arXiv:2601.07228v1 [math.ST] 12 Jan 2026

# 利用矩方法研究相关数据经验测度的 Wasserstein 集中

ArashA. Amini LucianoVinas
aaamini@g.ucla.edu lucianovinas@g.ucla.edu

2026年1月13日

**摘要**

我们建立了一个一般性的集中结果，研究一列随机变量的经验测度与其期望之间的 1-Wasserstein 距离。不同于依赖独立性（例如 Sanov 定理）或特定混合条件的标准结果，我们的结果仅要求满足两个条件：(1) 控制经验矩的方差；(2) 满足一种灵活的尾部条件，我们称之为 $\Psi_{r_n}$-次高斯性。该方法允许变量之间存在显著依赖，只要它们的代数矩表现可预测即可。证明采用矩方法，并借助 Jackson 核对 Lipschitz 函数进行多项式逼近，从而将矩集中转化为拓扑集中。

## 1 引言

经验测度 $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$ 收敛到确定性极限 $\bar{\mu}$ 是概率论与统计学中的一个基本问题。当变量 $Y_i$ 独立同分布（i.i.d.）时，Sanov 定理或 Varadarajan 定理等经典结果可以提供有力的保证。然而，在许多现代高维情形中——例如随机矩阵的谱分布、相互作用粒子系统或图神经网络中的特征传播——变量 $Y_i$ 呈现出复杂的依赖结构。

在这类具有依赖性的情形中，要确立独立性或混合性质可能难以处理。然而，通常可以通过组合技巧（例如迹方法或游走计数）计算系统的矩 $\mathbb{E}[n^{-1}\sum Y_i^k]$。经典的矩方法[4, 第 2 节。5]利用这些代数量证明弱收敛。

本文给出一个“衔接”定理，将代数矩控制提升为对 1-Wasserstein 度量（$W_1$）下的集中性。具体而言，我们证明，若经验矩的方差渐近趋于零，且分布的尾部衰减足够快，则经验测度集中于 $W_1$。

一项关键的技术贡献是对检验函数的处理。矩收敛自然能够控制多项式，而 $W_1$ 度量则是通过 Lipschitz 函数定义的。为弥合这一差距，我们利用 Chebyshev-Jackson 核对 Lipschitz 函数进行多项式逼近。这样既能控制多项式系数的大小，又能达到 Lipschitz 函数逼近的最优速率；这两点都是证明所必需的。

## 2 预备知识

令$\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_{i,\ n}}$ 为实值随机变量三角阵$\{Y_{i,n}\}_{i=1}^n$的经验测度。我们将期望测度记为 $\bar{\mu}_n = \mathbb{E}[\mu_n]$。两个

概率测度 $\mu$ 和 $\nu$ 在 $\mathbb{R}$ 上的 1-Wasserstein 距离定义为其对偶形式：

$$W_1(\mu,\nu) = \sup_{f\in \mathrm{Lip}(1)} \left| \int f d\mu - \int f d\nu \right|, \tag{1}$$

其中，Lip(1) 表示 1-Lipschitz 函数的集合 $f\colon \mathbb{R}\to\mathbb{R}$。

### 2.1 Ψr 范数与渐近次高斯性

我们分析中的一个核心特征是采用了比一致次高斯性更宽松的尾部条件。我们考虑一列参数 $r_n$，它可能随 $n$ 增大。

**定义 1 (**截断 $\Psi_r$ 范数**)**。对于实数 $r \geq 2$，令 $\Psi_r(x) = \sum_{j=1}^{\lfloor r/2\rfloor} \frac{x^{2j}}{j!}$。随机变量 $X$ 的 $\Psi_r$ 范数定义为：

$$\|X\|_{\Psi_r} = \inf\left\{K>0 : \mathbb{E}\left[\Psi_r(|X|/K)\right] \leq 1\right\}. \tag{2}$$

