arXiv:2611.32604v1 [math.ST] 24 Nov 2026

Oja中位数及相关单纯形中位数的递归估计

Antoine Godichon-Baggioni[1] Guillaume Sallé[2]

[1]索邦大学、巴黎西岱大学、法国国家科学研究中心、概率论、统计学与建模实验室、LPSM，法国巴黎 F-75005

[2]鲁昂诺曼底国立应用科学学院，国立应用科学学院数学实验室，76800 鲁夫赖圣艾蒂安，法国

摘要

我们研究由最小化随机单纯形期望体积所定义的多元位置泛函的递归估计。这一框架同时涵盖空间中位数和仿射等变 Oja 中位数。我们引入滑动窗口随机梯度算法及其平均版本，并证明了它们的强相合性和渐近正态性。进一步地，我们通过与简单的不相交分块随机梯度算法及离线方法进行比较，表明滑动窗口方法在统计精度和计算成本之间实现了更有利的权衡；统计比较基于渐近协方差。数值实验展示了所提出方法在计算方面的优势及其有限样本精度。

关键词： Oja中位数；空间中位数；随机逼近；平均化；随机单纯形；U-统计量。

1 引言

多元位置的估计是非参数统计中的经典问题。均值向量具有仿射等变性：对一个分布应用非奇异仿射变换时，其均值也以相同方式变换。然而，均值对极端观测值的敏感性促使人们采用多元中位数。空间中位数通过最小化期望欧几里得距离来定义，是一元中位数的自然推广；参见 Haldane（1948年）、Brown（1983年）和 Kemperman（1987年）。它在平移、正交变换和统一尺度变换下具有等变性，但在一般非奇异仿射变换下不具有等变性。Oja中位数由 Oja（1983年）提出，具有仿射等变性。它通过最小化一个满维单纯形的期望体积来定义，该单纯形的顶点由候选位置和来自该分布的独立观测值构成。

这两个中位数属于 Dürre 和 Paindav- eine（2022a）研究的位置泛函族。对于 $p \geq 1$ 和 $1 \leq \ell \leq d$，其总体目标函数为

$$\mu \longmapsto \mathbb{E}\left[m_\ell\left(\mathrm{Simpl}\left(X_1, \ldots, X_\ell, \mu\right)\right)^p\right],$$

其中，X_1、……、X_ℓ 是具有共同分布的独立随机向量，定义在 $\mathbb{R}^d$，m_ℓ 表示ℓ维体积，Simpl 表示其参数所构成的凸包。对于 $p = 1$，取值$\ell = 1$ 和 $\ell = d$ 分别得到空间中位数和 Oja中位数。 ℓ 的中间取值则给出基于低维单纯形的相关位置泛函。我们将这些泛函称为单纯形中位数。我们研究中位数情形 $p = 1$ 下的递归估计，其中$1 \leq \ell \leq d$取遍整个范围。在此情形下，次梯度的范数关于 μ 一致有界，其上界为由

ℓ 个观测值构成的单纯形体积除以 ℓ。这一性质使我们能够通过该单纯形体积的矩来控制连续更新。

给定 n 个观测值，完整离线估计量的经验目标函数是一个阶数为 ℓ 的 U 统计量，通过对观测值的所有 $\binom{n}{\ell}$ 子集进行平均化得到（Dürre 和 Paindaveine，2022a，第 2.1 节）。对该求和中的所有项进行求值，在 d 和 ℓ 固定时，计算成本为 n^{ℓ} 阶；当 n 很大且 ℓ 不小时，计算成本会变得难以承受，尤其是对于 Oja 中位数而言，此时 $\ell = d$（Fischer 等，2020）。为了降低这一成本，Dürre 和 Paindaveine（2022a，第 2.4 节）考虑了一种基于从观测样本中随机抽取元组的不完整 U 统计量，大幅减少了经验目标函数中的项数。然而，这些离线方法仍需基于固定数据集反复计算经验目标函数。本文提出一种在线随机梯度算法，该算法使用滑动窗口，在每当有新观测值到达时更新估计量，并且只需固定的存储空间。

随机逼近为此提供了一种自然的方法（Robbins 和 Monro，1951年）。对于空间中位数，Cardot 等人（2013年）研究了希尔伯特空间中随机梯度估计量的几乎必然收敛和均方收敛速率，并建立了其平均版本的渐近正态性。Cardot 等人（2017年）研究了非渐近置信球。对迭代值进行平均化采用了 Ruppert（1988年）以及 Polyak 和 Juditsky（1992年）的方法；另请参见 Pelletier（2000年）。由于每个随机梯度依赖于 ℓ个观测值，使用独立数据块进行随机梯度下降时，每次更新前都需要等待 ℓ 个新的观测值。然而，在 n个观测值之后，这样只能进行 $\lfloor n/\ell \rfloor$ 次更新。相反，我们使用滑动窗口随机梯度估计量，每获得一个新的观测值就执行一次更新，并使用该观测值及其之前的 $\ell - 1$ 个观测值。因此，在单次遍历 n 个观测值的过程中，该算法执行 $n - \ell + 1$ 次更新。如下文所述，在重叠窗口中重复使用观测值还会降低平均估计量的渐近方差。

重叠窗口的使用打破了当前迭代值与用于计算下一步随机梯度的观测值之间的独立性，因此，无法直接应用基于条件无偏梯度的标准结果。关于相关观测值下的随机逼近已有大量文献；有关研究可参见 Benveniste 等（1990）；Kushner 和 Yin（2003），关于平均化则可参见 Kushner 和 Yang（1993）。这里的困难在于：尽管单纯形体积并非处处可微，但仍需在关于观测值共同分布的假设下控制随机梯度的变化 X_i。

对于 Oja中位数这一特殊情形，由观测值确定的超平面的每一侧，单纯形体积的梯度都是常数，并且两侧的取值相反。在关于观测值密度 X_i的适当假设下，我们控制该梯度在平均意义下的变化，并在递归估计量及其平均版本的收敛速率和渐近正态性的证明中使用所得的界。该情形对应于（p, ℓ）$=$(1, d)，在 Dürre 和 Paindaveine（2022a，定理2.5）的离线估计量极限定理中被排除在外。在多元正态性下，关于样本 Oja中位数的更早渐近结果由 Oja和 Niinimaa（1985年）获得。

