (g, y)，定义

$$
\phi_{g,y}(x) := \frac{q_T(x, y)}{D_g(x)}.
$$

Then

$$\widetilde{\mathcal{C}}_N[g](y) = \frac{1}{N}\sum_{i=1}^{N}\phi_{g,y}(X_i), \qquad \mathcal{C}[g](y) = \mathbb{E}_{X\sim\rho_0}[\phi_{g,y}(X)].$$

根据 (B.9) 以及 q_T 在 (Q)上的高斯上界，再结合 $B(x_0, R_0)$的紧性，存在$F_\infty < \infty$(取$F_\infty = c_+/D$)，使得

$$|\phi_{g,y}(x)| \le F_\infty, \quad x \in B(x_0, R_0),\ y \in \mathbb{R}^d,\ g \in \mathcal{G}.$$

固定 $g,\ h \in \mathcal{G}$和 $x \in B(x_0,\ R_0)$。根据引理 B.1，

$$|D_g(x) - D_h(x)| \le L_{D,\infty}\|g - h\|_\infty .$$

因此，映射 $(g,\ y) \mapsto \phi_{g,\ y}(x)$ 在 $\mathcal{G} \times K$ 上关于 $\|\cdot\|_\infty$（在 g 中）和欧几里得范数（在 y 中）是利普希茨的。对于 y 的依赖性，这是由 q_T 的光滑性（或利普希茨连续性）以及 $1/D_g(x)$ 在 $B(x_0,\ R_0)$ 上的有界性得出的。因此，对于所有 $x \in B(x_0,\ R_0)$、$g,\ h \in \mathcal{G}$ 和 $y,\ y' \in K$，

$$|\phi_{g,y}(x) - \phi_{h,y'}(x)| \le L_g\,\|g - h\|_{L^\infty(\mathbb{R}^d)} + L_y\,\|y - y'\|, \tag{B.10}$$

其中

$$L_g := \frac{c_+\,L_{D,\infty}}{\underline{D}^2}, \qquad L_y := \frac{L_q}{\underline{D}}. \tag{B.11}$$

Let

$$\Phi_K := \{x \mapsto \phi_{g,y}(x) : g \in \mathcal{G}, y \in K\}.$$

注意，

$$\log \mathcal{N}_{[]}(\Phi_K, L^2(\rho_0), \varepsilon) \ \le\ \log \mathcal{N}_{[]}\Big(\mathcal{G}, \|\cdot\|_\infty, \frac{\varepsilon}{2L_g}\Big) \ +\ C_d\,\log\Big(\frac{L_y\,\mathrm{diam}(K)}{\varepsilon}\Big), \tag{B.12}$$

其中 $C_d \le d$是一个维数常数（源于对 K 在欧几里得范数下的覆盖）。接着由Gin´e 和 Nickl [2016],中的定理 3.5.13 得出

$$\sup_{g\in\mathcal{G}}\sup_{y\in\mathbb{R}^d}|(\mathrm{B})| \lesssim \frac{1}{\sqrt{N}}\int_0^{8F_\infty}\sqrt{\log \mathcal{N}_{[]}\Big(\mathcal{G}, \|\cdot\|_\infty, \frac{\varepsilon}{2L_g}\Big)}\,d\varepsilon + \frac{1}{\sqrt{N}}\sqrt{\log\Big(\frac{L_y\,\mathrm{diam}(K)}{F_\infty}\Big)}. \tag{B.13}$$

此外，对于固定的 $(g,\ y)$,

$$(\mathrm{A}) = \frac{1}{N}\sum_{i=1}^{N} q_T(X_i, y)\left(\frac{1}{\widehat{D}_g(X_i)} - \frac{1}{D_g(X_i)}\right).$$

根据（B.9），$u \mapsto 1/u$ 在 $[D, D]$ 上是利普希茨连续的，常数为 $L_D :\ = D^2/D^2$，因此

$$\left|\frac{1}{\mathcal{T}_{[\underline{D},\overline{D}]}[\widehat{D}_g](X_i)} - \frac{1}{D_g(X_i)}\right| \le L_D\big|\widehat{D}_g(X_i) - D_g(X_i)\big|.$$

根据*(Q)*，我们有

$$q_T(x,y) \le c_+ \quad \text{for all } x,y \in \mathbb{R}^d.$$

因此，对于所有 $X_i \in B(x_0,\ R_0)$ 以及所有 $y \in \mathbb{R}^d$,

$$q_T(X_i,y) \le c_+.$$

因此,

$$\sup_{y\in K} |(\mathrm{A})| \le \frac{c_+ L_D}{N} \sum_{i=1}^{N} \big|\widehat{D}_g(X_i) - D_g(X_i)\big|.$$

对 $g \in \mathcal{G}$ 取上确界并求期望,

$$\mathbb{E}\left[\sup_{g\in\mathcal{G}} \sup_{y\in K} |(\mathrm{A})|\right] \le c_+ L_D\, \mathbb{E}\left[\sup_{g\in\mathcal{G}} \frac{1}{N}\sum_{i=1}^{N} \big|\widehat{D}_g(X_i) - D_g(X_i)\big|\right].$$

现在，对于每个固定的 x 和 g,

$$D_g(x) = \mathbb{E}_{Z\sim\rho_T}\big[\psi_g(x,Z)\big], \qquad \widehat{D}_g(x) = \frac{1}{M}\sum_{k=1}^{M} \psi_g(x,Y_k),$$

with

$$\psi_g(x,z) := q_T(x,z)\,\frac{1}{g(z)}.$$

Let

$$\Psi := \big\{(x,z) \mapsto \psi_g(x,z) : x \in B(x_0,R_0),\ g \in \mathcal{G}\big\}.$$

由于 *(G)*,

$$g(z) \ge c_{\mathcal{G}} e^{-a_{\mathcal{G}}\|z\|^2}, \qquad \forall z \in \mathbb{R}^d,$$

并且对于所有 $x \in B(x_0,\ R_0)$ 以及所有 $g \in \mathcal{G}$,

$$\psi_g(x,z) \le F(z) := \frac{c_+}{c_{\mathcal{G}}}\, e^{a_{\mathcal{G}}\|z\|^2}.$$

Since ρ_T 满足高斯上界 **(RT)**，且有 $b_T^+ > 2a_{\mathcal{G}}$，我们得到 $F \in L^2(\rho_T)$ and

$$\|F\|_{L^2(\rho_T)} \le \frac{c_+}{c_{\mathcal{G}}} \sqrt{c_T^+}\left(\frac{\pi}{b_T - 2a_{\mathcal{G}}}\right)^{d/4}.$$

Gin´e and Nickl [2016] 给出

$$\mathbb{E}\Big[\sup_{x\in B(x_0,R_0)} \sup_{g\in\mathcal{G}} \big|\widehat{D}_g(x) - D_g(x)\big|\Big] \lesssim \frac{1}{\sqrt{M}} \int_0^{8\|F\|_{L^2(\rho_T)}} \sqrt{\log(2N_{[]}(\Psi, L^2(\rho_T), \varepsilon))}\, d\varepsilon. \tag{B.14}$$

