arXiv:2301.02168v2 [math.ST] 7 Jan 2024

# 高斯变分推断的近似精度

Anya Katsevich Philippe Rigollet
akatsevi@mit.edu rigollet@math.mit.edu

2024年1月9日

**摘要**

贝叶斯推断中的主要计算挑战，是计算针对高维后验分布的积分。过去几十年来，变分推断（VI）已成为计算这些积分的一种易处理近似方法，也是更成熟的马尔可夫链蒙特卡罗范式的一种可行替代方案。然而，人们对变分推断的近似精度知之甚少。本文根据维数和样本量，对高斯变分推断的全变差误差以及均值和协方差近似误差进行了界定。我们的误差分析依赖于对数后验的厄米特级数展开，其前几项恰好被与高斯变分推断优化问题相关的一阶最优性条件抵消。

## 1 引言

贝叶斯推断中的一个核心挑战是计算关于后验分布的积分 $\pi$，定义在 $\mathbb{R}^d$上。经典方法是通过采样来获取后验分布，即模拟一个旨在收敛到 $\pi$ 的马尔可夫链，并运行足够长的时间。然而，MCMC的计算开销可能很大，而且众所周知，难以为算法确定明确的停止准则 [6]。另一种通常计算成本较低的方法是变分推断(VI) [4]。VI的思路是在某个参数化族 $\mathcal{P}$中的所有测度里，找到与 $\pi$ 最接近的测度。自VI问世以来，人们提出了多种不同的接近程度度量 [7, 8], ，本文采用KL散度，这是迄今为止最常见的选择。通常， $\mathcal{P}$ 选择该参数族时，会确保关于族中测度的积分可以直接求得，或者易于计算。本文考虑高斯分布族。我们定义

$$\hat{\pi} = \mathcal{N}(\hat{m}, \hat{S}) \in \operatorname*{argmin}_{p \in \mathcal{P}_{\text{Gauss}}} \mathrm{KL}(p \,||\, \pi), \tag{1.1}$$

其中， $\mathcal{P}_{\text{Gauss}}$ 表示 $\mathbb{R}^d$上的非退化高斯分布族。得到高斯近似$\hat{}\pi$ 后，我们现在可以利用 $\int f d\hat{}\pi$近似 $\int g d\pi$ 。对于某些 $g$ （例如$\hat{}$多项式），后一个积分可以用$\hat{}m$ 和 $S$以闭式计算。否则，可以采用高斯采样，其计算成本远低于针对原始测度进行MCMC。

MCMC 与 VI 的一个关键区别在于，只要运行时间足够长，无偏马尔可夫链蒙特卡罗算法就能从 $\pi$ 中生成精度任意高的样本。另一方面，完美的 VI 算法输出的是 $\hat{}\pi$，而它本身也只是对 $\pi$ 的近似。因此，VI 中一个根本问题是理解近似 $\hat{}\pi \approx \pi$ 的质量。在这项工作中，我们针对一大类函数 $g$，以及若干特定

的指标，界定近似误差 $\|\hat{m}-\bar{m}\|$：全变差误差 TV($\pi$, $\hat{\pi}$)，以及均值和协方差误差 $\|\hat{m}-\bar{m}\|$ 和 $\|S-\Sigma\|$；其中 $\bar{m}$, $\Sigma$ 分别是 $\pi$ 的真实均值和协方差。

当然，我们不能期望任意一个可能呈多峰的 $\pi$ 都能被高斯分布很好地近似。然而，在贝叶斯推断的背景下，伯恩斯坦-冯·米塞斯定理（BvM）保证，在满足某些正则性条件时，当样本量趋于无穷大，后验分布会收敛到一个高斯密度 [27, 第 10]章。为了理解其中的原因，考虑一个一般的后验分布 $\pi=\pi_n$，其概率密度具有如下形式

$$\pi_n(\theta \mid x_{1:n}) \propto \nu(\theta)\prod_{i=1}^{n} p_\theta(x_i) \tag{1.2}$$

这里，$\nu$是先验，$p_\theta$是模型密度，而 $x_{1:n}=x_1$、.....、$x_n$是独立同分布观测。若$\nu$ 和 $p_\theta$ 处处为正，则可将 $\pi_n$ 写作

$$\pi_n(\theta) \propto e^{-nv_n(\theta)}, \qquad v_n(\theta) := -\frac{1}{n}\sum_{i=1}^{n}\log p_\theta(x_i) - \frac{1}{n}\log \nu(\theta). \tag{1.3}$$

随着 $n$ 增大，我们预期 $v_n$ 会趋近于 $v_\infty$，即与 $n$无关的总体负对数似然，也称为交叉熵。因此，如果 $n$ 较大且 $v_\infty$ 具有唯一的全局最小值，那么 $\pi_n$ 的大部分概率质量将集中在该点的一个邻域内。换言之，$\pi_n$ 实际上是单峰的，因此在这种情况下，高斯近似是合理的。这一思路催生了另一种近似后验分布的策略：令 $\pi_n$ 为一个以 $v_n$的全局极小值点为中心的高斯分布。尽管已有许多研究探讨拉普拉斯近似的准确性（包括近期的研究，见下文），但高斯变分推断尚无可比的结果。本文的主要目标是建立此类近似误差界。

主要贡献 在上述对 $\pi=\pi_n$ 进行高斯近似的启发式论证中，我们考虑了一列函数 $v_n$，当 $n\to\infty$ 时收敛到某个 $v_\infty$。然而，我们的结果并不要求采用这种设定。相反，我们只考虑具有如下形式的目标测度 $\pi \propto e^{-nv}$，其中 $v\in C^4(\mathbb{R}^d)$，并对三元组（$v$, $d$, $n$）施加要求。具体而言，我们假设 $a_3(v)d/\sqrt{n}$ 小于某个常数，且 $a_3(v)d/\sqrt{n}+a_4(v)d^2/n\le 1$；其中，$a_3(v)$ 和 $a_4(v)$ 分别界定了 $v$ 的三阶导数和四阶导数的大小（精确定义见假设 2）。我们的误差界与对（$v$, $d$, $n$）的这一条件相吻合，也就是说，只要 $a_3(v)d/\sqrt{n}+a_4(v)d^2/n$ 足够小，误差就会很小。除这一限制外，$v$ 可以同时依赖于 $d$ 和 $n$。

另外，假设 $v$ 有唯一全局极小点 $m^*$，$\|\nabla^3 v\|$ 和 $\|\nabla^4 v\|$ 离开 $m^*$ 后至多呈多项式增长，且 $v$ 离开 $m^*$ 后至少呈线性增长，我们为一大类函数 $g$ 给出如下误差界：

$$\left|\int g d\pi - \int g d\hat{\pi}\right| \lesssim_v \sqrt{1\vee \mathrm{Var}_{\hat{\pi}}(g)}\begin{cases} d/\sqrt{n}, \\ (d/\sqrt{n})^2, & g \text{ even about } \hat{m}, \\ (d/\sqrt{n})^3, & g \text{ linear}. \end{cases} \tag{1.4}$$

为清晰起见，我们将依赖于 $v$的因子 $a_3(v)$ $a_4(v)$ 隐藏在 $\lesssim_v$ 符号中，尽管原则上这些因子可能会影响上界的数量级。定理2.1给出了结果的精确表述；请注意，后验均值，即贝叶斯估计量，按这三种速率中最快的速率得到近似。(1.4) 中的第二个界来自对误差的首阶展开。具体而言，我们证明

$$\int g d\pi - \int g d\hat{\pi} = \int gQ d\hat{\pi} + \mathrm{Rem}_g, \qquad |\mathrm{Rem}_g| \lesssim_v \sqrt{1\vee \mathrm{Var}_{\hat{\pi}}(g)}\left(\frac{d}{\sqrt{n}}\right)^2, \tag{1.5}$$

其中， $Q$ 是一个关于$\hat{}m$为奇函数的三次多项式。因此，如果 $g$ 关于$\hat{}m$为偶函数，那么首阶项 $\int gQd\hat{}\pi$ 便消失。类似地，(1.4) 中的第三个界来自对误差作更高阶展开。我们指出，推导出首阶项不仅能深入揭示误差，还为修正高斯测度$\hat{}\pi$ 以获得更精确的近似 $(1+Q)\hat{}\pi$铺平了道路。不过，本文不深入探讨这一可能性，而将其留待未来研究。通过乘法修正改进高斯后验近似的做法已在 [19], 中得到研究，但针对的是另一种对后验分布的拉普拉斯近似（定义见 (1.8)）。

这一主要结果（1.4）给出了 $\int gd\pi - \int gd\hat{}\pi$ 的一个界，该界针对可观测量 $g$进行了量身定制。我们还可以应用这一结果，推导常见贝叶斯感兴趣量的误差界。例如，以下全变差界量化了使用 $\hat{}\pi$ 为 $\pi$构造近似可信集时产生的误差：

$$\mathrm{TV}(\pi, \hat{\pi}) \lesssim_v \frac{d}{\sqrt{n}}. \tag{1.6}$$

我们还对均值和协方差近似误差给出了界。后验均值在贝叶斯推断中尤为重要，因为它是针对平方误差风险而言未知参数的贝叶斯最优估计量。后验协方差是衡量后验分布中不确定性的另一种方式，可与可信集一并使用。为了处理 $\pi \propto e^{-nv}$ 的方差趋于零（其阶数为 $n^{-1}$）这一情况，我们对下文的近似误差作了适当缩放：

$$\sqrt{n}\|\hat{m} - \bar{m}\| \lesssim_v \left(\frac{d}{\sqrt{n}}\right)^3, \qquad n\|\hat{S} - \Sigma\| \lesssim_v \left(\frac{d}{\sqrt{n}}\right)^2. \tag{1.7}$$

综合界（1.4）、（1.6）和（1.7），我们便能全面了解在贝叶斯推断中计算所有相关感兴趣量时，将 $\hat{}\pi$ 用作 $\pi$ 的替代量所引入的误差。在第 3节中，我们将一般理论应用于高斯设计下逻辑回归所得的后验分布，并为其 VI 近似误差推导高概率界。

**拉普拉斯近似作为比较基准** 如上所述，拉普拉斯方法是另一种广泛应用于实践中的 $\pi$ 高斯近似，因为它计算简便。我们将其作为基准，以便更好地理解上述误差界。 $\pi \propto e^{-nv}$ 的拉普拉斯近似为

$$\pi \approx \mathcal{N}(m^*, (n\nabla^2 v(m^*))^{-1}), \tag{1.8}$$

其中 $m^*$ 是 $v$的全局极小值点。该近似只需将 $v$ 替换为其在 $m^*$处的二阶泰勒展开。近期的工作 [19], 中，第一作者证明，拉普拉斯近似的全变差误差和协方差误差，与变分推断的相应误差具有相同的阶数，就 $d$ 和 $n$ 依赖性而言均如此。然而，变分推断均值误差比拉普拉斯近似小两个数量级： $(d/\sqrt{n})^3$ 对应变分推断，而 $d/\sqrt{n}$ 对应拉普拉斯近似。关于这两种方法近似精度的更详细比较，请参见第2.3 节。

我们强调，本研究的目的并非对高斯变分推断与拉普拉斯近似进行优劣评判。事实上，方法的精度并非衡量其整体表现的唯一维度。全面比较这两种方法是一项微妙的工作，既涉及计算方面的考量，也因下游统计任务而异。全面的比较分析超出了本研究的范围，但这些近似结果在某些情境下无疑具有参考价值。

相反，我们工作的主要贡献是准确阐明高斯变分推断的近似精度；此前人们对此几乎一无所知，远远落后于文献中的算法发展

。我们之所以将变分推断与拉普拉斯近似进行比较，仅仅是因为拉普拉斯近似是另一种常用方法，可通过高斯拟合简化后验计算。

拉普拉斯近似与变分推断相关的另一种方式是，第一作者在 [18]中发展的用于界定拉普拉斯近似误差的证明技术，在本文中同样发挥着关键作用。具体而言，注意到，只要$d^2 \ll n$，界（1.4）、（1.6）和（1.7）都很小。这种对维数的依赖性改进了长期以来关于贝叶斯—冯·米塞斯定理和拉普拉斯近似精度的研究中所要求的条件 $d^3 \ll n$ 。 [13, 24, 23, 10, 15, 25]此后，近期的研究 [16] 和 [18] 发展了新的证明技术，将这种依赖性降低到 $d^2 \ll n$，不过，只有后者的证明技术足够灵活，能够推导出关于$d/\sqrt{n}$的首阶渐近。

本文采用了相同的技术。此外，我们的证明建立在对为何最小化 KL目标函数（1.1）能够带来良好近似性质这一问题的重要新见解之上。下面我们简要概述这些见解。

**一阶最优性条件与厄米特级数展开** 我们对高斯变分推断近似精度的分析，建立在一阶最优性条件与势函数 $v$ 的厄米特级数展开之间一种显著的相互作用之上。

此后，我们将 $\theta$ 替换为 $x$ 并令 $V = nv$， $\pi \propto e^{-V}$。本文的重点是关于最小化（1.1）的一阶最优性方程:

$$\nabla_{m,S}\mathrm{KL}(\mathcal{N}(m,S) \mid\mid \pi)\big|_{(m,S)=(\hat{m},\hat{S})} = 0.$$

这也等价于在 $p = \mathcal{N}(\hat{m}, \hat{S})$ 处将 KL$(p \mid\mid \pi)$ 的 Bures-Wasserstein梯度设为零，如 [20]所示。具体而言，我们得到，$(m, S) = (\hat{m}, \hat{S})$ 是以下方程的解:

$$\mathbb{E}\left[\nabla V(m + S^{1/2}Z)\right] = 0, \qquad \mathbb{E}\left[\nabla^2 V(m + S^{1/2}Z)\right] = S^{-1}, \tag{1.9}$$

其中 $Z \sim \mathcal{N}(0, I_d)$ 和 $S^{1/2}$ 是 $S$的正定对称平方根；有关此计算，请参见 [20]。从某种意义上说，$\hat{\pi} = \mathcal{N}(\hat{m}, \hat{S})$ 将 KL散度最小化至 $\pi$ 这一事实，并不能解释为何 $\hat{\pi}$ 能够很好地近似 $\pi$ 。真正的原因与不动点方程（1.9）解的性质有关。

要了解其中的原因，考虑函数 $V_0(x) = V(\hat{m} + \hat{S}^{1/2}x)$。如果 $\pi \propto e^{-V}$接近$\mathcal{N}(\hat{m}, \hat{S})$ 的概率密度，那么 $\rho \propto e^{-V_0}$ 应当接近 $\mathcal{N}(0, I_d)$ 的概率密度。换句话说，我们应有 $V_0(x) \approx$ const $. + \|x\|^2/2$。这由一阶最优性方程 (1.9) 保证。事实上，注意到 (1.9) 可用 $V_0$ 表示为

$$\mathbb{E}\left[\nabla V_0(Z)\right] = 0, \qquad \mathbb{E}\left[\nabla^2 V_0(Z)\right] = I_d. \tag{1.10}$$

正如我们在第 4.2 节所解释的，方程（1.10）将 $V_0$ 的厄米级数展开中的一阶和二阶系数分别设为 0 和 $I_d$。因此， $V_0(x) - \|x\|^2/2 =$ 常数$. + r_3(x)$ ，其中 $r_3$ 是只包含三阶及更高阶厄米多项式的厄米级数。高斯变分推断近似在积分线性和二次函数时具有较高精度，原因在于埃尔米特余项 $r_3$ 的阶数为 $r_3 \sim 1/\sqrt{n}$，并且 $r_3$ 关于高斯测度与线性和二次函数正交。关于这种基于埃尔米特级数的误差分析的概要，请参见第 4.2 节。

**相关工作** 关于变分推断的文献大致可分为统计和算法两类。统计方面的研究主要关注在大 $n$ （样本量）情形下，变分后验围绕真实参数的收缩。（我们用“变分后验”作为后验分布的变分近似的简称。）例如， [28]证明了变分后验的伯恩斯坦-冯·米塞斯定理的一个类比， [30] 研究了非参数设定下变分后验围绕真实参数的收缩率，另有 [2]研究了高维情形下温度化后验的变分近似的收缩率。

这些研究与我们的研究之间的一个关键区别在于，我们考察的是变分后验的统计量与后验分布本身的统计量有多吻合，而不是与某个极限（$n \to \infty$）分布的统计量有多吻合。据我们所知，只有另一项研究探讨了如何量化变分推断的后验近似精度。在 [14], 中，作者考虑在贝叶斯潜变量模型中使用平均场变分后验的均值$\hat{}m$ 作为最大似然估计的替代。他们证明了$\sqrt{n}\|\hat{}m - \text{MLE}\| \lesssim 1/n^{1/4}$。

关于高斯变分推断的算法方面，请读者参阅以下研究及其中的参考文献： [20, 12] 这些研究与我们的分析相辅相成，因为它们为若干求解优化问题（1.1）的算法提供了严格的收敛保证。另请参阅第2.3节的讨论。

接下来，我们介绍一种密切相关的近似方法，称为期望传播（EP）。在高斯EP中，目标是在所有高斯测度$\pi \mid\mid p$上最小化KL散度 KL($p$。（注意，相较于（ $\pi$ 和 $p$ ）中的写法，1.1）里的顺序已调换。）最小化该目标相当于使 $p$ 的一阶和二阶矩与 $\pi$的一阶和二阶矩相匹配。若不计算关于 $\pi$ 的积分，就无法直接最小化该目标。因此，EP通过迭代构造高斯分布与 $\pi$之间的混合分布，并使混合分布（其矩更容易计算）的矩与 $\pi$的矩相匹配。研究 [9] 和 [11] 考察了EP相对于 $n$的均值和协方差近似精度。有趣的是，作者证明，EP的均值和协方差精度与本文所示传统高斯变分推断的结果完全相同： $\|\hat{}m - \bar{m}\| \lesssim n^{-2}$和 $\|\hat{}S - \Sigma\| \lesssim n^{-2}$（参见（1.7））。这究竟是巧合，还是两种方法之间存在更深层联系的结果，是一个值得探索的有趣方向。

最后，我们注意到 [29]也使用厄米多项式来近似后验分布。该近似由埃奇沃思级数给出，实质上是对高斯测度的扰动。级数的系数取决于后验分布的累积量。因此，写出该级数需要计算累积量，而累积量来自对后验分布进行积分。因此，该展开本身不能直接用作近似后验分布积分的工具，而这正是我们工作的目标。

## 论文结构

本文其余部分安排如下。第 2 节陈述关于高斯变分推断近似精度的假设和主要结果。第 3节将一般理论应用于逻辑回归后验分布。第 4节概述主要定理的证明。最后，在第 5节，我们证明一阶最优性条件（1.9）的解存在且唯一。省略的证明见附录。

## 符号表示

记号 $a \lesssim b$ 表示存在常数 $C > 0$ ，使得 $a \leq Cb$。除非另有说明，省略的常数均为绝对常数。我们用 $\gamma$ 表示 $d$维标准高斯分布的概率密度$\mathcal{N}(0, I_d)$，并用 $\int f d\gamma$、 $\mathbb{E}[f(Z)$或 $\gamma(f)$表示 $f$的期望

在 $\gamma$下。我们用 $Z$ 表示 $\mathbb{R}^d$中的标准多元高斯随机变量 $Z \sim \gamma$ 。我们令 $\mathrm{Var}_{\hat{\pi}}(f) = \int (f - \hat{\pi}(f))^2 d\hat{\pi}$, 且 $\|f\|_2 = (\int f^2 d\gamma)^{1/2}$。

阶数为 $k$ 的张量 $T$ 是一个数组 $T = (T_{i_1 i_2 \ldots i_k})_{i_1,\ldots,i_k=1}^{d}$。对于两个 $k$ 阶张量 $T$ 和 $S$，我们令 $\langle T, S\rangle$ 表示逐项内积。若对 $i_1 \ldots i_k$ 的所有排列 $j_1 \ldots j_k$，都有 $T_{i_1 \ldots i_k} = T_{j_1 \ldots j_k}$，则称 $T$ 为对称的。

设 $H$ 为对称正定矩阵。对于向量 $x \in \mathbb{R}^d$，我们用 $\|x\|_H$ 表示 $\|x\|_H = \sqrt{x^T H x}$。对于阶数为 $k \geq 2$ 的对称张量 $T$，我们将 $T$ 的 $H$-加权算子范数定义为

$$\|T\|_H := \sup_{\|u\|_H = 1} \langle T, u^{\otimes k} \rangle. \tag{1.11}$$

根据 [31], 关于对称张量的定理 2.1，定义（1.11）与标准算子范数定义一致：

$$\sup_{\|u_1\|_H = \cdots = \|u_k\|_H = 1} \langle T, u_1 \otimes \cdots \otimes u_k \rangle = \|T\|_H = \sup_{\|u\|_H = 1} \langle T, u^{\otimes k} \rangle.$$

当 $H = I_d$ 时，范数 $\|T\|_{I_d}$ 就是通常的算子范数，此时我们省略下标。

## 2 主要结果陈述

鉴于不动点方程 (1.9) 的核心地位，本节开篇重新定义 $\hat{}(\hat{}m, S)$：将其定义为方程 (1.9) 的解，而非 KL散度目标函数 (1.1) 的极小化点。只有在 $V$ 不强凸的情况下，这些定义才会有所不同。事实上，若 $V$ 强凸，则 $\mathrm{KL}(\cdot \,||\, \pi)$ 在高斯分布子流形上是强测地凸的；例如，参见 [20]。因此，在这种情况下，$\mathrm{KL}(\cdot \,||\, \pi)$ 在该子流形上存在唯一最小值点 $\hat{\pi}$ ，对应于方程 (1.9) 的唯一解 $(\hat{}m, \hat{}S) \in \mathbb{R}^d \times \mathbf{S}_{++}^d$。然而一般而言，一阶方程 (1.9) 可能存在一些近似性质较差的伪解 $(m, S)$。为说明这一点，回顾 $v = V/n$，将方程写成如下形式：

$$\mathbb{E}\left[\nabla v(m + S^{1/2} Z)\right] = 0, \qquad S\,\mathbb{E}\left[\nabla^2 v(m + S^{1/2} Z)\right] = \frac{1}{n} I_d. \tag{2.1}$$

设 $x \neq m_*$ 是 $v$ 的临界点，即 $\nabla v(x) = 0$，并考虑数对 $(m, S) = (x, 0)$。对于这个 $(m, S)$，我们有

$$\mathbb{E}\left[\nabla v(m + S^{1/2} Z)\right] = \nabla v(x) = 0, \qquad S\,\mathbb{E}\left[\nabla^2 v(m + S^{1/2} Z)\right] = 0 \approx \frac{1}{n} I_d.$$

因此，$(x, 0)$ 是 (2.1) 的一个近似解。根据连续性，我们预期附近存在一个精确解。换言之，对于每个临界点 $x$， $v$ 都对应一个解 $(m, S) \approx (x, 0)$，它是 (2.1) 的解。因此，我们所关注的 $(m, S)$ 这个 (2.1) 的解，是靠近 $(m_*, 0)$ 的那个。下面的引理 2.1 将这一想法形式化；我们证明在集合

$$\mathcal{R}_V = \left\{ (m, S) \in \mathbb{R}^d \times \mathbf{S}_{++}^d \;:\; S \preceq 2H_V^{-1},\ \|H_V^{1/2} S^{1/2}\|^2 + \|H_V^{1/2}(m - m_*)\|^2 \leq 8 \right\}, \tag{2.2}$$

其中 $H_V = \nabla^2 V(m_*) = n \nabla^2 v(m_*)$。请注意，由于 $H_V$ 相对于 $n$的缩放，集合 $\mathcal{R}_V$是 $(m_*, 0)$ 的一个小邻域。我们将 $\mathcal{R}_V$中的这一唯一解 $(m, S)$ 称为方程 (1.9) 的“典范”解。我们预计，对应于该典范解的高斯分布是 (1.1) 的极小值，尽管我们尚未证明这一点。无论这一结论是否成立，

我们将重新定义 $(\hat{m}, \hat{S})$ 来表示典范解。事实上，对于近似 $\pi$ 而言，$\hat{\pi} = \mathcal{N}(\hat{m}, \hat{S})$ 究竟是否使 KL散度达到极小值，还是仅为一个局部极小值点，都无关紧要。

在本节余下部分，我们将陈述关于 $v$的假设、一个保证存在规范$m$的引理，以及关于高斯变分推断近似精度的主要结果。$S$，针对（1.9）。

## 2.1 关于势函数的假设

假设1（正则性与唯一严格极小值）。势函数 $v \in C^4$ 的唯一全局极小点为 $x = m^*$，且 $H_v = \nabla^2 v(m^*) \succ 0$。

假设2（ $\|\nabla^k v\|_{H_v}$、 $k = 3$ 的多项式增长，4。）。存在某些 $q$， $a_3$， $a_4 > 0$，使得

$$\|\nabla^k v(x)\|_{H_v} \le a_k \left(1 \vee \sqrt{n/d}\|x - m^*\|_{H_v}\right)^q \quad \forall x \in \mathbb{R}^d, \quad k = 3, 4, \tag{2.3}$$

$$a_3 \frac{d}{\sqrt{n}} \le C(q), \qquad a_3 \frac{d}{\sqrt{n}} + a_4 \frac{d^2}{n} \le 1 \tag{2.4}$$

对于某个常数 $C(q) < 1$，它仅取决于 $q$。

其中的$\sqrt{n/d}$ 因子（2.3）用于对 $\sqrt{d/n}$ 尺度进行“放大”，该尺度是测度 $\pi \propto e^{-nv}$ 集中的尺度。

假设 3（ $v$的增长下界）。 存在 $c_0 > 0$ 使得

$$c_0\sqrt{d/n}\|x - m^*\|_{H_v} \le v(x) - v(m^*) \qquad \forall\, x : \|x - m^*\|_{H_v} \ge \frac{1}{2}\sqrt{d/n}. \tag{2.5}$$