我们首先在不作任何密度假设的情况下证明几乎处处收敛，随后在密度满足多项式上界以及适当的正则性和步长条件下，分别为递归估计量及其平均版本建立几乎处处收敛速率和渐近正态性。平均估计量的平方误差为 $O\left(n^{-1}\log\log\ n\right)$ 几乎处处，并且在 $\sqrt{n}$ 尺度下具有渐近正态性。对于 $\ell < d$，在不要求密度的替代假设下，两个估计量的收敛速率和极限定理同样成立。平均滑动窗口估计量的渐近协方差矩阵在 Loewner序下介于完整离线估计量和平均不相交分块估计量的渐近协方差矩阵之间；当 $\ell \geq 2$ 且分布具有密度时，不等式严格成立。两种递推的计算成本处于同一数量级。

本文组织如下。第2节介绍统计设定和假设。第3节定义递归估计量，陈述收敛结果，并比较它们的极限协方差矩阵和计算成本。第4节包含模拟研究。证明和辅助结果置于附录中。

2 理论框架

记号。 我们用 $\|\cdot\|$ 表示欧几里得范数，用 $\|\cdot\|_{\mathrm{op}}$ 表示诱导算子范数。对于 $\mathbb{R}^d$中的点的有限元组 $\mathbf{x}$ ，分别用 Simpl(x) 和 aff(x) 表示其凸包和仿射张成。对于 $j \in \{0, \ldots, d\}$，我们用 m_j 表示 j 维体积。我们用 dist(u, E) 表示 u 到集合 E的欧几里得距离。

2.1 单纯形中位数及其经验对应项

设 P 是 $\mathbb{R}^d$上的概率测度，其中 $d \geq 2$，并固定 $1 \leq \ell \leq d$。设 $(X_i)_{i \geq 1}$ 为一列服从共同分布 P 的独立随机向量，并记 $\mathbf{X} \coloneqq (X_1, \ldots, X_\ell)$。对于$\mathbf{x} \coloneqq (x_1, \ldots, x_\ell) \in (\mathbb{R}^d)^\ell$ 且 $\mu \in \mathbb{R}^d$，我们将 $g(\mathbf{x}; \mu)$ 定义为顶点为 $x_1, \ldots, x_\ell, \mu$ 的单纯形的 ℓ维体积:

$$g(\mathbf{x}; \mu) \coloneqq m_\ell(\mathrm{Simpl}(x_1, \ldots, x_\ell, \mu)).$$

我们研究 μ^* 的极小化点

$$G(\mu) \coloneqq \mathbb{E}[g(\mathbf{X}; \mu)].$$

对于 $\ell = 1$， $G(\mu) = \mathbb{E}[\|X_1 - \mu\|]$ 和 μ^* 是空间中位数。对于 $\ell = d$，它是 Oja中位数。

如果点 $x_1, \ldots, x_\ell$ 仿射无关，则令 $\Gamma(\mathbf{x})$表示到 span $(x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell)$的正交补上的正交投影。对于这样的元组以及任意 $\mu \in \mathbb{R}^d$，我们有 $\|\Gamma(\mathbf{x})(\mu - x_\ell)\| = \mathrm{dist}(\mu, \mathrm{aff}(\mathbf{x}))$。对于 $\ell = 1$，我们令 $\Gamma(x) \coloneqq I_d$，并令 $m_0(\mathrm{Simpl}(x)) \coloneqq 1$。对于 $\ell > 1$，我们在仿射相关的元组上令 $\Gamma(\mathbf{x}) \coloneqq 0$；此时 $g(\mathbf{x}; \mu)$和 $m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))$均为零。底高公式（Dürre 和 Paindaveine，2022b，引理 S.1.1）给出

$$g(\mathbf{x}; \mu) = \frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell} \|\Gamma(\mathbf{x})(\mu - x_\ell)\|. \tag{2.1}$$

这一表 示表明， $g(\mathbf{x}; \cdot)$ 是凸函数。我们采用其次梯度的如下选 t:

$$\nabla_\mu g(\mathbf{x}; \mu) \coloneqq \begin{cases} \dfrac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell} \dfrac{\Gamma(\mathbf{x})(\mu - x_\ell)}{\|\Gamma(\mathbf{x})(\mu - x_\ell)\|}, & \text{if } \Gamma(\mathbf{x})(\mu - x_\ell) \neq 0, \\ 0, & \text{if } \Gamma(\mathbf{x})(\mu - x_\ell) = 0. \end{cases} \tag{2.2}$$

当通常梯度存在时，该记号表示通常梯度。在不可微点处，所选取的值是一个次梯度。

对于 $n \geq \ell$，离线估计量的经验目标函数为

$$\widehat{G}_n(u) \coloneqq \binom{n}{\ell}^{-1} \sum_{1 \leq i_1 < \cdots < i_\ell \leq n} g(X_{i_1}, \ldots, X_{i_\ell}; u). \tag{2.3}$$

当其极小化点集合非空时，令 $\widehat{\mu}_n$ 表示任意一个可测的极小化点选择。在下面的密度假设下，对于 $n \geq \ell + 1$，该集合几乎处处非空、紧且凸。其存在性与可测性在附录 B.1 中定理 3.4 的证明中得以确立。

如果 P 关于 a 中心对称，且极小化点唯一，那么 $\mu^* = a$。事实上，将所有顶点关于 a 反射不会改变单纯形体积，因此 $G(a + u) = G(a - u)$。对于 $\ell = d$，可逆仿射变换 $x \mapsto Ax + b$ 会将每个单纯形体积乘以 $|\det A|$。因此，唯一的总体最小化元会变换为 $A\mu^* + b$；参见 Dürre 和 Paindaveine（2022a，第2.2节）。经验极小化器的集合具有相同的变换性质。

2.2 分布假设

为了得到几乎必然收敛，我们只需要对 P作出以下条件要求。
假设 A1. 函数 G 有唯一极小元 $\mu^* \in \mathbb{R}^d$，且 $\mathbb{E}\left[m_{\ell-1}(\text{Simpl}(\mathbf{X}))^2\right] < \infty$。