固定 $x \in B(x_0, R_0)$ 和 $z \in \mathbb{R}^d$。设

$$\psi_g(x,z) - \psi_h(x,z) = q_T(x,z)\Big(\frac{1}{g(z)} - \frac{1}{h(z)}\Big).$$

我们有

$$\frac{1}{g(z)} \le \frac{1}{c_{\mathcal{G}}} e^{a_{\mathcal{G}}\|z\|^2}, \qquad \frac{1}{h(z)} \le \frac{1}{c_{\mathcal{G}}} e^{a_{\mathcal{G}}\|z\|^2}.$$

因此，利用 $|1/u - 1/v| = |u-v|/(uv)$,

$$\Big|\frac{1}{g(z)} - \frac{1}{h(z)}\Big| \le \frac{|g(z)-h(z)|}{c_{\mathcal{G}}^2}\, e^{2a_{\mathcal{G}}\|z\|^2}.$$

两边乘以 $|q_T(x,z)| \le c_+$，得到

$$|\psi_g(x,z) - \psi_h(x,z)| \le \frac{c_+}{c_{\mathcal{G}}^2}\, e^{2a_{\mathcal{G}}\|z\|^2}\, \|g-h\|_\infty.$$

取 $L^2(\rho_T)$ 范数，得到

$$\|\psi_g(x,\cdot) - \psi_h(x,\cdot)\|_{L^2(\rho_T)} \le \frac{c_+}{c_{\mathcal{G}}^2} \Big(\int e^{4a_{\mathcal{G}}\|z\|^2} \rho_T(z)\, dz \Big)^{1/2} \|g-h\|_\infty.$$

因此,

$$\|\psi_g(x,\cdot) - \psi_h(x,\cdot)\|_{L^2(\rho_T)} \;\le\; \frac{c_+}{c_{\mathcal{G}}^2}\, M_\rho\, \|g-h\|_\infty$$

因此我们得到,

$$\log \mathcal{N}_{[]}\big(\Psi, L^2(\rho_T), \varepsilon\big) \;\le\; \log \mathcal{N}_{[]}\Big(\mathcal{G}, \|\cdot\|_\infty, \frac{\varepsilon}{C_g}\Big) \;+\; d\, \log\Big(\frac{C\, R_0}{\varepsilon}\Big),$$

其中 $C_g := \frac{c_+}{c_{\mathcal{G}}^2} M_\rho$. 。我们选择 $K = B(0,R)$，使得 $R > R_0$. 。固定 $g \in \mathcal{G}$，并取满足 $\|y\| \ge R$ 的 $y \in \mathbb{R}^d$ 。对于任意 $x \in B(0,R_0)$，，我们有

$$\|x-y\| \;\ge\; \|y\| - \|x\| \;\ge\; R - R_0,$$

因此，根据核的上界

$$q_T(x,y) \le c_+ \exp\big(-a_+\|x-y\|^2\big) \le c_+ \exp\big(-a_+(R-R_0)^2\big).$$

利用 $D_g(x) \ge \underline{D}$，并且由于 ρ_0 是概率密度,

$$\begin{aligned} |\mathcal{C}[g](y)| &= \Big| \int_{B(0,R_0)} \rho_0(x)\, \frac{q_T(x,y)}{D_g(x)}\, dx \Big| \le \frac{1}{\underline{D}} \int_{B(0,R_0)} \rho_0(x)\, q_T(x,y)\, dx \\ &\le \frac{1}{\underline{D}} \sup_{x \in B(0,R_0)} q_T(x,y) \le \frac{c_+}{\underline{D}} \exp\big(-a_+(R-R_0)^2\big). \end{aligned}$$

此外，利用 $\mathcal{T}_{[\underline{D},\overline{D}]}[D_g](X_i) \ge \underline{D}$ 以及 $X_i \in B(0,\ R_0)$,

$$|\widehat{\mathcal{C}}_{N,M}[g](y)| \le \frac{1}{N} \sum_{i=1}^N \frac{q_T(X_i,y)}{\mathcal{T}_{[\underline{D},\overline{D}]}[\widehat{D}_g](X_i)} \le \frac{1}{\underline{D}} \sup_{x \in B(0,R_0)} q_T(x,y) \le \frac{c_+}{\underline{D}} \exp\big(-a_+(R-R_0)^2\big).$$

因此，对于所有 $g \in \mathcal{G}$ 以及所有 $\|y\| \ge R$,

$$\big|\widehat{\mathcal{C}}_{N,M}[g](y) - \mathcal{C}[g](y)\big| \le |\widehat{\mathcal{C}}_{N,M}[g](y)| + |\mathcal{C}[g](y)| \le \frac{2c_+}{\underline{D}} \exp\big(-a_+(R-R_0)^2\big).$$

那么，对于每个 $R > R_0$,

$$\mathbb{E}\left[\sup_{g\in\mathcal{G}}\sup_{y\notin B(0,R)}\left|\widehat{\mathcal{C}}_{N,M}[g](y)-\mathcal{C}[g](y)\right|\right] \le \frac{2c_+}{\underline{D}}\,\exp\bigl(-a_+(R-R_0)^2\bigr). \tag{B.15}$$

□

C 逼近误差

引理 C.1。 假设 (Q)、(R0)、(RT) 和(G) 成立。进一步假设 $b_T^+ > 4a_{\mathcal{G}}$。则存在常数 $L_{\mathcal{C},\,2} > 0$使得

$$\|\mathcal{C}[g]-\mathcal{C}[h]\|_{L^2(\rho_T)} \le L_{\mathcal{C},2}\,\|g-h\|_{L^2(\rho_T)}. \tag{C.1}$$

证明。 类似于（B.5），

$$|D_g(x)-D_h(x)| \le (c_{\mathcal{G}})^{-2}\int q_T(x,z)\,\rho_T(z)\,\exp(2a_{\mathcal{G}}\|z\|^2)\,|g(z)-h(z)|\,dz.$$

关于 $\rho_T(z)\,dz$应用柯西-施瓦茨不等式：

$$|D_g(x)-D_h(x)| \le (c_{\mathcal{G}})^{-2}\Bigl(\int q_T^2(x-z)\exp(4a_{\mathcal{G}}\|z\|^2)\,\rho_T(z)\,dz\Bigr)^{1/2}\|g-h\|_{L^2(\rho_T)}.$$

对 $x \in B(x_0,\ R_0)$ 取上确界可得

$$\|D_g-D_h\|_{L^\infty(B(x_0,R_0))} \le L_{D,2}\,\|g-h\|_{L^2(\rho_T)}, \tag{C.2}$$

对于某个常数$L_{D,\,2} < \infty$若

$$\sup_{x\in B(x_0,R_0)}\int q_T^2(x,z)e^{4a_{\mathcal{G}}\|z\|^2}\,\rho_T(z)\,dz$$

是有限的。现在我们在*(Q)* 和*(RT)*中的上尾条件下检验这种有限性。根据*(Q)*,

$$q_T^2(x,z) \le c_+^2\exp(-2a_+\|x-z\|^2) \le c_+^2\exp(2a_+R_\star^2)\exp(-a_+\|z\|^2),$$