若 $r\to\infty$, $\Psi_r(x)$ 趋于 $\exp(x^2)-1$，则恢复标准的 $\psi_2$（次高斯）范数。然而，对于有限的 $r$，$\|X\|_{\Psi_r}<\infty$ 仅意味着直到 $r$ 阶的矩表现得如同次高斯变量的矩。这使我们能够处理固定 $n$ 时可能并非严格次高斯、但其矩在不断增长的阶数 $r_n$ 以内表现良好的变量。

**引理 2**（矩与尾部）。设 $r\geq 2$ 以及 $\|X\|_{\Psi_r}\leq K$。则：

（*a*）（矩界）对于所有 $p\in[2,2\lfloor r/2\rfloor]$，$(\mathbb{E}|X|^p)^{1/p}\leq C_1 K\sqrt{p}$。

（*b*）（尾界）存在常数 $c_0$、$c_1>0$，使得对于 $t\geq c_0 K$：

$$\mathbb{P}(|X|\geq t)\leq \exp\left(-c_1\min\left\{\frac{t^2}{K^2}, \lfloor r/2\rfloor\right\}\right).$$

（a）部分的证明遵循标准的奥利奇范数论证。对于（b）部分，请参见 [3, 引理 25]。

## 3 主要结果

我们的主要结果建立了基于矩方差和尾部衰减的 $W_1$ 集中性。

**定理 3**。设 $\{Y_{i,n}\}_{i=1}^n$ 为一个实值随机变量三角阵。设 $\mu_n$ 为其经验测度，且 $\bar{\mu}_n=\mathbb{E}[\mu_n]$。假设存在一个序列 $r_n=\omega(1)$ 使得：

***(a)*** 一致尾部控制：这些变量一致为 $\Psi_{r_n}$ 次高斯。也就是说，存在$\zeta>0$ 使得 $\sup_{i\in[n]}\|Y_{i,n}\|_{\Psi_{r_n}}\leq\zeta$.

***(b)*** 矩集中：对于每个固定整数 $k\geq 1$，经验矩的方差趋于零：

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_{i,n}^k\right)\to 0 \quad as\ n\to\infty.$$

那么，期望瓦瑟斯坦距离趋于零：

$$\mathbb{E}\big[W_1(\mu_n,\bar{\mu}_n)\big]\to 0 \quad as\ n\to\infty.$$

当变量独立同分布时，条件 (b) 显然成立；但对于许多相关系统，只要相关性衰减得足够快，条件 (b) 也成立（例如随机矩阵的谱统计量、稀疏图中的节点特征）。条件 (a) 确保变量虽不必有界，但其概率质量不会过快地逃逸到无穷远。

### 3.1 扩展到随机向量

矩方法在高维情形下尤其强大，因为此时计算投影矩 $\mathbb{E}\langle\theta,\ Y\rangle^k$ 通常是可行的。我们将定理 3 推广到 $\mathbb{R}^d$中的随机向量。

我们将测度 $\mu$ 在 $\mathbb{R}^d$ 上的绝对一阶矩记为 $M_1(\mu) = \int \|x\| d\mu(x)$。

**定理 4.** 设 $Y_{i,n}$ 是 $\mathbb{R}^d$中的随机向量。设 $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{Y_{i,n}}$ 和 $\bar{\mu}_n = \mathbb{E}\mu_n$。假设存在一个序列 $r_n = \omega(1)$，使得对每个单位向量 $\theta \in S^{d-1}$:

*(a)* 一致投影尾部：投影一致地服从 $\Psi_{r_n}$ 次高斯分布：

$$\sup_{i\in[n]} \|\langle\theta, Y_{i,n}\rangle\|_{\Psi_{r_n}} \le \zeta(\theta) < \infty.$$

*(b)* 投影矩集中：对于每个固定的 $k \ge 1$:

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \langle\theta, Y_{i,n}\rangle^k\right) \to 0 \quad as\ n \to \infty.$$