由于 G 有唯一极小元，它并非恒等于无穷大，因此 $G(\mu^*) < \infty$。由 (2.1) 和假设 A1 中的矩条件可知，G 在 $\mathbb{R}^d$ 上有限且连续。根据 (C.1)，只要 $\mathbb{E}\left[\|X_1\|^2\right] < \infty$，假设 A1 中的矩条件就成立。

对于收敛速率和渐近分布，我们对 $\ell < d$ 和 $\ell = d$ 均采用以下假设。

假设 A2. 分布 P 关于勒贝格测度具有博雷尔可测密度 f，且存在 $\nu > 2$ 和有限常数 $C_f > 0$，使得对于每个 $x \in \mathbb{R}^d$,

$$0 \le f(x) \le \frac{C_f}{(1+\|x\|)^{d+\nu}}.$$

假设 A2 蕴含假设 A1；见命题 F.1。它对每个 $0 < q < \nu$ 给出 $\mathbb{E}[\|X_1\|^q] < \infty$ 。同一命题还表明，G 有限、严格凸、强制且可微，并且其梯度是 Lipschitz 的。

假设 A3. 在假设 A1 中定义的极小化点 μ^* 处，函数 G 二阶可微，且海森矩阵 H 正定:
$= \nabla^2 G(\mu^*)$。此外，存在一个指数α，满足 $0 < \alpha \le 1$，以及有限常数 $r_\alpha, C_\alpha > 0$，使得对于任意满足 $\|u\| \le r_\alpha$的 $u \in \mathbb{R}^d$，∇G 在 μ^* 处的一阶泰勒余项满足

$$\|\nabla G(\mu^* + u) - Hu\| \le C_\alpha \|u\|^{1+\alpha}.$$

备注. 假设 A3 中的余项界可由海森矩阵在 μ^*处的赫尔德连续性验证。更确切地说，假设 G 在 μ^*的某个邻域内二阶可微，并且存在一个指数 α，满足 $0 < \alpha \le 1$，以及有限常数 $L_\alpha, r > 0$，使得对于任意满足 $\|u\| \le r$的 $u \in \mathbb{R}^d$，

$$\left\|\nabla^2 G(\mu^* + u) - H\right\|_{\text{op}} \le L_\alpha \|u\|^\alpha.$$

由于 $\nabla G(\mu^*) = 0$，沿着从 μ^* 到 $\mu^* + u$ 的线段积分可得

$$\nabla G(\mu^* + u) - Hu = \int_0^1 \left(\nabla^2 G(\mu^* + tu) - H\right) u \, dt.$$

因此，余项由 $L_\alpha \|u\|^{1+\alpha}/(1+\alpha)$ 有界，所以假设 A3 中的界在 $r_\alpha = r$ 和 $C_\alpha = L_\alpha/(1+\alpha)$ 下成立。

Polyak 和 Juditsky（1992年，假设 3.2）采用了带有 $\|u\|^{1+\alpha}$ 阶余项的局部线性化。对于 $\alpha = 1$, 假设 A3 中的余项界是 Godichon-Baggioni 和 Werge（2025年，假设 5）中的局部二次余项条件。

2.3 局部展开的充分条件

以下命题给出了一个可直接根据密度及其一阶导数检验的假设 A3 充分条件。

命题 2.1。 假设假设 *A2* 成立，且密度为 $f \in C^1\left(\mathbb{R}^d\right)$。对于假设 *A2* 中的指数 ν，假设存在有限常数 $C_\nu > 0$，使得对于每个 $x \in \mathbb{R}^d$,

$$\|\nabla f(x)\| \le \frac{C_\nu}{(1+\|x\|)^{d+\nu}}.$$

那么，对于每个 $1 \le \ell \le d$，函数 G 二次连续可微，其海森矩阵是 *Lipschitz* 的，并且在 $\mathbb{R}^d$ 的每一点都正定。特别地，假设 *A3* 以 $\alpha = 1$ 成立。

证明见附录F.2。导数界用于通过对密度进行平移来控制总体海森矩阵的变化。

这些条件允许存在多项式尾部。非退化高斯密度及其有限混合对每个 $\nu > 2$ 都满足命题 2.1 的假设。具有紧支撑的 C^1 密度同样满足这些条件。对于具有 $\nu > 2$ 自由度的椭圆 Student密度，在 ν 取该值时，相关假设成立。这些断言将在附录F.2末尾得到验证。

2.4 当 $\ell < d$ 时的其他假设

对于 $p = 1$，Dürre 和 Paindaveine（2022a，定理2.5）在满足矩条件和几何条件的情况下，建立了 $1 \le \ell < d$ 离线估计量的渐近分布，而无需假设存在密度。对于具有 $1 \le \ell < d$ 的递归估计量，假设 A2 可由以下条件替代，同时需另行施加假设 A1 和 A3。

假设 A4。 分布 P 是（$\ell - 1$）-光滑的，这意味着它对每个维数为 $\ell - 1$ 的仿射子空间赋予的概率均为零，并且存在 $q > 2$ 使得 $\mathbb{E}\left[\|X_1\|^q\right] < \infty$。

假设 A5。 对于 $2 \le \ell < d$，

$$\sup_{u \in \mathbb{R}^d} \mathbb{E}\left[\frac{m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{X}\right)\right) \prod_{i=1}^{\ell}\left(1 + \|X_i\|\right)}{\ell \left\|\Gamma\left(\mathbf{X}\right)\left(u - X_\ell\right)\right\|}\right] < \infty.$$

假设 A4 中的（$\ell - 1$）-光滑性条件源自 Dürre 和 Paindaveine（2022a，第2.3节）。对于 $p = 1$，他们的定理2.5要求具有有限二阶矩，而假设 A4 要求具有 $q > 2$ 阶矩。对于 $2 \le \ell < d$，假设 A5 强化了他们的逆距离条件（2.6）；该条件仅在 $u = \mu^*$ 处施加，且不包含因子$\prod_{i=1}^{\ell}(1 + \|X_i\|)$。此处，额外因子以及关于 u 的一致界用于控制在依赖于重叠观测值的迭代值处计算的梯度之差；参见（A.10）。