关于这些假设的讨论，请参见第 2.3节。

## 2.2 主要结果

现在我们可以陈述主要结果了。在本节中，我们使用记号 $\epsilon = d/\sqrt{n}$。首先，我们刻画高斯变分推断参数 $(\hat{m}, S)$:

**引理 2.1**。 设满足假设 *1* 和 *2*，并设 $H_V = \nabla^2 V(m_*) = n\nabla^2 v(m_*)$。那么存在唯一的 $(m, S) = (\hat{m}, S)$ 属于集合

$$\mathcal{R}_V = \left\{(m, S) \in \mathbb{R}^d \times \mathbf{S}^d_{++} : S \preceq 2H_V^{-1},\ \|H_V{}^{1/2} S^{1/2}\|^2 + \|H_V{}^{1/2}(m - m_*)\|^2 \le 8\right\}$$

它满足

$$\mathbb{E}\left[\nabla V(m + S^{1/2}Z)\right] = 0, \qquad \mathbb{E}\left[\nabla^2 V(m + S^{1/2}Z)\right] = S^{-1}.$$

此外， $S$ 满足

$$\frac{2}{3}H_V^{-1} \preceq \hat{S} \preceq 2H_V^{-1}. \tag{2.6}$$

关于此引理的证明，请参见第 5 节。现在我们陈述使用 $\hat{\pi} = \mathcal{N}(\hat{m}, \hat{S})$ 近似真实测度 $\pi$ 下期望时的误差界。主定理如下。

**定理 2.1**。 假设 $v$ 满足假设 *1*、*2*和 *3*，并令 $\hat{m}$， $S$ 如引理 *2.1*所述。设$g$ 为一个函数，对某个 $R_g > 0$ 满足以下不等式:

$$|g(x) - \int g d\hat{\pi}| \le \exp\left(c_0\sqrt{d}\|x - \hat{m}\|_{\hat{S}^{-1}}/4\right), \quad \forall\, x : \|x - \hat{m}\|_{\hat{S}^{-1}} \ge R_g\sqrt{d}. \tag{2.7}$$

*Then*

$$\left|\int g d\pi - \int g d\hat{\pi}\right| \lesssim \left(1 + \mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}}\right)\left(a_3\epsilon + a_4\epsilon^2\right), \tag{2.8}$$

此外，如果 $g$ 关于 $\hat{}m$ 为偶函数，则

$$\left|\int g d\pi - \int g d\hat{\pi}\right| \lesssim \left(1 + \mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}}\right)(a_3^2 + a_4)\epsilon^2. \tag{2.9}$$

最后，如果 $g$ 是线性的，那么

$$\left|\int g d\pi - \int g d\hat{\pi}\right| \lesssim \left(1 + \mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}}\right)\left(\left(a_3^3 + a_3 a_4\right)\epsilon^3 + a_4^2\epsilon^4\right). \tag{2.10}$$

在上述所有界中，被省略的常数是关于 $q$、 $c_0^{-1}$ 和 $R_g$ 的递增函数。

改进后的界（2.9）实际上适用于更广泛的一类函数 $g$；参见注 2.1。关于条件（2.7），请参见第 2.3 节。

**定理 2.2**（变分推断近似误差的首阶项）。 假设 $v$ 满足假设 *1*、*2* 和 *3*。定义函数

$$Q(x) = \left\langle \mathbb{E}_{X\sim\hat{\pi}}[\nabla^3 V(X)],\ \frac{1}{2}(x-\hat{m})\otimes\hat{S} - \frac{1}{6}(x-\hat{m})^{\otimes 3}\right\rangle. \tag{2.11}$$

如果 $g$ 满足 （2.7），那么

$$\left|\int g d\pi - \int g d\hat{\pi} - \int g Q d\hat{\pi}\right| \lesssim \left(1 + \mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}}\right)(a_3^2 + a_4)\epsilon^2. \tag{2.12}$$

被省略的常数是关于 $q$、 $c_0^{-1}$ 和 $R_g$ 的递增函数。

关于首阶项的更多见解，请参见第2.3节 $\int gQd\hat{}\pi$。

备注2.1。 注意，定理 2.1中的界（2.9 是定理 2.2的直接推论，因为 $\int gQd\hat{}\pi = 0$ 当 $g$ 关于 $\hat{}m$对称时。

接下来，我们界定全变差误差。令 $\mathcal{B}(\mathbb{R}^d)$ 为 $\mathbb{R}^d$ 的所有博雷尔集构成的集合，令 $\mathcal{S}_{\hat{m}}(\mathbb{R}^d)$ 为关于 $\hat{}m$ 对称的博雷尔集构成的集合。

**推论 2.1**（全变差误差）。 假设 $v$ 满足假设 *1*、*2*和*3*。那么

$$\begin{aligned}\sup_{A\in\mathcal{B}(\mathbb{R}^d)} |\pi(A) - \hat{\pi}(A)| &\lesssim a_3\epsilon + a_4\epsilon^2,\\ \sup_{A\in\mathcal{S}_{\hat{m}}(\mathbb{R}^d)} |\pi(A) - \hat{\pi}(A)| &\lesssim \left(a_3^2 + a_4\right)\epsilon^2.\end{aligned} \tag{2.13}$$

被省略的常数是 $q$ 和 $c_0^{-1}$ 的增函数。

**推论 2.2**（均值和协方差误差）。 令 $\bar{m} = \int x d\pi$（$x$）为 $\pi$的均值，且 $\Sigma = \int$（$x-\bar{m}$）（$x-\bar{m}$）$^T d\pi$（$x$）为 $\pi$的协方差。假设 $v$满足假设 *1*、*2*和*3*。那么

$$\begin{aligned}\|\hat{S}^{-1/2}(\bar{m} - \hat{m})\| &\lesssim \left(a_3^3 + a_3 a_4\right)\epsilon^3 + a_4^2\epsilon^4,\\ \|\hat{S}^{-1/2}(\Sigma - \hat{S})\hat{S}^{-1/2}\| &\lesssim \left(a_3^2 + a_4\right)\epsilon^2.\end{aligned} \tag{2.14}$$

被省略的常数是 $q$ 和 $c_0^{-1}$ 的递增函数。

注释 2.2. 回顾式（2.6）可知，$\hat{}S$ 等价于 $H^{-1}$，这里采用正定矩阵序。因此，给定下界 $n\alpha \preceq H$，我们可以将推论 2.2中的界转换为相应的不加权界：

$$\|\bar{m} - \hat{m}\| \lesssim \frac{1}{\sqrt{n\alpha}} \left[ \left(a_3^3 + a_3 a_4\right) \frac{d^3}{n\sqrt{n}} + a_4^2 \frac{d^4}{n^2} \right],$$
$$\|\Sigma - \hat{S}\| \lesssim \frac{1}{n\alpha} \left[ \left(a_3^2 + a_4\right) \frac{d^2}{n} \right]$$

## 2.3 讨论

关于 $v$ 和 $g$ 的假设。回顾一下，在贝叶斯推断中，$v$ 是归一化负对数后验，形式如 (1.3)，其中 $x_i$ 独立同分布。在典型情况下，我们预期当 $n$ 较大时，先验的影响可以忽略，且 $v(\theta) \approx -\mathbb{E} \log p_\theta(X)$，此时 $v$ 仅取决于 $d$。然而，我们的结果并不只适用于这种典型情形。函数 $v$ 可以同时依赖于 $n$ 和 $d$，因此假设中的常数—— $a_3$、$a_4$、$q$、$c_0$ ——也可能同时依赖于 $n$ 和 $d$。只要这些量以及 $d$ 和 $n$ 满足 (2.4) 的约束，我们的结果就成立。不过，对于逻辑回归示例，我们将看到 $a_3$、$a_4$、$q$、$c_0$ 都可以取为绝对常数。

当 $v$ 是凸函数时，假设 3 可由 (2.4) 自动推出。事实上，有如下引理。

**引理 2.2。** 设 $v$ 为凸函数，且假设 $a_3(1+(1/2)^q)\sqrt{d/n} \le 3$。则 $v$ 满足假设 3，其中 $c_0 = 1/8$。

证明见附录 A 。若 $v$ 不是凸函数，注意条件（2.5）仅对较大的 $x$才是非平凡要求。为此，固定最大的 $r$ 使得 $a_3 r(1+r^q)\sqrt{d/n} \le 3/2$。接着，直接进行泰勒展开可得

$$v(x) - v(m^*) \ge \frac{1}{4}\|x - m^*\|_{H_v}^2 \ge \frac{1}{8}\sqrt{\frac{d}{n}}\|x - m^*\|_{H_v}$$

对所有满足 $x$ $(1/2)\sqrt{d/n} \le \|x - m^*\|_{H_v} \le r\sqrt{d/n}$的 $\|x - m^*\|_{H_v} \ge r\sqrt{d/n}$。因此，只需对以下情况检验条件（2.5）：$\|x - m^*\|_{H_v} \ge r\sqrt{d/n}$。

接下来，我们讨论定理 2.7中关于函数 $g$ 的条件2.1。当 $g$ 具有以下形式时，最容易检验该条件： $g(x) = f(\hat{}S^{-1/2}(x - \hat{}m))$。在这种情况下，我们有以下充分条件。

**引理 2.3。** 令 $g(x) = f(\hat{}S^{-1/2}(x - \hat{}m))$ 并假设 $f$满足

$$|f(y) - f(0)| \le e^{C_f\sqrt{d}\|y\|^\alpha} \quad \forall y \in \mathbb{R}^d, \tag{2.15}$$

对于某个 $0 \le \alpha < 1$ 和 $C_f > 0$。那么，对于某个函数 $R_g = R_g(C_f, \alpha, c_0^{-1})$，该函数关于各变量均单调递增，条件 (2.7) 成立。

请参见附录G 中的证明。

**主导阶项：变分推断与拉普拉斯近似** 我们可以通过考察差异 $\int g d\pi - \int g d\hat{\pi}$ 的相应主导阶项（LOT），深入了解拉普拉斯近似与变分推断近似之间的差异，其中 $\hat{\pi}$ 可以是拉普拉斯高斯近似或变分推断高斯近似。为进行这一比较，我们先注意变分推断首阶项$\int gQd\hat{\pi}$的另一种表示。令

$$V_0(x) = V(\hat{m} + \hat{S}^{1/2}x), \qquad f(x) = g(\hat{m} + \hat{S}^{1/2}x), \tag{2.16}$$

并令 $\mathrm{H}_3$ 为三阶厄米多项式张量。关于维数为 $d \geq 1$ 的厄米多项式简介，请参见附录 E。令 $\mathrm{A}_3(f) = \mathbb{E}\left[f(Z)\mathrm{H}_3(Z)\right]$，以及 $\mathrm{A}_3(V_0) = \mathbb{E}\left[V_0(Z)\mathrm{H}_3(Z)\right] = \mathbb{E}\left[\nabla^3 V_0(Z)\right]$，，其中第二种表示利用高斯分部积分得到。（我们不将 $\mathrm{A}_3(f)$ 写成这种形式，因为 $f$ 不一定是 $C^3$。）可以证明（参见定理 4.1 和注 4.1），

$$\int gQd\hat{\pi} = -\frac{1}{3!}\left\langle \mathbf{A}_3(V_0),\ \mathbf{A}_3(f)\right\rangle = -\frac{1}{6}\left\langle \mathbb{E}\left[\nabla^3 V_0(Z)\right], \mathbb{E}\left[f(Z)\mathbf{H}_3(Z)\right]\right\rangle.$$

相反，[19] 文中证明，差值 $\int g d\pi - \int g d\hat{\pi}$ 的拉普拉斯近似 $\hat{\pi} = \mathcal{N}(m^*,\ H_V^{-1})$ 的首阶项形式如下：$-\frac{1}{6}\langle \nabla^3 V_0(0)\, \mathbb{E}\left[f(Z)Z^{\otimes 3}\right]\rangle$ 其中现在$V_0$，$f$ 定义为

$$V_0(x) = V(m^* + H_V^{-1/2}x), \qquad f(x) = g(m^* + H_V^{-1/2}x). \tag{2.17}$$

总而言之，非正式地说，我们有以下误差：

$$\int g d\pi - \int g d\hat{\pi} = -\frac{1}{6}\left\langle \mathbb{E}\left[\nabla^3 V_0(Z)\right], \mathbb{E}\left[f(Z)\mathbf{H}_3(Z)\right]\right\rangle + \mathcal{O}\left(\frac{d^2}{n}\right) \tag{VI}$$

$$\int g d\pi - \int g d\hat{\pi} = -\frac{1}{6}\left\langle \nabla^3 V_0(0),\ \mathbb{E}\left[f(Z)Z^{\otimes 3}\right]\right\rangle + \mathcal{O}\left(\frac{d^2}{n}\right) \tag{Laplace}$$

这里，$V_0$，$f$ 在第一行由式（2.16）给出，在第二行由式（2.17）给出。余项也取决于 $v$ 和 $g$，但为简洁起见，我们将 $\mathcal{O}(d^2/n)$ 记作。现在，我们比较各种函数 $f$对应的两种近似误差。最自然的做法是比较同一函数 $f$对应的两个 LOT，同时注意到，对于这两种方法，这个 $f$ 分别对应两个不同的函数$g$ 。

一般而言，这两个 LOT 的阶数均为 $d/\sqrt{n}$。特别是，如果 $f$ 是指示函数，那么两个 LOT 都不会消失。这解释了为什么对于 VI 和拉普拉斯，$\mathrm{TV}(\pi, \hat{\pi}) \lesssim d/\sqrt{n}$ 。当 $f$ 为偶函数时，VI 和拉普拉斯的 LOT 都会消失。特别是，这解释了为什么拉普拉斯和 VI 的协方差误差均有 $d^2/n$ 的上界。

最后，拉普拉斯与变分推断的关键区别在于，对于与所有三阶厄米多项式正交但并非偶函数的函数 $f$，变分推断的 LOT 为零，而拉普拉斯的 LOT 不为零。此类函数中可能最重要的例子是 $f(x) = x$。这解释了变分推断具有更高的均值近似精度。（误差为 $\mathcal{O}((d/\sqrt{n})^3)$ 而非 $\mathcal{O}((d/\sqrt{n})^2)$，是因为下一阶项发生了抵消。）此类$f$ 的另一个例子是五阶埃尔米特多项式。

我们指出，LOT 不仅可以用于证明上界，也可以证明下界。在 [19], 中，我们证明，对于源自逻辑回归的后验分布，拉普拉斯均值近似的误差具有下界 $d/\sqrt{n}$，从而确凿地证明，通常情况下变分推断的均值近似误差更小。上述拉普拉斯误差界的精确表述见 [19] 第 2.2 节。

**一阶最优性条件的解**（1.9）**与方程**（1.1）的解。回顾一下，在我们的结果中，（$\hat{m}$, $S$）是方程（1.9）的一个解，其中 $\hat{m}$ 位于 $m^*$的邻域内，即后验分布的众数 $\pi$。如果后验分布是多峰的，（$\hat{m}$, $S$）可能不是 KL目标函数（1.1）的全局极小值点。例如，后验分布景观可能有一个单一的、“正确的”主导众数，同时还有一个密度值更高的“错误”次级众数，但它仍处于低概率区域。这种低概率众数可能由数据中的离群值引起。在这种情况下，方程（1.9）的解将位于错误众数附近，而方程（1.1）的解则位于正确众数附近。因此，在这种情形下，研究 KL目标函数的全局极小值点，而不是像本文这样研究最高众数附近的临界点，将会很有意义。此外，这种情形下变分推断可能优于拉普拉斯近似。直观而言，我们预期变分推断对低概率扰动具有稳定性，而拉普拉斯近似则不然。

我们将变分推断在此类多峰情形下的表现留待未来研究。请注意，只要负对数后验在远离最小值处呈线性增长，我们的假设就允许出现多峰。我们的结果表明，在这种情况下，靠近众数的 (1.9) 的解是后验分布的良好近似，无论该解是否与 KL目标函数的全局极小值点重合。

**计算** 尽管高斯变分推断的计算方面并非本文重点，为求完整起见，我们总结已知的高斯变分推断算法的计算保证。典型算法会根据某种优化过程，迭代更新均值和协方差矩阵。例如， [20] 研究了一个常微分方程的离散化，该方程演化 $(m_t, \Sigma_t)$，对应于在正态分布子流形上约束的 $\mathrm{KL}(\cdot \,||\, \pi)$ 的瓦瑟斯坦梯度流。作者证明，如果 $V$ 是强凸的，那么梯度流（及其离散化）会以指数速度在瓦瑟斯坦距离下收敛到 $(\pi$ 1.1) 的 $\hat{\pi}$ ，此时它是高斯分布上 $\mathrm{KL}(\cdot \,||\, \pi)$ 的唯一最小值点。在非凸情形下， [12] 为一种涉及梯度步和近端步的“前向-后向”算法给出了驻点保证。 [20] 的更新步骤需要求 $d \times d$ 矩阵的逆，而 [12] 的更新步骤需要计算 $d \times d$ 矩阵的平方根。两种算法的更新步骤还都需要计算高斯期望，或者抽取一个高斯样本；后一种情况下，算法是随机的。

一种计算成本更低、复杂度随维数线性增长的算法是平均场高斯变分推断，其中目标函数（1.1）仅在具有对角协方差矩阵的高斯分布上进行最小化。参见 [5] ， 了解从对角协方差矩阵到完整协方差矩阵等一系列高斯变分推断变体的计算复杂度讨论。然而，目前尚不清楚高斯平均场变分推断的准确性。

## 3 使用高斯设计的逻辑回归

在逻辑回归中，我们生成 $n$ 个特征向量 $X_i \in \mathbb{R}^d$ ，并观测其对应的标签$Y_i \in \{0, 1\}$。给定特征时标签的分布建模为

$$p(Y_i \mid X_i, \theta) = \exp\left(Y_i X_i^T \theta - \psi(X_i^T \theta)\right), \tag{3.1}$$

其中 $\psi(t) = \log(1 + e^t)$ 和 $\theta$ 是未知的系数向量。注意，(3.1) 只是 $\mathrm{Bern}(\sigma(X_i^T \theta))$ 的概率分布，其中 $\sigma(t) = \psi'(t) = (1 + e^{-t})^{-1}$ 是逻辑函数。我们假设模型设定正确，因此存在一个真实的 $\theta_0 \in \mathbb{R}^d$ 使得

$$Y_i \sim \mathrm{Bern}(\sigma(X_i^T \theta_0)).$$

我们假设特征是从标准高斯分布中生成的 :

$$X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_d), \quad i = 1, \ldots, n.$$

我们考虑系数向量 $\theta$上的高斯先验 $\theta \sim \mathcal{N}(0, \Sigma)$。给定标签 $Y_i$ 时， $\theta$ 的后验分布为 $\pi(\theta) \propto e^{-nv(\theta)}$, 其中

$$\begin{aligned} v(\theta) &= \ell(\theta) + \frac{1}{2n}\theta^T \Sigma^{-1} \theta, \\ \ell(\theta) &:= -\frac{1}{n}\sum_{i=1}^{n} \log p(Y_i \mid X_i, \theta) = -\theta^T \left[\frac{1}{n}\sum_{i=1}^{n} Y_i X_i\right] + \frac{1}{n}\sum_{i=1}^{n} \psi(X_i^T \theta) \end{aligned} \tag{3.2}$$

这里， $\ell$ 是负归一化对数似然， $\theta^T \Sigma^{-1} \theta / 2n$ 是对数先验的贡献。形式上， $\Sigma^{-1} = 0$ 对应于平坦先验的情形。注意， $X_i$ 的分布不会影响后验分布。

## 3.1 检验假设

为验证假设1，我们需要证明存在唯一的全局极小值点 $\theta^*$ ，使其满足 $v$。为此，我们首先证明存在唯一的全局极小值点 $\theta_\ell^*$ ，使其满足 $\ell$。容易验证，只要特征 $X_i$ 张成 $\mathbb{R}^d$， $\ell$ 就是严格凸的，而这一条件以概率 1 成立。因此，如果$\nabla \ell(\theta_\ell^*) = 0$对某个 $\theta_\ell^*$成立，那么该点必定是 $\ell$的唯一全局极小值点。为证明 $\ell$存在临界点，我们将使用以下引理，其中 $f = \ell$ 以及 $\theta = \theta_0$。

**引理 3.1** (推论*F.6*，其中 $A = I_d$)。 设 $f \in C^2(\mathbb{R}^d)$。假设存在一点 $\theta$ ，使得对于某个 $\lambda$， $s > 0$，以下条件成立:

$$\begin{aligned} &\nabla^2 f(\theta) \succeq \lambda I_d, \\ &\|\nabla f(\theta)\| \leq 2\lambda s, \\ &\|\nabla^2 f(\theta') - \nabla^2 f(\theta)\| \leq \frac{\lambda}{4} \quad \forall \|\theta' - \theta\| \leq s. \end{aligned} \tag{3.3}$$

那么存在 $\theta_f^*$使得 $\nabla f(\theta_f^*) = 0$且 $\|\theta_f^* - \theta\| \leq s$。

当然， $\ell$ 是一个随机函数，因此对于 3.3)，条件 ( $f = \ell$3.3) 不可能确定性地成立。相反，我们给出条件 (3.3) 成立概率的下界。

**引理3.2**。 假设 $d/n \leq 1/10$ 且 $n \geq 13$。则事件

$$E_1 = \{\|\nabla \ell(\theta_0)\| \leq 8(d/\sqrt{n})^{1/2}\}$$

在真实分布下的概率至少为 $1 - e^{-n/4} - n^{-d/4}$: $X_i \sim \mathcal{N}(0, I_d)$ $Y_i \mid X_i \sim$伯努利分布($(\sigma(X_i^T \theta_0))$, 且 $(X_i, Y_i)$、 $i = 1$、...、 $n$ 独立同分布。

该引理的证明与 [17], 中的引理 4.2 的证明类似，但为完整起见，我们将证明收录于附录B。

**引理 3.3** （改编自引理 7，第 3, [26]章）。设 $d < n$。那么偶 $t$

$$E_2 = \left\{ \inf_{\|\theta\| \leq c} \lambda_{\min}\left(\nabla^2 \ell(\theta)\right) \geq \tau(c) \quad \forall c \geq 0 \right\} \tag{3.4}$$

其概率至少为 $1 - 4e^{-Cn}$ ，其中 $C$ 是一个绝对常数。这里， $\tau$: $[0, \infty)$ $\to$ $(0, \infty)$ 是一个通用的非增函数。

我们注意到， $\ell$ 引理 7 中实际考虑的函数 [26]与我们的 $\ell$ 之间仅相差一个 $\sqrt{n}$缩放因子。我们已将结果的表述转换为关于我们的函数 $\ell$。

**引理 3.4**。 假设 $d \leq n$。那么存在绝对常数 $C_3$、 $C_4$、 $C$ ， 使得事件

$$E_3 = \left\{ \sup_{\theta \in \mathbb{R}^d} \|\nabla^k \ell(\theta)\| \leq C_k, \quad k = 3, 4 \right\} \tag{3.5}$$

发生的概率至少为 $1 - \exp(-C(nd)^{1/9})$。

参见附录 B 中的证明，该证明几乎可以直接通过应用 [1] 中的定理 4.2 得到。现在我们可以应用引理 3.1 来证明 $\ell$ 具有唯一的全局极小点 $\theta^*_\ell$。由此再应用一次引理 3.1还可推出， $v$ 在附近具有唯一的全局极小点。引理 3.2、3.3、3.4 还为我们检验其他假设提供了必要的工具。

**推论 3.1**。 假设 $\|\theta_0\| \leq C$、 $\|\Sigma^{-1}\| \leq C$ ， 且 $d/\sqrt{n}$ 和 $n^{-1}$ 小于足够小的绝对常数。那么，以至少 $1 - \exp(-C(nd)^{1/9}) - n^{-d/4} - 5e^{-Cn}$ 的概率，以下陈述全部成立：

1. 函数 $\ell$ $v$ $\theta^*_\ell$ $\theta^*$ $\|\theta^*_\ell - \theta_0\| \lesssim$和分别具有唯一全局极小点 和 。其中 和 。特别地，假设 $\sqrt{d/n}$ $\|\theta^* - \theta^*_\ell\| \lesssim n^{-1}$ 成立。
2. 我们有 $\|\theta^*\| \leq C$ 和 $\lambda_{\min}(H_v) = \lambda_{\min}(\nabla^2 v(\theta^*)) \geq \tau(C)$。
3. 我们有 上确界$_{\theta \in \mathbb{R}^d} \|\nabla^k v(\theta)\|_{H_v} \leq C$ 对于 $k = 3, 4$。特别地，假设 (2.3) 2 的条件在 $q = 0$ 和 $a_3 = a_4 = C$下成立。此外，假设 (2.4) 2 的条件也成立，前提是 $d/\sqrt{n}$ 小于某个固定的确定性绝对常数。
4. 假设 3 在 $c_0 = 1/8$下成立。

参见附录 B中的证明。

### 3.2 第 2 节结果在逻辑回归中的应用

我首 先计算首阶项中的函数 $Q$ 。现在可以验证

$$\nabla^3 v(\theta) = \frac{1}{n} \sum_{i=1}^n \psi'''(\theta^T X_i) X_i^{\otimes 3},$$

并回忆起 $V = nv$。因此，

$$\mathbb{E}_{\theta \sim \hat{\pi}}[\nabla^3 V(\theta)] = \sum_{i=1}^n b_i(\theta) X_i^{\otimes 3},$$

其中

$$b_i(\theta) = \mathbb{E}_{\theta \sim \hat{\pi}} \left[ \psi'''(\theta^T X_i) \right]. \tag{3.6}$$

Usi根据定义（2.11）的 $Q$ 以及上述关于 $\mathbb{E}_{\theta \sim \hat{\pi}}[\nabla^3 V$（$\theta$）$]$, 我们得 et

$$Q(\theta) = \sum_{i=1}^n b_i(\theta) \left[ \frac{1}{2} X_i^T (\theta - \hat{m}) X_i^T \hat{S} X_i - \frac{1}{6} (X_i^T (\theta - \hat{m}))^3 \right] \tag{3.7}$$

现在，我们总结第2 节中应用于逻辑回归的结果。我们假设采用本节开头的设定：设定正确的模型、独立同分布高斯设计、有界的真实参数 $\|\theta_0\| \leq C$，以及高斯先验或平坦先验 $\mathcal{N}(0, \Sigma)$，其中 $\|\Sigma^{-1}\| \leq C$。在此设定下，推论3.1表明，假设1、2、3 均得到满足，并且 $q = 0$， $a_3 = a_4 = C$。

**定理 3.1**。假设 $d/\sqrt{n}$ 和 $n^{-1}$ 小于某些绝对常数。在概率至少为 $1-\exp(-C(nd)^{1/9})-5e^{-Cn}-n^{-d/4}$ 的事件上，以下结果成立。

$\hat{}$在区域 $\hat{}m, S$ 存在唯一解满足 (1.9) $\mathcal{R}_V$。此外，如果 $g$，并且 $c_0 = 1/8$满足，那么 (2.7)

$$\begin{aligned}\left|\int g d\pi - \int g d\hat{\pi}\right| &\lesssim_{R_g} (1+\mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}})\frac{d}{\sqrt{n}},\\ \left|\int g d\pi - \int g d\hat{\pi} - \int gQ d\hat{\pi}\right| &\lesssim_{R_g} (1+\mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}})\left(\frac{d}{\sqrt{n}}\right)^2,\end{aligned} \tag{3.8}$$

其中 $Q$ 定义于 (3.7) 和 (3.6)。如果 $g$ 关于 $\hat{}m$是偶对称的，那么

$$\left|\int g d\pi - \int g d\hat{\pi}\right| \lesssim_{R_g} (1+\mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}})\left(\frac{d}{\sqrt{n}}\right)^2. \tag{3.9}$$

如果 $g$ 是线性的，那么

$$\left|\int g d\pi - \int g d\hat{\pi}\right| \lesssim_{R_g} (1+\mathrm{Var}_{\hat{\pi}}(g)^{\frac{1}{2}})\left(\frac{d}{\sqrt{n}}\right)^3. \tag{3.10}$$

接下来，我们有

$$\begin{aligned}\sup_{A\in\mathcal{B}(\mathbb{R}^d)} |\pi(A)-\hat{\pi}(A)| &\lesssim \frac{d}{\sqrt{n}},\\ \sup_{A\in\mathcal{B}_{s,\hat{m}}(\mathbb{R}^d)} |\pi(A)-\hat{\pi}(A)| &\lesssim \left(\frac{d}{\sqrt{n}}\right)^2.\end{aligned} \tag{3.11}$$

最后，令 $\bar{\theta} = \int \theta d\pi(\theta)$ 为 $\pi$ 的均值，$\Sigma = \int(\theta-\bar{\theta})(\theta-\bar{\theta})^T d\pi(\theta)$ 为 $\pi$的协方差。则

$$\begin{aligned}\sqrt{n}\|\bar{m}-\hat{m}\| &\lesssim \left(\frac{d}{\sqrt{n}}\right)^3,\\ n\|\Sigma-\hat{S}\| &\lesssim \left(\frac{d}{\sqrt{n}}\right)^2\end{aligned} \tag{3.12}$$

在 (3.11) 和 (3.12)中，所有被省略的常数均为绝对常数。

该证明可直接由推论 3.1 以及第 2 节中的结果得出 .