*(c)* 有界一阶矩：上确界$_{n\ge1}\, M_1(\bar{\mu}_n) < \infty$。
于是，$\mathbb{E}\big[W_1(\mu_n, \bar{\mu}_n)\big] \to 0$ 当 $n \to \infty$ 时。

# 4 多项式逼近

为了证明主要结果，我们需要关于 Lipschitz 函数多项式逼近的结果。标准的魏尔斯特拉斯逼近给出的多项式逼近速率为 $m^{-1/2}$，其中多项式次数为 $m$，这太慢了。Lipschitz 函数的最优逼近速率为 $m^{-1}$，这已足够。不过，我们还需要控制多项式系数的大小。我们采用切比雪夫–杰克逊逼近，它能在这两者之间取得平衡。

我们先给出切比雪夫多项式系数的一个界。

**引理 5.** 设 $T_k$ 为第 $k$ 个切比雪夫多项式，并设 $[T_k]_j$ 为 $T_k(x)$中 $x^j$ 的系数。则，$|[T_k]_0| \le 1$ 且

$$\max_{1\le j\le k} |[T_k]_j| \le (1+\sqrt{2})^k \le 3^k.$$

证明。第一部分显然，因为 $[T_k]_0 \in \{0, 1\}$。对于第二部分，由递推关系 $T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$，可得

$$|[T_{k+1}]_j| \le 2|[T_k]_{j-1}| + |[T_{k-1}]_j|.$$

假设对于某个常数 $c$ 和所有 $T_r$，$r \le k$，当 $\max_{1\le j\le k} |[T_k]_j| \le c^k$ 时该结论成立，则有 $|[T_{k+1}]_j| \le 2\cdot c^k + c^{k-1}$。因此，若 $2c^k + c^{k-1} \le c^{k+1}$，结论便可由归纳法得出。该不等式对 $c \ge 1+\sqrt{2}$ 成立。证明完毕。 □

**引理 6** (切比雪夫–杰克逊逼近)。设 $B \ge 3$。那么，对于任意满足 $f(0) = 0$ 的 $f$：$[-B, B] \to \mathbb{R}$1-利普希茨函数，都存在一个多项式 $P(x) = \sum_{j=0}^m c_j x^j$，满足 $m \in 4\mathbb{N}$，使得

$$\sup_{x\in[-B,B]} |f(x) - P(x)| \le \frac{18B}{m}, \quad and \quad |c_j| \le 6B\cdot 3^{m-j}, \quad for\ all\ j \ge 0.$$

证明。考虑定义在 $[-1,1]$ 上的 $L$-利普希茨函数 $g$，满足 $g(0)=0$。那么，对于每个 $m\in 4\mathbb{N}$，都存在一个形如以下形式的多项式

$$Q_m(x)=\sum_{k=0}^{m}\lambda_{k,m}a_k(g)T_k(x)$$

其中 $\lambda_{k,\,m}$ 由 Jackson 核导出，满足 $0\le\lambda_{k,\,m}\le 1$，而 $a_k(g)$ 是 $g$ 的切比雪夫系数，使得

$$\sup_{x\in[-1,1]}|g(x)-Q_m(x)|\le\frac{18L}{m},\quad |a_k(g)|\le\frac{\sqrt{8/\pi}L}{k},\quad k\ge 1.$$

参见 [2] 中的事实3.2和3.3。切比雪夫系数为

$$a_k(g)=\frac{2}{\pi}\int_{-1}^{1}\frac{g(x)T_k(x)}{\sqrt{1-x^2}}dx,\quad k\ge 1,$$

对于 $k=0$，将 $2/\pi$ 替换为 $1/\pi$ 后，同一公式仍然成立。对于 $k=0$，利用 $g(0)=0$，使得对所有 $x\in[-1,1]$，$|g(x)|\le L|x|$，且 $T_0(x)=1$，我们有

$$|a_0(g)|\le\frac{1}{\pi}\int_{-1}^{1}\frac{L|x|}{\sqrt{1-x^2}}dx=\frac{2L}{\pi}.$$