在假设 A2 下，假设 A5 中的界可由引理 C.2 中的 (C.7) 得出。当$\ell = 1$时，各个连续窗口相互独立，因此命题 3.3 中不需要假设 A5。

3 递归估计量与渐近结果

3.1 滑动窗口递推

设 μ_0 为取值于 $\mathbb{R}^d$ 的随机向量，设c_γ 为严格为正的随机变量。假设 (μ_0, c_γ) 与 $(X_i)_{i \ge 1}$ 独立，且 $\mathbb{E}\left[\|\mu_0\|^2 + c_\gamma^2\right] < \infty$。对于 $0 \le j < \ell$，令$\mu_j := \mu_0$。对于 $n \ge \ell$，定义

$$\mu_n := \mu_{n-1} - \gamma_n \nabla_\mu g\left(X_{n-\ell+1}, \ldots, X_n; \mu_{n-1}\right), \tag{3.1}$$

其中步长由 $\gamma_n := c_\gamma n^{-\gamma}$ 给出，指数为固定指数 $1/2 < \gamma < 1$。每一步都会有一个观测值进入窗口，同时移除最早的观测值。对于每个 $n \ge \ell$，平均估计量定义为

$$\bar{\mu}_n := \frac{1}{n - \ell + 1} \sum_{k=\ell}^{n} \mu_k, \tag{3.2}$$

并且对于每个 $n \geq \ell$，都可以递归计算为

$$\bar{\mu}_{n+1} = \bar{\mu}_n + \frac{\mu_{n+1} - \bar{\mu}_n}{n - \ell + 2}, \quad \text{and} \quad \bar{\mu}_\ell = \mu_\ell.$$

对于 $\ell \geq 2$ 和 $n > \ell$，观测值 $X_{n-\ell+1}, \ldots, X_{n-1}$ 已用于计算μ_{n-1}。即使 G 可微，条件期望

$$\mathbb{E}\left[\nabla_\mu g\left(X_{n-\ell+1}, \ldots, X_n; \mu_{n-1}\right) \mid \mu_0, c_\gamma, X_1, \ldots, X_{n-1}\right]$$

也未必等于 $\nabla G(\mu_{n-1})$。在证明中，我们将 μ_{n-1} 与 $\mu_{n-\ell}$进行比较，其中 $\mu_{n-\ell}$与 $(X_{n-\ell+1}, \ldots, X_n)$ 独立。由此产生的误差由 (A.10) 和 (A.14) 控制。

3.2 收敛、收敛速率与极限分布

我们首先陈述几乎必然收敛的结果。该结果仅用到假设 A1；特别地，不需要密度假设。

定理 3.1. 在假设 *A1* 下，递推 (3.1) 满足

$$\sup_n \mathbb{E}\left[\|\mu_n - \mu^*\|^2\right] < \infty \qquad \textit{and} \qquad \mu_n \xrightarrow[n\to\infty]{} \mu^* \quad \textit{a.s.}$$

附录 A.1 中的证明利用 $g(x;\cdot)$ 的凸性，在证明几乎必然收敛之前建立一致二阶矩界。

接下来我们给出这两个估计量的几乎处处收敛速率和渐近分布。

定理 3.2。 假设假设 *A2* 和 *A3* 成立，且 $\gamma\nu > 2$。令 $H := \nabla^2 G(\mu^*)$。对于 $j \in \{0, \ldots, \ell - 1\}$，定义

$$\Gamma_j := \mathbb{E}\left[\nabla_\mu g\left(X_1, \ldots, X_\ell; \mu^*\right) \nabla_\mu g\left(X_{1+j}, \ldots, X_{\ell+j}; \mu^*\right)^\top\right].$$

我们还定义

$$\Sigma := \Gamma_0 + \sum_{j=1}^{\ell-1} \left(\Gamma_j + \Gamma_j^\top\right) \quad \textit{and} \quad V_H := \int_0^\infty e^{-tH} \Sigma e^{-tH}\, dt. \tag{3.3}$$

Then

$$\|\mu_n - \mu^*\|^2 = O\left(n^{-\gamma} \log n\right) \quad \textit{a.s.} \quad \textit{and} \quad \gamma_n^{-1/2}\left(\mu_n - \mu^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d\left(0, V_H\right).$$

此外，如果 $\gamma(1+\alpha) > 1$，则

$$\|\bar{\mu}_n - \mu^*\|^2 = O\left(\frac{\log\log n}{n}\right) \quad \textit{a.s.} \quad \textit{and} \quad \sqrt{n}\left(\bar{\mu}_n - \mu^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d\left(0, H^{-1}\Sigma H^{-1}\right).$$

对 γ 的限制分别有其作用。界 $\gamma > 1/2$ 保证步长平方具有可求和性，并使累积的条件偏差除以 $\sqrt{n}$ 后可以忽略。条件 $\gamma < 1$ 控制分部求和产生的边界项。矩限制 $\gamma\nu > 2$ 允许在几乎处处收敛速率的证明中选择 $2/\gamma < q < \nu$。最后，$\gamma(1+\alpha) > 1$ 使累积的泰勒余项在相同尺度下可以忽略。

矩阵 V_H 是 $HV_H + V_H H = \Sigma$ 的唯一对称解。在归一化 $n^{\gamma/2}$ 下，未平均迭代量在给定 c_γ 条件下的极限分布为 $\mathcal{N}_d(0,\ c_\gamma V_H)$。$\sqrt{n}\,(\bar{\mu}_n - \mu^*)$ 的极限协方差 $H^{-1}\Sigma H^{-1}$ 在给定 c_γ 条件下和无条件下相同，并且不依赖于步长中的乘法因子 c_γ。