我们得到

$$q_T^2(x,z)e^{4a_{\mathcal{G}}\|z\|^2} \le c_+^2\exp(2a_+R_\star^2)\,\exp\Bigl(-\bigl(a_+-4a_{\mathcal{G}}+b_T^+\bigr)\|z\|^2\Bigr).$$

这是可积的 如果 $a_+ + b_T^+ - 4a_{\mathcal{G}} > 0$。因此 $L_{D,\,2} < \infty$ 且（C.2）成立。由（C.2）可知

$$|\mathcal{C}[g](y)-\mathcal{C}[h](y)| \le \frac{L_{D,2}}{\underline{D}^2}\|g-h\|_{L^2(\rho_T)}\underbrace{\int_{K_0}\rho_0(x)\,q_T(x,y)\,dx}_{=:m_0(y)}.$$

现在在 y 中取 $L^2(\rho_T)$ 范数：

$$\|\mathcal{C}[g]-\mathcal{C}[h]\|_{L^2(\rho_T)} \le \frac{L_{D,2}}{\underline{D}^2}\,\|m_0\|_{L^2(\rho_T)}\,\|g-h\|_{L^2(\rho_T)}.$$

最后，$\|m_0\|_{L^2(\rho_T)} < \infty$ 在 *(Q)*、*(R0)*、*(RT)* 条件下，由于 $m_0(y)$ 在 y (一个紧支撑的次高斯核混合分布) 中是次高斯的，并且 ρ_T 具有次高斯尾部。因此，C.1) 成立，其中

$$L_{\mathcal{C},2} := \frac{L_{D,2}}{\underline{D}^2} \|m_0\|_{L^2(\rho_T)} < \infty.$$

□

推论 C.2。 设 $g^\star \in \mathcal{G}$ 是 $\mathcal{C}$的不动点，即 $g^\star = \mathcal{C}[g^\star]$。那么在引理 C.1的假设下，对所有 $g \in \mathcal{G}$,

$$\big|\mathcal{R}(g) - \mathcal{R}(g^\star)\big| \le L_\ell\big(1 + L_{\mathcal{C},2}\big) \|g - g^\star\|_{L^2(\rho_T)}. \tag{C.3}$$

证明。 利用 *(L)* 以及不动点性质 $\mathcal{C}[g^\star] = g^\star$，我们几乎必然有（其中 $Y \sim \rho_T$）

$$\Big|\ell\big(g(Y), \mathcal{C}[g](Y)\big) - \ell\big(g^\star(Y), g^\star(Y)\big)\Big| \le L_\ell\Big(|g(Y) - g^\star(Y)| + |\mathcal{C}[g](Y) - g^\star(Y)|\Big).$$

取期望并应用柯西-施瓦茨不等式可得

$$\big|\mathcal{R}(g) - \mathcal{R}(g^\star)\big| \le L_\ell\Big(\|g - g^\star\|_{L^2(\rho_T)} + \|\mathcal{C}[g] - g^\star\|_{L^2(\rho_T)}\Big).$$

由于 $g^\star = \mathcal{C}[g^\star]$,

$$\|\mathcal{C}[g] - g^\star\|_{L^2(\rho_T)} = \|\mathcal{C}[g] - \mathcal{C}[g^\star]\|_{L^2(\rho_T)} \le L_{\mathcal{C},2} \|g - g^\star\|_{L^2(\rho_T)}$$

由引理 C.1。结合这些不等式可得（C.3）。 □

C.1 Bargmann变换

对于 $f \in L^2(\mathbb{R}^d)$，(Segal-)Bargmann变换定义为

$$(\mathcal{B}f)(z) := \pi^{-d/4} \int_{\mathbb{R}^d} \exp\Big(-\tfrac{1}{2}|y|^2 + \sqrt{2}\, z \cdot y - \tfrac{1}{2}\, z \cdot z\Big) f(y)\, dy, \qquad z \in \mathbb{C}^d,$$

其中 $z \cdot y = \sum_{j=1}^d z_j y_j$ 和 $z \cdot z = \sum_{j=1}^d z_j^2$。映射 $\mathcal{B}$ 是从 $L^2(\mathbb{R}^d)$ 到由整函数构成的 Fock空间$\mathcal{F}^2(\mathbb{C}^d)$ 的酉映射。定义 Hermite多项式如下:

$$\mathrm{He}_n(x) := (-1)^n e^{x^2} \frac{d^n}{dx^n}\Big(e^{-x^2}\Big), \qquad n \in \mathbb{N}_0.$$

那么，相应的 $L^2(\mathbb{R})$–归一化厄米函数为

$$\psi_n(x) := \frac{1}{2^{n/2}\pi^{1/4}\sqrt{n!}}\, \mathrm{He}_n(x)\, e^{-x^2/2}, \qquad x \in \mathbb{R},\ n \in \mathbb{N}_0. \tag{C.4}$$

直接计算（使用Hermite多项式的生成函数）可得，对于所有 $n \in \mathbb{N}_0$,

$$(\mathcal{B}\psi_n)(z) = \frac{z^n}{\sqrt{n!}}, \qquad z \in \mathbb{C}.$$

事实上，根据定义，

$$(\mathcal{B}\psi_0)(z) = \pi^{-1/4}\int_{\mathbb{R}} e^{-\frac{1}{2}y^2+\sqrt{2}zy-\frac{1}{2}z^2}\ \pi^{-1/4}e^{-y^2/2}\,dy = \pi^{-1/2}\int_{\mathbb{R}} e^{-y^2+\sqrt{2}zy-\frac{1}{2}z^2}\,dy.$$

配方：

$$-y^2+\sqrt{2}zy = -\big(y-\tfrac{z}{\sqrt{2}}\big)^2+\tfrac{1}{2}z^2,$$

因此

$$-y^2+\sqrt{2}zy-\tfrac{1}{2}z^2 = -\big(y-\tfrac{z}{\sqrt{2}}\big)^2.$$

因此，

$$(\mathcal{B}\psi_0)(z) = \pi^{-1/2}\int_{\mathbb{R}} e^{-(y-\frac{z}{\sqrt{2}})^2}\,dy = \pi^{-1/2}\int_{\mathbb{R}} e^{-u^2}\,du = 1.$$

设 $f \in \mathcal{S}(\mathbb{R})$ 为施瓦茨函数，以确保所有分部积分均有依据。记

$$K(y,z) := \exp\Big(-\tfrac{1}{2}y^2+\sqrt{2}zy-\tfrac{1}{2}z^2\Big).$$

Then

$$(\mathcal{B}(a^*f))(z) = \pi^{-1/4}\int_{\mathbb{R}} K(y,z)\,\frac{1}{\sqrt{2}}\Big(y-\frac{d}{dy}\Big)f(y)\,dy = \frac{1}{\sqrt{2}}\pi^{-1/4}\Big(I_1-I_2\Big),$$