因此，对于所有 $k\ge 0$，一个适用的粗略上界为 $|a_k(g)|\le 2L$。

令 $a_{k,\,m}=\lambda_{k,\,m}a_k(g)$，并注意到根据上述讨论，对所有 $k\ge 0$，$|a_{k,\,m}|\le 2L$。重写$Q_m(x)=\sum_{j=0}^{m}b_jx^j$，可得 $b_j=\sum_{k=j}^{m}a_{k,\,m}[T_k]_j$，其中 $[T_k]_j$ 是 $T_k(x)$ 中 $x^j$ 的系数。由引理 5 可知

$$|b_j|\le\sum_{k=j}^{m}2L\cdot 3^k\le 2L\cdot 3^m\sum_{k=j}^{m}3^{k-m}\le 2L\cdot 3^m\frac{1}{1-3^{-1}}\le 6L\cdot 3^m$$

对所有 $j\ge 0$。

如果 $f$ 在 $[-B,B]$ 上是 1-Lipschitz 的，且 $f(0)=0$，那么 $g(x)=f(Bx)$ 在 $[-1,1]$ 上是 $B$-Lipschitz 的，且 $g(0)=0$。令 $Q_m$ 为关于 $g$的上述多项式，并令 $P(x)=Q_m(x/B)=\sum_{j=0}^{m}(b_j/B^j)x^j=:\sum_{j=0}^{m}c_jx^j$。于是，

$$|c_j|\le 6B\frac{3^m}{B^j}\le 6B\cdot 3^{m-j}$$

假设 $B\ge 3$。我们还有 $\sup_{x\in[-B,B]}|f(x)-P(x)|=\sup_{x\in[-1,1]}|g(x)-Q_m(x)|\le\frac{18B}{m}$。证明完毕。□

# 5 主要结果的证明（标量情形）

定理 *3* 的证明。我们将 $\mathrm{Lip}(f)=\sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|}$记作 $f$的 Lipschitz常数。考虑函数集合

$$\mathcal{L}=\{f:\mathbb{R}\to\mathbb{R}\mid \mathrm{Lip}(f)\le 1,f(0)=0\},\quad \mathcal{L}_B=\{f\mathbb{1}_{|x|\le B}\mid f\in\mathcal{L},\ B>0\}.$$

令 $\varpi_n:=\mu_n-\bar{\mu}_n$。根据 $W_1$ 的对偶刻画，我们有

$$W_1(\mu_n,\bar{\mu}_n)\le\sup_{f\in\mathcal{L}}|\varpi_nf|.$$

通过拆分 $f = f\mathbb{1}_{|x|\le B} + f\mathbb{1}_{|x|>B}$，我们有

$$W_1(\mu_n, \bar{\mu}_n) \le \sup_{f\in\mathcal{L}_B} |\varpi_n f| + \sup_{f\in\mathcal{L}} |\varpi_n(f\mathbb{1}_{|x|>B})|. \tag{3}$$

固定 $\epsilon \in (0,\ 1)$，先考虑第二项。对于任意可积的 $f$，有

$$\begin{aligned} |\varpi_n(f\mathbb{1}_{|x|>B})| &\le |\mu_n(f\mathbb{1}_{|x|>B})| + |\bar{\mu}_n(f\mathbb{1}_{|x|>B})| \\ &\le \mu_n(|f|\mathbb{1}_{|x|>B}) + \bar{\mu}_n(|f|\mathbb{1}_{|x|>B}). \end{aligned} \tag{4}$$

对于 $f \in \mathcal{L}$，有 $|f(x)| = |f(x) - f(0)| \le |x - 0|$。

于是，有

$$|\varpi_n(f\mathbb{1}_{|x|>B})| \le \mu_n(|x|\mathbb{1}_{|x|>B}) + \bar{\mu}_n(|x|\mathbb{1}_{|x|>B})$$