对于 μ_n 的几乎必然收敛速率，假设 A3 仅通过 G 在 μ^* 处的二阶可微性以及其海森矩阵的正定性而被使用；证明见附录 A.2。

备注 （仿射变换）。 假设 $\ell = d$ ，并且定理 3.2 中关于 $\bar{\mu}_n$ 的假设成立。令 A 为一个可逆 $d \times d$ 矩阵，令 $b \in \mathbb{R}^d$，并设 $c \coloneqq |\det A|$。单纯形体积恒等式给出

$$g\left(Ax_1 + b, \ldots, Ax_d + b; A\mu + b\right) = c\, g\left(x_1, \ldots, x_d; \mu\right).$$

根据链式法则和零次梯度约定，变换后各点处所选取的梯度等于 $c\left(A^{-1}\right)^{\top} \nabla_{\mu} g(\mathbf{x}; \mu)$。由于 $c\left(A^{-1}\right)^{\top}$ 通常不等于 A，采用相同步长的递推 (3.1) 不具有仿射等变性。

对于变换后的分布，总体最小化元为 $A\mu^* + b$，矩阵 H 和 Σ 分别变为

$$c\left(A^{-1}\right)^{\top} H A^{-1} \quad \text{and} \quad c^2\left(A^{-1}\right)^{\top} \Sigma A^{-1}.$$

通过变量代换和范数等价性，假设 A2 和 A3 在指数 ν 和 α不变的情况下仍然成立。因此，定理 3.2 适用于基于 $(AX_i + b)_{i \geq 1}$ 的递推，该递推以 $A\mu_0 + b$ 为初值并采用相同步长。其平均值以 $A\mu^* + b$ 为中心，在 $\sqrt{n}$ 尺度下渐近服从正态分布，极限协方差为

$$A\left(H^{-1}\Sigma H^{-1}\right)A^{\top}.$$

下一个命题给出了在关于 $\ell < d$ 的另一组假设下的收敛速率和渐近分布。

命题 3.3。 假设 $1 \leq \ell < d$ ，且假设 *A1*、*A3* 和 *A4* 成立，其中$\gamma q > 2$， q 为假设 *A4* 中的矩指数。若 $\ell \geq 2$，则还假设假设 *A5* 成立。于是，定理 *3.2* 中陈述的 μ_n 的几乎处处速率和渐近正态性仍然成立。此外，若 $\gamma(1+\alpha) > 1$，则该定理中关于 $\bar{\mu}_n$ 的结论也成立。

3.3 与离线估计量和不相交分块估计量的比较

我们将式 (3.2) 中定义的平均滑动窗口估计量 $\bar{\mu}_n$ 与离线估计量 $\widehat{\mu}_n$（即式 (2.3) 中 G_n 的一个可测最小化元），以及一种基于大小为 ℓ的不相交区块的平均随机梯度估计量进行比较。我们在 $\sqrt{n}$ 尺度下比较协方差，其中 n 为观测值的数量。令

$$\zeta_1 \coloneqq \operatorname{Var}\left(\mathbb{E}\left[\nabla_{\mu} g\left(X_1, \ldots, X_{\ell}; \mu^*\right) \mid X_1\right]\right).$$

对于 $p = 1$ 和 $\ell < d$，Dürre 和 Paindaveine（2022a，定理2.5）证明了离线估计量的渐近正态性，其所需的矩条件和几何条件不要求存在密度。下面的定理在递推所采用的假设下证明了该估计量的渐近正态性，其中包括 Oja 情形 $\ell = d$。

定理 3.4。 假设假设 *A3* 成立。进一步假设假设 *A2* 成立，或者 $\ell < d$ 和假设 *A1* 与 *A4* 成立。则离线估计量 $\widehat{\mu}_n$（定义为 G_n 在 (2.3) 中的一个$\widehat{\ }$可测最小化元）满足

$$\sqrt{n}\left(\widehat{\mu}_n - \mu^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d\left(0, \ell^2 H^{-1}\zeta_1 H^{-1}\right). \tag{3.4}$$

为便于比较，我们设定 $\mu_0^{\mathrm{SGD}} \coloneqq \mu_0$ ，并对 $j \geq 1$ 定义

$$\mu_j^{\mathrm{SGD}} \coloneqq \mu_{j-1}^{\mathrm{SGD}} - \gamma_j \nabla_{\mu} g\left(X_{(j-1)\ell+1}, \ldots, X_{j\ell}; \mu_{j-1}^{\mathrm{SGD}}\right),$$

其中，$\gamma_j \coloneqq c_{\gamma} j^{-\gamma}$ 按区块数量编索引，具有与式 (3.1) 中相同的 c_{γ} 和固定指数 γ。经过 $n \geq \ell$ 个观测值后，该递推已执行 $\lfloor n/\ell \rfloor$ 次更新，其平均估计量为

$$\bar{\mu}_n^{\mathrm{ASGD}} \coloneqq \frac{1}{\lfloor n/\ell \rfloor} \sum_{j=1}^{\lfloor n/\ell \rfloor} \mu_j^{\mathrm{SGD}}.$$

在假设 A2 和 A3 下，对于 $\gamma\nu > 2$ 和 $\gamma\,(1+\alpha) > 1$，区块的独立性使得给定先前区块以及 (μ_0, c_γ) 时，每个梯度都是条件无偏的。因此，定理 3.2 证明中使用的平均化论证表明，在 $\sqrt{\lfloor n/\ell \rfloor}$ 尺度下的极限协方差为 $H^{-1}\Gamma_0 H^{-1}$。由于 $n/\lfloor n/\ell \rfloor \to \ell$，可得

$$\sqrt{n}\left(\bar{\mu}_n^{\mathrm{ASGD}} - \mu^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d\left(0, \ell H^{-1}\Gamma_0 H^{-1}\right).$$

在命题 3.3 中关于 $\bar{\mu}_n$ 的假设下，通过对独立区块进行平均化的论证，可得相同的极限协方差。在相同尺度下，定理 3.4 和 3.2 分别给出 $\widehat{\mu}_n$ 和 $\bar{\mu}_n$ 的协方差 $\ell^2 H^{-1}\zeta_1 H^{-1}$ 和 $H^{-1}\Sigma H^{-1}$。它们的比较见下一个引理。