### 3.3 数值模拟

我们通过数值模拟验证部分理论结果，具体而言是界 (3.12)。数值结果如图 1 所示。我们取 $d=2$ 和 $n=100,\ 200,\ldots,1000$，并对 $\theta$采用平坦先验。对于每个 $n$，我们从 $\mathcal{N}(0, I_d)$ 中抽取十组协变量 $x_i, i=1,\ldots,n$，从而得到十个后验分布 $\pi_n(\cdot\mid x_{1:n})$。随后，我们针对每个 $n$ 以及给定 $n$时的十个后验分布，计算拉普拉斯和变分推断的均值和协方差近似误差。图 1 中的实线表示每个 $n$下十个分布的平均近似误差。阴影区域表示十个近似误差中间六个的分布范围。模拟细节请参见附录 B。

图1 左侧面板描绘了均值误差，其中最佳拟合直线的斜率分别为 $-1.04$和 $-2.02$，对应拉普拉斯方法和高斯变分推断。对于右侧面板中的协方差误差，最佳拟合直线的斜率分别为 $-2.09$ 和 $-2.12$，对应拉普拉斯方法和高斯变分推断。这证实，我们关于变分推断的界以及拉普拉斯界，在其 $n$ 依赖关系上都是紧的。

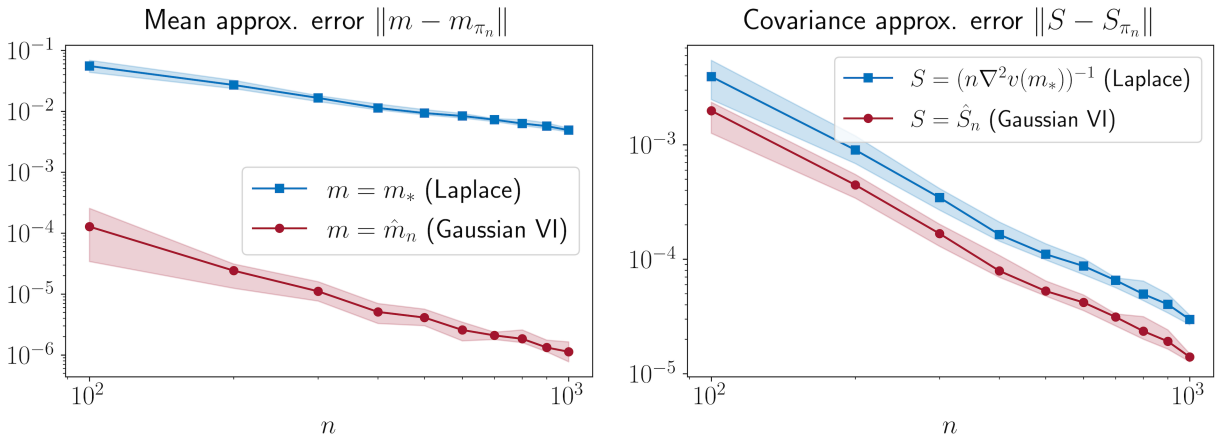


图 1：与拉普拉斯近似相比，高斯变分推断能够得到更准确的均值估计，而两者的协方差估计处于相同阶数。此处，$\pi_n$ 是给定$n$ 个观测值时，维数为 $d = 2$ 的逻辑回归似然。对于左图，拉普拉斯近似和高斯变分推断的最佳拟合线斜率分别为 $-1.04$ 和 $-2.02$ 。对于协方差：拉普拉斯近似和变分推断的最佳拟合线斜率分别为 -2.09 和 -2.12。

# 4 定理 2.1 和 2.2 的证明

本节将介绍定理2.1 和2.2证明中的主要步骤。我们将引理2.1作为已知结论；其证明概要见第5节。

## 4.1 化归为与标准高斯分布的比较

回顾一下，$\hat{}\pi = \mathcal{N}(\hat{}m, \hat{}S)$，其中 $\hat{}m$，$\hat{}S$ 是引理 2.1 所给出的方程（1.9）的解。我们关注的是差值 $\int g d\pi - \int g d\hat{}\pi$，但我们总可以通过双射 $T$进行变量变换，将该差值写成如下形式：

$$\int g d\pi - \int g d\hat{\pi} = \int f dT_{\#}\pi - \int f dT_{\#}\hat{\pi}, \qquad f = g \circ T^{-1}. \tag{4.1}$$

为方便起见，我们选择 $T$，使得 $T_{\#}\hat{}\pi = \gamma$ 为标准高斯分布。具体而言，令

$$T(x) = \hat{S}^{-1/2}(x - \hat{m}), \tag{4.2}$$

并定义 $\rho = T_{\#}\pi \propto e^{-V_0}$，其中

$$V_0(x) := V(\hat{m} + \hat{S}^{1/2}x) = V(T^{-1}(x)). \tag{4.3}$$

接下来，我们重新表述定理 2.1 和 2.2。为此，定义 $\mathbf{A}_3 = \mathbb{E}\left[\nabla^3 V_0(Z)\right]$ 和

$$p_3(x) = \frac{1}{6}\left\langle \mathbf{A}_3, \mathbf{H}_3(x)\right\rangle = \left\langle \mathbf{A}_3, \frac{1}{6}x^{\otimes 3} - \frac{1}{2}x \otimes I_d \right\rangle.$$

这里，$\mathbf{H}_3(x)$）是三阶多元厄米多项式的 $d \times d \times d$ 张量；详见第 4.2节和附录 E以了解更多详情。

**定理 4.1**。 设 $f = g \circ T^{-1}$，其中 $T$ 如 (4.2)所示，并设 $g$ 满足 (2.7)。此外，设 $\rho \propto e^{-V_0}$如 (4.3)所示，且 $\Delta_f = \int f d\rho - \int f d\gamma$. 。那么在定理 *2.1*的条件下，有

$$|\Delta_f| \lesssim (1 + \|f_0\|_2) \left(a_3\epsilon + a_4\epsilon^2\right), \tag{4.4}$$

$$\left|\Delta_f - \int f(-p_3)d\gamma\right| \lesssim (1 + \|f_0\|_2)(a_3^2 + a_4)\epsilon^2. \tag{4.5}$$

若 $f$ 与所有三阶厄米多项式正交（特别是当 $f$ 为偶函数时），则

$$|\Delta_f| \lesssim (1 + \|f_0\|_2)(a_3^2 + a_4)\epsilon^2. \tag{4.6}$$

若 $f$ 是线性的，则

$$|\Delta_f| \lesssim (1 + \|f_0\|_2) \left(\left(a_3^3 + a_3a_4\right)\epsilon^3 + a_4^2\epsilon^4\right). \tag{4.7}$$

在上述所有界中，被省略的常数是 $q$、$c_0^{-1}$、$R_g$ 的增函数。

界 (4.4)、(4.6) 和 (4.7) 分别对应于定理 2.1 中的 (2.8)、(2.9) 和 (2.10)。这可直接由 (4.1) 以及 $\mathrm{Var}_{\hat{\pi}}(g) = \mathrm{Var}_\gamma(f) = \|f_0\|_2^2$ 这一事实推出。界 (4.5) 对应于定理 2.2 中的 (2.12)。这是因为 $\int f(-p_3)d\gamma = \int gQd\hat{}\pi$，如附录 G 中的简短计算所示。

## 4.2 证明概要

现在我们概述等价定理4.1的证明。该证明基于几个关键观察。首先，注意到最优性方程 (1.9) 和 4.3 的定义 $V_0$ 蕴含

$$\mathbb{E}\left[\nabla V_0(Z)\right] = 0, \qquad \mathbb{E}\left[\nabla^2 V_0(Z)\right] = I_d. \tag{4.8}$$

该证明通过另外六项观察利用了（4.8）。

1）最优性条件（4.8）意味着 $V_0$ 的厄米特级数展开具有以下形式$V_0$（$x$）$=$常数. $+ \frac{1}{2}\|x\|^2 + r_3$（$x$），其中 $\quad -$

$$r_3(x) = \sum_{k\geq 3} \frac{1}{k!}\langle \mathbf{A}_k, \mathbf{H}_k(x)\rangle. \tag{4.9}$$

特别地， $d\rho \propto e^{-V_0} \propto e^{-r_3}d\gamma$。

2）由于 $r_3$ 只包含三阶及更高阶的厄米多项式，因此，相对于高斯测度， $r_3$ 与所有二次或更低次多项式正交。换言之，对于所有二次 $f$， $\int f r_3 d\gamma$ 。作为一个特例（取 $f = 1$），我们看到 $\int r_3 d\gamma = 0$。

3）令 $p_3 = \frac{1}{3!}\langle \mathbf{A}_3, \mathbf{H}_3\rangle$ 为对 $V_0$的三阶厄米多项式贡献项，并写作$r_3 = p_3 + r_4$，其中 $r_4$ 是四阶 Hermite 级数余项。通过变换 $T$ 进行变量替换，可确保 $r_3 \sim p_3 \sim n^{-1/2}$ 和 $r_4 \sim n^{-1}$。

4) 令 $f_0 = f - \gamma(f)$，使得 $\Delta_f = \int f_0 d\rho$。利用 $d\rho \propto e^{-r_3}d\gamma$、$r_3 \sim n^{-1/2}$ 以及 $\int f_0 d\gamma = \int r_3 d\gamma = 0$，可得

$$\begin{aligned}\Delta_f = \int f_0 d\rho = \frac{\int f_0 e^{-r_3}d\gamma}{\int e^{-r_3}d\gamma} &= \frac{\int f_0\left(1 - r_3 + r_3^2/2 + \mathcal{O}(n^{-3/2})\right)d\gamma}{\int \left(1 - r_3 + \mathcal{O}(n^{-1})\right)d\gamma} \\ &= \frac{\int f_0\left(-r_3 + r_3^2/2 + \mathcal{O}(n^{-3/2})\right)d\gamma}{1 + \mathcal{O}(n^{-1})}\end{aligned} \tag{4.10}$$

5) 对于一般的 $f$， $\Delta_f$ 的阶数等于第二行分子中第一个剩余项的阶数： $\Delta_f \sim \int f_0 r_3 d\gamma \sim n^{-1/2}$。但如果 $f$ 是线性或二次的，那么 $\int f_0 r_3 d\gamma = 0$，因为 $r_3$ 与二次多项式正交。因此， $\Delta_f$ 的阶数至多为 $\int f_0 r_3^2 d\gamma \sim n^{-1}$。

6）通过写出 $r_3 = p_3 + r_4 = \mathcal{O}(n^{-1/2}) + \mathcal{O}(n^{-1})$，从而得到 $r_3^2 = p_3^2 + \mathcal{O}(n^{-3/2})$，我们可以获得更精细的估计。具体而言，有

$$\Delta_f \sim \int f_0 \left(-r_3 + \frac{1}{2} r_3^2\right) d\gamma = \underbrace{\int f_0(-p_3) d\gamma}_{\mathcal{O}(n^{-1/2})} + \underbrace{\int f_0 \left(-r_4 + \frac{1}{2} p_3^2\right) d\gamma}_{\mathcal{O}(n^{-1})} + \mathcal{O}(n^{-3/2}).$$

由此分解可知，若 $f$ 与三阶厄米多项式正交，则 $\int f_0 p_3 d\gamma = 0$，因此在这种情况下 $\Delta_f \sim \mathcal{O}(n^{-1})$。若 $f(x) = a^T x + b$，则 $0 = \int f_0 p_3 d\gamma = \int f_0 r_4 d\gamma$自动成立；但我们还有 $\int f_0 p_3^2 d\gamma = 0$，因为 $f_0(x) = a^T x$ 为奇函数，而 $p_3^2$ 为偶函数。因此， $\mathcal{O}(n^{-1/2})$ 项和 $\mathcal{O}(n^{-1})$ 项均消失，从而得到 $\Delta_f \sim n^{-3/2}$。

这就是证明的核心。既然我们已经概述了上述要点，下面进一步讨论其中的几个细节。为解释 1)，回顾一下， $V_0$的厄米特级数展开定义为

$$V_0(x) = \sum_{k=0}^{\infty} \frac{1}{k!} \langle \mathbf{A}_k(V_0), \mathbf{H}_k(x) \rangle, \qquad \mathbf{A}_k(V_0) := \mathbb{E}\left[V_0(Z)\mathbf{H}_k(Z)\right] \tag{4.11}$$

这里， $\mathbf{A}_k(V_0)$ 和 $\mathbf{H}_k(x)$ 是 $(\mathbb{R}^d)^{\otimes k}$ 中的张量。具体而言， $\mathbf{H}_k(x)$ 是所有 $k$阶厄米多项式组成的张量，枚举为 $H_k^{(\alpha)}$，$\alpha \in [d]^k$，其中有些项会重复。对于 $k = 0,1,2$，厄米特张量定义为

$$\mathbf{H}_0(x) = 1, \quad \mathbf{H}_1(x) = x, \quad \mathbf{H}_2(x) = xx^T - I_d. \tag{4.12}$$

详见附录E，其中进一步介绍了厄米级数和多项式。不同的厄米多项式关于高斯权重彼此正交。利用表示式

$$\mathbf{H}_k(x) e^{-\|x\|^2/2} = (-1)^k \nabla^k e^{-\|x\|^2/2}, \tag{4.13}$$

由此可得一个有用的恒等式，即适用于 $k$ 次可微函数 $f$ 的“高斯分部积分”恒等式。具体而言，

$$\mathbb{E}\left[f(Z)\mathbf{H}_k(Z)\right] = \mathbb{E}\left[\nabla^k f(Z)\right]. \tag{4.14}$$

这是 Stein 恒等式 $\mathbb{E}\left[Z_i f(Z)\right] = \mathbb{E}\left[\partial_i f(Z)\right]$ 的推广。由于 $V_0$ 四次可微，我们可以利用高斯分部积分将 $\mathbf{A}_1$,$\mathbf{A}_2$ 写成

$$\begin{aligned} \mathbf{A}_1(V_0) &:= \mathbb{E}\left[V_0(Z)\mathbf{H}_1(Z)\right] = \mathbb{E}\left[\nabla V_0(Z)\right] = 0, \\ \mathbf{A}_2(V_0) &:= \mathbb{E}\left[V_0(Z)\mathbf{H}_2(Z)\right] = \mathbb{E}\left[\nabla^2 V_0(Z)\right] = I_d, \end{aligned} \tag{4.15}$$

其中每行最后一个等式均由最优性条件（4.8）得出。因此， $V_0$ 的厄米级数展开形式为

$$\begin{aligned} V_0(x) &= \langle A_0(V_0), H_0 \rangle + \langle 0, \mathbf{H}_1(x) \rangle + \frac{1}{2} \langle I_d, \mathbf{H}_2(x) \rangle + r_3(x) \\ &= \text{const.} + \frac{\|x\|^2}{2} + r_3(x), \end{aligned} \tag{4.16}$$

其中 $r_3$ 是三阶余项（4.9）。这里，我们由（4.12）回顾可知 $\mathbf{H}_2(x) = xx^T - I_d$。这说明了第 1 点，而第 2 点则是不同厄米多项式正交性的直接应用。

为了对第 3) 点给出启发式证明，回顾一下 $V_0(x) = V(\hat{m} + \hat{S}^{1/2}x)$， $V = nv$，以及 $\hat{S}^{1/2} \sim n^{-1/2}$。因此，根据链式法则， $\nabla^3 V_0 \sim n^{-1/2}$，并且 $\nabla^4 V_0 \sim n^{-1}$。于是

$$p_3 = \langle \mathbf{A}_3, \mathbf{H}_3 \rangle = \langle \mathbb{E}\left[\nabla^3 V_0(Z)\right], \mathbf{H}_3 \rangle \sim n^{-1/2}$$

以及 f 对于 $k \geq 4$，我们可以应用“部分”高斯分部积分，将 $\mathbf{A}_k$ a 表示为 s

$$\mathbf{A}_k = \mathbb{E}\left[\mathbf{H}_k(Z) V_0(Z)\right] = \mathbb{E}\left[\mathbf{H}_{k-4}(Z) \otimes \nabla^4 V_0(Z)\right] \sim n^{-1}.$$

因此 $r_4 \sim n^{-1}$。

备注 4.1。 回顾定理 4.1的陈述，我们将 $\mathbf{A}_3$ 定义为 $\mathbf{A}_3 = \mathbb{E}\left[\nabla^3 V_0(Z)\right]$。现在我们看到， $\mathbf{A}_3 = \mathbf{A}_3(V_0)$ 恰好是 $V_0$的厄米特级数展开中的三阶系数张量。此外，现在我们可以解释第 2.3节中给出的首阶项的表达形式。令 $\mathbf{A}_3(f)$ 为 $f$的厄米特级数展开中的三阶系数张量，我们可以写成

$$\int f p_3 d\gamma = \frac{1}{3!} \left\langle \mathbf{A}_3(V_0), \int f \mathbf{H}_3 d\gamma \right\rangle = \frac{1}{3!} \left\langle \mathbf{A}_3(V_0), \mathbf{A}_3(f) \right\rangle.$$

### 4.2.1 维数依赖性

到目前为止，我们只讨论了 $n$ 界的依赖性。现在我们也简要讨论 $d$依赖性。正如第 6 点所示，在典型情况下，首阶项为 $\Delta_f \sim \int f_0 p_3 d\gamma$，由柯西-施瓦茨不等式可得其上界为 $\|f_0\|_2 \|p_3\|_2$。在 [18]，中，我们证明了对于对称 $d \times d \times d$张量 $T$，有

$$\mathbb{E}\left[\langle T, \mathbf{H}_3(Z)\rangle^p\right] \leq C_k \mathbb{E}\left[\langle T, \mathbf{H}_3(Z)\rangle^2\right]^{p/2} \leq C_p \left(d\|T\|\right)^p \tag{4.17}$$

对于偶数 $p$。因此，如果 $\|T\| \sim 1/\sqrt{n}$，那么我们就得到了所需的关于 $\epsilon = d/\sqrt{n}$ 的依赖性。

另一个难点是如何对 $\|r_4\|_p$ 进行上界估计。直观而言， $r_4(x)$ 主要由四阶多项式 $\langle \mathbf{A}_4, \mathbf{H}_4(x)\rangle$ 所支配，因此我们预期

$$\mathbb{E}\left[|r_4(Z)|^p\right] \sim \mathbb{E}\left[|\langle \mathbf{A}_4, \mathbf{H}_4(Z)\rangle|^p\right] \sim (d^2/n)^p, \tag{4.18}$$

因为 $\mathbf{A}_4 \sim n^{-1}$ 和 $\mathbb{E}\left[\|Z\|^{4p}\right] \sim d^{2p}$。然而，仅凭 $\|r_4\|_p$ 的级数表示，很难将这一直觉转化为形式化的界。因此，我们推导出 Hermite 级数余项的以下显式公式：

**命题 4.1**。 假设 $V_0 \in C^K$。令 $V_0(x) = \sum_{j=0}^{\infty} \frac{1}{j!} \langle \mathbf{A}j, \mathbf{H}j(x)\rangle$为 $V_0$的厄米特级数展开，并定义

$$r_k(x) = V_0(x) - \sum_{j=0}^{k-1} \frac{1}{j!} \langle \mathbf{A}_j, \mathbf{H}_j(x)\rangle. \tag{4.19}$$

那么，对所有 $k \leq K$，均有

$$r_k(x) = \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} \mathbb{E}\left[\left\langle \nabla^k V_0\left((1-t)Z + tx\right), \mathbf{H}_k(x) - Z \otimes \mathbf{H}_{k-1}(x)\right\rangle\right] dt. \tag{4.20}$$

注意，(4.20) 与泰勒级数余项的积分形式类似。命题的证明见附录 E.3 中的引理 E.3。公式 (4.20) 在一维情形下是已知的；见 [22]中的第 4.15 节。然而，我们在文献中没有找到它的多维版本，因此在此给出了证明。利用这个公式，我们证明启发式论证 (4.18) 基本正确，因为它给出了关于 $d$ 和 $n$的正确尺度关系。

### 4.3 关键引理

在非正式地概述了证明步骤之后，我们现在转向严格证明。我们将给出一系列引理，由此可以推出定理4.1 。省略的证明见附录 D。

首先，我们定义几个量。

定义 4.1。 设 $r_3$ 如（4.19）中所定义，并设 $E := \int \left(e^{-r_3} - 1 + r_3\right) d\gamma$. 此外，对于函数$f \in L^2$（$\gamma$），令

$$L(f) = \int f_0 \left(-r_3 + \frac{r_3^2}{2}\right) d\gamma, \quad E(f) = \int f_0 \left(e^{-r_3} - 1 + r_3 - \frac{r_3^2}{2}\right) d\gamma.$$

在下一个引理中，我们将 $\int f d\rho - \int f d\gamma$启发式地写成 $L(f)$ 加上一个余项。

**引理 4.1**。成立

$$\left| \int f d\rho - \int f d\gamma - L(f) \right| \leq |E| \, |L(f)| + |E(f)|. \tag{4.21}$$

证明。 回顾证明概要可知， $\int f d\rho - \int f d\gamma = \int f_0 e^{-r_3} d\gamma / \int e^{-r_3} d\gamma$，且 $\int f_0 d\gamma = \int r_3 d\gamma = 0$。因此，我们得到

$$\int f d\rho - \int f d\gamma = \frac{L(f) + E(f)}{1 + E}. \tag{4.22}$$

此外， $1 + E = \int e^{-r_3} d\gamma \geq 1$ 根据詹森不等式，由于 $\int (-r_3) d\gamma = 0$。将 $L(f)$ 从 (4.22) 的两边同时减去，并对所得右端进行界定，可得 (4.21)。 □

考察 $L(f)$ $E(f)$ 的定义以及 $E$。由于 $r_3 \sim n^{-1/2}$，我们预期 $L(f) \sim n^{-1/2}$,$E \sim n^{-1}$，且 $E(f) \sim n^{-3/2}$。下一个引理将证实这一点。

**引理 4.2**。 设 $f = g \circ T^{-1}$，如定理 *4.1*中所述。则在定理 *2.1*的条件下，有

$$\begin{aligned} |L(f)| &\lesssim \|f_0\|_2 \left(a_3\epsilon + a_4\epsilon^2\right), \\ |E| &\lesssim \left(a_3\epsilon + a_4\epsilon^2\right)^2, \\ |E(f)| &\lesssim (1 + \|f_0\|_2) \left(a_3\epsilon + a_4\epsilon^2\right)^3, \end{aligned} \tag{4.23}$$