对 $f \in \mathcal{L}$ 取上确界，然后取期望，有

$$\mathbb{E}\sup_{f\in\mathcal{L}} |\varpi_n(f\mathbb{1}_{|x|>B})| \ \le\ 2\bar{\mu}_n(|x|\mathbb{1}_{|x|>B}) = \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\big(|Y_{i,n}|\mathbb{1}_{\{|Y_{i,n}|>B\}}\big).$$

取足够大的 $n$，使得

$$r_n \ge 2\Big(\frac{B^2}{\zeta^2} + 1\Big) \tag{5}$$

我们将在最后验证这一点。此外，取 $B \ge B_0(\zeta) := c_0\zeta$，其中 $c_0$ 是引理 2(b) 中的常数。于是，根据该引理，有 $\mathbb{P}(|Y_{i,\ n}| > B) \le \exp(-c_1B^2/\zeta^2)$；根据引理 2(a)，有$\mathbb{E}[Y_{i,n}^2] \le 2C_1^2\zeta^2$。然后，由柯西-施瓦茨不等式，有

$$\mathbb{E}\big(|Y_{i,n}|\mathbb{1}_{\{|Y_{i,n}|>B\}}\big) \le \sqrt{\mathbb{E}[|Y_{i,n}|^2]\cdot\mathbb{P}(|Y_{i,n}| > B)} \le \sqrt{2}C_1\zeta\cdot\exp(-cB^2/2\zeta^2).$$

取足够大的 $B \ge B_1(\zeta)$，使得 $B_1(\zeta)$ 足够大，便可使右端任意小，从而得到

$$\mathbb{E}\sup_{f\in\mathcal{L}} |\varpi_n(f\mathbb{1}_{|x|>B})| \le 2\epsilon.$$

现在考虑式 (3) 中的第一项。将 $\mathcal{L}_B$ 视为 $(C_b([-B,B])\ \|\cdot\|_\infty)$ 的一个子空间，限制到 $[-B,B]$ 后，$\mathcal{L}_B$ 一致有界且等度连续，因此由阿尔泽拉-阿斯科利定理，它在上确界范数拓扑下相对紧。这进而意味着 $\mathcal{L}_B$ 是全有界的。因此，存在 $f_1,\dots,f_M \in \mathcal{L}_B$，它们构成 $\mathcal{L}_B$ 在上确界范数下的一个 $\epsilon$-网，其中某个 $M = M(\epsilon, B) < \infty$。也就是说，对于任意 $f \in \mathcal{L}_B$，存在 $f_\ell$ 使得 $\|f - f_\ell\|_\infty \le \epsilon$，因此

$$\begin{aligned} |\varpi_n f| \ &\le\ |\varpi_n(f - f_\ell)| + |\varpi_n f_\ell| \\ &\le\ \|\varpi_n\|_{\mathrm{TV}}\cdot\|f - f_\ell\|_\infty + |\varpi_n f_\ell| \ \le\ 2\epsilon + |\varpi_n f_\ell|. \end{aligned}$$

对 $f \in \mathcal{L}_B$ 取上确界，得到

$$\sup_{f\in\mathcal{L}_B} |\varpi_n f| \ \le\ 2\epsilon + \sup_{\ell\in[M]} |\varpi_n f_\ell|.$$

取 $B \ge 36$，$f_\ell$ $\sum_{j=0}^{m} c_{j\ell}x^j$ $m = 4\lceil C_2B/\epsilon\rceil \in 4\mathbb{N}$ $C_2 = 18$ $Q_\ell\ x\ =\ \mathbb{1}_{\{|x|\le B\}}\cdot$。根据引理，每个对象都存在一个（截断）多项式 ()，其中（可取），使得

$$\|f_\ell - Q_\ell\|_\infty \le \epsilon,$$

以及 $|c_{j\ell}| \le 6B\cdot 3^{m-j} =:$ 对于所有 $j \ge 0$ 和 $\ell \in [M]$，均有 $a_j$。我们有

$$|\varpi_n f_\ell| \le \|\varpi_n\|_{\mathrm{TV}}\cdot\|f_\ell - Q_\ell\|_\infty + |\varpi_n Q_\ell|.$$