引理 3.5。在假设 *A1* 和 *A3* 下，

$$\ell^2\zeta_1 \preceq \Sigma \preceq \ell\Gamma_0,$$

其中，$A \preceq B$ 表示 $B - A$ 为半正定；当 $B - A$ 为正定时，我们记作 $A \prec B$。此外，若假设 *A2* 成立且 $\ell \geq 2$，则两个不等式都严格成立。

证明见附录 B.2。令 V_{offline}、V_{SW} 和 V_{ASGD} 分别表示在共同的 $\sqrt{n}$ 归一化下，$\widehat{\mu}_n$、$\bar{\mu}_n$ 和 $\bar{\mu}_n^{\mathrm{ASGD}}$ 的渐近协方差矩阵。因此，

$$\underbrace{\ell^2 H^{-1}\zeta_1 H^{-1}}_{V_{\mathrm{offline}}} \preceq \underbrace{H^{-1}\Sigma H^{-1}}_{V_{\mathrm{SW}}} \preceq \underbrace{\ell H^{-1}\Gamma_0 H^{-1}}_{V_{\mathrm{ASGD}}}.$$

因此，平均滑动窗口估计量在计算成本和渐近方差之间取得了折中。与不相交分块估计量一样，在 d 和 ℓ 固定时，其计算成本随 n 线性增长，且存储空间不依赖于 n；参见第 3.4 节。在假设 A2 下，对于 $\ell \geq 2$，重用重叠窗口中的观测值可使每个方向上的渐近方差都严格小于不相交分块估计量的渐近方差，且无需最小化离线经验目标函数。

备注。对于 $\ell = 1$，滑动窗口递推和不相交分块递推分别给出 $\mu_n = \mu_n^{\mathrm{SGD}}$ 和 $\bar{\mu}_n = \bar{\mu}_n^{\mathrm{ASGD}}$，对每个 $n \geq 1$ 均成立。由于 $\zeta_1 = \Gamma_0 = \Sigma$，这些平均估计量的极限协方差与经验空间中位数相同。

对于 $\ell \geq 2$，由上述公式定义的矩阵 V_{offline}、V_{SW} 和 V_{ASGD} 在假设 A1 和 A3 下两两不同。更确切地说，对于满足 $v^\top V_{\mathrm{ASGD}} v > 0$ 的每个 $v \in \mathbb{R}^d$，

$$v^\top V_{\mathrm{offline}} v < v^\top V_{\mathrm{SW}} v < v^\top V_{\mathrm{ASGD}} v.$$

此外，它们的差满足

$$\frac{1}{\ell}\left(V_{\mathrm{ASGD}} - V_{\mathrm{offline}}\right) \preceq V_{\mathrm{SW}} - V_{\mathrm{offline}} \preceq \frac{2\ell - 1}{3\ell}\left(V_{\mathrm{ASGD}} - V_{\mathrm{offline}}\right).$$

特别地，对于 $\ell = 2$、$V_{\mathrm{SW}} = \frac{1}{2}\left(V_{\mathrm{offline}} + V_{\mathrm{ASGD}}\right)$。本备注中的断言已在引理 3.5 的证明中得到证实。

我们还考虑一种递推：对每个不相交的数据块进行固定次数 $r \geq 1$ 的连续更新，在每次更新后重新计算梯度，并对所有迭代值取平均。Agarwal 等（2020）也研究了在同一数据块上重复进行梯度更新的情形，并为光滑凸目标函数建立了超额风险界。在定理 3.2 中平均估计量所满足的假设下，根据命题 B.1，其极限协方差仍为 V_{ASGD}。取 $r = \ell$ 时，需要进行 $\ell\lfloor n/\ell \rfloor$ 次梯度计算；相比之下，滑动窗口需要 $n - \ell + 1$ 次；这两个数渐近等价于 n。因此，$\ell \geq 2$ 的严格改进源于使用不同的元组，而不仅仅是在相同元组上重复更新。

3.4 计算成本

现在我们描述滑动窗口递推（3.1）的存储需求和算术成本。在 $n \geq \ell$ 个观测值之后，它已进行 $n-\ell+1$ 次梯度计算。存储当前窗口需要 $O(d\ell)$ 的内存。当前迭代值及其平均值还需要 $O(d)$ 的额外内存，而平均化每次更新的成本为 $O(d)$次操作。

对于 $2 \leq \ell \leq d$，梯度可根据定义基单纯形的 $\ell - 1$ 差分向量的格拉姆矩阵计算。只要窗口中的观测值仿射无关，该矩阵就是非奇异的；在假设 A2 下，这几乎处处成立；当 $\ell < d$ 时，在假设 A4 下也成立。替换一个观测值会使格拉姆矩阵发生秩至多为二的矩阵变化。经过初始计算（耗费 $O(d\ell^2)$）后，更新其逆矩阵和行列式，每次更新需要 $O(d\ell)$ 次运算。因此，总成本为 $O(nd\ell)$，存储空间为 $O(d\ell)$。对于 $\ell = 1$，每次梯度计算需要 $O(d)$ 次运算，因此具有相同的界。相关公式和运算计数见附录 G。

前一小节定义的不相交分块递推执行 $\lfloor n/\ell \rfloor$ 次更新。分别计算每个分块需要 $O(d\ell^2)$ 次运算，因此总成本为 $O(nd\ell)$。由此可见，两种递推的计算成本阶数相同，而滑动窗口递推会在每个新观测值到来后更新估计量。对于 Oja中位数，$\ell = d$，这些界给出 $O(nd^2)$ 次运算和 $O(d^2)$ 的存储空间。

相比之下，离线估计量的经验目标函数（2.3）包含 $\binom{n}{\ell}$ 个单纯形体积。对于该函数的一次评估，若从其 Gram 矩阵分别计算这些体积，则需要 $O(d\ell^2\binom{n}{\ell})$ 次运算，而数值最小化通常需要重复进行评估。对于 Oja中位数，项数为 $\binom{n}{d}$，因此即使样本量和维度适中，直接枚举也可能变得代价高昂。不完全离线估计量通过最小化基于随机选取的元组集合的目标函数，减少了项数（Dürre 和 Paindaveine，2022a，第2节 2.4）。第4节中的模拟采用了这一不完全版本。