其中，被省略的常数是关于 $q$、$c_0^{-1}$ 和 $R_g$ 的增函数。

引理 4 1 4 2 $\int f d\rho - \int f d\gamma =$。

$L$ $f$ $+ \mathcal{O}$ $n^{-3/2}$

和 。立即得到以下推论，表明 ( ) ( )。

**推论 4.1**。 在与引理*4.2*相同的设定下，有

$$\left| \int f d\rho - \int f d\gamma - L(f) \right| \lesssim (1 + \|f_0\|_2) \left(a_3\epsilon + a_4\epsilon^2\right)^3, \tag{4.24}$$

其中省略的常数是关于 $q$、$c_0^{-1}$ 和 $R_g$ 的增函数。

因此，接下来只需研究项 $L(f)$。通过令 $r_3 = p_3 + r_4$，我们将 $L(f)$ 分解为一个 $\mathcal{O}(n^{-1/2})$ 阶的项、一个 $\mathcal{O}(n^{-1})$ 阶的项，以及一个 $\mathcal{O}(n^{-3/2})$ 阶的余项。

**引理 4.3**。 设 $p_3(x) = \frac{1}{3!}\langle \mathbf{A}_3, \mathbf{H}_3(x)\rangle$ 且 $r_4 = r_3 - p_3$。在与引理 4.2相同的条件下，有

$$L(f) = \int f_0(-p_3)d\gamma + \int f_0\left(-r_4 + \frac{1}{2}p_3^2\right)d\gamma + R,$$

其中

$$\left|\int f_0(-p_3)d\gamma\right| \lesssim_q \|f_0\|_2 a_3\epsilon, \tag{4.25}$$

$$\left|\int f_0\left(-r_4 + \frac{1}{2}p_3^2\right)d\gamma\right| \lesssim_q \|f_0\|_2(a_3^2 + a_4)\epsilon^2, \tag{4.26}$$

$$|R| \lesssim_q \|f_0\|_2\left(a_3a_4\epsilon^3 + a_4^2\epsilon^4\right). \tag{4.27}$$

结合引理 4.3与推论 4.1，现在可以证明定理 4.1。

定理 4.1 的证明。为得到 (4.4)，我们将界 (4.24) 与 $|L(f)|$ 的界相加，后者由 (4.25)、(4.26) 和 (4.27) 的求和给出。为得到 (4.5)，我们将 (4.24)、(4.27) 和 (4.26) 相加，省略界 (4.25)，因为它已并入左侧。同样地，如果 $f$ 与所有三阶厄米多项式正交，则 $\int f_0 p_3 d\gamma = 0$，因此界 (4.6) 源于将 (4.24)、(4.26) 和 (4.27) 相加，同样省略 (4.25)。最后，如果 $f(x) = a^T x + b$，则 $\int f_0 p_3 d\gamma = \int f_0 r_4 d\gamma = 0$。此外，$f_0(x) = a^T x$，这是奇函数，因此 $\int f_0 p_3^2 d\gamma = 0$。所以，通过将 (4.24) 和 (4.27) 相加，并省略 (4.25) 和 (4.26)，即可得到 (4.7)。 □

# 5 平稳条件下唯一解的存在性

与定理 2.1的证明类似，证明引理 2.1 的第一步是将引理改写为与尺度无关的形式。我们将采用以下记号

$$\begin{aligned} B_r(0,0) &= \{(m,\sigma) \in \mathbb{R}^d \times \mathbb{R}^{d\times d} \ : \ \|\sigma\|^2 + \|m\|^2 \le r^2\}, \\ B_r &= \{\sigma \in \mathbb{R}^{d\times d} \ : \ \|\sigma\| \le r\}, \\ S_{c_1,c_2} &= \{\sigma \in \mathbf{S}_+^d \ : \ c_1 I_d \preceq \sigma \preceq c_2 I_d\}. \end{aligned} \tag{5.1}$$

特别注意，$S_{0,\ r} \subset B_r$。

## 5.1 坐标变换

回想一下，$m^*$ 是 $V = nv$的唯一最小值点，且 $H_V = \nabla^2 V(m^*)$。定义

$$W(x) = V(m^* + H_V^{-1/2}x). \tag{5.2}$$

注 5.1。 注意，这种变量变换不同于第 4 节中的变量变换。该节中的变量变换（4.3）已预设了 $\hat{}$ $m$ 和 $S$ 的存在性与唯一性，而这正是我们将在本节证明的内容。式（ $V \mapsto W$ 5.2）中的变换仅将最小值点移至零点，并将该点处的海森矩阵重新缩放为单位矩阵。

**引理 5.1.** 设 $W$ 如 (5.2) 中所示，对 $V = nv$成立，其中 $v$ 满足假设 1 和2。则存在唯一一对$(m,\sigma) \in B_{2\sqrt{2}}(0,0) \cap \mathbb{R}^d \times S_{0,\sqrt{2}}$ 满足

$$\mathbb{E}\left[\nabla W(m + \sigma Z)\right] = 0, \tag{5.3}$$

$$\mathbb{E}\left[\nabla^2 W(m + \sigma Z)\right] = (\sigma\sigma^T)^{-1}. \tag{5.4}$$

并且这对元素满足 $\sigma \in S_{\sqrt{2/3},\ \sqrt{2}}$。

利用这个引理，通过缩放论证即可轻松证明引理 2.1；参见附录 G。

ˆ注：5.2。 $(m,\ \sigma)$ 与 $(\hat{}m,\ S)$ 之间的关系如下：

$$\hat{m} = m^* + H_V^{-1/2} m, \qquad \hat{S} = H_V^{-1/2} \sigma\sigma^T H_V^{-1/2}. \tag{5.5}$$

## 5.2 引理 5.1的证明

我们首先概述引理 5.1 的证明，然后正式陈述由此可得出该结果的引理。这些引理的证明见附录F。

设 $f\colon \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}^d$ 由 $f(m,\sigma) = \mathbb{E}[\nabla W(m+\sigma Z)]$.给出。注意到 $f(0,0) = 0$，因此根据隐函数定理，存在一个映射 $m(\sigma)$，其定义在 $\sigma = 0$的邻域内，使得 $f(m(\sigma)\ \sigma) = 0$。在引理 5.3 中，我们对这一陈述进行定量化，证明对于$r = 2\sqrt{2}$ ，有如下结果：对于每个 $\sigma \in B_{r/2}$ ，存在唯一的 $m = m(\sigma)$，使得 $(m,\sigma) \in B_r(0,0)$，且 $f(m,\sigma) = 0$。由于 $S_{0,r/2} \subset B_{r/2}$，因此特别地，区域 $B_r(0,0)\cap \mathbb{R}^d \times S_0$ 中的任意解 $(m,\sigma)_{r/2}$均具有 $(m(\sigma)\ \sigma)$ 的形式。因此，只需证明存在唯一的解 $\sigma \in S_{0,r/2}$，满足方程 $\mathbb{E}[\nabla^2 W(m(\sigma) + \sigma Z)] = (\sigma\sigma^T)^{-1}$。

为此，我们将该方程改写为 $F(\sigma) = \sigma$，其中

$$F(\sigma) = \mathbb{E}[\nabla^2 W(m(\sigma) + \sigma Z)]^{-1/2}.$$

我们在引理 5.4 中证明， $F$ 在 $S_{0,r/2}$ 上定义良好，是一个压缩映射，并且满足 $F(S_{0,r/2}) \subset S_{c_1,}$ $_{c_2} \subset S_{0,r/2}$。因此，根据压缩映射定理，存在唯一的 $\sigma \in S_{0,r/2}$ 满足$F(\sigma) = \sigma$。但是，由于 $F$ 将 $S_{0,}$ $_{r/2}$ 映射到 $S_{c_1,c_2}$，不动点 $\sigma$ 必然属于 $S_{c_1,c_2}$。证明完毕。

利用 [21], 中给出的反函数定理的定量表述，下面的引理确定了映射 $m(\sigma)$ 定义所在邻域的大小。

**引理 5.2**。 设 $f = (f_1, \ldots, f_d)\colon \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}^d$ 为 $C^3$，其中 $\mathbb{R}^{d\times d}$ 是矩阵集合，并赋予标准矩阵算子范数。假设对于所有 $m, \sigma$， $f(0,0) = 0$, $\nabla_\sigma f(0,0) = 0$, $\nabla_m f(m,\sigma)$均为对称矩阵，且 $\nabla_m f(0,0) = I_d$。设 $r > 0$ 满足

$$\sup_{(m,\sigma)\in B_r(0,0)} \|\nabla f(m,\sigma) - \nabla f(0,m_*)\|_{op} \le \frac{1}{4}. \tag{5.6}$$

于是，对于每个 $\sigma \in \mathbb{R}^{d\times d}$ 使得 $\|\sigma\| \le r/2$ 存在唯一的 $m = m(\sigma) \in \mathbb{R}^d$ ，使得$f(m$ $(\sigma)\ \sigma) = 0$，且 $(m(\sigma)\ \sigma) \in B_r(0,0)$。此外，映射 $\sigma \mapsto m(\sigma)$ 是 $C^2$，并且

$$\begin{aligned} \frac{1}{2} I_d \preceq \nabla_m f(m,\sigma)\big|_{m=m(\sigma)} \preceq \frac{3}{2} I_d, \\ \|\nabla_\sigma m(\sigma)\|_{op} \le 1. \end{aligned} \tag{5.7}$$

参见附录 F ，其中对上文出现的范数给出了严谨的定义，并给出了该引理的证明。

**引理 5.3**。 设 $f\colon \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}^d$ 由下式给出： $f(m,\ \sigma) = \mathbb{E}[\nabla W(\sigma Z + m)]$。则引理 *5.2* 的所有条件均得到满足；特别地，(5.6) 在$r = 2\sqrt{2}$下得到满足。因此，引理 *5.2* 的结论对这一¯$r$选择成立。

**引理 5.4**。 设 $r = 2\sqrt{2}$ 和 $\sigma \in S_{0,\ r/2} \mapsto m(\sigma) \in \mathbb{R}^d$ 是引理 5.2 和 5.3 所给映射在对称非负矩阵上的限制。则函数 $F$ 定义为

$$F(\sigma) = \mathbb{E}\left[\nabla^2 W(m(\sigma) + \sigma Z)\right]^{-1/2}$$

在 $S_{0,r/2}$上定义良好且是严格压缩映射。此外，

$$F(S_{0,r/2}) \subseteq S_{c_1,c_2} \subseteq S_{0,r/2}, \qquad where \ \ c_1 = \sqrt{2/3}, \ \ c_2 = \sqrt{2} = r/2.$$

由此引理可完成引理 5.1 的证明，因为根据压缩映射定理， $F$ 存在唯一的不动点 $\sigma \in S_{0,r/2}$ ，而 $F(\sigma) = \sigma$ 只是对第二个最优性方程（5.4）的改写。我们知道 $\sigma$ 必须属于 $S_{c_1,c_2}$ ，因为 $F$ 将 $S_{0,r/2}$ 映射到该集合。

## 致谢

A. Katsevich 获得美国国家科学基金会资助，项目编号为 DMS-2202963。P. Rigollet 获得美国国家科学基金会资助，项目编号为 IIS-1838071、DMS-2022448 和 CCF-2106377。

我们回顾附录中通篇使用的一些记号。对于满足 $\int |f|^p d\gamma < \infty$ 的函数 $f : \mathbb{R}^d \to \mathbb{R}$ ，我们定义

$$\|f\|_p = \left(\int |f|^p d\gamma\right)^{\frac{1}{p}}.$$

对于任意其他测度 $\lambda$，我们明确标出 $p$-范数对 $\lambda$ 的依赖。特别地，令

$$\|f\|_{L^p(\lambda)} = \left(\int |f|^p d\lambda\right)^{\frac{1}{p}}, \quad \mathrm{Var}_\lambda(f) = \int (f - \lambda(f))^2 d\lambda.$$

## 第2节的证明

推论2.1的证明。令 $g = \mathbb{1}_A$。注意到 $|g(x) - \hat{}\pi(g)| \le 1$ 对所有 $x$，因此 $g$ 满足（2.7），其中 $R_g = 0$。此外，$\mathrm{Var}_{\hat{\pi}}(g) \le 1$。现在应用定理2.8和（2.9）中的2.1 即可得出结论。 □

推论2.2的证明。固定 $\|u\| = 1$，并令 $g_u(x) = u^T \hat{}S^{-1/2}(x - \hat{}m)$。我们有

$$\|\hat{S}^{-1/2}(\bar{m} - \hat{m})\| = \sup_{\|u\|=1} \int g_u d\pi - \int g_u d\hat{\pi}.$$

现在， $g_u(x) = f_u(\hat{}S^{-1/2}(x - \hat{}m))$，其中 $f_u(y) = u^T y$。我们有

$$|f_u(y) - f(0)| \le \|y\| \le e^{2\|y\|^{1/2}} \quad \forall y \in \mathbb{R}^d,$$

因此， $f$ 满足带有 $\alpha = 1/2$ 和 $C_{f_u} = 2$的条件 (2.15)。由引理 2.3 可知， $g_u$ 满足带有 $R_g = C(c_0^{-1})$的条件 (2.7)（注意， $C_{f_u}$ 和 $\alpha$ 是绝对常数，因此我们不将它们列出）。最后，注意对于所有 $\|u\| = 1$，$\mathrm{Var}_{\hat{\pi}}(g_u) = \mathrm{Var}(u^T Z) = 1$ 。由定理 2.1 中的 (2.10) 可得均值误差的界。由于 $R_g$ 仅取决于 $c_0^{-1}$，被省略的常数仅取决于 $c_0^{-1}$和 $q$。

为界定协方差误差，我们考虑函数 $g_u^2$，它关于$\hat{m}$是偶函数。现在，

$$\begin{aligned}\|\hat{S}^{-1/2}(\Sigma-\hat{S})\hat{S}^{-1/2}\| &= \sup_{\|u\|=1}|\mathrm{Var}_{X\sim\pi}(g_u(X))-1| \\ &\le \sup_{\|u\|=1}\left|\mathbb{E}_{X\sim\pi}\left[g_u(X)^2\right]-1\right| + \sup_{\|u\|=1}\mathbb{E}\left[g_u(X)\right]^2 \\ &= \sup_{\|u\|=1}\left|\int g_u^2 d\pi - \int g_u^2 d\hat{\pi}\right| + \|\hat{S}^{-1/2}(\bar{m}-\hat{m})\|^2\end{aligned} \tag{A.1}$$

对于 $\|\hat{S}^{-1/2}(\bar{m}-\hat{m})\|^2$ 的界，我们使用刚刚证明的均值误差。对于 $\int g_u^2 d\pi - \int g_u^2 d\hat{\pi}$ 的界，我们利用 $g_u^2(x) = f_u(\hat{S}^{-1/2}(x-\hat{m}))$，其中 $f_u(y) = (u^T y)^2$。我们有

$$|f_u(y)-f(0)| \le \|y\|^2 \le e^{4\|y\|^{1/4}} \quad \forall y\in\mathbb{R}^d,$$

因此 $f$ 满足 (2.15)，其中 $\alpha = 1/4$ 和 $C_{f_u} = 4$。所以根据引理 2.3，我们有 $g_u^2$ 满足 (2.7)，其中 $C_{g_u^2} = C(c_0^{-1})$（注意 $C_{f_u}$ 和 $\alpha$ 是绝对常数，因此我们不把它们计入）。最后，注意对所有 $\|u\|=1$，$\mathrm{Var}_{\hat{\pi}}(g_u^2) = \mathrm{Var}((u^T Z)^2) \le 3$。现在由定理 2.1 的 (2.9) 可得 $\int g_u^2 d\pi - \int g_u^2 d\hat{\pi}$ 的界。由于 $C_{g_u^2}$ 仅取决于 $c_0^{-1}$，省略的常数仅取决于 $c_0^{-1}$ 和 $q$。 □

引理的证明 2.2。设 $x$ 满足 $\|x-m^*\|_{H_v} = (1/2)\sqrt{d/n}$ 对某个 $r > 0$成立。对 $v$ 在 $m^*$ 附近进行泰勒展开，得到

$$\begin{aligned}v(x)-v(m^*) &= \frac{1}{2}\|x-m^*\|_{H_v}^2 + \frac{1}{3!}\langle\nabla^3 v(\xi),(x-m^*)^{\otimes 3}\rangle \\ &\ge \frac{1}{2}\|x-m^*\|_{H_v}^2\left(1-\frac{1}{3}\|\nabla^3 v(\xi)\|_{H_v}\|x-m^*\|_{H_v}\right) \\ &\ge \frac{1}{2}\|x-m^*\|_{H_v}^2\left(1-\frac{a_3}{6}\sqrt{d/n}(1+(1/2)^q)\right) \ge \frac{1}{4}\|x-m^*\|_{H_v}^2\end{aligned} \tag{A.2}$$

usi利用假设 $a_3(1+(1/2)^q)\sqrt{d/n} \le 3$ 得到最后一个不等式。因此 e,

$$\inf_{\|x-m^*\|_{H_v}=(1/2)\sqrt{d/n}} \frac{v(x)-v(m^*)}{\|x-m^*\|_{H_v}} \ge \frac{1}{8}\sqrt{d/n} \tag{A.3}$$

现在，由于 $v$ 是凸的，我们有

$$\frac{v(y)-v(m^*)}{\|y-m^*\|_{H_v}} \ge \inf_{\|x-m^*\|_{H_v}=(1/2)\sqrt{d/n}} \frac{v(y=x)-v(m^*)}{\|x-m^*\|_{H_v}} \ge \frac{1}{8}\sqrt{d/n}$$

对于所有 $\|y-m^*\|_{H_v} \ge (1/2)\sqrt{d/n}$。 □

引理的证明2.3，参见第G节。

# B 逻辑回归示例

### 数值模拟的详细信息

图 1所示的数值模拟中，我们取 $d=2$和 $n = 100, 200, \ldots, 1000$。对于每个 $n$，我们抽取十组协变量 $x_i, i=1,\ldots,n$，这些协变量取自 $\mathcal{N}(0,\lambda^2 I_d)$，其中 $\lambda=\sqrt{5}$，由此得到十个

后验分布 $\pi_n(\cdot \mid x_{1:n})$。对于每个 $\pi_n$，我们通过直接计算积分来求得真实均值和协方差，所用方法是等间距网格（在二维情况下可行）。通过标准优化过程求得 $m_*$ 的 $\pi_n$ 众数，并使用 $m$，$S$ 根据 [20]中所述的过程计算高斯变分推断估计值 ^^。我们使用了该算法的作者的实现，代码见 https://github.com/marc-h-lambert/W-VI。随后，我们针对每个 $n$ 以及给定 $n$下的十个后验分布，计算拉普拉斯和变分推断的均值和协方差近似误差。图 1中的实线表示每个 $n$下十个分布的平均近似误差。阴影区域表示十个近似误差中间八个的分布范围。

**验证假设**

引理 *3.2* 的证明。设 $S = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i \mid X_i])$ 和 $\mathcal{N}$是 1/2-网，位于 $\mathbb{R}^d$中，因此 $\inf_{u\in\mathcal{N}} \|u - w\| \le 1/2$ 对所有 $\|w\| = 1$均成立。标准论证表明，我们可以取 $\mathcal{N}$使其至多包含 $5^d$ 个元素。对于某个 $s \ge 0$，有

$$\mathbb{P}(\|S\| \ge 2s\sqrt{d/n}) \le \mathbb{P}(\sup_{u\in\mathcal{N}} u^T S \ge s\sqrt{d/n}) \le 5^d \sup_{\|u\|=1} \mathbb{P}(u^T S \ge s\sqrt{d/n}) \tag{B.1}$$

usi应用并集界以及 $|\mathcal{N}| \le 5^d$ 这一事实。对于固定的单位向量 $u$，我们有 e

$$\begin{aligned}
\mathbb{P}(u^T S \ge s\sqrt{d/n}) &= \mathbb{E}\left[\mathbb{P}\left(u^T S \ge s\sqrt{d/n} \mid \{X_i\}_{i=1}^n\right)\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i \mid X_i])u^T X_i \ge s\sqrt{nd} \mid \{X_i\}_{i=1}^n\right)\right] \\
&\le \mathbb{E}\left[\exp\left(-\frac{2s^2 nd}{\sum_{i=1}^{n}(u^T X_i)^2}\right)\right] = \mathbb{E}\left[\exp\left(-\frac{2s^2 nd}{\|Z\|^2}\right)\right].
\end{aligned} \tag{B.2}$$

在最后一行中，$Z \sim \mathcal{N}(0,\ I_n)$。为得到第三行，我们使用了霍夫丁不等式。现在，我们有

$$\mathbb{E}\left[\exp\left(-\frac{2s^2 nd}{\|Z\|^2}\right)\right] \le \mathbb{P}(\|Z\| \ge 2\sqrt{n}) + \exp\left(-s^2 d/2\right) \le e^{-n/2} + e^{-s^2 d/2}. \tag{B.3}$$

合并（B.1）、（B.2）、（B.3）可得

$$\mathbb{P}(\|S\| \ge 2s\sqrt{d/n}) \le e^{d\log 5 - n/2} + e^{d(\log 5 - s^2/2)} \tag{B.4}$$

取 $s = \log n$ 得到

$$\begin{aligned}
\mathbb{P}(\|S\| \ge 2\log n\sqrt{d/n}) &\le e^{d\log 5 - n/2} + e^{d(\log 5 - \frac{1}{2}\log^2 n)} \\
&\le e^{-n/4} + e^{-d(\log n)^2/4} \le e^{-n/4} + n^{-d/4}
\end{aligned}$$

利用当 $n \ge 13$ 时 $\log 5 - \frac{1}{2}\log^2 n \le -\frac{1}{4}\log^2 n$，以及当 $d/n \le 0.1$ 时 $d\log 5 - n/2 < -n/4$。最后，注意到

$$2\log n\sqrt{d/n} = \frac{2\log n}{n^{1/4}}\left(\frac{d}{\sqrt{n}}\right)^{\frac{1}{2}} = \frac{8\log(n^{1/4})}{n^{1/4}}\left(\frac{d}{\sqrt{n}}\right)^{\frac{1}{2}} \le 8\left(\frac{d}{\sqrt{n}}\right)^{\frac{1}{2}}.$$

□

引理3.4的证明。首先，注意到 $\nabla^k v(\theta) = \frac{1}{n}\sum_{i=1}^n \psi^{(k)}(X_i^T\theta)X_i^{\otimes k}$当 $k=3$时，4。因此，

$$\|\nabla^k v(\theta)\| \lesssim 1 + \sup_{\|u\|=1} \frac{1}{n}\sum_{i=1}^n |u^T X_i|^4, \quad k = 3, 4. \tag{B.5}$$

同时利用 $\|\psi^{(k)}\|_\infty \lesssim 1$。现在，我们应用以下结果。

**定理 B.1**（改编自 [1]中的定理 4.2）。设 $X_1, \ldots, X_n$ 为独立同分布的标准高斯分布。存在绝对常数 $C$, $C'$，使得若 $Ct^8[\log(2t^2)]^6 \le n/d^2$ 则

$$\sup_{\|u\|=1} \frac{1}{n}\sum_{i=1}^n |u^T X_i|^4 \le 4 \tag{B.6}$$

以至少 $1 - e^{-C't\sqrt{d}}$ 的概率。

我们将该定理应用于 $t = c(n/d^2)^{1/9}$。设 $M = n/d^2$，并假设 $M \ge 1$。条件 $Ct^8[\log(2t^2)]^6 \le n/d^2$ 因而可写为 $Cc^8\log(2c^2M^{2/9}) \le M^{1/9}$，取 $c$ 为足够小的绝对常数即可满足该条件。因此，(B.6) 成立的概率为 $1 - \exp(-C''(nd)^{1/9}$，其中 $C'' > 0$ 是另一个绝对常数。结合 (B.6) 和 (B.5)，证明得证。

□

推论 3.1 的证明。我们断言，这些结论在事件 $E = E_1 \cap E_2 \cap E_3$ 上成立，其中 $E_1$、$E_2$、$E_3$ 分别在引理 3.2、3.3、3.4 中定义。注意，$\mathbb{P}(E)$ 满足所需的下界。现在，在 $E$ 上，我们有 $\|\nabla\ell(\theta_0)\| \le 8(d/\sqrt{n})^{1/2}$，并且 $\nabla^2\ell(\theta_0) \succeq \tau(\|\theta_0\|)I_d$。此外，若 $\|\theta' - \theta_0\| \le s$，则 $\|\nabla^2\ell(\theta') - \nabla^2\ell(\theta_0)\| \le C_3 s$. 令 $\lambda = \tau(\|\theta_0\|)$ 且 $s = 4(d/\sqrt{n})^{1/2}/\lambda$。若 $(d/\sqrt{n})^{1/2} \le \lambda^2/(16C_3) = \tau(\|\theta_0\|)^2/(16C_3)$，则对于 $f = \ell$ 和 $\theta = \theta_0$，不等式 (3.3) 均成立。因此我们得出结论：$\ell$ 存在唯一全局极小点 $\theta_\ell^*$，它满足

$$\|\theta_\ell^* - \theta_0\| \le (4/\tau(\|\theta_0\|))(d/\sqrt{n})^{1/2} \le C(d/\sqrt{n})^{1/2}$$

接下来，注意在 $E$ 上，我们有

$$\|\nabla v(\theta_\ell^*)\| = \|\nabla\ell(\theta_\ell^*) + n^{-1}\Sigma^{-1}\theta_\ell^*\| = n^{-1}\|\Sigma^{-1}\theta_\ell^*\| \lesssim n^{-1}\|\theta_\ell^*\|$$

and

$$\nabla^2 v(\theta_\ell^*) \succeq \nabla^2\ell(\theta_\ell^*) \succeq \tau(C)I_d,$$