由此可得

$$\sup_{\ell\in[M]}|\varpi_n f_\ell| \le 2\epsilon + \sup_{\ell\in[M]}|\varpi_n Q_\ell|$$

并且我们有

$$\begin{aligned}\sup_{\ell\in[M]}|\varpi_n Q_\ell| &\le \sup_{\ell\in[M]}\Big|\sum_{j=0}^m c_{j\ell}\,\varpi_n(x^j\mathbb{1}_{|x|\le B})\Big| \\ &\le \sum_{j=0}^m\big(\sup_{\ell\in[M]}|c_{j\ell}|\big)\cdot|\varpi_n(x^j\mathbb{1}_{|x|\le B})| \le \sum_{j=0}^m a_j\,|\varpi_n(x^j\mathbb{1}_{|x|\le B})|\end{aligned}$$

我们有

$$|\varpi_n(x^j\mathbb{1}_{|x|\le B})| \le |\varpi_n(x^j)| + |\varpi_n(x^j\mathbb{1}_{|x|>B})|.$$

然后，对于第二项，利用（4），对于所有 $j\in[m]$，我们有

$$\begin{aligned}|\varpi_n(x^j\mathbb{1}_{|x|>B})| &\le \mu_n(|x^j|\mathbb{1}_{|x|>B}) + \bar\mu_n(|x^j|\mathbb{1}_{|x|>B}) \\ &\le \mu_n(|x^m|\mathbb{1}_{|x|>B}) + \bar\mu_n(|x^m|\mathbb{1}_{|x|>B}).\end{aligned}$$

对 $j\in[m]$ 取最大值，再取期望，我们有

$$\mathbb{E}\sup_{j\in[m]}|\varpi_n(x^j\mathbb{1}_{|x|>B})| \le 2\bar\mu_n(|x|^m\mathbb{1}_{|x|>B}) = \frac{2}{n}\sum_{i=1}^n\mathbb{E}\big(|Y_{i,n}|^m\mathbb{1}_{\{|Y_{i,n}|>B\}}\big).$$

取 $n$ 足够大，使得

$$r_n \ge 2m = 8\lceil C_2B/\epsilon\rceil, \tag{6}$$

我们将在最后验证这一点。于是，根据引理 2(a)，我们有 $\mathbb{E}[|Y_{i,n}|^{2m}] \le (C_1\zeta)^{2m}(2m)^m = (2C_1^2\zeta^2 m)^m$。然后，根据柯西-施瓦茨不等式，我们有

$$\begin{aligned}\mathbb{E}\big(|Y_{i,n}|^m\mathbb{1}_{\{|Y_{i,n}|>B\}}\big) &\le \sqrt{\mathbb{E}[|Y_{i,n}|^{2m}]\cdot\mathbb{P}(|Y_{i,n}|>B)} \\ &\le (2C_1^2\zeta^2 m)^m\cdot\exp(-cB^2/2\zeta^2)\end{aligned}$$

利用 $a_j = 6B\cdot 3^{m-j}$，我们得到 $\sum_{j=0}^m a_j \le 9B\cdot 3^m$。由此可知

$$\begin{aligned}\mathbb{E}\Big[\sum_{j=0}^m a_j|\varpi_n(x^j\mathbb{1}_{|x|>B})|\Big] &\le \Big(\sum_{j=0}^m a_j\Big)\cdot\mathbb{E}\sup_{j\in[m]}|\varpi_n(x^j\mathbb{1}_{|x|>B})| \\ &\le 9B\cdot 3^m\cdot 2(2C_1^2\zeta^2 m)^m\cdot\exp(-cB^2/2\zeta^2) \\ &\le 18\exp\Big(\log B + m\log(6C_1^2\zeta^2 m) - cB^2/2\zeta^2\Big) \\ &\le 18\exp\Big(\log B + 4\lceil C_2B/\epsilon\rceil\log\Big(24C_1^2\zeta^2\lceil C_2B/\epsilon\rceil\Big) - cB^2/2\zeta^2\Big).\end{aligned}$$