其中

$$I_1 := \int_{\mathbb{R}} K(y,z)\,y\,f(y)\,dy, \qquad I_2 := \int_{\mathbb{R}} K(y,z)\,f'(y)\,dy$$

and

$$a^* := \frac{1}{\sqrt{2}}\Big(y-\frac{d}{dy}\Big).$$

对 I_2 进行分部积分（边界项为零，因为 $K(\cdot,z)$ 具有高斯衰减，且 f 是施瓦茨函数）：

$$I_2 = -\int_{\mathbb{R}} \partial_y K(y,z)\,f(y)\,dy.$$

计算 $\partial_y K$：

$$\partial_y K(y,z) = (-y+\sqrt{2}z)\,K(y,z).$$

因此

$$I_2 = -\int_{\mathbb{R}} (-y+\sqrt{2}z)\,K(y,z)\,f(y)\,dy = \int_{\mathbb{R}} y\,K(y,z)\,f(y)\,dy - \sqrt{2}z\int_{\mathbb{R}} K(y,z)\,f(y)\,dy.$$

因此，

$$I_1 - I_2 = \sqrt{2}z\int_{\mathbb{R}} K(y,z)\,f(y)\,dy,$$

所以

$$(\mathcal{B}(a^*f))(z) = \frac{1}{\sqrt{2}}\pi^{-1/4}\cdot\sqrt{2}z\int_{\mathbb{R}} K(y,z)\,f(y)\,dy = z\,(\mathcal{B}f)(z).$$

因此我们已经证明

$$\mathcal{B}\circ a^* = (\text{multiplication by } z)\circ\mathcal{B} \qquad \text{on } \mathcal{S}(\mathbb{R}). \tag{C.5}$$

根据 ψ_n 的定义，我们有 $\psi_n := \frac{1}{\sqrt{n}} a^* \psi_{n-1}$， $n \geq 1$ 且（C.5）蕴含

$$(\mathcal{B}\psi_n)(z) = \frac{1}{\sqrt{n}} (\mathcal{B}(a^*\psi_{n-1}))(z) = \frac{1}{\sqrt{n}} z (\mathcal{B}\psi_{n-1})(z).$$

从 $(\mathcal{B}\psi_0)(z) = 1$ 出发，我们递归地得到

$$(\mathcal{B}\psi_n)(z) = \frac{z^n}{\sqrt{n!}}, \qquad n \in \mathbb{N}_0.$$

这一恒等式体现了这样一个事实：Bargmann变换将谐振子对角化：Hermite函数被映射为单项式。 d 维 Hermite函数可分解为

$$\psi_\alpha(x) = \prod_{j=1}^{d} \psi_{\alpha_j}(x_j), \qquad \alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d.$$

同样，Bargmann 核按坐标分解为：

$$e^{-\frac{1}{2}|y|^2 + \sqrt{2}\, z\cdot y - \frac{1}{2} z\cdot z} = \prod_{j=1}^{d} e^{-\frac{1}{2} y_j^2 + \sqrt{2}\, z_j y_j - \frac{1}{2} z_j^2}.$$

应用 Fubini定理和一维恒等式可得

$$(\mathcal{B}\psi_\alpha)(z) = \prod_{j=1}^{d} (\mathcal{B}\psi_{\alpha_j})(z_j) = \prod_{j=1}^{d} \frac{z_j^{\alpha_j}}{\sqrt{\alpha_j!}} = \frac{z^\alpha}{\sqrt{\alpha!}}, \tag{C.6}$$

其中 $z^\alpha = \prod_{j=1}^{d} z_j^{\alpha_j}$ and $\alpha! = \prod_{j=1}^{d} \alpha_j!$。由于 $\{\psi_\alpha\}_{\alpha \in \mathbb{N}_0^d}$ 是 $L^2(\mathbb{R}^d)$ 的正交归一基，每个 $f \in L^2(\mathbb{R}^d)$ 都具有 Hermite展开

$$f = \sum_{\alpha \in \mathbb{N}_0^d} \langle f, \psi_\alpha \rangle \, \psi_\alpha \quad (\text{in } L^2(\mathbb{R}^d)).$$

由于 $\mathcal{B}$ 是酉的，我们可以逐项应用它：

$$\mathcal{B}f = \sum_{\alpha \in \mathbb{N}_0^d} \langle f, \psi_\alpha \rangle \, \mathcal{B}\psi_\alpha.$$

利用（C.6），我们得到幂级数表示

$$(\mathcal{B}f)(z) = \sum_{\alpha \in \mathbb{N}_0^d} \langle f, \psi_\alpha \rangle \, \frac{z^\alpha}{\sqrt{\alpha!}}, \qquad z \in \mathbb{C}^d.$$

因此，Bargmann变换将 f 的Hermite展开转换为整函数 $\mathcal{B}f$的Taylor展开。在这种对应关系中，

$$\langle f, \psi_\alpha \rangle = \frac{\partial^\alpha (\mathcal{B}f)(0)}{\sqrt{\alpha!}},$$

这解释了为什么关于 $\mathcal{B}f$ 增长的界能够立即得到 Hermite系数的衰减估计。

C.2 通过 Hermite多项式进行近似

由于我们关注的是 $T = 1$ 这种情况，因此在 q的记号中省略 T。于是

$$q(z) := (2\pi)^{-d/2} \exp\Big(- \frac{\|z\|^2}{2} \Big).$$

命题 C.3。 令 $d \geq 1$ *let* $w:\ \mathbb{R}^d \to [0,\infty)$ 具有紧支撑，满足

$$\operatorname{supp}(w) \subseteq \{x \in \mathbb{R}^d :\ \|x\| \leq R\}, \qquad M_0 := \int_{\mathbb{R}^d} w(x)\, dx < \infty.$$

定义

$$g^\star(y) := (w * q)(y) = \int_{\mathbb{R}^d} w(x)\, q(x-y)\, dx.$$

令 $\{\psi_\alpha\}_{\alpha \in \mathbb{N}_0^d}$ 为在 $L^2(\mathbb{R}^d)$ 中正交归一的 d 维张量积厄米基，并定义 *Hermite*系数

$$c_\alpha := \langle g^\star, \psi_\alpha \rangle_{L^2(\mathbb{R}^d)}, \qquad \alpha \in \mathbb{N}_0^d.$$

那么，对于每个满足 $m :\ = |\alpha| \geq 1$的多重指标 $\alpha \in \mathbb{N}_0^d$，

$$|c_\alpha| \ \leq\ C_d\, m^{1/4} \Big(\frac{e^{1/2}\beta}{\sqrt{m}} \Big)^m, \qquad C_d = \pi^{-d/4} 2^{-d/2} M_0, \qquad \beta := R\sqrt{\frac{d}{2}}. \tag{C.7}$$

等价地，

$$|c_\alpha| \ \leq\ C_d\, m^{1/4} \exp\Big(- \frac{m}{2} \log m + m \log(e^{1/2}\beta) \Big). \tag{C.8}$$

特别地，若 $m \geq (e^{1/2}\beta)^2 = e\beta^2$，则 $\log(e^{1/2}\beta) \ \leq \frac{1}{2} \log m$ 且

$$|c_\alpha| \ \leq\ C_d\, m^{1/4} \exp\Big(- \frac{m}{4} \log m \Big). \tag{C.9}$$