4 模拟研究

本节评估所提出的滑动窗口估计量的有限样本性能和计算效率。尽管第3节的框架适用于所有 $1 \leq \ell \leq d$，但实验聚焦于 Oja中位数，对应于 $\ell = d$。我们考虑关于 $\mu^* = 0$ 球对称的分布。模拟使用 Python 实现，并在计时前使用 Numba 编译主要数值例程。该实现更新观测值矩阵的行列式，并使用 Sherman–Morrison公式更新其逆矩阵；参见附录 G。

4.1 在线收敛和平均化

我们比较基于不相交分块的随机梯度估计量（随机梯度下降）和滑动窗口估计量（滑动窗口估计量），以及它们的平均版本，分别记为平均随机梯度下降和 A-SW。随机梯度下降在每处理完 d 个观测值后更新一次，而滑动窗口估计量从第 d 个观测值开始，在每个新观测值之后更新一次。

对于每个 $d \in \{3、10、50\}$，我们从 $\mathcal{N}_d(0、 I_d)$ 中生成 20 个包含 10^7 个观测值的独立样本。所有估计量均初始化为全 1 向量，并且每个样本只进行一次遍历。滑动窗口算法在观测值 X_n 处使用步长 $c_\gamma n^{-\gamma}$，持续 $n \geq d$，其中 $\gamma = 0.75$ 和 $c_\gamma := \sqrt{d!}$ 。选择 c_γ 可得到 $c_\gamma^2 \Gamma_0 = I_d$，因为在 $\mu^* = 0$ 处的梯度具有协方差 $\Gamma_0 = I_d/d!$ ；参见附录 H。不相交区块算法使用相同的步长序列，并按更新次数进行索引。

图 1 展示了根据这些样本估计的均方误差。在 10^7 个观测值之后，对于 d 的三个取值，A-SW 的误差均为最小。平均化改善了最终

SW 的误差。在不相交分块的情况下，当 $d = 3$ 和 $d = 10$ 时，它改善了最终误差；而当 $d = 50$ 时，SGD 和平均随机梯度下降的误差相近。对于 $d = 50$，不相交区块算法执行 $2 \cdot 10^5$ 次更新；早期迭代值的影响可能仍会限制平均化带来的益处。

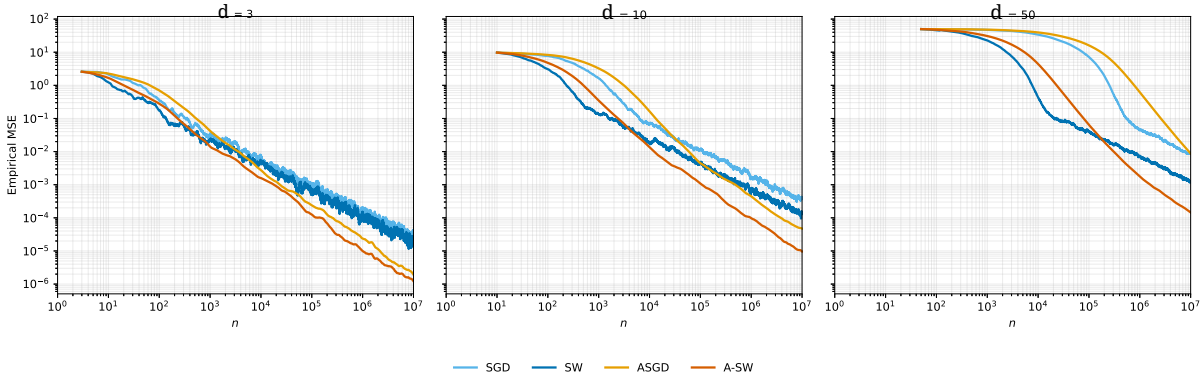


图 1：基于 20 个独立样本估计的均方误差，作为观测值数量 n 的函数，其中 $d \in \{3、10、50\}$。

4.2 与不完全离线估计量的比较

我们在包含 2000年 个观测值的数据集上，将滑动窗口估计量及其平均版本与不完全离线估计量进行比较。与 Dürre 和 Paindaveine（2022a）的表1一样，我们考虑维度 $d \in \{2、3、4、6、10\}$ 以及以下分布：标准高斯分布、自由度为 $\nu \in \{3、6、9、12\}$ 的标准多元 Student 分布，以及形状参数为 $\eta \in \{2、3、4、6\}$ 且密度与 $\exp\left(-\|x\|^{2\eta}/2\right)$ 成正比的幂指数分布。(2.3) 中的完整求和包含 $\binom{2000}{d}$ 项。按照 Dürre 和 Paindaveine（2022a）第2节 2.4 节中的不完全构造，我们将该求和替换为由$M := 4000$ 个项构成的和，这些项均匀地进行有放回抽取，并在整个优化过程中保持所选项不变。记相应的元组为 $\mathbf{X}^{(s)}$、$1 \le s \le M$，我们最小化

$$\sum_{s=1}^{M} g\left(\mathbf{X}^{(s)}; u\right)$$

关于 $u \in \mathbb{R}^d$。每一项都是关于 u 的仿射函数的绝对值。我们预先计算这些函数的系数，因此联合计算目标函数及其梯度的代价为 $O(Md)$ 次运算。

对于滑动窗口估计量及其平均版本，我们采用步长 $c_\gamma n^{-\gamma}$，其中$\gamma = 0.75$，并使用由前 100 个观测值组成的预备样本来选择初始值 μ_0 和常数 c_γ 。初始值 μ_0 是该样本的逐坐标边际中位数向量。我们从该样本中 $\binom{100}{d}$个由 d 个不同观测值组成的可能元组中有放回地均匀抽取 50 个元组，对每个元组在 μ_0 处计算梯度，并将 c_γ 设为梯度范数平均值的逆矩阵。随后，算法处理其余 1900 个观测值；这些观测值与预备样本相互独立，索引从 $n = 1$开始。

优化采用 BFGS，并使用满足 Armijo 条件和弱 Wolfe 条件的线搜索，其常数分别为 $c_1 = 10^{-4}$ 和 $c_2 = 0.9$。初始值设为完整数据集中各坐标边际中位数构成的向量。停止规则使用根据所选元组通过对偶性得到的目标函数值误差数值上界。将目标函数除以预先计算的线性系数的范数之和后，绝对容差为 10^{-10}