由于 $B^2$ 的增长速度快于 $B\log B$，对于足够大的某个 $B_2(\zeta,\epsilon)$，右端可以设为 $\le\epsilon$，使其适用于 $B\ge B_2(\zeta,\epsilon)$。对于这个 $B$，我们有

$$\begin{aligned}\mathbb{E}\sup_{\ell\in[M]}|\varpi_n Q_\ell| &\le \sum_{j=0}^m a_j\mathbb{E}|\varpi_n(x^j)| + \epsilon \\ &\le \sum_{j=0}^m a_j\sqrt{\mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^n Y_{i,n}^j\Big)} + \epsilon.\end{aligned}$$

根据定理的假设 (b),

$$\max_{0\le j\le m}\operatorname{Var}\Big(\frac{1}{n}\sum_{i=1}^{n}Y_{i,n}^{j}\Big)\le\epsilon^{2}/(\sum_{j=0}^{m}a_{j})^{2} \tag{7}$$

当 $n$ 充分大时。这给出 $\mathbb{E}\sup_{\ell\in[M]}|\varpi_n Q_\ell|\le 2\epsilon$。综合各部分可得

$$\mathbb{E}\sup_{f\in\mathcal{L}_B}|\varpi_n f|\le 2\epsilon+2\epsilon+2\epsilon=6\epsilon.$$

总之，取 $B=\max\{3,B_0(\zeta)\ B_1(\zeta)\ B_2(\zeta,\ \epsilon)\}$，并令 $n$ 足够大，使得对于所选的 $B$，(5) 和 (6) 成立，且 (7) 也成立，我们得到 $\mathbb{E}W_1(\mu_n,\ \bar{\mu}_n)\le 8\epsilon$。证明完毕。 □

# 6 向量扩展的证明（定理 4）

为证明向量情形，我们依赖这样一个事实：一维投影的收敛（我们已在定理 3 中证明）蕴含整个测度的收敛，前提是一阶矩受到控制。

我们用记号 $\eta_\theta$ 表示测度 $\eta$ 在 $\mathbb{R}^d$ 上经投影$x\mapsto\langle\theta,\ x\rangle$得到的推前。

**引理 7 (**投影的利普希茨连续性)。设 $\eta$ 是 $\mathbb{R}^d$ 上的概率测度。那么，对任意 $\theta_1$、$\theta_2\in\mathbb{R}^d$,

$$W_1(\eta_{\theta_1},\eta_{\theta_2})\le\|\theta_1-\theta_2\|M_1(\eta).$$

证明。令 $X\sim\eta$。利用 $\mathbb{R}$ 上测度的 $W_1$ 对偶表述:

$$\begin{aligned}W_1(\eta_{\theta_1},\eta_{\theta_2})&=\sup_{f\in\mathrm{Lip}(1)}\big|\mathbb{E}f(\langle\theta_1,X\rangle)-\mathbb{E}f(\langle\theta_2,X\rangle)\big|\\&\le\sup_{f\in\mathrm{Lip}(1)}\mathbb{E}\big|f(\langle\theta_1,X\rangle)-f(\langle\theta_2,X\rangle)\big|\\&\le\mathbb{E}|\langle\theta_1-\theta_2,X\rangle|\le\|\theta_1-\theta_2\|\cdot\mathbb{E}\|X\|.\end{aligned}$$

证毕。 □

以下命题使我们能够将标量收敛提升为向量收敛。

**命题 8**。设 $\{\mu_n\}$ 和 $\{\eta_n\}$ 是 $\mathbb{R}^d$ 上期望分别为 $\bar{\mu}_n$ 和$\bar{\eta}_n$的随机概率测度。假设:

*1.* 上确界$_{n\ge1}(M_1(\bar{\mu}_n)+M_1(\bar{\eta}_n))<\infty$。*2.* 对任意 $\theta\in S^{d-1}$, $\mathbb{E}\big[W_1(\mu_{n,\theta},\eta_{n,\theta})\big]\to0$ 当 $n\to\infty$ 时。