备注 C.4. 考虑方差为 $T > 1$的高斯核，

$$q_T(x-y) = (2\pi T)^{-d/2} \exp\Big(- \frac{|x-y|^2}{2T} \Big), \qquad g^\star = w * q_T.$$

虽然 $g^\star$ 的未缩放巴格曼变换在复变量 z中呈现高斯（二次）增长，但通过对Hermite基进行适当缩放可以补偿这一点。更准确地说，取 $\lambda \in (0,1)$，使得

$$\lambda^2 = \frac{1}{T}, \qquad \text{equivalently} \qquad \lambda = T^{-1/2}.$$

若将 $g^\star$ 按缩放厄米基展开

$$\psi_\alpha^{(1/\lambda)}(x) = \lambda^{d/2} \prod_{j=1}^d \psi_{\alpha_j}(\lambda x_j), \qquad \alpha \in \mathbb{N}_0^d,$$

并考虑相应的缩放的巴格曼变换，则巴格曼表示的指数前因子中的二次项会被缩放完全抵消。因此， $g^\star$ 在缩放变量下的巴格曼变换满足如下形式的线性指数增长界

$$\sup_{\max_j |z_j|\le r} \big|(\mathcal{B}g^\star)(z)\big| \;\le\; C_{d,T}\,\exp\big(b_{d,T}\,r\big), \qquad r>0,$$

其中常数 $C_{d、T}$、 $b_{d、T}>0$ 仅取决于 d、 T 以及 w 的支撑半径，但不取决于 r。特别是，在适当选择缩放参数 λ 后，Bargmann变换呈现出与临界情形 $T=1$ 中相同的线性指数增长，而所得的缩放的厄米系数会随总次数 $|\alpha|$ 超几何地衰减。这一观察说明，缩放 $\lambda=T^{-1/2}$ 对于将 $T=1$ 系数估计推广到一般方差 $T>1$ 是自然且最优的。

证明。 Bargmann变换将 Hermite函数映射为归一化单项式:

$$(\mathcal{B}\psi_\alpha)(z)=\frac{z^\alpha}{\sqrt{\alpha!}}, \qquad \alpha\in\mathbb{N}_0^d,$$

因此，对于 $g^\star\in L^2$（$\mathbb{R}^d$），

$$(\mathcal{B}g^\star)(z)=\sum_{\alpha\in\mathbb{N}_0^d} c_\alpha\,\frac{z^\alpha}{\sqrt{\alpha!}}, \qquad z\in\mathbb{C}^d.$$

特别是,

$$c_\alpha=\frac{\partial^\alpha(\mathcal{B}g^\star)(0)}{\sqrt{\alpha!}}. \tag{C.10}$$

对于 $r>0$，设闭多圆盘 $D_r:=\{z\in\mathbb{C}^d:\ \max_j|z_j|\le r\}$。根据多变量柯西估计,

$$|\partial^\alpha F(0)|\le \alpha!\,r^{-|\alpha|}\sup_{z\in D_r}|F(z)|, \qquad \alpha\in\mathbb{N}_0^d.$$

将此应用于 $F=\mathcal{B}g^\star$ 并结合（C.10）：

$$|c_\alpha|\le\sqrt{\alpha!}\,r^{-m}\sup_{z\in D_r}|(\mathcal{B}g^\star)(z)|, \qquad m:=|\alpha|. \tag{C.11}$$

将（C.19）（见命题C.6）代入（C.11）：

$$|c_\alpha|\le C_0\sqrt{\alpha!}\,r^{-m}e^{\beta r}, \qquad r>0.$$

最小化 $\phi(r):=\beta r-m\log r$，其中 $r>0$。由于 $\phi'(r)=\beta-\frac{m}{r}$，唯一的极小值点为

$$r_*=\frac{m}{\beta}.$$

因此

$$|c_\alpha|\le C_0\sqrt{\alpha!}\left(\frac{\beta}{m}\right)^m e^{\beta(m/\beta)}=C_0\sqrt{\alpha!}\left(\frac{e\beta}{m}\right)^m. \tag{C.12}$$

由于 $\alpha!\le m!$ 对于 $m=|\alpha|$，我们有 $\sqrt{\alpha!}\le\sqrt{m!}$. 根据斯特林估计，存在一个绝对常数$C>0$ 使得

$$\sqrt{m!}\;\le\; C\,m^{1/4}\Big(\frac{m}{e}\Big)^{m/2}, \qquad m\ge 1. \tag{C.13}$$

结合（C.12）和（C.13）可得

$$|c_\alpha| \le C\, C_0\, m^{1/4} \Big(\frac{m}{e}\Big)^{m/2} \Big(\frac{e\beta}{m}\Big)^{m}.$$

化简：

$$\Big(\frac{m}{e}\Big)^{m/2} \Big(\frac{e\beta}{m}\Big)^{m} = m^{m/2} e^{-m/2} \cdot e^{m} \beta^{m} m^{-m} = e^{m/2} \beta^{m} m^{-m/2} = \Big(\frac{e^{1/2}\beta}{\sqrt{m}}\Big)^{m}.$$

因此

$$|c_\alpha| \le C\, C_0\, m^{1/4} \Big(\frac{e^{1/2}\beta}{\sqrt{m}}\Big)^{m},$$

即（C.7）。 □

命题 C.5。 设 $d \ge 1$，并设 $\{\psi_\alpha\}_{\alpha \in \mathbb{N}_0^d}$ 为 $L^2(\mathbb{R}^d)$ 中的 d维 *Hermite*基。令 Π_n^F表示到

$$\operatorname{span}\{\psi_\alpha : |\alpha| \le n\}.$$

在命题 *C.3* 的假设下，对于每个 $n \in \mathbb{N}_0$ 满足 $n+1 > 2K^2$，有

$$\|g^\star - \Pi_n^F g^\star\|_{L^2(\mathbb{R}^d)} \;\le\; \widetilde{C}_d\, C_d\, (n+1)^{\frac{d}{2}-\frac{1}{4}} \Big(\frac{K}{\sqrt{n+1}}\Big)^{n+1} \tag{C.14}$$

其中 C_d 由下式给出：(C.7) 并且

$$\widetilde{C}_d = \left(1 + 2^{d-1/2}\, \frac{\Gamma(d+1/2)}{(\ln 2)^{d+1/2}}\right)^{1/2}.$$

证明。 根据 $\{\psi_\alpha\}$ 的正交归一性以及 Π_n^F的定义，

$$\|g^\star - \Pi_n^F g^\star\|_{L^2(\mathbb{R}^d)}^2 = \sum_{|\alpha|>n} |c_\alpha|^2. \tag{C.15}$$

利用（C.7），取 $m := |\alpha|$ 可得

$$\sum_{|\alpha|>n} |c_\alpha|^2 \le C_d^2 \sum_{|\alpha|>n} m^{1/2} \Big(\frac{K^2}{m}\Big)^{m}.$$

按总次数分组 $m = |\alpha|$并令

$$N_d(m) := \#\{\alpha \in \mathbb{N}_0^d : |\alpha| = m\} = \binom{m+d-1}{d-1}.$$