因为 $\|\theta_\ell^*\| \le C(d/\sqrt{n})^{1/2} + \|\theta_0\| \le C$。此外，$\sup_\theta \|\nabla^3 v(\theta)\| \le C_3$，因为 $\nabla^3 v = \nabla^3\ell$。因此，由于假设 $n^{-1}$ 足够小，我们可以使用引理 3.1 得出 $v$ 有一个临界点（因而有一个唯一全局极小点）$\theta^*$，且 $\|\theta^* - \theta_\ell^*\| \lesssim n^{-1}$。根据 $\|\theta_0\|$、$\|\theta_\ell^* - \theta_0\|$、$\|\theta^* - \theta_\ell^*\|$的界可知，$\|\theta^*\| \le C$ 有界。此外，根据事件 $E_2$ 的定义，我们有$\lambda_{\min}(H_v) \ge \lambda_{\min}(\nabla^2\ell(\theta^*)) \ge \tau(C)$。为证明第 3 点，我们结合引理 3.4，以及 $\nabla^k v = \nabla^k\ell$、$k \ge 3$ 这一事实，利用第 2 点中 $\lambda_{\min}(H_v)$ 的下界，得到

$$\|\nabla^k v(\theta)\|_{H_v} \le \lambda_{\min}(H_v)^{-k/2}\|\nabla^k v(\theta)\| \le C \tag{B.7}$$

在事件 $E$上。最后，第 4 点由 $v$的凸性得到，利用引理2.2 以及 $a_3$ 的有界性（用于验证引理的假设）。

□

## C $W$ 与 $V_0$的性质

本节中，我们证明重缩放函数 $W(x) = V(m^* + H_V^{-1/2}x)$ 和 $V_0(x) = V(\hat{m} + \hat{S}^{1/2}x) = V(T^{-1}(x))$ 的若干关键性质。这些性质将在第 4节和第 5节的证明中使用。首先注意，对于定义在 $\mathbb{R}^d$上的测度 $\lambda$，有

$$\left\| \|\nabla^k W\| \right\|_{L^p(\lambda)} = \left( \int \|\nabla^k W(x)\|^p d\lambda(x) \right)^{1/p},$$

其中，积分内的范数是标准张量算子范数

**引理 C.1**。 若 $v$ 满足假设 *1*、*2*、*3* 则 *0* 是 $W$ 的唯一全局极小点，其中$\nabla^2 W(0) = I_d$，并且

$$\begin{aligned} \|\nabla^3 W(x)\| &\le \frac{a_3}{\sqrt{n}}(1 + \|x/\sqrt{d}\|^q), \quad \forall x \in \mathbb{R}^d \\ \|\nabla^4 W(x)\| &\le \frac{a_4}{n}(1 + \|x/\sqrt{d}\|^q), \quad \forall x \in \mathbb{R}^d. \end{aligned} \tag{C.1}$$

此外，

$$W(x) - W(0) \lesssim 1 + \|x\|^{3+q} \quad \forall x \in \mathbb{R}^d, \tag{C.2}$$

$$W(x) - W(0) \ge c_0\sqrt{d}\|x\| \quad \forall \|x\| \ge \sqrt{\frac{2}{3}}\sqrt{d}, \tag{C.3}$$

最后，对于 $m \in \mathbb{R}^d$、 $\sigma \in \mathbb{R}^{d\times d}$，有

$$\begin{aligned} \left\| \|\nabla^3 W\| \right\|_{L^p(\mathcal{N}(m,\sigma\sigma^T))} &\lesssim_{p,q} \frac{a_3}{\sqrt{n}} \left(1 + \|m/\sqrt{d}\|^q + \|\sigma\|^q\right), \\ \left\| \|\nabla^4 W\| \right\|_{L^p(\mathcal{N}(m,\sigma\sigma^T))} &\lesssim_{p,q} \frac{a_4}{n} \left(1 + \|m/\sqrt{d}\|^q + \|\sigma\|^q\right). \end{aligned} \tag{C.4}$$

证明见本节末尾。接下来，我们总结定理 4.1 的证明中所用函数 $V_0(x) = V(\hat{m} + \hat{S}^{1/2}x) = V(T^{-1}(x))$ 的性质。这里，$\hat{m}$、$\hat{S}$ 是引理 2.1 中一阶最优性条件的解。我们先将 $V_0$ 改写为关于 $W$ 的形式。为此，令 $\hat{m}_0 = m$、$\hat{\sigma}_0 = \sigma$，其中 $(m, \sigma)$ 是引理 5.1 给出的 (5.3)、(5.4) 的唯一解（这正是经过重新缩放的最优性条件）。利用注 5.2 中给出的 $(m, \sigma) = (\hat{m}_0, \hat{\sigma}_0)$ 与 $(\hat{m}, \hat{S}^{1/2})$ 之间的关系 (5.5)，可以直接证明

$$V_0(x) = V(\hat{m} + \hat{S}^{1/2}x) = W(\hat{m}_0 + \hat{\sigma}_0 x). \tag{C.5}$$

此外，回顾引理5.1 可知

$$\|\hat{m}_0\| \le 2\sqrt{2}, \qquad \sqrt{\frac{2}{3}} I_d \preceq \hat{\sigma}_0 \preceq \sqrt{2} I_d. \tag{C.6}$$

现在我们得到 $V_0$ 的以下性质。

**引理 C.2**。成立

$$|V_0(x) - V_0(0)| \lesssim_q 1 + \|x\|^{3+q} \quad \forall x \in \mathbb{R}^d, \tag{C.7}$$

$$V_0(x) - V_0(0) \ge \frac{c_0\sqrt{d}}{2}\|x\| - C(q) \quad \forall \|x\| \ge \sqrt{d} + 8\sqrt{2}. \tag{C.8}$$

此外，如果 $0 < s \leq 1$，那么

$$\left\| \|\nabla^3 V_0\| \right\|_{L^p(\mathcal{N}(\mu, s^2 I_d))} \lesssim_q \frac{a_3}{\sqrt{n}} \left(1 + \|\mu/\sqrt{d}\|^q\right), \tag{C.9}$$

$$\left\| \|\nabla^4 V_0\| \right\|_{L^p(\mathcal{N}(\mu, s^2 I_d))} \lesssim_q \frac{a_4}{n} \left(1 + \|\mu/\sqrt{d}\|^q\right). \tag{C.10}$$

特别是，

$$d \left\| \|\nabla^3 V_0\| \right\|_{L^p(\mathcal{N}(0, s^2 I_d))} \lesssim_q a_3 \epsilon, \tag{C.11}$$

$$d^2 \left\| \|\nabla^4 V_0\| \right\|_{L^p(\mathcal{N}(0, s^2 I_d))} \lesssim_q a_4 \epsilon^2, \tag{C.12}$$

其中 $\epsilon = d/\sqrt{n}$。

现在证明引理C.1和C.2。

引理*C.1*的证明。 首先注意到

$$\|\nabla^k W(x)\| = n^{1-k/2} \|\nabla^k v(m^* + H_v^{-1/2} x/\sqrt{n})\|_{H_v}. \tag{C.13}$$

现在，利用假设2，可得

$$\begin{aligned} \|\nabla^k W(x)\| &= n^{1-k/2} \|\nabla^k v(m^* + H_v^{-1/2} x)\|_{H_v} \\ &\leq \frac{a_k}{n^{\frac{k}{2}-1}} \left[1 + \left(\sqrt{n/d} \|m^* + H_v^{-1/2} x/\sqrt{n} - m^*\|_{H_v}\right)^q\right] \\ &= \frac{a_k}{n^{\frac{k}{2}-1}} \left[1 + \|x/\sqrt{d}\|^q\right], \end{aligned} \tag{C.14}$$

这正是所需的。为证明（C.2），我们对 $W$ 在 0 处作泰勒展开，并应用（C.1）。得到

$$\begin{aligned} W(x) - W(0) &= \frac{\|x\|^2}{2} + \frac{1}{3!} \langle \nabla^3 W(tx), x^{\otimes 3} \rangle \\ &\leq \frac{\|x\|^2}{2} + \frac{\|x\|^3}{6} (a_3/\sqrt{n})(1 + \|x/\sqrt{d}\|^q) \\ &\lesssim 1 + \|x\|^{3+q}, \end{aligned} \tag{C.15}$$

回忆 $a_3/\sqrt{n} \leq 1$ 由（2.4）。不等式（C.3）由 $W$ 的定义以及（C.13）得出。为证明（C.4），我们利用（C.1）得到

$$\begin{aligned} \mathbb{E}\left[\|\nabla^3 W(m + \sigma Z)\|^p\right] &\leq \left(a_3/\sqrt{n}\right)^p \mathbb{E}\left[1 + (\|m + \sigma Z\|/\sqrt{d})^{pq}\right] \\ &\lesssim_{p,q} \left(a_3/\sqrt{n}\right)^p \left(1 + \|m/\sqrt{d}\|^{pq} + \|\sigma\|^{pq}\right). \end{aligned} \tag{C.16}$$

我们取 $p$ 次方根即可得出结论。 $\mathbb{E}\left[\|\nabla^4 W(m + \sigma Z)\|^p\right]$ 上的界可用类似方法证明。 □

证明 引理 *C.2*。 回忆表达式（C.5）中关于 $V_0$ 的 $W$表达式。我们首先对$W(\hat{m}_0) - W(0)$ 进行界定。利用泰勒展开、 $\|\hat{m}_0\| \leq 2\sqrt{2}$这一事实以及（2.4），可得

$$0 \leq W(\hat{m}_0) - W(0) \leq \frac{\|\hat{m}_0\|^2}{2} + \frac{\|\hat{m}_0\|^3}{3!} \frac{a_3(1 + \|\hat{m}_0/\sqrt{d}\|^q)}{\sqrt{n}} \leq C(q).$$

现在证明（C.7）。利用（C.2）和（C.6），我们有

$$\begin{aligned}|V_0(x)-V_0(0)| &= |W(\hat{m}_0+\hat{\sigma}_0 x)-W(\hat{m}_0)| \le |W(\hat{m}_0+\hat{\sigma}_0 x)-W(0)|+C(q)\\ &\lesssim 1+\|\hat{m}_0+\hat{\sigma}_0 x\|^{3+q}+C(q)\\ &\lesssim_q 1+\|x\|^{3+q}.\end{aligned} \tag{C.17}$$

为证明（C.8），固定 $\|x\| \ge \sqrt{d}+9$。那么

$$\|\hat{m}_0+\hat{\sigma}_0 x\| \ge \|\hat{\sigma}_0 x\|-\|\hat{m}_0\| \ge \lambda_{\min}(\hat{\sigma}_0)\|x\|-\|\hat{m}_0\| \ge \sqrt{\frac{2}{3}}\|x\|-2\sqrt{2} \ge \frac{1}{2}\sqrt{d}.$$

因此

$$\begin{aligned}V_0(x)-V_0(0) &= W(\hat{m}_0+\hat{\sigma}_0 x)-W(\hat{m})\\ &= W(\hat{m}_0+\hat{\sigma}_0 x)-W(0)+W(0)-W(\hat{m}_0)\\ &\ge c_0\sqrt{d}\|\hat{m}_0+\hat{\sigma}_0 x\|+W(0)-W(\hat{m}_0)\\ &\ge c_0\sqrt{d}\left(\sqrt{\frac{2}{3}}\|x\|-2\sqrt{2}\right)-C(q)\\ &\ge \frac{c_0}{2}\sqrt{d}\|x\|-C(q).\end{aligned} \tag{C.18}$$

为得到最后一行，我们利用了以下事实：当 $\|x\| \ge 9$时，$\sqrt{2/3}\|x\|-2\sqrt{2} \ge \|x\|/2$。至此，完成了（C.8）的证明。为证明（C.9），利用 $\|\hat{}\sigma_0\|$ 的有界性，可得

$$\|\nabla^k V_0(y)\| \lesssim \|\nabla^k W(\hat{m}_0+\hat{\sigma}_0 y)\|, \quad k=3,4. \tag{C.19}$$

因此，

$$\mathbb{E}\left[\|\nabla^3 V_0(Y)\|^p\right] \lesssim \mathbb{E}\left[\|\nabla^3 W(\hat{m}_0+\hat{\sigma}_0 Y)\|^p\right]. \tag{C.20}$$

现在，$\hat{}m_0+\hat{}\sigma_0 Y = \hat{}m_0+\hat{}\sigma_0(\mu+S^{1/2}Z) = m+\sigma Z$，其中 $m=\hat{}m_0+\hat{}\sigma_0\mu$ 且 $\sigma=\hat{}\sigma_0 S^{1/2}$。注意到 $\|m\| \le 2\sqrt{2}+\sqrt{2}\|\mu\|$ 且 $\|\sigma\| \le \sqrt{2}s \le \sqrt{2}$。因此，应用引理C.4中的（C.1），得到

$$\left\|\|\nabla^3 V_0\|\right\|_{L^p(\mathcal{N}(\mu,s^2 I_d))} \lesssim_q \frac{a_3}{\sqrt{n}}\left(1+\|\mu/\sqrt{d}\|^q\right).$$

关于 C.10 的界 $\left\|\|\nabla^4 V_0\|\right\|_{L^p(\mathcal{N}(\mu,S))}$ 是类似的。界 C.11、(C.12) 可由 C.9) 和 (C.10) 立即得到。

□

# D 第 4 节的证明

关于 4.5) 与 (2.12) 的等价性，参见 G 节。此处我们证明第 4.3 节的关键引理。首先，我们提出两个观察：第一，若 $g$满足 (2.7) 且 $f(x) = g(T^{-1}(x)) = g(\hat{}m+\hat{}S^{1/2}x)$，则 $f$满足

$$|f_0(x)| = |f(x)-\gamma(f)| \le \exp\left(\frac{c_0}{4}\sqrt{d}\|x\|\right) \quad \forall \|x\| \ge R_g\sqrt{d}. \tag{D.1}$$

其次，利用4.16证明概要中的），注意到

$$V_0(x)-V_0(0) = \delta + r_3(x) - \|x\|^2/2$$

对于某个常数 $\delta$。将 $x=0$ 代入此方程，可得 $0=\delta+r_3(0)$，即 $\delta=-r_3(0)$。但是$r_3(0)=\frac{1}{3!}\langle \mathbf{A}_3,\mathbf{H}_3(0)\rangle+r_4(0)=r_4(0)$。因此，$\delta=-r_4(0)$，即我们有

$$V_0(x)-V_0(0)=-r_4(0)+r_3(x)+\frac{\|x\|^2}{2}. \tag{D.2}$$

现在，为了证明引理 4.2，我们先证明若干补充引理。整体证明结构与 [18] 的证明相似。该文第 4.3 节给出的提纲或许有助于理解下面的证明。

**引理 D.1**。设 $g,R_g$ 如定理 *2.1*所述，且 $f=g\circ T$。设 $R\geq\max(R_g,1+8\sqrt{2/d})$ 和$\mathcal{U}(R)=\{\|x\|\leq R\sqrt{d}\}$ 那么

$$\begin{aligned}|E(f)|&\lesssim_q\left(1+\left\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\right\|_4\right)\|f_0\|_2\|r_3\|_{12}^3+I(R,9+3q,c_0/4),\\|E|&\lesssim_q\left(1+\left\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\right\|_2\right)\|r_3\|_4^2+I(R,6+2q,c_0/2),\end{aligned} \tag{D.3}$$

其中

$$I(R,p,b)=\frac{1}{(2\pi)^{d/2}}\int_{\|x\|\geq R\sqrt{d}}^{\infty}\|x\|^pe^{-b\sqrt{d}\|x\|}dx. \tag{D.4}$$

证明见本节末尾。我们看到，需要对 $I(R,p,b),\left\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\right\|_4$ 和 $\|r_3\|_p$进行界定。接下来的两个引理给出了前两个量的界。关于 $\|r_3\|_p$ 的界见第 E.6节的引理E.4。

**引理 D.2** （尾积分的界）。存在某个 $C_p$，仅依赖于 $p$，使得若 $R\geq C_p(1\vee b^{-1})^2$，则 $I(R,p,b)\lesssim_p e^{-Rbd/2}$. 特别地，

$$I(R,9+3q,c_0/4)\vee I(R,6+2q,c_0/2)\lesssim_q e^{-Rc_0d/8}\quad if\quad R\geq C(q)(1\vee c_0^{-1})^2$$

对于某个 $C(q)$，它仅依赖于 $q$。

证明见第 G 节。

**引理 D.3** （$\mathcal{U}(R)$ 中指数积分的界）。令 $R\geq 1$。那么

$$\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\|_4\leq\exp\left(C(q)R^{4+q}\omega\right),$$

其中 $\omega=a_3\epsilon+a_4\epsilon^2$。

证明见本节末尾。有了这些引理，以及下文引理$\|r_3\|_p$ 中的界 E.6，现在我们可以证明引理 4.2。

引理的证明 *4.2*。我们有

$$|L(f)|\leq\|f_0\|_2(\|r_3\|_2+\|r_3\|_4^2)\lesssim_q\|f_0\|_2\omega, \tag{D.5}$$

利用引理E. 6 $\omega\leq 1$ 2 4 $R\geq\bar{C}=$ $R_g$ $1+8\sqrt{2/d}$ $C$ $q$ $1\vee$ 以及由（.）可知的事实。接下 $c_0^{-1}$ 2 3 1 $|E|$ $|E$ $f$ $|$ 来，令：max(,,( )( ) )。于是我们可以用引理D.中的（D.）来界定 和 ( )。此外，对于这一选择 $R$，我们可以用引理D.2得到尾积分 $I(R,9+3q,c_0/4)$ 和 $I(R,6+$

$2q, c_0/2$）均有上界 $e^{-Rc_0 d/8}$。最后，引理D.3 给出了$\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\|_4$的上界，而引理E.6给出了 $\|r_3\|_p$ 的上界。综合所有这些界，（D.3）可化简为

$$\begin{aligned} |E| &\lesssim_q \omega^2 \exp\left(C(q)R^{4+q}\omega\right) + e^{-\frac{Rc_0}{8}d} \\ |E(f)| &\lesssim_q \|f_0\|_2\omega^3 \exp\left(C(q)R^{4+q}\omega\right) + e^{-\frac{Rc_0}{8}d}. \end{aligned} \tag{D.6}$$

现在，若 $\omega = 0$则得到 $|E|$，$|E(f)| \lesssim e^{-\frac{Rc_0}{8}d}$，并且可以取 $R \to \infty$从而得出 $|E|$，$|E(f)| = 0$。若 $\omega > 0$ 则由（2.4），可知 $0 < \omega \leq 1$。令

$$R = \max\left(\bar{C}, \frac{8}{c_0 d}\log(\omega^{-3})\right), \tag{D.7}$$

使得 $\exp(-(Rc_0/8)\, d) \leq \omega^3$。此外，利用 $\log(1/\omega)^{4+q}\omega \leq C(q)$，因为 $\omega \leq 1$，可知对于 $R$ 如（D.7）中的情形，指数项 $C(q)R^{4+q}\omega$ 在（D.6）中可由某个固定函数从上界定$C(R_g, c_0^{-1}, q)$。利用这些界，得到

$$|E| \leq C(q, c_0^{-1}, R_g)\omega^2, \qquad |E(f)| \leq C(q, c_0^{-1}, R_g)(1 + \|f_0\|_2)\omega^3 \tag{D.8}$$

注意，若 $\omega = 0$，此界仍然成立。最后，回想一下 $\omega$ 的定义，即可完成证明。

□

接下来，我们证明引理4.3，这可直接由引理 E.6 和若干代数运算推出。

引理的证明 *4.3*。我们有

$$\begin{aligned} L(f) + \int f_0 p_3 d\gamma + \int f_0\left(r_4 - \frac{1}{2}p_3^2\right)d\gamma &= \int f_0\left(\left[-r_3 + \frac{1}{2}r_3^2\right] + r_3 - \frac{1}{2}p_3^2\right)d\gamma \\ = \frac{1}{2}\int f_0((p_3 + r_4)^2 - p_3^2)d\gamma &= \int f_0 p_3 r_4 d\gamma + \frac{1}{2}\int f_0 r_4^2 d\gamma. \end{aligned}$$

接下来，

$$\begin{aligned} \left|\int f_0 p_3 r_4 d\gamma + \frac{1}{2}\int f_0 r_4^2 d\gamma\right| &\leq \|f_0\|_2\left(\|p_3\|_4\|r_4\|_4 + \|r_4\|_4^2\right) \\ &\lesssim \|f_0\|_2\left(a_3\epsilon a_4\epsilon^2 + (a_4\epsilon^2)^2\right) \\ &\lesssim \|f_0\|_2\left(a_3 a_4\epsilon^3 + a_4^2\epsilon^4\right). \end{aligned}$$

接下来，我们有

$$\left|\int f_0\left(r_4 - \frac{1}{2}p_3^2\right)d\gamma\right| \leq \|f_0\|_2\left(\|r_4\|_2 + \|p_3\|_4^2\right) \lesssim \|f_0\|_2(a_3^2 + a_4)\epsilon^2.$$

最后，

$$\left|\int f p_3 d\gamma\right| \leq \|f_0\|_2\|p_3\|_2 \lesssim \|f_0\|_2 a_3\epsilon.$$

□

引理的证明*D.1*。 泰勒余项论证表明

$$\left|e^{-r_3}-1+r_3-r_3^2/2\right| \le |r_3|^3+|r_3|^3e^{-r_3}.$$

因此，

$$\begin{aligned}|E(f)| &\le \int |f_0||r_3|^3 d\gamma + \int |f_0||r_3|^3 e^{-r_3} d\gamma \\ &= \int |f_0||r_3|^3 d\gamma + \int |f_0| r_3^3 (e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}) d\gamma + \int_{\mathcal{U}(R)^c} |f_0||r_3|^3 e^{-r_3} d\gamma \\ &\le \|f_0\|_2 \|r_3\|_{12}^3 (1+\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\|_4) + \int_{\mathcal{U}(R)^c} |f_0||r_3|^3 e^{-r_3} d\gamma.\end{aligned} \tag{D.9}$$

在第 t 三行中，我们使用了幂次为 2¨的广义 H、4、4。接下来，由式（D.2）可得 e

$$e^{-r_3(x)} d\gamma(x) = e^{-r_4(0)} (2\pi)^{-d/2} e^{V_0(0)-V_0(x)} dx.$$

现在，我们在区域 C.8上应用界（ $V_0(0) - V_0(x)$, $\|x\| \ge R\sqrt{d} \ge \sqrt{d}+8\sqrt{2}$从而得到以下尾积分上界：

$$\begin{aligned}\int_{\mathcal{U}(R)^c} |f_0||r_3|^3 e^{-r_3} d\gamma(x) &= \frac{e^{-r_4(0)}}{(2\pi)^{\frac{d}{2}}} \int_{\mathcal{U}(R)^c} |f_0||r_3|^3 e^{V_0(0)-V_0(x)} dx \\ &\lesssim_q \frac{e^{-r_4(0)}}{(2\pi)^{\frac{d}{2}}} \int_{\mathcal{U}(R)^c} |f_0||r_3|^3 e^{-\frac{c_0}{2}\sqrt{d}\|x\|} dx.\end{aligned} \tag{D.10}$$

接下来，由式（D.2）可得 $|r_3(x)| \le |V_0(x)-V_0(0)|+|r_4(0)|+\|x\|^2$。因此，由引理 C.2中的第一个界可得

$$\begin{aligned}|r_3(x)|^3 &\lesssim |V_0(x)-V_0(0)|^3 + |r_4(0)|^3 + \|x\|^6 \\ &\lesssim_q 1 + \|x\|^{9+3q} + \|x\|^6 + |r_4(0)|^3 \\ &\lesssim_q e^{|r_4(0)|}\|x\|^{9+3q}.\end{aligned} \tag{D.11}$$

我们用到了 $\|x\| \ge R\sqrt{d} \ge 1$。最后，注意到 $|f_0(x)| \le \exp(c_0\sqrt{d}\|x\|/4)$，当 $x \bar{\in} \mathcal{U}(R)$ 时，因为 $\|x\| \ge R\sqrt{d} \ge R_g\sqrt{d}$。将此界和式（D.11）代入式（D.10），再将所得的界代入式（D.9），得到

$$|E(f)| \lesssim_q \|f_0\|_2\|r_3\|_{12}^3(1+\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\|_4) + e^{2|r_4(0)|}(2\pi)^{-\frac{d}{2}} \int_{\mathcal{U}(R)^c} \|x\|^{9+3q} e^{-\frac{c_0}{4}\sqrt{d}\|x\|} dx. \tag{D.12}$$

接下来，注意到利用引理 E.5 和（2.4），我们有 $|r_4(0)| \le C(q)a_4\epsilon^2 \le C(q)$ 。将此界代入（D.12），再结合 $I(R,\ p,\ b)$ 在 $p = 9+3q$和 $b = c_0/4$ 下的定义，即可完成证明。关于 $|E|$ 的界也可用类似方法证明。 □

引理的证明 *D.3*。 回顾一下， $\mathbf{A}_3 = \mathbb{E}[\nabla^3 V_0(Z)]$, 并且 $r_3 = \frac{1}{3!}\langle \mathbf{A}_3, \mathbf{H}_3(x)\rangle + r_4(x)$。因此

$$\begin{aligned}\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\|_4 &\le \sup_{\|x\|\le R\sqrt{d}} e^{|r_4(x)|} \|e^{-\frac{1}{3!}\langle \mathbf{A}_3, \mathbf{H}_3\rangle}\mathbb{1}_{\mathcal{U}(R)}\|_4 \\ &\le e^{C(q)R^{4+q}a_4\epsilon^2} \|e^{-\frac{1}{3!}\langle \mathbf{A}_3, \mathbf{H}_3\rangle}\mathbb{1}_{\mathcal{U}(R)}\|_4,\end{aligned} \tag{D.13}$$

利用引理 $|r_4|$ E.5 中的逐点界，得到第二行。为界定 $\|e^{-\frac{1}{3}!\langle \mathbf{A}_3, \mathbf{H}^3\rangle}\mathbb{1}_{\mathcal{U}(R)}\|_4$，我们利用如下事实：若 $f$ 在开集 $\mathcal{U}$ 上的限制是 Lipschitz 连续的，且 Lipschitz 常数为 $L$，则

$$\left|\int_{\mathcal{U}} e^f d\gamma\right| \leq \exp\left(CL^2 + L\|f\|_1\right)$$