那么，$\mathbb{E}\big[W_1(\mu_n,\eta_n)\big]\to0$。

证明。令 $L_n=M_1(\mu_n)+M_1(\eta_n)$。根据引理 7 和三角不等式，映射 $\theta\mapsto W_1(\mu_{n,\ \theta},\ \eta_{n,\ \theta})$ 在球面上是 $L_n$-利普希茨的。具体而言，

$$\begin{aligned}W_1(\mu_{n,\theta_1},\eta_{n,\theta_1})-W_1(\mu_{n,\theta_2},\eta_{n,\theta_2})&\le W_1(\mu_{n,\theta_1},\mu_{n,\theta_2})+W_1(\eta_{n,\theta_1},\eta_{n,\theta_2})\\&\le L_n\|\theta_1-\theta_2\|.\end{aligned}$$

Bayraktar 和 Guo [1] 证明，$\mathbb{R}^d$ 中的 $W_1$ 距离受投影上的上确界控制：存在一个常数 $C(d)$, 使得

$$W_1(\mu_n,\eta_n)\le C(d)\sup_{\theta\in S^{d-1}}W_1(\mu_{n,\theta},\eta_{n,\theta}).$$

设 $\{\theta_1, \ldots, \theta_N\}$ 是 $S^{d-1}$的一个 $\epsilon$-网。对于任意 $\theta$，令 $\theta_j$为其最近邻。于是 :

$$W_1(\mu_{n,\theta}, \eta_{n,\theta}) \leq W_1(\mu_{n,\theta_j}, \eta_{n,\theta_j}) + L_n\epsilon.$$

取上确界的期望,

$$\mathbb{E}[W_1(\mu_n, \eta_n)] \leq C(d) \left( \epsilon\, \mathbb{E}[L_n] + \sum_{j=1}^{N} \mathbb{E} W_1(\mu_{n,\theta_j}, \eta_{n,\theta_j}) \right).$$

注意到 $\mathbb{E}[L_n] = M_1(\bar{\mu}_n) + M_1(\bar{\eta}_n)$，根据假设，它一致有界。根据假设（条件 2），求和项当 $n \to \infty$ 时趋于零。取 $n \to \infty$，则 $\epsilon \to 0$，即可得到结论。 □

定理 *4* 的证明。我们对 $\eta_n = \bar{\mu}_n$ 应用命题 8。根据假设 (c)，条件 1 成立。对于条件 2，固定 $\theta \in S^{d-1}$。投影变量 $Z_{i,\ n} = \langle \theta,\ Y_{i,\ n} \rangle$ 满足定理 3 的标量假设（具有一致的 $\Psi_{r_n}$ 尾部且矩方差趋于零）。注意到 $\bar{\mu}_{n,\ \theta} = \mathbb{E}[\mu_{n,\theta}]$。因此，定理 3 蕴含 $\mathbb{E} W_1(\mu_{n,\ \theta},\ \bar{\mu}_{n,\ \theta}) \to 0$。结论立即得出。 □

# 参考文献

[1] Erhan Bayraktar 和 Gaoyue Guo。Wasserstein 型度量之间的强等价性。*Electronic Communications in Probability*，26：1–13，2021。[2] Vladimir Braverman、Aditya Krishnan 和 Christopher Musco。亚线性时间谱密度估计。载于第 *54* 届 *ACM SIGACT* 计算理论研讨会论文集，第 1144–1157 页，2022。[3] Luciano Vinas 和 Arash A. Amini。多项式图神经网络的紧确界及图噪声的影响。*arXiv* 预印本 *arXiv:2407.19567*，2024。[4] A. W. van der Vaart，《渐近统计学》，剑桥统计与概率数学系列，剑桥大学出版社，剑桥，1998。