Then

$$\sum_{|\alpha|>n} |c_\alpha|^2 \le C_d^2 \sum_{m=n+1}^{\infty} N_d(m)\, m^{1/2} \Big(\frac{K^2}{m}\Big)^{m}. \tag{C.16}$$

使用标准估计 $N_d(m) \le \widetilde{C}_d\, m^{d-1}$，我们得到

$$\sum_{|\alpha|>n} |c_\alpha|^2 \le \widetilde{C}_d\, C_d^2 \sum_{m=n+1}^{\infty} m^{d-\frac{1}{2}} \Big(\frac{K^2}{m}\Big)^m. \tag{C.17}$$

Set

$$q := \frac{K^2}{n+1} \in (0,1).$$

对于 $m \ge n+1$，我们有 $\frac{K^2}{m} \le q$，因此

$$\Big(\frac{K^2}{m}\Big)^m \le q^m.$$

因此，

$$\begin{aligned}\sum_{m=n+1}^{\infty} m^{d-\frac{1}{2}} \Big(\frac{K^2}{m}\Big)^m &\le q^{n+1}(n+1)^{d-1/2} \sum_{k=0}^{\infty} \Big(1+\frac{k}{n+1}\Big)^{d-1/2} q^k \\ &\le q^{n+1}(n+1)^{d-1/2} \left(1 + 2^{d-1/2}\, \frac{\Gamma(d+1/2)}{(-\ln q)^{d+1/2}}\right)\end{aligned}$$

将其代入（C.17），然后开平方：

$$\|g^\star - \Pi_n^F g^\star\|_2 \le \widetilde{C}_d\, C_d\, (n+1)^{\frac{d}{2}-\frac{1}{4}}\, q^{(n+1)/2}.$$

最后，$q^{(n+1)/2} = \Big(\frac{K}{\sqrt{n+1}}\Big)^{n+1}$ 以及 $1-q = 1 - \frac{K^2}{n+1}$，从而得到（C.14）。 □

命题 C.6。设命题*C.3* 的假设成立。对于每个 $z \in \mathbb{C}^d$，

$$(\mathcal{B}g^\star)(z) = \pi^{-d/4}\, 2^{-d/2} \int_{\mathbb{R}^d} w(x)\, \exp\Big(-\frac{|x|^2}{4} + \frac{1}{\sqrt{2}}\, x\cdot z\Big)\, dx. \tag{C.18}$$

此外，对于每个 $r > 0$，

$$\sup_{\max_j |z_j| \le r} |(\mathcal{B}g^\star)(z)| \;\le\; \pi^{-d/4}\, 2^{-d/2}\, M_0\, \exp\Big(\frac{R}{\sqrt{2}}\, \sqrt{d}\, r\Big). \tag{C.19}$$

证明。根据定义和 Fubini 定理，

$$(\mathcal{B}g^\star)(z) = \pi^{-d/4} e^{-\frac{1}{2}z\cdot z} \int_{\mathbb{R}^d} w(x)\, (2\pi)^{-d/2} \int_{\mathbb{R}^d} \exp\Big(-\tfrac{1}{2}|y|^2 + \sqrt{2}\, z\cdot y - \tfrac{1}{2}|x-y|^2\Big)\, dy\, dx.$$

展开 $|x-y|^2 = |x|^2 - 2x\cdot y + |y|^2$ 得到

$$-\tfrac{1}{2}|y|^2 - \tfrac{1}{2}|x-y|^2 + \sqrt{2}\, z\cdot y = -|y|^2 + (x+\sqrt{2}\, z)\cdot y - \tfrac{1}{2}|x|^2.$$

配方：

$$-|y|^2 + (x+\sqrt{2}z)\cdot y = -\Big|y - \frac{x+\sqrt{2}z}{2}\Big|^2 + \frac{|x+\sqrt{2}z|^2}{4}.$$

因此，

$$\int_{\mathbb{R}^d} \exp\Big(-|y|^2 + (x+\sqrt{2}z)\cdot y\Big)\, dy = \pi^{d/2} \exp\Big(\frac{|x+\sqrt{2}z|^2}{4}\Big).$$

因此

$$\begin{aligned}(2\pi)^{-d/2} &\int_{\mathbb{R}^d} \exp\Big(-\tfrac{1}{2}|y|^2 + \sqrt{2}\, z\cdot y - \tfrac{1}{2}|x-y|^2\Big)\, dy \\ &= (2\pi)^{-d/2}\, \pi^{d/2}\, \exp\Big(-\frac{|x|^2}{2} + \frac{|x+\sqrt{2}z|^2}{4}\Big) = 2^{-d/2} \exp\Big(-\frac{|x|^2}{4} + \frac{1}{\sqrt{2}} x\cdot z + \frac{1}{2} z\cdot z\Big),\end{aligned}$$

其中我们使用了 $|x+\sqrt{2}z|^2 = |x|^2 + 2\sqrt{2}\, x\cdot z + 2\, z\cdot z$。将 Bargmann 核中的前因子 $e^{-\frac{1}{2}z\cdot z}$ 相乘后，$z\cdot z$ 中的二次项被抵消，从而得到 (C.18)。由 (C.18) 和 $e^{-|x|^2/4} \le 1$,

$$|(\mathcal{B}g^\star)(z)| \le \pi^{-d/4}\, 2^{-d/2} \int_{\mathbb{R}^d} |w(x)| \left| \exp\Big(\frac{1}{\sqrt{2}} x\cdot z\Big)\right| dx = \pi^{-d/4}\, 2^{-d/2} \int_{\mathbb{R}^d} |w(x)|\, \exp\Big(\frac{1}{\sqrt{2}} x\cdot \operatorname{Re} z\Big)\, dx.$$

若 $\operatorname{supp}(w) \subseteq \{|x| \le R\}$，则 $x\cdot \operatorname{Re}\, z \le |x|\, |\operatorname{Re} z| \le R\, |\operatorname{Re}\, z|$，因此

$$|(\mathcal{B}g^\star)(z)| \le \pi^{-d/4}\, 2^{-d/2}\, M_0\, \exp\Big(\frac{R}{\sqrt{2}} |\operatorname{Re} z|\Big).$$

在多圆盘 $\max_j |z_j| \le r$ 上，我们有 $|\operatorname{Re}\, z| \le |z| \le \sqrt{d}\, r$，因此

$$\sup_{\max_j |z_j| \le r} |(\mathcal{B}g^\star)(z)| \le \pi^{-d/4}\, 2^{-d/2}\, M_0\, \exp\Big(\frac{R}{\sqrt{2}} \sqrt{d}\, r\Big),$$

这就是 (C.19)。 □

引理 C.7。设 $\mathcal{F}_B = \{f_c(x) := \sum_{|\alpha|\le n} c_\alpha \psi_\alpha(x)\ \sum_{|\alpha|\le n} |c_\alpha|^2 \le B\}$。则存在某个绝对常数 $C > 0$，使得

$$\log \mathcal{N}_{[]}(\mathcal{F}_B, \|\cdot\|_\infty, \varepsilon) \le p \log\left(1 + \frac{BC\sqrt{p}}{\varepsilon}\right),$$