对于某个绝对常数 $C$。这由 [3] 中的命题 5.4.1以及 $\gamma$ 满足具有绝对常数的对数 Sobolev 不等式这一事实得出。我们将此结果应用于 $f = -\frac{1}{3}!\langle \mathbf{A}_3, \mathbf{H}_3\rangle$。我们有

$$\|\nabla f(x)\| = \frac{1}{2}\|\langle \mathbf{A}_3, xx^T - I_d\rangle\| \lesssim R^2 d\|\mathbf{A}_3\| \lesssim_q R^2 a_3\epsilon, \quad \|x\| \leq R\sqrt{d}.$$

关于 $d\|\mathbf{A}_3\|$ 的界由引理 C.11 中的 (C.2) 给出。因此，$f$ 在限制于 $L = C(q)R^2 a_3\epsilon$ 时的 Lipschitz 常数为 $\mathcal{U}(R)$。此外，$\|\langle A_3, \mathbf{H}_3\rangle\|_1 \lesssim d\|A_3\| \lesssim a_3\epsilon$ 由 (E.32) 给出。因此，

$$\|e^{-\frac{1}{3!}\langle \mathbf{A}_3, \mathbf{H}_3\rangle}\mathbb{1}_{\mathcal{U}(R)}\|_4 \leq \exp\left(C(q)R^4(a_3\epsilon)^2\right).$$

将此界代入 (D.13)，得到

$$\|e^{-r_3}\mathbb{1}_{\mathcal{U}(R)}\|_4 \leq \exp\left(C(q)\left[R^{4+q}a_4\epsilon^2 + R^4(a_3\epsilon)^2\right]\right) \leq \exp\left(C(q)R^{4+q}\omega\right), \tag{D.14}$$

其中 $\omega = a_3\epsilon + a_4\epsilon^2$。这里我们用到了 $R^4 \leq R^{4+q}$ 以及根据 $(a_3\epsilon)^2 \leq a_3\epsilon$ 由 (2.4) 得到。 □

# E Hermite 级数余项

## E.1 简要入门

下面简要介绍厄米多项式、多项式多重指标和级数展开。令 $H_k : \mathbb{R} \to \mathbb{R}$, $k = 0$、1、2、.....为 $k$ 阶概率论 Hermite 多项式。我们有 $H_0(x) = 1$，$H_1(x) = x$，$H_2(x) = x^2 - 1$，$H_3(x) = x^3 - 3x$。对于所有 $k \geq 1$，我们都可以通过递推关系生成 $H_{k+1}$

$$H_{k+1}(x) = xH_k(x) - kH_{k-1}(x), \quad k \geq 1. \tag{E.1}$$

特别地，$H_k(x)$ 是一个 $k$ 阶多项式，由与 $k$ 奇偶性相同的单项式求和得到。厄米多项式关于高斯测度正交；也就是说，我们有 $\mathbb{E}[H_k(Z)H_j(Z)] = k!\delta_{jk}$。我们还注意到，以供后续参考：

$$\mathbb{E}[ZH_k(Z)H_{k+1}(Z)] = \mathbb{E}\left[\left(H_{k+1}(Z) + kH_{k-1}(Z)\right)H_{k+1}(Z)\right] = (k+1)!, \tag{E.2}$$

使用递推关系（E.1）。

Hermite 多项式多重指标由厄米多项式的乘积给出，并由$\gamma \in \{0, 1, 2, \dots\}^d$索引。令 $\gamma = (\gamma_1$、.....、$\gamma_d)$，其中 $\gamma_j \in \{0, 1, 2, \dots\}$。那么

$$H_\gamma(x_1, \dots, x_d) = \prod_{j=1}^{d} H_{\gamma_j}(x_j),$$

其阶数为 $|\gamma| := \sum_{j=1}^{d} \gamma_j$。注意，如果 $|\gamma| = k$，那么 $H_\gamma(x)$ 可表示为若干与 $k$ 奇偶性相同的单项式之和。事实上，每个 $H_{\gamma_j}(x_j)$ 都是 $x_j^{\gamma_j - 2p}$、$p \leq \lfloor \gamma_j/2 \rfloor$的线性组合。

因此， $H_\gamma(x)$ 是形如 $\prod_{j=1}^d x_j^{\gamma_j-2p_j}$ 的单项式的线性组合，其总阶数为$k-2\sum_j p_j$。利用 $Z=(Z_1,\ldots,Z_d)$ 各项之间的独立性，我们有

$$\mathbb{E}\left[H_\gamma(Z)H_{\gamma'}(Z)\right]=\gamma!\prod_{j=1}^d \delta_{\gamma_j,\gamma_j'},$$

其中 $\gamma!:=\prod_{j=1}^d \gamma_j!$. $H_\gamma$ 也可以显式定义如下:

$$e^{-\|x\|^2/2}H_\gamma(x)=(-1)^{|\gamma|}\partial^\gamma\left(e^{-\|x\|^2/2}\right), \tag{E.3}$$

其中 $\partial^\gamma f(x)=\partial_{x_1}^{\gamma_1}\ldots\partial_{x_d}^{\gamma_d}f(x)$。由此可得实用的高斯分部积分恒等式 ,

$$\mathbb{E}\left[f(Z)H_\gamma(Z)\right]=\mathbb{E}\left[\partial^\gamma f(Z)\right],\quad \text{if } f\in C^{|\gamma|}(\mathbb{R}^d).$$

厄米多项式 $H_\gamma,\gamma\in\{0,1,\ldots\}^d$ 构成希尔伯特空间中函数 $f\colon\mathbb{R}^d\to\mathbb{R}$ 的完备正交基，其内积为 $\langle f,g\rangle=\mathbb{E}\left[f(Z)g(Z)\right]$。特别地，若 $f\colon\mathbb{R}^d\to\mathbb{R}$ 满足 $\mathbb{E}\left[f(Z)^2\right]<\infty$，则 $f$ 有如下厄米特展开:

$$f(x)=\sum_{\gamma\in\{0,1,\ldots\}^d}\frac{1}{\gamma!}a_\gamma(f)H_\gamma(x),\qquad a_\gamma(f):=\mathbb{E}\left[f(Z)H_\gamma(Z)\right]. \tag{E.4}$$

Let

$$r_k(x)=f(x)-\sum_{|\gamma|\le k-1}\frac{1}{\gamma!}a_\gamma(f)H_\gamma(x) \tag{E.5}$$

这是从 $f$ 的厄米特级数展开中去掉 $\le k-1$ 阶多项式后得到的余项。我们可以将 $r_k$ 写成 $f$ 对某个核的积分。具体而言，定义

$$K(x,y)=\sum_{|\gamma|\le k-1}\frac{1}{\gamma!}H_\gamma(x)H_\gamma(y). \tag{E.6}$$

注意

$$\mathbb{E}\left[f(Z)K(x,Z)\right]=\sum_{|\gamma|\le k-1}\frac{1}{\gamma!}a_\gamma(f)H_\gamma(x)$$

是截断的 厄米特级数展开。因此，余项 $r_k$ 可写为 s

$$r_k(x)=f(x)-\mathbb{E}\left[f(Z)K(x,Z)\right]=\mathbb{E}\left[(f(x)-f(Z))K(x,Z)\right],$$

利用 $\mathbb{E}\left[K(x,Z)\right]=1$。

## E.2 余项的精确表达式

**引理 E.1.** 设 $k\ge 1$ 和 $r_k$, $K$分别如 (E.5)、(E.6)所示。假设 $f\in C^1$，且 $\|\nabla f(x)\|\lesssim e^{c\|x\|^2}$ 对某个 $0\le c<1/2$成立。那么

$$\begin{aligned} r_k(x)&=\mathbb{E}\left[(f(x)-f(Z))K(x,Z)\right]\\ &=\int_0^1\sum_{i=1}^d\sum_{|\gamma|=k-1}\frac{1}{\gamma!}\mathbb{E}\left[\partial_i f((1-t)Z+tx)\left(H_{\gamma+e_i}(x)H_\gamma(Z)-H_{\gamma+e_i}(Z)H_\gamma(x)\right)\right]dt. \end{aligned} \tag{E.7}$$

证明依赖于以下恒等式：

**引理 E.2.** 对于每个 $i = 1, \ldots, d$，均有

$$K(x,y) = \frac{1}{x_i - y_i} \sum_{|\gamma|=k-1} \frac{1}{\gamma!} \left( H_{\gamma+e_i}(x) H_\gamma(y) - H_{\gamma+e_i}(y) H_\gamma(x) \right). \tag{E.8}$$

该恒等式的证明见本节末尾。

引理 *E.1* 的证明。令

$$\begin{aligned} f(x) - f(Z) &= \int_0^1 (x - Z)^T \nabla f((1-t)Z + tx) dt \\ &= \sum_{i=1}^d \int_0^1 (x_i - Z_i) \partial_i f((1-t)Z + tx) dt, \end{aligned} \tag{E.9}$$

因此，利用（E.8），可得

$$\begin{aligned} &\mathbb{E}\left[ (f(x) - f(Z)) K(x,Z) \right] \\ &= \sum_{i=1}^d \sum_{|\gamma|=k-1} \frac{1}{\gamma!} \mathbb{E}\left[ \int_0^1 \partial_i f((1-t)Z + tx) \left( H_{\gamma+e_i}(x) H_\gamma(Z) - H_{\gamma+e_i}(Z) H_\gamma(x) \right) dt \right]. \end{aligned} \tag{E.10}$$

根据假设，

$$\begin{aligned} &\sup_{t\in[0,1]} |\partial_i f((1-t)Z + tx)| \left( |H_\gamma(Z)| + |H_{\gamma+e_i}(Z)| \right) \\ &\qquad \lesssim \exp\left( c\|Z\|^2 + 2c\|Z\|\|x\| \right) \left( |H_\gamma(Z)| + |H_{\gamma+e_i}(Z)| \right) \end{aligned} \tag{E.11}$$

对某个 $0 \le c < 1/2$成立。右侧关于高斯测度可积，因此我们可以交换（E.10）中的期望与积分。因此，

$$\begin{aligned} &\mathbb{E}\left[ (f(x) - f(Z)) K(x,Z) \right] \\ &= \int_0^1 \sum_{i=1}^d \sum_{|\gamma|=k-1} \frac{1}{\gamma!} \mathbb{E}\left[ \partial_i f((1-t)Z + tx) \left( H_{\gamma+e_i}(x) H_\gamma(Z) - H_{\gamma+e_i}(Z) H_\gamma(x) \right) \right] dt. \end{aligned} \tag{E.12}$$

□

引理的证明 *E.2*。不失一般性，设 $i = 1$。为简化证明，我们还假设 $d = 2$。读者可以验证，对于一般的 $d$，证明过程也完全相同。根据一维埃尔米特多项式的递推关系（E.1），我们得到

$$H_{\gamma_1+1,\gamma_2}(x) = x_1 H_{\gamma_1,\gamma_2}(x) - \gamma_1 H_{\gamma_1-1,\gamma_2}(x),$$

其中 $x = (x_1, x_2)$。将此方程乘以 $H_\gamma(y)$（其中 $y = (y_1, y_2)$），并交换 $x$ 和 $y$，得到以下两个方程

$$\begin{aligned} H_{\gamma_1+1,\gamma_2}(x) H_{\gamma_1,\gamma_2}(y) &= x_1 H_{\gamma_1,\gamma_2}(x) H_{\gamma_1,\gamma_2}(y) - \gamma_1 H_{\gamma_1-1,\gamma_2}(x) H_{\gamma_1,\gamma_2}(y), \\ H_{\gamma_1+1,\gamma_2}(y) H_{\gamma_1,\gamma_2}(x) &= y_1 H_{\gamma_1,\gamma_2}(x) H_{\gamma_1,\gamma_2}(y) - \gamma_1 H_{\gamma_1-1,\gamma_2}(y) H_{\gamma_1,\gamma_2}(x). \end{aligned} \tag{E.13}$$

Let

$$S_{\gamma_1,\gamma_2} = H_{\gamma_1+1,\gamma_2}(x)H_{\gamma_1,\gamma_2}(y) - H_{\gamma_1+1,\gamma_2}(y)H_{\gamma_1,\gamma_2}(x).$$

将 (E.13) 的第二个方程从第一个方程中相减，并使用 $S_{\gamma_1,\ \gamma_2}$ 记法，可得

$$S_{\gamma_1,\gamma_2} = (x_1 - y_1)H_{\gamma_1,\gamma_2}(x)H_{\gamma_1,\gamma_2}(y) + \gamma_1 S_{\gamma_1-1,\gamma_2} \tag{E.14}$$

因此

$$\frac{S_{\gamma_1,\gamma_2}}{\gamma_1!\gamma_2!} = (x_1 - y_1)\frac{H_{\gamma_1,\gamma_2}(x)H_{\gamma_1,\gamma_2}(y)}{\gamma_1!\gamma_2} + \frac{S_{\gamma_1-1,\gamma_2}}{(\gamma_1-1)!\gamma_2!}.$$

将这个递归关系迭代 $\gamma_1 - 1$ 次，可得

$$\frac{S_{\gamma_1,\gamma_2}}{\gamma_1!\gamma_2!} = (x_1 - y_1)\sum_{j=0}^{\gamma_1-1}\frac{H_{\gamma_1-j,\gamma_2}(x)H_{\gamma_1-j,\gamma_2}(y)}{(\gamma_1-j)!\gamma_2!} + \frac{S_{0,\gamma_2}}{0!\gamma_2!}. \tag{E.15}$$

现在，我们有

$$\begin{aligned} S_{0,\gamma_2} &= H_{1,\gamma_2}(x)H_{0,\gamma_2}(y) - H_{1,\gamma_2}(y)H_{0,\gamma_2}(x) \\ &= H_1(x_1)H_{\gamma_2}(x_2)H_{\gamma_2}(y_2) - H_1(y_1)H_{\gamma_2}(y_2)H_{\gamma_2}(x_2) \\ &= (x_1 - y_1)H_{\gamma_2}(x_2)H_{\gamma_2}(y_2) \\ &= (x_1 - y_1)H_{0,\gamma_2}(x)H_{0,\gamma_2}(y). \end{aligned} \tag{E.16}$$

因此,

$$\frac{S_{0,\gamma_2}}{0!\gamma_2!} = (x_1 - y_1)\frac{H_{\gamma_1-j,\gamma_2}(x)H_{\gamma_1-j,\gamma_2}(y)}{(\gamma_1-j)!\gamma_2!}, \qquad j = \gamma_1$$

因此（E.15）可写为

$$\frac{S_{\gamma_1,\gamma_2}}{\gamma_1!\gamma_2!} = (x_1 - y_1)\sum_{j=0}^{\gamma_1}\frac{H_{\gamma_1-j,\gamma_2}(x)H_{\gamma_1-j,\gamma_2}(y)}{(\gamma_1-j)!\gamma_2!}$$

因此

$$\begin{aligned} \frac{1}{x_1 - y_1}\sum_{\gamma_1+\gamma_2=k-1}\frac{S_{\gamma_1,\gamma_2}}{\gamma_1!\gamma_2!} &= \sum_{\gamma_1+\gamma_2=k-1}\sum_{j=0}^{\gamma_1}\frac{H_{\gamma_1-j,\gamma_2}(x)H_{\gamma_1-j,\gamma_2}(y)}{(\gamma_1-j)!\gamma_2!} \\ &= \sum_{\gamma_1+\gamma_2\le k-1}\frac{H_{\gamma_1,\gamma_2}(x)H_{\gamma_1,\gamma_2}(y)}{\gamma_1!\gamma_2!} = K(x,y), \end{aligned} \tag{E.17}$$

利用这一观察结果:

$$\begin{aligned} &\{(\gamma_1 - j, \gamma_2)\ : \gamma_1 + \gamma_2 = k-1,\ 0 \le j \le \gamma_1\} \\ &= \{(\tilde{\gamma}_1, \gamma_2)\ : \tilde{\gamma}_1 + \gamma_2 \le k-1\}. \end{aligned} \tag{E.18}$$

代回$S_{\gamma_1}$的定义，$_{\gamma_2}$ 即可得到所需结果。 □

### E.3 张量形式的厄米级数余项

利用（E.7），很难得到 $|r_k(x)|$的上界，因为我们需要对所有 $\gamma$ 阶数为 $k-1$的 $r_k$ 求和。本节将 $k$表示为$k$ -张量的标量积，从而得到一种更紧凑的表示。随后，我们利用一个非常有用的表示：阶数为 $r_k$ 的厄米多项式张量可表示为向量外积的期望。这样，我们便能用算子范数而非弗罗贝尼乌斯范数来界定 $d$ 公式中的标量积（后者会导致更大的依赖）。

首先，我们将所有唯一的 $k$ 阶厄米多项式放入一个含有 $d^k$ 个元素的张量中，其中有些元素会重复，并由多重指标 $\alpha=(\alpha_1、\ldots\ldots、\alpha_k)\in[d]^k$ 枚举。这里，$[d]=\{1,\ldots,d\}$。具体如下：给定 $\alpha\in[d]^k$，定义 $\gamma(\alpha)=(\gamma_1(\alpha)\ldots\ldots、\gamma_d(\alpha))$ 为

$$\gamma_j(\alpha)=\sum_{\ell=1}^{k}\mathbb{1}\{\alpha_\ell=j\},$$

也就是说，$\gamma_j(\alpha)$ 统计指标 $j$ 在 $\alpha$ 中出现的次数。因此，我们用计数指标一词指称形式为 $\gamma=(\gamma_1、\ldots\ldots、\gamma_d)\in\{0,1,2,\ldots\}^d$ 的指标，而用标准术语“多重指标”指称 $\alpha$。请注意，如果 $\alpha\in[d]^k$，则自动有 $|\gamma(\alpha)|=k$。现在，对于 $x\in\mathbb{R}^d$，定义 $\mathbf{H}_0(x)=1$ 和 $\mathbf{H}_k(x)$，$k\geq 1$ 为张量

$$\mathbf{H}_k(x)=\{H_{\gamma(\alpha)}(x)\}_{\alpha\in[d]^k},\quad x\in\mathbb{R}^d.$$

枚举$\mathbf{H}_k$的条目时，我们用 $H_k^{(\alpha)}$ 表示 $H_{\gamma(\alpha)}$。注意，对于每个 $\gamma$ 满足$|\gamma|=k$，都存在 $\binom{k}{\gamma}$ $\alpha$个，使得 $\gamma(\alpha)=\gamma$。

例 E.1。考虑张量$\mathbf{H}_6(x)$的 $\alpha=(i,\ j,\ j,\ k,\ k,\ k)$条目，其中 $i,\ j,\ k\in[d]$ 两两不同。我们计数得到，$i$ 出现一次，$j$ 出现两次，$k$ 出现三次。因此

$$H_6^{(i,j,j,k,k,k)}(x)=H_1(x_i)H_2(x_j)H_3(x_k)=x_i(x_j^2-1)(x_k^3-3x_k).$$

前两个张量$\mathbf{H}_1$,$\mathbf{H}_2$ 可以明确写出。对于$\mathbf{H}_1$, 的条目，我们只需写成$H_1^{(i)}(x)=H_1(x_i)=x_i$，即$\mathbf{H}_1(x)=x$。对于$\mathbf{H}_2$, 的条目，我们有 $H_2^{(i,i)}(x)=H_2(x_i)=x_i^2-1$以及 $H_2^{(i,j)}(x)=H_1(x_i)H_1(x_j)=x_ix_j$, $i\neq j$。因此$\mathbf{H}_2(x)=xx^T-I_d$。

现在，我们根据阶数 $|\gamma|$，将厄米特级数展开（E.4）中的各项分组。考虑求和中所有满足 $\gamma$ 的 $|\gamma|=k$。我们断言

$$\sum_{|\gamma|=k}\frac{1}{\gamma!}a_\gamma(f)H_\gamma(x)=\frac{1}{k!}\sum_{\alpha\in[d]^k}a_{\gamma(\alpha)}(f)H_k^{(\alpha)}(x).\tag{E.19}$$

事实上，对于固定的 $\gamma$，满足 $|\gamma|=k$，在 $[d]^k$ 中有 $\binom{k}{\gamma}$ $\alpha$ 个使得 $\gamma(\alpha)=\gamma$，而右侧求和中与这些 $\alpha$ 相对应的被加数都相同，均等于 $a_\gamma(f)H_\gamma(x)$。因此，我们得到 $a_\gamma(f)H_\gamma(x)$ 的 $\binom{k}{\gamma}$ 个副本，只需注意 $\binom{k}{\gamma}/k!=1/\gamma!$。

类似于 $\mathbf{H}_k(x)$，定义张量 $\mathbf{A}_k\in(\mathbb{R}^d)^{\otimes k}$，其第 $\alpha$ 个分量为

$$\mathbf{A}_k^{(\alpha)}=a_{\gamma(\alpha)}=\mathbb{E}\left[f(Z)H_{\gamma(\alpha)}(Z)\right]=\mathbb{E}\left[f(Z)H_k^{(\alpha)}(Z)\right].$$

于是我们看到，求和（E.19）可以写成 $\frac{1}{k!}\langle\mathbf{A}_k,\ \mathbf{H}_k(x)\rangle$，因此 $f$ 的级数展开可以写成

$$f(x)=\sum_{k=0}^{\infty}\frac{1}{k!}\langle\mathbf{A}_k(f),\ \mathbf{H}_k(x)\rangle,\qquad\mathbf{A}_k(f):=\mathbb{E}\left[f(Z)\mathbf{H}_k(Z)\right].\tag{E.20}$$

**引理 E.3**。 设 $f$ 满足引理 *E.1* 的假设，并且进一步假设 $f \in C^k$。则余项 $r_k$，即引理（E.7）中给出的*E.1*，也可以写成如下形式

$$r_k(x) = \int_0^1 \frac{(1-t)^{k-1}}{(k-1)!} \mathbb{E}\left[\left\langle \nabla^k f\left((1-t)Z + tx\right),\, \mathbf{H}_k(x) - Z \otimes \mathbf{H}_{k-1}(x)\right\rangle\right] \tag{E.21}$$

证明。 回想 $\partial^\gamma := \partial_1^{\gamma_1} \dots \partial_d^{\gamma_d}$，并且

$$H_\gamma(z) e^{-\|z\|^2/2} = (-1)^{|\gamma|} \partial^\gamma (e^{-\|z\|^2/2}).$$

于是，对于 $|\gamma| = k-1$，有

$$\begin{aligned} \mathbb{E}\left[\partial_i f((1-t)Z + tx) H_\gamma(Z)\right] &= (1-t)^{k-1} \mathbb{E}\left[\partial^{\gamma+e_i} f\right], \\ \mathbb{E}\left[\partial_i f((1-t)Z + tx) H_{\gamma+e_i}(Z)\right] &= (1-t)^{k-1} \mathbb{E}\left[\partial^{\gamma+e_i} f Z_i\right], \end{aligned} \tag{E.22}$$

利用 $f \in C^k$。为简洁起见，我们省略了右侧的自变量 $(1-t)Z + tx$。为了得到第二个方程，我们只将 $\gamma$ 个 $\gamma + e_i$ 导数从 $e^{-\|z\|^2/2}$ 移到$\partial_i f$上，保留 $-\partial_i(e^{-\|z\|^2/2}) = z_i$。将这两个方程代入 (E.7)，得到

$$\begin{aligned} &\mathbb{E}\left[(f(x) - f(Z))K(x, Z)\right] \\ &= \int_0^1 (1-t)^{k-1} \sum_{i=1}^d \sum_{|\gamma|=k-1} \frac{1}{\gamma!} \mathbb{E}\left[\left(\partial^{\gamma+e_i} f\right)\left(H_{\gamma+e_i}(x) - Z_i H_\gamma(x)\right)\right] dt \\ &= \frac{1}{(k-1)!} \int_0^1 (1-t)^{k-1} \sum_{i=1}^d \sum_{|\gamma|=k-1} \binom{k-1}{\gamma} \mathbb{E}\left[\left(\partial^{\gamma+e_i} f\right)\left(H_{\gamma+e_i}(x) - Z_i H_\gamma(x)\right)\right] dt \end{aligned} \tag{E.23}$$

现在，定义集合

$$\begin{aligned} A &= \{(i, \gamma + e_i) \,:\, i = 1, \dots, d,\ \gamma \in \{0, 1, \dots\}^d,\ |\gamma| = k-1\}, \\ B &= \{(i, \tilde{\gamma}) \,:\, \tilde{\gamma} \in \{0, 1, \dots\}^d,\ |\tilde{\gamma}| = k,\ \tilde{\gamma}_i \geq 1\}. \end{aligned} \tag{E.24}$$

显然， $A = B$。因此，

$$\begin{aligned} &\sum_{i=1}^d \sum_{|\gamma|=k-1} \binom{k-1}{\gamma} \mathbb{E}\left[\partial^{\gamma+e_i} f\right] H_{\gamma+e_i}(x) \\ &\qquad = \sum_{|\tilde{\gamma}|=k} \sum_{i:\, \tilde{\gamma}_i \geq 1} \binom{k-1}{\tilde{\gamma} - e_i} \mathbb{E}\left[\partial^{\tilde{\gamma}} f\right] H_{\tilde{\gamma}}(x) \\ &\qquad = \sum_{|\tilde{\gamma}|=k} \sum_{i:\, \tilde{\gamma}_i \geq 1} \binom{k}{\tilde{\gamma}} \frac{\tilde{\gamma}_i}{k} \mathbb{E}\left[\partial^{\tilde{\gamma}} f\right] H_{\tilde{\gamma}}(x) \\ &\qquad = \sum_{|\tilde{\gamma}|=k} \binom{k}{\tilde{\gamma}} \mathbb{E}\left[\partial^{\tilde{\gamma}} f\right] H_{\tilde{\gamma}}(x) = \left\langle \mathbb{E}\left[\nabla^k f\right],\, \mathbf{H}_k(x)\right\rangle \end{aligned} \tag{E.25}$$