其中 $p = \binom{n+d}{d}$。

证明。将系数集合记为 $\mathcal{C} := \{c = (c_\alpha, |\alpha| \le n) \in \mathbb{R}^p$，其中 $\sum_{|\alpha|\le n} |c_\alpha|^2 \le B\}$。注意到

$$\mathcal{N}(\mathcal{C}, \|\cdot\|_2, \varepsilon) \le \left(1 + \frac{2B}{\varepsilon}\right)^p,$$

设 $f_c, f_{c'} \in \mathcal{F}_B$。则

$$\|f_c - f_{c'}\|_\infty \le \sqrt{p} \max_{|\alpha|\le n} \|\psi_\alpha\|_\infty \|c - c'\|_2 \le C^d \sqrt{p} \|c - c'\|_2,$$

其中，C 是某个绝对常数。因此，

$$\mathcal{N}_{[]}(\mathcal{F}_B, \|\cdot\|_\infty, \varepsilon) \le \left(1 + \frac{4BC\sqrt{p}}{\varepsilon}\right)^p$$

□

命题 C.8。 令 H_n 表示由生成函数定义的物理学家厄米多项式

$$\sum_{n=0}^{\infty} H_n(x)\,\frac{t^n}{n!} = e^{2xt-t^2}, \qquad x,t\in\mathbb{R}.$$

定义（归一化的）*Hermite*函数

$$\psi_n(x) := \frac{1}{(2^n n!\sqrt{\pi})^{1/2}}\,H_n(x)\,e^{-x^2/2}, \qquad x\in\mathbb{R},\ n\in\mathbb{N}_0, \tag{C.20}$$

以及产生算符

$$a^* := \frac{1}{\sqrt{2}}\Big(x-\frac{d}{dx}\Big).$$

那么对每个 $n\geq 1$,

$$a^*\psi_{n-1} = \sqrt{n}\,\psi_n, \qquad \textit{equivalently} \qquad \psi_n = \frac{1}{\sqrt{n}}\,a^*\psi_{n-1}. \tag{C.21}$$

证明。 根据生成函数，对 x求导:

$$\sum_{n=0}^{\infty} H_n'(x)\,\frac{t^n}{n!} = \partial_x\big(e^{2xt-t^2}\big) = 2t\,e^{2xt-t^2} = 2t\sum_{n=0}^{\infty} H_n(x)\,\frac{t^n}{n!} = \sum_{n=1}^{\infty} 2n\,H_{n-1}(x)\,\frac{t^n}{n!}.$$

比较 t^n 的系数，得到

$$H_n'(x) = 2n\,H_{n-1}(x), \qquad n\geq 1. \tag{C.22}$$

接下来，对生成函数关于 t求导:

$$\sum_{n=0}^{\infty} H_{n+1}(x)\,\frac{t^n}{n!} = \partial_t\big(e^{2xt-t^2}\big) = (2x-2t)e^{2xt-t^2} = 2x\sum_{n=0}^{\infty} H_n(x)\,\frac{t^n}{n!} - 2\sum_{n=0}^{\infty} H_n(x)\,\frac{t^{n+1}}{n!}.$$

将最后一项改写为 $\sum_{n=0}^{\infty} 2n\,H_{n-1}(x)\,\frac{t^n}{n!}$ 并比较系数:

$$H_{n+1}(x) = 2x\,H_n(x) - 2n\,H_{n-1}(x), \qquad n\geq 1. \tag{C.23}$$

利用乘积法则,

$$\frac{d}{dx}\big(H_{n-1}(x)e^{-x^2/2}\big) = H_{n-1}'(x)e^{-x^2/2} - xH_{n-1}(x)e^{-x^2/2}.$$

因此

$$\begin{aligned}\Big(x-\frac{d}{dx}\Big)\big(H_{n-1}e^{-x^2/2}\big) &= xH_{n-1}e^{-x^2/2} - \Big(H_{n-1}'e^{-x^2/2} - xH_{n-1}e^{-x^2/2}\Big)\\ &= \big(2xH_{n-1} - H_{n-1}'\big)e^{-x^2/2}.\end{aligned} \tag{C.24}$$

现在将 C.22 与 $n-1$一起使用:

$$H_{n-1}'(x) = 2(n-1)H_{n-2}(x).$$

将其代入 C.24，并对索引 $n-1$ 应用 C.23：

$$2xH_{n-1} - H'_{n-1} = 2xH_{n-1} - 2(n-1)H_{n-2} = H_n.$$

因此我们已证明

$$\Big(x - \frac{d}{dx}\Big)\big(H_{n-1}(x)e^{-x^2/2}\big) = H_n(x)e^{-x^2/2}. \tag{C.25}$$

Let

$$N_n := (2^n n!\sqrt{\pi})^{-1/2}, \quad \text{so that} \quad \psi_n = N_n\, H_n\, e^{-x^2/2}.$$

利用（C.25），

$$a^*\psi_{n-1} = \frac{1}{\sqrt{2}}\Big(x - \frac{d}{dx}\Big)\big(N_{n-1}H_{n-1}e^{-x^2/2}\big) = \frac{N_{n-1}}{\sqrt{2}}\, H_n\, e^{-x^2/2}.$$

接下来只需比较 $\frac{N_{n-1}}{\sqrt{2}}$ 与 $\sqrt{n}\, N_n$：

$$\frac{N_{n-1}}{\sqrt{2}} = \frac{1}{\sqrt{2}}\,(2^{n-1}(n-1)!\sqrt{\pi})^{-1/2} = (2^n(n-1)!\sqrt{\pi})^{-1/2},$$

and

$$\sqrt{n}\, N_n = \sqrt{n}\,(2^n n!\sqrt{\pi})^{-1/2} = (2^n(n-1)!\sqrt{\pi})^{-1/2}.$$

因此 $\frac{N_{n-1}}{\sqrt{2}} = \sqrt{n}\, N_n$，从而

$$a^*\psi_{n-1} = \sqrt{n}\, N_n\, H_n\, e^{-x^2/2} = \sqrt{n}\,\psi_n,$$

即为（C.21）。 □

D 数值实验的进一步细节

本节将详细介绍第 4 节所呈现数值实验的相关细节。本节安排如下：附录 D.1 讨论一般实现细节。附录D.2介绍所用指标的信息，并描述该算法。附录 D.3 补充介绍瑞士卷到 S 曲线二维问题实验的相关信息。附录 D.4 补充介绍 25 个高斯混合外推实验的详细信息。附录 D.5 补充介绍生物数据实验的详细信息。最后，附录 D.6列出实验所用的全部最终超参数值。

D.1 一般实现细节

在我们的实验中，我们高度重视超参数的选择。鉴于该算法的特性，仔细调整超参数对于稳定训练和生成至关重要。我们使用 Optuna [Akiba 等人， 2019]。对于 25 个高斯实验中切片 $\mathbb{W}_1$ 指标的 Optuna 试验优化，我们进行了 50 次；对于单细胞数据中的切片 $\mathbb{W}_1$ 指标，我们进行了 100 次 Optuna 试验优化。

所有计算均在 Nvidia T4 GPU 上完成。对于 SinkhornBridge 算法，我们采用了官方代码库中提供的超参数值，因为我们无法对该算法进行广泛测试并自行选择其超参数。

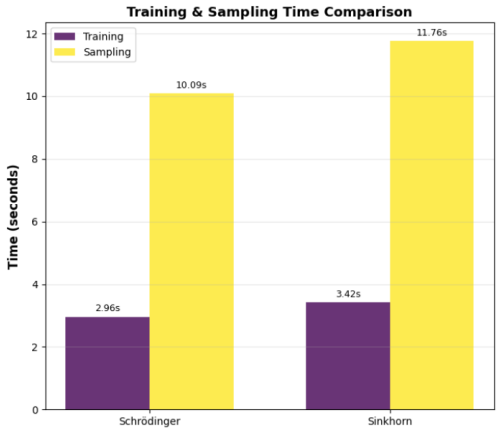


图 3：在瑞士卷到 S 曲线的转换任务中，ERM-Bridge 和 SinkhornBridge 的训练时间与采样时间比较。两种算法均使用 2000 个训练点。

D.2 实验中使用的指标

我们采用切片Wasserstein距离作为主要指标，因为它能够提供定量且有理论依据的分布相似性度量，并且已广泛应用于研究论文和实际应用中。

$$\text{Sliced } \mathbb{W}_1(X,Y) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{W}_1(\{\langle x_i, \theta_n\rangle\}_{i=1}^{M}, \{\langle y_i, \theta_n\rangle\}_{i=1}^{M}),$$

其中，$\{\theta_n\}_{n=1}^{N}$ 是在 $\mathbb{S}^{d-1}$上均匀分布的随机投影。

对于切片Wasserstein距离，使用了 100 个随机投影。

我们还提供了所提出算法的训练算法描述1 和采样2。所提供的伪代码使读者能够直观地评估我们的方法与 Sinkhorn-Bridge 有何不同。

D.3 瑞士卷到 S 曲线实验的评估细节

在本实验中，我们在生成建模中的一个常见二维问题上，验证了算法在各个中间时间点的性能和采样结果。由于该问题并不复杂，且对于任何合理的训练和采样参数，两种算法都能很好地解决该问题，因此 Optuna 未对其进行超参数优化。

我们还比较了本算法与 SinkhornBridge 在该问题上的运行时间，结果表明采样速度显著提升。

D.4 高斯混合评估详情

在本实验中，我们旨在清晰展示所提出的算法以及 Sinkhorn-Bridge 在采样数据集相对于训练数据集存在显式数据偏差时的行为。为此，我们创建了一个

合成示例，使用二维空间中由 25 个高斯分布构成的均匀网格，以及截断正态分布：训练数据集中为 $[-10, 10]$，测试集中为 $[-1, 1], [-3, 3], [-5, 5], [-10, 10]$。

通过目视检查这些图，可以明显看出 SinkhornBridge 离散性质所带来的问题。该实验的价值在于，合成示例中展示的类似问题也会出现在真实数据中。例如，我们的算法在单细胞数据上的表现更好D.5，其中初始分布和最终分布都只能以样本的形式获得。

D.5 单细胞数据评估详情

在此实验中，我们旨在展示算法在数据到数据的转换任务中的表现。连续对数势函数不仅取得了最佳指标值，而且更易于优化，所需的采样时间也更短。

LightSB 模型的结果取自论文 [Korotin 等人， 2024]。

D.6 最终超参数值

以下列出了所有实验的最终超参数值。

Swiss-Roll 到 S 曲线实验：

- 实验参数：批量大小 $= 1000$，学习率 $= 2e-3$，训练轮数 $= 1500$，σ 结束 =0。5，损失缩放 $= 1.0$，隐藏维度 $= 64$.

高斯混合实验：

- 实验参数：批量大小 $= 64$，学习率 $= 5e-4$，训练轮数 $= 140$，σ结束 $= 0.9$，损失缩放 =0。11 隐藏维度 $= 128$。

单细胞实验：

- 实验参数：批量大小 $= 2048$，学习率 $= 1e-4$，训练轮数 $= 141$，σ 结束 =0。4216，损失缩放 $= 196.5431$，隐藏维度 -= 2048.

算法 1ERM-Bridge 的训练

输入： 数据集 $\mathcal{X}, \mathcal{Y}$、核带宽 σ、学习率 η、批量大小 B、迭代次数 N_{steps}

初始化： 神经网络参数 θ，用于 ϕ_θ

对于 i，从 1 到 N_{steps} **执行**

  采样批次 $X_b \sim \mathcal{X}$ 和 $Y_b \sim \mathcal{Y}$

  对于 $y \in Y_b$ **执行**

    $V(y) \leftarrow$ 对数 $\phi_\theta(y)$

  结束循环

  对于 $x \in X_b$ **执行**

    项$(y) \leftarrow -\frac{\|x-y\|^2}{2\sigma^2} - V(y)$　　对于所有 $y \in Y_b$

    对数 $D(x) \leftarrow$ 对数 $\sum_{y \in Y_b} \exp(\text{term}(y))$

  结束循环

  对于 $y \in Y_b$ **执行**

    term$(x) \leftarrow -\frac{\|y-x\|^2}{2\sigma^2} -$ 对数 $D(x)$　　对于所有 $x \in X_b$

    log$(C\phi)(y) \leftarrow$ 对数 $\sum_{x \in X_b} \exp(\text{term}(x))$

  结束循环

  $\Delta(y) \leftarrow V(y) - \log(C\phi)(y)$

  $\mathcal{L}(\theta) \leftarrow \frac{1}{|Y_b|} \sum_{y \in Y_b} (\Delta(y) - \text{均值}(\Delta))^2$

  $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}(\theta)$

结束循环

输出： 学习得到的势参数 θ^*

Algorithm 2 Sampling via Learn[ed Co] ntinuous Drif

t 1：输入：训练模型 ϕ_{θ^*}，初始样本 x_0，参考目标样本 $\mathcal{Y}_{ref}$，总时间 T，步数 K2：初始化： $x \leftarrow x_0, t \leftarrow 0, \Delta t \leftarrow T/K$3：4：对于 $k = 0$ 到 $K - 1$ 执行 5：设置噪声水平 σ_t（例如，通过余弦调度）6：$\nu_t \leftarrow \sigma_t^2(T - t)$ 7：8：对于 $j = 1 \ldots |\mathcal{Y}_{ref}|$ 执行 9：$L_j \leftarrow -\frac{\|x-y_j\|^2}{2\nu_t} - \log \phi_{\theta^*}(y_j)$ 10：结束循环 11：$h(x) \leftarrow \log \sum_j \exp(L_j)$ 12：$g \leftarrow \nabla_x h(x)$ 13：$u(x,t) \leftarrow \sigma_t^2 \cdot g$ 14：15：采样噪声 $\xi \sim \mathcal{N}(0, I)$ 16：$x \leftarrow x + u(x,t)\Delta t + \sigma_t \sqrt{\Delta t} \cdot \xi$17：$t \leftarrow t + \Delta t$18：结束循环 19：返回：输运样本 x_T