接下来，请注意

$$\sum_{|\gamma|=k-1} \binom{k-1}{\gamma} \partial^\gamma \partial_i f H_\gamma(x) = \langle \nabla^{k-1} \partial_i f,\, \mathbf{H}_{k-1}(x)\rangle,$$

因此

$$\sum_{i=1}^{d}\sum_{|\gamma|=k-1}\binom{k-1}{\gamma}\mathbb{E}\left[(\partial^{\gamma+e_i}f)Z_i\right]H_\gamma(x)=\mathbb{E}\left[\langle\nabla^k f, Z\otimes\mathbf{H}_{k-1}(x)\rangle\right]. \tag{E.26}$$

将（E.25）和（E.26）代入（E.23）可得

$$\begin{aligned} r_k(x)&=\mathbb{E}\left[(f(x)-f(Z))K(x,Z)\right]\\ &=\int_0^1\frac{(1-t)^{k-1}}{(k-1)!}\mathbb{E}\left[\left\langle\nabla^k f\left((1-t)Z+tx\right),\,\mathbf{H}_k(x)-Z\otimes\mathbf{H}_{k-1}(x)\right\rangle\right] \end{aligned} \tag{E.27}$$

□

在下一节中，我们将在 $|r_k(x)|$ 的情形下，得到 $|r_k(x)|$ 的逐点上界 $f=V_0$。为了使该上界关于 $d$的依赖关系足够紧，我们需要一个关于厄米特张量内积的补充结果。为说明这一补充结果的动机，考虑用两侧张量的弗罗贝尼乌斯范数之积来界定（E.21）中的内积。粗略地说，$\|\nabla^k f\|_F\sim d^{k/2}\|\nabla^k f\|$，其中回顾一下，$\|\nabla^k f\|$ 是 算子范数 $\nabla^k f$。因此，我们更希望将内积用 $\|\nabla^k f\|$ 来界定，从而得到关于 $d$的更紧上界。然而，先验地看，这似乎是不可能的，因为 $\mathbf{H}_k(x)$ 并非由 $k$ 向量的外积给出。但下面对 $k$ 阶厄米多项式的表示将使这一点成为可能。

$$\mathbf{H}_k(x)=\mathbb{E}\left[(x+iZ)^{\otimes k}\right], \tag{E.28}$$

其中 $Z\sim\mathcal{N}(0,I_d)$。利用（E.28），我们可以用 $\langle\nabla^k f,\mathbf{H}_k(x)\rangle$ 和$\langle\nabla^k f, Z\otimes\mathbf{H}_{k-1}(x)\rangle$ 的算子范数来界定这类标量积 $\nabla^k f$。更一般地，我们有以下引理。

**引理 E.4**。设 $T\in(\mathbb{R}^d)^{\otimes k}$ 是一个 $k$-张量，且 $v\in\mathbb{R}^d$。那么对所有 $0\le\ell\le k$，均有

$$|\langle T,\,v^{\otimes\ell}\otimes\mathbf{H}_{k-\ell}(x)\rangle|\lesssim_k\|T\|\|v\|^\ell(\|x\|^{k-\ell}+d^{\frac{k-\ell}{2}}).$$

证明。利用（E.28），可得

$$\langle T,\,v^{\otimes\ell}\otimes\mathbf{H}_{k-\ell}(x)\rangle=\mathbb{E}\left[\langle T,\,v^{\otimes\ell}\otimes(x+iZ)^{\otimes(k-\ell)}\rangle\right]$$

因此

$$\begin{aligned} |\langle T,\,v^{\otimes\ell}\otimes\mathbf{H}_{k-\ell}(x)\rangle|&\le\mathbb{E}\left|\langle T,\,v^{\otimes\ell}\otimes(x+iZ)^{\otimes k-\ell}\rangle\right|\\ &\le\|T\|\|v\|^\ell\mathbb{E}\left[\|x+iZ\|^{k-\ell}\right]\\ &\lesssim_k\|T\|\|v\|^\ell(\|x\|^{k-\ell}+\sqrt{d}^{k-\ell}). \end{aligned} \tag{E.29}$$

□

## E.4 将厄米特余项公式应用于 $V_0$

现在考虑 $f=V_0$的情形，其中 $V_0$ 定义见（C.5），并满足引理 C.2中所述的性质。回顾一下，

$$r_k(x)=V_0(x)-\sum_{j=0}^{k-1}\frac{1}{j!}\left\langle\mathbf{A}_j,\mathbf{H}_j(x)\right\rangle, \tag{E.30}$$

其中 $\mathbf{A}_j=\mathbb{E}\left[V_0(Z)\mathbf{H}j(Z)\right]=\mathbb{E}\left[\nabla^j V_0(Z)\right],$，当且仅当 $V_0\in C^j$ 时第二个等式成立。我们首先证明 $r_4$ 的逐点界，然后给出 $r_3$ 和 $r_4$ 的 $L^p(\gamma)$ 范数界。

**引理 E.5**。成立

$$|r_4(x)| \lesssim_q \frac{a_4}{n}\left(\|x\|^4 + d^2 + \sqrt{d}\|x\|^3\right)(1 + \|x/\sqrt{d}\|^q).$$

证明。令 $\nabla^k V_0$ 作为 $\nabla^k V_0((1-t)Z + tx)$ 的简写。利用 E.21式处理 $f = V_0$，可得

$$\begin{aligned}
|r_4(x)| &\le \int_0^1 \mathbb{E}\left[\left|\left\langle \nabla^4 V_0,\ \mathbf{H}_4(x) - Z \otimes \mathbf{H}_3(x)\right\rangle\right|\right] dt \\
&\lesssim \int_0^1 \mathbb{E}\left[\|\nabla^4 V_0\|\left(\|x\|^4 + d^2 + \|Z\|(\|x\|^3 + d^{3/2})\right)\right] dt \\
&\lesssim \left(\|x\|^4 + d^2 + \sqrt{d}\|x\|^3\right)\int_0^1 \sqrt{\mathbb{E}\left[\|\nabla^4 V_0((1-t)Z + tx)\|^2\right]} dt \\
&= \left(\|x\|^4 + d^2 + \sqrt{d}\|x\|^3\right)\int_0^1 \left\|\|\nabla^4 V_0\|\right\|_{L^2(\mathcal{N}(x,(1-t)^2 I_d))} dt \\
&\lesssim_q \frac{a_4}{n}\left(\|x\|^4 + d^2 + \sqrt{d}\|x\|^3\right)(1 + \|x/\sqrt{d}\|^q)
\end{aligned} \tag{E.31}$$

第二行用到了引理 E.4，第三行用到了柯西-施瓦茨不等式以及 $\mathbb{E}\left[\|Z\|^k\right] \lesssim_k d^{k/2}$这一事实。要得到第五行，我们使用了引理 C.2的 C.10式 □

下一引理中关于 $d$-依赖的 $\|r_3\|_p$ 界依赖于 [18]的引理 2.1、C.2 和 C.1。结合这三个引理可知，对于对称的 $d \times d \times d$ 张量 $T$，有

$$\|\langle T, \mathbf{H}_3\rangle\|_{2k} \lesssim_k \|\langle T, \mathbf{H}_3\rangle\|_2 \lesssim \|T\|_F \lesssim d\|T\| \tag{E.32}$$

**引理 E.6**。令 $p_3(x) = \frac{1}{3!}\langle \mathbf{A}_3, \mathbf{H}_3\rangle(x)$。成立

$$\begin{aligned}
\|r_4\|_p &\lesssim_{p,q} a_4\epsilon^2 \\
\|p_3\|_p &\lesssim_{p,q} a_3\epsilon, \\
\|r_3\|_p &\lesssim_{p,q} a_3\epsilon + a_4\epsilon^2 =: \omega,
\end{aligned}$$

其中 $\epsilon = d/\sqrt{n}$。

证明。对于 $\|r_4\|_p$ 的界，我们使用引理 E.5，它给出

$$\mathbb{E}\left[|r_4(Z)|^p\right] \lesssim_{p,q} \left(\frac{a_4}{n}\right)^p \mathbb{E}\left[\left(\|Z\|^4 + d^2 + \sqrt{d}\|Z\|^3\right)^p (1 + \|Z/\sqrt{d}\|^q)^p\right] \lesssim_{p,q} \left(\frac{a_4}{n}\right)^p d^{2p}.$$

取 $p$ 次方根，得到$\|r_4\|_p \lesssim_{p,\ q} a_4 d^2/n = a_4\epsilon^2$。为了界定 $\|p_3\|_p$，我们使用引理 E.32 和 C.11 中的公式，参见引理 C.2:

$$\|p_3\|_p \le \|\langle \mathbf{A}_3, \mathbf{H}_3\rangle\|_p \lesssim_p d\|\mathbf{A}_3\| \le d\mathbb{E}\left[\|\nabla^3 V_0(Z)\|\right] \lesssim_q a_3\epsilon. \tag{E.33}$$

最后，注意到 $r_3 = p_3 + r_4$，即可得到 $\|r_3\|_p$ 的界。 □

# F 第 5 节的证明：驻定条件解的存在性与唯一性

引理 2.1 的证明5.1，请参见第 G节。本节中，我们通过第 5.1节概述的辅助引理来证明引理 5.2。证明依赖于张量-矩阵和张量-向量标量积。下面回顾这类标量积的运算规则，以及如何界定这些量的

算子范数。设 $v \in \mathbb{R}^d$、 $A \in \mathbb{R}^{d\times d}$ 和 $T \in \mathbb{R}^{d\times d\times d}$。我们将向量$\langle T,\ A\rangle \in \mathbb{R}^d$ 以及矩阵 $\langle T,\ v\rangle \in \mathbb{R}^{d\times d}$ 定义为

$$\begin{aligned}\langle T, A\rangle_i &= \sum_{j,k=1}^{d} T_{ijk}A_{jk}, \quad i = 1,\dots,d,\\ \langle T, v\rangle_{ij} &= \sum_{k=1}^{d} T_{ijk}v_k, \quad i,j = 1,\dots,d.\end{aligned} \tag{F.1}$$

我们始终对张量最后两个或最后一个指标求和。注意，矩阵 $\langle T,\ v\rangle$ 的范数为 $\|\langle T,\ v\rangle\| = \sup_{\|u\|=\|w\|=1} u^T\langle T,\ v\rangle w$，并且有

$$u^T\langle T, v\rangle w = \sum_{i,j=1}^{d} u_i w_j \sum_{k=1}^{d} T_{ijk}v_k = \langle T, u\otimes w\otimes v\rangle \le \|T\|\|v\|.$$

因此， $\|\langle T,\ v\rangle\| \le \|T\|\|v\|$。

我们还回顾函数导数的算子范数这一概念，并指出这种算子范数与标准张量算子范数之间的区别。具体而言，考虑一个 $C^2$ 函数 $f = (f_1,\ \dots\dots,\ f_d) : \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}^d$，其中 $\mathbb{R}^{d\times d}$ 赋有标准矩阵范数。那么， $\nabla_\sigma f(m,\ \sigma)$ 是从 $\mathbb{R}^{d\times d}$ 到 $\mathbb{R}^d$的线性泛函，我们令 $\langle \nabla_\sigma f(m,\ \sigma)\ A\rangle \in \mathbb{R}^d$表示将 $\nabla_\sigma f(m,\ \sigma)$ 作用于 $A$。注意，我们可以用 $d\times d\times d$ 张量 $(\nabla_{\sigma_{jk}} f_i)_{i,\ j,\ k=1}^{d}$表示 $\nabla_\sigma f$，因此 $\langle \nabla_\sigma f(m,\ \sigma)\ A\rangle$ 与上文给出的张量-矩阵标量积定义一致。然而， $\|\nabla_\sigma f\|_{\mathrm{op}}$ 并非标准张量算子范数。相反，

$$\|\nabla_\sigma f\|_{\mathrm{op}} = \sup_{A\in\mathbb{R}^{d\times d},\|A\|=1} \|\langle \nabla_\sigma f, A\rangle\| = \sup_{\substack{A\in\mathbb{R}^{d\times d},\|A\|=1,\\ u\in\mathbb{R}^d,\|u\|=1}} \langle \nabla_\sigma f, A\otimes u\rangle.$$

我们继续用 $\|\nabla_\sigma f\|$ 表示标准张量算子范数，即

$$\|\nabla_\sigma f\| = \sup_{\substack{u,v,w\in\mathbb{R}^d,\\ \|u\|=\|v\|=\|w\|=1}} \langle \nabla_\sigma f, u\otimes v\otimes w\rangle.$$

另请注意， $\nabla_m f \in \mathbb{R}^{d\times d}$ 是一个矩阵，并且

$$\begin{aligned}&\max\left(\|\nabla_\sigma f(m,\sigma)\|_{\mathrm{op}},\ \|\nabla_m f(m,\sigma)\|_{\mathrm{op}}\right)\\ &\qquad \le \|\nabla f(m,\sigma)\|_{\mathrm{op}} \le \|\nabla_\sigma f(m,\sigma)\|_{\mathrm{op}} + \|\nabla_m f(m,\sigma)\|_{\mathrm{op}}.\end{aligned} \tag{F.2}$$

最后，回顾一下记号

$$\begin{aligned}B_r(0,0) &= \{(m,\sigma)\in\mathbb{R}^d\times\mathbb{R}^{d\times d} \ :\ \|\sigma\|^2 + \|m\|^2 \le r^2\},\\ B_r &= \{\sigma\in\mathbb{R}^{d\times d} \ :\ \|\sigma\| \le r\},\\ S_{c_1,c_2} &= \{\sigma\in\mathbf{S}_+^d \ :\ c_1 I_d \preceq \sigma \preceq c_2 I_d\}.\end{aligned} \tag{F.3}$$

引理 5.2的证明用到了以下引理。

**引理 F.1 (**《第十四章》中的引理 1.3，出自 [21])。 设 $U$ 是巴拿赫空间 $E$中的开集，并设$f$: $U \to E$属于 $C^1$类。假设 $f(0) = 0$且 $f'(0) = I$。设 $r > 0$满足 $\bar{B}_r(0) \subset U$。若

$$|f'(z) - f'(x)| \le s, \qquad \forall z, x \in \bar{B}_r(0)$$

对于某个 $s \in (0,\ 1)$，则 $f$ 将 $B_r(0)$ 双射映射到 $B_{(1-s)r}(0)$。

引理的证明 5.2。设 $\phi:\ \mathbb{R}^d \times \mathbb{R}^{d\times d} \to \mathbb{R}^d \times \mathbb{R}^{d\times d}$ 由下式给出 $\phi(m,\ \sigma) = (f(m,\ \sigma)\ \sigma)$，因此 $\phi(0,\ 0) = (0,\ 0)$，并且

$$\nabla\phi(m,\sigma) = \begin{pmatrix} \nabla_m f(m,\sigma) & \nabla_\sigma f(m,\sigma) \\ 0 & I_d \end{pmatrix}. \tag{F.4}$$

对于每个 $(m,\sigma),(m',\sigma') \in B_r(0,0)$，我们有

$$\begin{aligned} \|\nabla\phi(m,\sigma) - \nabla\phi(m',\sigma')\|_{\mathrm{op}} &= \|\nabla f(m,\sigma) - \nabla f(m',\sigma')\|_{\mathrm{op}} \\ &\le 2 \sup_{(m,\sigma)\in B_r(0,0)} \|\nabla f(m,\sigma) - \nabla f(0,0)\|_{\mathrm{op}} \le \frac{1}{2}. \end{aligned} \tag{F.5}$$

还要注意，$\nabla\phi(0,0)$ 是恒等算子。因此，根据引理 F.1，我们有 $\phi$ 是从 $B_r(0,0)$ 到 $B_{r/2}(\phi(0,0)) = B_{r/2}(0,0)$ 的双射。现在，任取满足 $\|\sigma\| \le r/2$ 的 $\sigma \in \mathbb{R}^{d\times d}$。那么 $(0,\sigma) \in B_{r/2}(0,0)$，因此存在唯一的 $(m,\sigma') \in B_r(0,0)$，使得 $(0,\sigma) = \phi(m,\sigma') = (f(m,\sigma')\ \sigma')$。于是 $\sigma = \sigma'$ 且 $f(m,\sigma) = 0$。换言之，对于每个满足 $\|\sigma\| \le r/2$ 的 $\sigma$，存在唯一的 $m = m(\sigma)$，使得 $(m(\sigma)\ \sigma) \in B_r(0,0)$，且满足 $0 = f(m,\sigma)$。

映射 $\sigma \mapsto m(\sigma)$ 可由标准的隐函数定理论证得到 $C^2$。为了证明不等式（5.7）的第一个不等式成立，注意到

$$\|\nabla_m f(m(\sigma),\sigma) - \nabla_m f(0,0)\|_{\mathrm{op}} \le \|\nabla f(m(\sigma),\sigma) - \nabla f(0,0)\|_{\mathrm{op}} \le 1/4 \le 1/2$$

由（5.6）可得，因为我们知道 $(m(\sigma)\ \sigma) \in B_r(0,\ 0)$。因此，

$$\begin{aligned} I_d = \nabla^2 W(0) = \nabla_m f(0,0) \\ \implies \frac{1}{2} I_d \preceq \nabla_m f(m(\sigma),\sigma) \preceq \frac{3}{2} I_d. \end{aligned} \tag{F.6}$$

为了证明（5.7）中的第二个不等式，我们首先需要以下补充界

$$\|\nabla_\sigma f(m(\sigma),\sigma)\|_{\mathrm{op}} = \|\nabla_\sigma f(m(\sigma),\sigma) - \nabla_\sigma f(0,0)\|_{\mathrm{op}} \le 1/2 \tag{F.7}$$

根据与上文相同的推理，这一结论成立。现在，

$$\partial_{\sigma_{jk}} m = -\nabla_m f(m,\sigma)^{-1} \partial_{\sigma_{jk}} f(m,\sigma) \in \mathbb{R}^d$$

根据标准的隐函数定理论证，其中 $\nabla_m f(m,\sigma)$ 是一个矩阵，$\partial_{\sigma_{jk}} f(m,\sigma)$ 是一个向量，而 $\nabla_\sigma m$，$\nabla_\sigma f$ 是从 $\mathbb{R}^{d\times d}$ 到 $\mathbb{R}^d$的线性映射。因此，结合（5.7）中的第一个不等式和（F.7），我们得到

$$\begin{aligned} \|\nabla_\sigma m(\sigma)\|_{\mathrm{op}} &= \sup_{\|A\|=1} \|\langle \nabla_\sigma m(\sigma), A\rangle\| \\ &= \sup_{\|A\|=1} \|\nabla_m f(m,\sigma)^{-1} \sum_{j,k=1}^{d} \partial_{\sigma_{jk}} f(m,\sigma) A_{jk}\| \\ &= \sup_{\|A\|=1} \|\nabla_m f(m,\sigma)^{-1} \langle \nabla_\sigma f, A\rangle\| \\ &\le \|\nabla_m f(m,\sigma)^{-1}\| \|\nabla_\sigma f\|_{\mathrm{op}} \le 2 \times \frac{1}{2} = 1. \end{aligned} \tag{F.8}$$

□

引理 5.3 的证明。注意到 $f$ 是 $C^2$，这是因为 $W$ 是 $C^3$，且根据假设 2，$\nabla W$ 呈多项式增长（因为三阶导数张量呈多项式增长）。于是我们立即得到$f(0,0) = \nabla W(0) = 0, \nabla_m f(m,\sigma) = \mathbb{E}[\nabla^2 W(m+\sigma Z)]$对所有 $m,\sigma$都是对称的，并且 $\nabla_m f(0,0) = \nabla^2 W(0) = I_d$。为证明 $\nabla_\sigma f(0,0) = 0$，我们计算该张量的 $i,j,k$ 项:

$$\partial_{\sigma_{jk}} f_i = \partial_{\sigma_{jk}} \mathbb{E}[\partial_i W(m+\sigma Z)] = \mathbb{E}[\partial^2_{ij} W(m+\sigma Z) Z_k],$$

使得 $\partial_{\sigma_{jk}} f_i(0,0) = \mathbb{E}[\partial^2_{i,j} W(0) Z_k] = 0$。接下来只需证明，对于 $r = 2\sqrt{2}$，我们有

$$\sup_{(m,\sigma)\in B_r(0,0)} \|\nabla f(m,\sigma) - \nabla f(0,0)\|_{\mathrm{op}} \le \frac{1}{4}. \tag{F.9}$$

首先，注意到

$$\sup_{(m,\sigma)\in B_r(0,0)} \|\nabla f(m,\sigma) - \nabla f(0,0)\|_{\mathrm{op}} \le r \sup_{(m,\sigma)\in B_r(0,0)} \|\nabla^2 f(m,\sigma)\|_{\mathrm{op}}, \tag{F.10}$$

其中 $\nabla^2 f(m,\sigma)$ 是 $(\mathbb{R}^d \times \mathbb{R}^{d\times d})^2$, 上的双线性形式，并且我们有

$$\|\nabla^2 f(m,\sigma)\|_{\mathrm{op}} \le \|\nabla^2_\sigma f(m,\sigma)\|_{\mathrm{op}} + 2\|\nabla_\sigma \nabla_m f(m,\sigma)\|_{\mathrm{op}} + \|\nabla^2_m f(m,\sigma)\|_{\mathrm{op}}. \tag{F.11}$$

对于 $f(m,\sigma) = \mathbb{E}[\nabla W(\sigma Z + m)],$，这些二阶导数由下式给出

$$\begin{aligned}
\partial^2_{m_i,m_j} f(m,\sigma) &= \mathbb{E}[\partial^2_{i,j} \nabla W(m+\sigma Z)], \\
\partial_{m_i}\partial_{\sigma_{jk}} f(m,\sigma) &= \mathbb{E}[\partial^2_{i,j} \nabla W(m+\sigma Z) Z_k], \\
\partial^2_{\sigma_{jk},\sigma_{\ell p}} f(m,\sigma) &= \mathbb{E}[\partial^2_{j,\ell} \nabla W(m+\sigma Z) Z_k Z_p],
\end{aligned} \tag{F.12}$$

每个都是 $\mathbb{R}^d$ 中的一个向量。由第一行可得

$$\|\nabla^2_m f(m,\sigma)\|_{\mathrm{op}} \le \mathbb{E}\|\nabla^3 W(m+\sigma Z)\| = \big\|\|\nabla^3 W\|\big\|_{L^1(\mathcal{N}(m,\sigma\sigma^T))} \tag{F.13}$$

其中 $\|\nabla^3 W\|$ 是标准张量范数。由第二行可得

$$\begin{aligned}
&\|\nabla_m \nabla_\sigma f(m,\sigma)\|_{\mathrm{op}} \\
&\quad= \sup_{\|A\|=1,\|x\|=1} \left\| \mathbb{E}\left[ \sum_{i,j,k=1}^d \partial^2_{i,j} \nabla W(m+\sigma Z) Z_k x_i A_{jk} \right] \right\| \\
&\quad= \sup_{\|A\|=1,\|x\|=1} \left\| \mathbb{E}\left[ \sum_{i,j=1}^d \partial^2_{i,j} \nabla W(m+\sigma Z) x_i (AZ)_j \right] \right\| \\
&\quad= \sup_{\|A\|=1,\|x\|=1} \left\| \mathbb{E}\left[ \left\langle \nabla^3 W(m+\sigma Z), x \otimes AZ \right\rangle \right] \right\| \\
&\quad\le \sup_{\|A\|=1,\|x\|=1} \mathbb{E}\left[ \|x\|\|AZ\|\|\nabla^3 W(m+\sigma Z)\| \right] \\
&\quad\le \sqrt{d}\,\big\|\|\nabla^3 W\|\big\|_{L^2(\mathcal{N}(m,\sigma\sigma^T))}
\end{aligned} \tag{F.14}$$

类似的计算可得

$$\begin{aligned}
\|\nabla^2_\sigma f(m,\sigma)\|_{\mathrm{op}} &\le \sup_{\|A\|=1,\|B\|=1} \mathbb{E}[\|AZ\|\|BZ\|\|\nabla^3 W(m+\sigma Z)\|] \\
&\le 2d\,\big\|\|\nabla^3 W\|\big\|_{L^2(\mathcal{N}(m,\sigma\sigma^T))}
\end{aligned} \tag{F.15}$$

将（F.13）、（F.14）和（F.15）代入（F.11）可得

$$\begin{aligned}\|\nabla^2 f(m,\sigma)\|_{\mathrm{op}} &\le (2d+2\sqrt{d}+1)\left\|\|\nabla^3 W\|\right\|_{L^2(\mathcal{N}(m,\sigma\sigma^T))}\\ &\le 5d\sup_{(m,\sigma)\in B_r(0,0)}\left\|\|\nabla^3 W\|\right\|_{L^2(\mathcal{N}(m,\sigma\sigma^T))}.\end{aligned}\tag{F.16}$$

将（F.16）代入（F.10，并令 $r=2\sqrt{2}$，然后使用引理 C.1的（C.4），可得

$$\begin{aligned}&\sup_{(m,\sigma)\in B_{2\sqrt{2}}(0,0)}\|\nabla f(m,\sigma)-\nabla f(0,0)\|_{\mathrm{op}}\\ &\le 10\sqrt{2}d\sup_{(m,\sigma)\in B_{2\sqrt{2}}(0,0)}\left\|\|\nabla^3 W\|\right\|_{L^2(\mathcal{N}(m,\sigma\sigma^T))}\\ &\le C(q)a_3\frac{d}{\sqrt{n}}\le\frac{1}{4}.\end{aligned}\tag{F.17}$$

最后一行成立的条件是 $a_3 d/\sqrt{n}\le 1/4C$（$q$），而这由（2.4）保证。 □

引理 5.4 的证明。首先，令 $G(\sigma)=\mathbb{E}[\nabla^2 W(m(\sigma)+\sigma Z)]$，并令 $f(m,\sigma)=\mathbb{E}[\nabla W(m+\sigma Z)]$，如引理 5.3 中所示。注意，$\nabla_m f(m,\sigma)=\mathbb{E}[\nabla^2 W(\sigma Z+m)],$，因此 $G(\sigma)=\nabla_m f(m,\sigma)|_{m=m(\sigma)}$，于是由引理 5.2 的（5.7）可得

$$\frac{1}{2}I_d\preceq G(\sigma)\preceq\frac{3}{2}I_d,\qquad\forall\sigma\in S_{0,r/2}.\tag{F.18}$$

但此时 $G$（$\sigma$）具有唯一的可逆对称正定平方根，我们将$F$（$\sigma$）$=G$（$\sigma$）$^{-1/2}$ 定义为该平方根的逆。此外，利用（F.18），可得

$$c_1 I_d\preceq F(\sigma)\preceq c_2 I_d,\qquad\forall\sigma\in S_{0,r/2},$$

其中 $c_1=\sqrt{2/3}$ 和 $c_2=\sqrt{2}^{-}=r/2.$。换言之，$F(S_{0,r/2})\subseteq S_{c_1,c_2}\subseteq S_{0,r/2}$。接下来需要证明 $F$ 在 $S_{0,r/2}$上是一个压缩映射。令 $\sigma_1,\sigma_2\in S_{0,\sqrt{2}}$。我们先对 $\|G(\sigma_1)-G(\sigma_2)\|$进行界定。我们有

$$\|G(\sigma_1)-G(\sigma_2)\|\le\|\sigma_1-\sigma_2\|\sup_{\sigma\in S_{0,\sqrt{2}}}\|\nabla_\sigma G(\sigma)\|_{\mathrm{op}},\tag{F.19}$$

and

$$\begin{aligned}\|\nabla_\sigma G(\sigma)\|_{\mathrm{op}}&=\sup_{\|A\|=1}\left\|\left\langle\nabla_\sigma G(\sigma),A\right\rangle\right\|\\ &=\sup_{\|A\|=1}\left\|\mathbb{E}\left\langle\nabla^3 W,\left\langle A,\ \nabla_\sigma\left(m(\sigma)+\sigma Z\right)\right\rangle\right\rangle\right\|.\end{aligned}\tag{F.20}$$

这里，右侧 $\|\cdot\|$ 内的量是矩阵。事实上，$\langle\nabla_\sigma G,\ A\rangle$ 表示将 $\nabla_\sigma G$ 作用于 $A$。由于 $G$ 将矩阵映射到矩阵，$\nabla_\sigma G$ 是一个也将矩阵映射到矩阵的线性泛函。在第三行中，$\nabla_\sigma(m(\sigma)+\sigma Z)$应解释为从 $\mathbb{R}^{d\times d}$到 $\mathbb{R}^d$的线性泛函，因此 $\langle A,\ \nabla_\sigma(m(\sigma)+\sigma Z)\rangle$ 是 $\mathbb{R}^d$中的一个向量，而该向量与 $d\times d\times d$张量 $\nabla^3 W$的内积是一个矩阵。利用 $\|\langle T,\ x\rangle\|\le\|T\|\|x\|$（如本节开头所述），我们得到

$$\begin{aligned}\left\|\left\langle\nabla^3 W,\left\langle A,\ \nabla_\sigma\left(m(\sigma)+\sigma Z\right)\right\rangle\right\rangle\right\|&\le\left\|\nabla^3 W\right\|\left\|\langle A,\nabla_\sigma\left(m(\sigma)+\sigma Z\right)\rangle\right\|\\ &\le\|\nabla^3 W\|\|\nabla_\sigma(m(\sigma)+\sigma Z)\|_{\mathrm{op}}\\ &=\|\nabla^3 W\|\|\nabla_\sigma m(\sigma)+Z\otimes I_d\|_{\mathrm{op}}\le\|\nabla^3 W\|(1+\|Z\|).\end{aligned}\tag{F.21}$$

为了得到最后一个界，我们用到了 $\|\nabla_\sigma m(\sigma)\|_{\mathrm{op}} \le 1$，如引理5.3所示。我们还用到了 $\|Z \otimes I_d\|_{\mathrm{op}} =$ 上确界$_{\|A\|=1} \|\langle A,\ Z \otimes I_d\rangle\| =$ 上确界$_{\|A\|=1} \|AZ\| = \|Z\|$这一事实。（请记住，由于 $Z \otimes I_d$ 是 $\nabla_\sigma m$的一部分，我们考虑的是 $Z \otimes I_d$作为矩阵上的算子，而不是 $d \times d \times d$张量，因此我们对矩阵 $A$取上确界。）

将（F.21）代回（F.19），得到

$$\begin{aligned}\|G(\sigma_1) - G(\sigma_2)\| &\le \|\sigma_1 - \sigma_2\| \sup_{\sigma \in S_{0,\sqrt{2}}} \mathbb{E}\left[\left\|\nabla^3 W\left(m(\sigma) + \sigma Z\right)\right\|(1 + \|Z\|)\right] \\ &\le \|\sigma_1 - \sigma_2\| C(q) a_3 \sqrt{\frac{d}{n}} \le \frac{1}{2\sqrt{2}} \|\sigma_1 - \sigma_2\|\end{aligned} \tag{F.22}$$

第二个不等式由柯西-施瓦茨不等式和引理 C.1中的（C.4）得出。若 $a_3\sqrt{d/n} \le 1/$（$2\sqrt{2}C$（$q$））， 则第三个不等式成立；根据假设 2中的（2.4），该条件得到满足。现在注意，由引理 5.3可知， $\lambda_{\min}$（$G$（$\sigma_1$））和 $\lambda_{\min}$（$G$（$\sigma_2$））都以下界 $1/2$为界。利用引理 F.2和（F.22），因此可得

$$\begin{aligned}\|F(\sigma_1) - F(\sigma_2)\| &\le \sqrt{2}\|G(\sigma_1) - G(\sigma_2)\| \\ &\le \frac{1}{2}\|\sigma_1 - \sigma_2\|.\end{aligned} \tag{F.23}$$

因此， $F$ 是一个严格压缩映射。 □

**引理 F.2**。 设 $A_0$ 和 $A_1$ 为半正定矩阵， $A_0^{1/2}$ 和 $A_1^{1/2}$ 分别为它们唯一的半正定平方根。不失一般性，假设 $\lambda_{\min}$（$A_0$）$\le \lambda_{\min}$（$A_1$）。那么

$$\|A_1^{-1/2} - A_0^{-1/2}\| \le \frac{\|A_1 - A_0\|}{2\lambda_{\min}(A_0)^{3/2}}.$$

证明。 首先注意到

$$A_1^{-1/2} - A_0^{-1/2} = A_1^{-1/2}(A_0^{1/2} - A_1^{1/2})A_0^{-1/2}$$

因此

$$\|A_1^{-1/2} - A_0^{-1/2}\| \le \|A_1^{-1/2}\|\|A_1^{-1/2}\|\|A_1^{1/2} - A_0^{1/2}\| \le \frac{\|A_1^{1/2} - A_0^{1/2}\|}{\lambda_{\min}(A_0)}.$$

现在，定义 $A_t = A_0 + t(A_1 - A_0)$，并令 $B_t = A_t^{1/2}$，其中 $B_t$ 是$A_t$的唯一半正定平方根。于是我们有 $\|A_1^{1/2} - A_0^{1/2}\| \le \sup_{t\in[0,1]} \|\dot{B}_t\|$。现在，我们将用 $\dot{A}_t$ 和 $B_t$来表示 $\dot{B}_t$。对 $B_t^2 = A_t$求导，得到

$$B_t\dot{B}_t + \dot{B}_t B_t = \dot{A}_t = A_1 - A_0. \tag{F.24}$$

现在可以验证，该方程的解 $B_t$ 为

$$\dot{B}_t = \int_0^\infty e^{-sB_t}(A_1 - A_0)e^{-sB_t}ds$$

因此

$$\|\dot{B}_t\| \le \|A_1 - A_0\| \int_0^\infty \|e^{-sB_t}\|^2 dt = \frac{\|A_1 - A_0\|}{2\lambda_{\min}(B_t)} = \frac{\|A_1 - A_0\|}{2\sqrt{\lambda_{\min}(A_t)}}.$$

现在注意到 $\lambda_{\min}(A_t) \ge \lambda_{\min}(A_0)$，因为 $A_t$ 只是 $A_0$ 和 $A_1$的凸组合。因此，对于所有 $t \in [0,1]$，都有$\|\dot{B}_t\| \le \|A_1 - A_0\|/2\sqrt{\lambda_{\min}(A_0)}$。结合以上所有估计可得

$$\|A_1^{-1/2} - A_0^{-1/2}\| \le \frac{\|A_1 - A_0\|}{2\lambda_{\min}(A_0)^{3/2}}.$$

□

## G 附加证明

我们首先证明论文中的两个变量替换结果，然后证明另外两个暂缓处理的引理；这些引理的证明涉及一些不太引人入胜的技术计算。

首先，我们证明 (4.5) 与 (2.12) 等价。只需证明 $\int f(-p_3)d\gamma = \int gQd\hat{}\pi$。回忆一下，$\hat{}\pi = T_{\#}\gamma$且 $g = f \circ T$。因此，$\int f(-p_3)d\gamma = -\int gp_3 \circ Td\hat{}\pi$，故只需证明$p_3 \circ T = -Q$。现在，

$$\begin{aligned} p_3(x) &= \left\langle \mathbb{E}\left[\nabla^3 V_0(Z)\right], \ \frac{1}{6}x^{\otimes 3} - \frac{1}{2}x \otimes I_d \right\rangle \\ &= \left\langle \mathbb{E}\left[\nabla^3 V(\hat{m} + \hat{S}^{1/2}Z)\right], \ \frac{1}{6}(\hat{S}^{1/2}x)^{\otimes 3} - \frac{1}{2}(\hat{S}^{1/2}x) \otimes \hat{S} \right\rangle \\ &= \left\langle \mathbb{E}_{X\sim\hat{\pi}}\left[\nabla^3 V(X)\right], \ \frac{1}{6}(\hat{S}^{1/2}x)^{\otimes 3} - \frac{1}{2}(\hat{S}^{1/2}x) \otimes \hat{S} \right\rangle. \end{aligned} \tag{G.1}$$

因此，

$$\begin{aligned} (p_3 \circ T)(x) &= p_3(\hat{S}^{-1/2}(x - \hat{m})) \\ &= \left\langle \mathbb{E}_{X\sim\hat{\pi}}\left[\nabla^3 V(X)\right], \ \frac{1}{6}(x - \hat{m})^{\otimes 3} - \frac{1}{2}(x - \hat{m}) \otimes \hat{S} \right\rangle = -Q(x), \end{aligned}$$

正如所需。

接下来，我们假设引理 5.1 的结论，以证明引理 2.1。

引理 *2.1* 的证明。 设 $m$， $\sigma$ 为引理 5.1中的解。则

$$\begin{aligned} 0 &= \mathbb{E}\left[\nabla W(m + \sigma Z)\right] = H_V^{-1/2}\mathbb{E}\left[\nabla V\left(m^* + H_V^{-1/2}(m + \sigma Z)\right)\right], \\ (\sigma\sigma^T)^{-1} &= \mathbb{E}\left[\nabla^2 W(m + \sigma Z)\right] = H_V^{-1/2}\mathbb{E}\left[\nabla V\left(m^* + H_V^{-1/2}(m + \sigma Z)\right)\right] H_V^{-1/2} \end{aligned}$$

Let

$$\hat{m} = m^* + H_V^{-1/2}m, \qquad \hat{S} = H_V^{-1/2}\sigma\sigma^T H_V^{-1/2}.$$

我们得出结论：

$$\begin{aligned} 0 &= \mathbb{E}\left[\nabla V(\hat{m} + H_V^{-1/2}\sigma Z)\right], \\ \hat{S}^{-1} &= \mathbb{E}\left[\nabla^2 V(\hat{m} + H_V^{-1/2}\sigma Z)\right] \end{aligned} \tag{G.2}$$

现在，我们可以将上述方程中的 $H_V^{-1/2}\sigma$ （$\hat{}S$ 的非对称平方根）替换为 $\hat{}S^{1/2}\hat{}$（ $S$ 的对称平方根），方程仍然成立。因此，$\hat{}m$， $S$ 满足（1.9）。此外，由于（$m$， $\sigma$） $\in B_{2\sqrt{2}}$（$\mathbf{0}$，$\mathbf{0}$）且 $\sigma \in S_{\sqrt{2/3},\ \sqrt{2}}$，我们得到

$$\|H_V^{1/2}\hat{S}H_V^{1/2}\| + \|H_V^{1/2}(\hat{m} - m^*)\|^2 = \|\sigma\sigma^T\| + \|m\|^2 = \|\sigma\|^2 + \|m\|^2 \le 8, \tag{G.3}$$

and

$$\frac{2}{3}H^{-1} \preceq \hat{S} = H_V^{-1/2}\sigma\sigma^T H_V^{-1/2} \preceq 2H^{-1}.$$

ˆ 还需证明 ˆ$m$， $S$ 是在区域 $\mathcal{R}_V$ 内满足（1.9）的唯一解。为证明这一点，注意，若（ $m$ $S$ 1 9 $\mathcal{R}_V$ $(H^{1/2}\ m-m^*\ ,\ H^{1/2}SH^{1/2}\ ^{1/2})\in$ ，）是区域 中（.）的一个解，$B_{2\sqrt{2}}\ 0,0\ \cap\mathbb{R}^d\times S_{0\ \sqrt{2}}$ 5 3 5 4
则 ()¯ () () ，是 (.) 、 (.¯) 的一个解。此外，映射

$$(m,S)\mapsto\left(H^{1/2}(m-m^*),(H^{1/2}SH^{1/2})^{1/2}\right)$$

是单射。由于我们知道5.3、5.4中的解在 $B_{2\sqrt{2}}(0,0)\cap\mathbb{R}^d\times S_{0,\sqrt{2}}$ 中由引理 5.1唯一确定，因此1.9中的解在 $\mathcal{R}_V$ 中也必定唯一。 □

现在我们转向技术计算。

引理 2.3 的证明。若 $g(x)=f(\hat{}S^{-1/2}(x-\hat{}m))$，则只要对所有 $\|y\|\geq R_g\sqrt{d}$ 都有 $|f(y)-\gamma(f)|\leq e^{\frac{c_0}{4}\sqrt{d}\|y\|}$，条件 (2.7) 就成立¯ 现在我们求出 $R_g$，以确保该不等式在 $f$ 上成立。不失一般性，设 $f$ (0) $=0$，且对所有 $y\in\mathbb{R}^d$ 都有 $|f(y)|\leq\exp(\bar{C}\sqrt{d}\|y\|^\alpha)$。我们先对 $\gamma(|f|)$ 作界。我们有

$$\gamma(|f|)\leq\max_{\|y\|\leq(4C_f+2\sqrt{2})\sqrt{d}}|f(y)|+(2\pi)^{-d/2}\int_{\|y\|\geq(4C_f+2\sqrt{2})\sqrt{d}}\exp\left(C_f\sqrt{d}\|y\|-\|y\|^2/2\right)dy. \tag{G.4}$$

现在，我们分别对这两项求和进行界定。对于第一项求和，有

$$\begin{aligned}K_1&:=\max_{\|y\|\leq(4C_f+2\sqrt{2})\sqrt{d}}|f(y)|\\&\leq\exp(C_f(4C_f+3)^\alpha(\sqrt{d})^{1+\alpha})\\&\leq\exp(C_f(4C_f+3)^\alpha d)=:e^{\bar{C}d},\end{aligned} \tag{G.5}$$

其中

$$\bar{C}:=C_f(4C_f+3)^\alpha.$$

对于（G.4）中的第二项求和，有

$$\begin{aligned}K_2&:=(2\pi)^{-d/2}\int_{\|y\|\geq(4C_f+2\sqrt{2})\sqrt{d}}e^{-\|y\|^2/4}dy\\&\leq 2^{\frac{d}{2}}(2\pi)^{-d/2}\int_{\|y\|\geq 2\sqrt{d}}e^{-\|y\|^2/2}dy\\&\leq 2^{\frac{d}{2}}e^{-\frac{d}{2}}\leq 1.\end{aligned} \tag{G.6}$$

在第二行中，我们利用了 $C_f\sqrt{d}\|y\|-\bar{}\|y\|^2/2\leq\|y\|^2/4-\|y\|^2/2=-\|y\|^2/4$ 当 $\|y\|\geq($ $4C_f+2\sqrt{2})\sqrt{d}$时。将（G.5）和（G.6）代入（G.4），得到

$$\gamma(|f|)\leq K_1+K_2\leq 2e^{\bar{C}d}. \tag{G.7}$$

现在，若 $R_g\geq\frac{4}{c_0}(\bar{C}+\log 4)$，则

$$|\gamma(f)|\leq 2e^{\bar{C}d}\leq\frac{1}{2}e^{\frac{c_0}{4}R_gd}\leq\frac{1}{2}e^{\frac{c_0}{4}\sqrt{d}\|x\|}\quad\forall\|x\|\geq R_g\sqrt{d}.$$

接下来，我们对 $|f(y)|$进行界定。注意，如果 $\|y\|\geq R_g\sqrt{d}$ 且 $R_g\geq 1$，那么

$$\begin{aligned}C_f\sqrt{d}\|y\|^\alpha+\log 2&=\left(\frac{C_f\sqrt{d}}{\|y\|^{1-\alpha}}+\frac{\log 2}{\|y\|}\right)\|y\|\\&\leq\left(\frac{C_f\sqrt{d}}{R_g^{1-\alpha}\sqrt{d}^{1-\alpha}}+\frac{\log 2}{R_g\sqrt{d}}\right)\|y\|\\&\leq\frac{(C_f+\log 2)\sqrt{d}}{R_g^{1-\alpha}}\|y\|\leq\frac{c_0}{4}\sqrt{d}\|y\|\end{aligned} \tag{G.8}$$

只要 $R_g \geq \left[\frac{4}{c_0}(C_f + \log 2)\right]^{\frac{1}{1-\alpha}}$。于是我们得到

$$|f(y)| \leq \exp(C_f\sqrt{d}\|y\|^\alpha) \leq \frac{1}{2}\exp\left(\frac{c_0}{4}\sqrt{d}\|y\|\right) \quad \forall \|y\| \geq R_g\sqrt{d}.$$

综合以上所有估计，我们得到

$$|f(y) - \gamma(f)| \leq |f(y)| + \gamma(|f|) \leq \exp\left(\frac{c_0}{4}\sqrt{d}\|y\|\right) \quad \forall \|y\| \geq R_g\sqrt{d}$$

只要

$$R_g \geq \max\left(\left[1 \vee \frac{4}{c_0}(C_f + \log 2)\right]^{\frac{1}{1-\alpha}}, \frac{4}{c_0}\left(C_f(4C_f+3)^\alpha + \log 4\right)\right).$$

这是 $C_f, \alpha, c_0$ 的函数，它随 $C_f, \alpha,$ 以及 $c_0^{-1}$ 递增，符合预期。 □

引理 *D.2*的证明。 引理 E.3 [18] 表明，若 $Rb > 1 + p/d$，则

$$I(R,p,b) \leq (eR)^p d^{\frac{p}{2}+1} \exp\left(\left[\frac{3}{2} + \log R - Rb\right] d\right) \tag{G.9}$$

整理这个上界后，我们得到

$$I(R,p,b) \lesssim_p R^p d^{\frac{p}{2}+1} \exp\left(\left[\frac{3}{2} + \log R - Rb\right] d\right) \leq d^{\frac{p}{2}+1} \exp\left(\left[\frac{3}{2} + (p+1)\log R - Rb\right] d\right).$$

现在，我们有

$$(p+1)\log R = (2p+2)\log\sqrt{R} \leq (2p+2)\sqrt{R} \leq \frac{Rb}{4}$$

满足 $R \geq (8p+8)^2 b^{-2}$。因此，

$$\frac{3}{2} + (p+1)\log R - Rb \leq \frac{3}{2} - \frac{3}{4}Rb \leq -1 - \frac{1}{2}Rb$$

对于 $R \geq 10b^{-1}$。因此 $I(R,\ p,\ b) \lesssim_p d^{\frac{p}{2}+1} e^{-d-Rbd/2} \lesssim_p e^{-Rbd/2}$。当满足以下条件时， $R$的下界得以达到：

$$R \geq \max\left((1+p)b^{-1}, (8p+8)^2 b^{-2}, 10b^{-1}\right),$$

这一条件 可通过取 $R = C_p(1 \vee b^{-1})^2$ 来实现，其中某个 $C_p$ 仅取决于 $p$ . □

# 参考文献


[1] Radoslaw Adamczak、Alexander Litvak、AlainPajor 和 NicoleTomczak-Jaegermann。《对对数凹系综中经验协方差矩阵收敛性的定量估计》。美国数学会杂志，23(2):535–561，2010年。

[2] Pierre Alquier 和 James Ridgway。温度化后验及其变分近似的集中性。统计年鉴，48(3):1475–1497，2020年。

[3] Dominique Bakry、Ivan Gentil、Michel Ledoux 等。《马尔可夫扩散算子分析与几何》[style id='2'已被保留]

[4] David M Blei、Alp Kucukelbir 和 Jon D McAuliffe。《变分推断：面向统计学家的综述》。《美国统计学会会刊》，112(518):859–877，2017年。[5] Edward Challis 和 David Barber。《高斯库尔贝克-莱布勒近似推断》。《机器学习研究杂志》，14(68):2239–2286，2013年。[6] Mary Kathryn Cowles 和 Bradley P. Carlin。《马尔可夫链蒙特卡罗收敛诊断：比较综述》。《美国统计学会会刊》，91(434):883–904，1996年。[7] Kam´elia Daudel 和 Randal Douc。《用于α-散度变分推断的混合权重优化》。载于《神经信息处理系统进展》，第34卷，第4397–4408页，2021年。[8] Kam´elia Daudel、Randal Douc 和 Francois Portier。《基于梯度的无限维α-散度最小化》。《统计年鉴》，49(4):2250–2270，2021年。[9] Guillaume Dehaene 和 Simon Barthelm´e。《大数据极限下的期望传播》。《英国皇家统计学会杂志B辑：统计方法学》，80(1):199–217，2018年。[10] Guillaume P Dehaene。《确定性且可计算的伯恩斯坦-冯·米塞斯定理》。《arXiv预印本 arXiv:1904.02505》，2019年。[11] Guillaume P Dehaene 和 Si mon Barthelm´e。《期望传播误差界》。《神经信息处理系统进展》，28，2015年。[12] Michael Diao、Krishnakumar Balasubramanian、Sinho Chewi 和 Adil Salim。《通过布雷斯-瓦瑟斯坦空间中的JKO进行前向-后向高斯变分推断》。《arXiv预印本 arXiv:2304.05398》，2023年。[13] Subhashis Ghosal。《参数数量趋于无穷时指数族后验分布的渐近正态性》。《多元分析杂志》，74(1):49–68，2000年。[14] Wei Han 和 Yun Yang。《平均场变分贝叶斯中的统计推断》。《arXiv预印本 arXiv:1911.01525》，2019年。[15] Tapio Helin 和 Remo Kretschmann。《贝叶斯逆问题中拉普拉斯近似的非渐近误差估计》。《数值数学》，150(2):521–549，2022年。[16] Mikolaj J Kasprzak、Ryan Giordano 和 Tamara Broderick。《你的高斯后验近似有多好？多种实用散度的有限样本可计算误差界》。《arXiv预印本 arXiv:2209.14992》，2022年。[17] Anya Katsevich。《通过新的拉普拉斯近似界改善伯恩斯坦-冯·米塞斯定理中的维数依赖性》。《arXiv预印本 arXiv:2308.06899》，2023年。[18] Anya Katsevich。《拉普拉斯近似的紧维度依赖性》。《arXiv预印本 arXiv:2305.17604》，2023年。[19] Anya Katsevich。《高维下拉普拉斯近似的紧偏斜校正》。《arXiv预印本 arXiv:2306.07262》，2023年。

[20] Marc Lambert、Sinho Chewi、Francis Bach、Silv`ere Bonnabel 和 Philippe Rigollet。通过 Wasserstein 梯度流进行变分推断。*arXiv* 预印本 *arXiv:2205.15902*，2022年。[21] Serge Lang。实分析与泛函分析。数学研究生教材。Springer New York，纽约，第 3 版，1993年。[22] N Lebedev。特殊函数及其应用，。Dover Publications，1972年。[23] Yulong Lu。关于高维非线性贝叶斯逆问题的伯恩斯坦-冯·米塞斯定理。*arXiv* 预印本 *arXiv:1706.00289*，2017年。[24] Vladimir Spokoiny。关于增长的参数维数的伯恩斯坦-冯·米塞斯定理。*arXiv* 预印本 *arXiv:1302.3430*，2013年。[25] Vladimir Spokoiny。高维拉普拉斯近似精度的与维数无关的非渐近界。*SIAM/ASA* 不确定性量化期刊，11(3):1044–1068，2023年。[26] Pragya Sur。高维逻辑回归的现代极大似然理论。博士论文，斯坦福大学，2019年。[27] Aad W Van der Vaart。渐近统计学，第 3 卷。剑桥大学出版社，2000年。[28] Yixin Wang 和 David M Blei。变分贝叶斯的频率学派一致性。美国统计学会会刊，114(527):1147–1161，2019年。[29] Ruby C Weng。通过 Stein 恒等式得到的贝叶斯 Edgeworth 展开。贝叶斯分析，5(4):741–764，2010年。[30] Fengshuo Zhang 和 Chao Gao。变分后验分布的收敛速率。统计年鉴，48(4):2180–2207，2020年。[31] Xinzhen Zhang、Chen Ling 和 Liqun Qi。对称张量的最佳秩一近似及相关的球面优化问题。*SIAM* 矩阵分析与应用期刊，33(3):806–821，2012年。