。所有 4500 个最终估计量均满足此容差。所有配置均使用相同的求解器设置。

表1报告了由100个独立样本估计得到的均方误差。不完整离线计算在各表中标记为“批次”。比较两种递归估计量，平均化在45种配置中的35种给出了更小的误差。其效果并不一致：例如，在维度 $d = 3$下、自由度为3的Student分布中，滑动窗口估计量的误差小于其平均版本。由于该递推执行的更新少于1900次，早期迭代值的贡献可能仍会限制平均化的收益。

表2报告了平均总计算时间。滑动窗口估计量及其平均版本是同时计算的，因此将二者共享的执行时间报告为“(A)-SW”。这些时间包括递归估计量的初始化和校准，以及离线估计量的初始化、元组采样、预计算和优化。对于每个维度 d，由于$\ell = d$，递归估计量在所有分布上执行相同次数的更新，且平均计算时间相近。

		$\nu = 3$	$\nu = 6$	$\nu = 9$	$\nu = 12$	$\mathcal{N}$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta = 6$
$d = 2$	批次	$3.4 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$1.2 \cdot 10^{-3}$	$9.7 \cdot 10^{-4}$	$7.9 \cdot 10^{-4}$
	SW	$7.7 \cdot 10^{-3}$	$7.0 \cdot 10^{-3}$	$5.8 \cdot 10^{-3}$	$6.6 \cdot 10^{-3}$	$6.3 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	$3.2 \cdot 10^{-3}$	$3.7 \cdot 10^{-3}$	$3.1 \cdot 10^{-3}$
	A-SW	$6.6 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	$3.6 \cdot 10^{-3}$	$4.4 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$
$d = 3$	批次	$6.4 \cdot 10^{-3}$	$4.5 \cdot 10^{-3}$	$3.7 \cdot 10^{-3}$	$3.9 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$
	SW	$1.9 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$	$1.2 \cdot 10^{-2}$	$9.5 \cdot 10^{-3}$	$9.7 \cdot 10^{-3}$	$6.8 \cdot 10^{-3}$	$5.8 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$	$4.4 \cdot 10^{-3}$
	A-SW	$2.1 \cdot 10^{-2}$	$9.6 \cdot 10^{-3}$	$9.0 \cdot 10^{-3}$	$7.8 \cdot 10^{-3}$	$7.2 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$
$d = 4$	批次	$1.3 \cdot 10^{-2}$	$6.9 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$	$5.7 \cdot 10^{-3}$	$5.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$
	SW	$4.7 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$7.9 \cdot 10^{-3}$	$5.5 \cdot 10^{-3}$	$5.4 \cdot 10^{-3}$	$6.6 \cdot 10^{-3}$
	A-SW	$5.5 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$2.3 \cdot 10^{-3}$
$d = 6$	批次	$4.3 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$7.6 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$1.7 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$1.1 \cdot 10^{-3}$
	SW	$6.4 \cdot 10^{0}$	$4.9 \cdot 10^{-2}$	$4.0 \cdot 10^{-2}$	$3.3 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$9.5 \cdot 10^{-3}$	$7.9 \cdot 10^{-3}$	$7.1 \cdot 10^{-3}$
	A-SW	$7.4 \cdot 10^{0}$	$5.0 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	$6.2 \cdot 10^{-3}$	$4.6 \cdot 10^{-3}$	$3.2 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$
$d = 10$	批次	$1.6 \cdot 10^{-1}$	$4.0 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$3.2 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$
	SW	$2.4 \cdot 10^{0}$	$1.4 \cdot 10^{-1}$	$1.2 \cdot 10^{-1}$	$1.8 \cdot 10^{-1}$	$5.5 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
	A-SW	$3.9 \cdot 10^{0}$	$1.6 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$2.9 \cdot 10^{-1}$	$6.3 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$	$7.2 \cdot 10^{-3}$	$5.7 \cdot 10^{-3}$	$5.0 \cdot 10^{-3}$

表1：根据100个独立样本，对所考虑的维度和分布，分别计算不完整离线计算（“批次”）、滑动窗口估计量和A-SW所得的均方误差。

		$\nu = 3$	$\nu = 6$	$\nu = 9$	$\nu = 12$	$\mathcal{N}$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta = 6$
$d = 2$	批次	$4.9 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$4.8 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$4.8 \cdot 10^{-2}$	$4.8 \cdot 10^{-2}$	$4.8 \cdot 10^{-2}$	$4.9 \cdot 10^{-2}$	$5.3 \cdot 10^{-2}$
	(A)-SW	$2.9 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$
$d = 3$	批次	$6.1 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$	$6.1 \cdot 10^{-2}$	$6.1 \cdot 10^{-2}$	$6.1 \cdot 10^{-2}$	$5.9 \cdot 10^{-2}$	$6.0 \cdot 10^{-2}$	$5.8 \cdot 10^{-2}$	$5.8 \cdot 10^{-2}$
	(A)-SW	$3.0 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$3.1 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$
$d = 4$	批次	$7.2 \cdot 10^{-2}$	$6.9 \cdot 10^{-2}$	$6.9 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$6.6 \cdot 10^{-2}$	$6.7 \cdot 10^{-2}$	$6.6 \cdot 10^{-2}$	$6.6 \cdot 10^{-2}$
	(A)-SW	$2.9 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$3.1 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$	$3.1 \cdot 10^{-3}$	$3.1 \cdot 10^{-3}$
$d = 6$	批次	$1.0 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$	$9.6 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$	$9.4 \cdot 10^{-2}$	$9.5 \cdot 10^{-2}$
	(A)-SW	$3.3 \cdot 10^{-3}$	$3.3 \cdot 10^{-3}$	$3.3 \cdot 10^{-3}$	$3.3 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$3.5 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$
$d = 10$	批次	$2.2 \cdot 10^{-1}$	$2.2 \cdot 10^{-1}$	$2.0 \cdot 10^{-1}$	$2.0 \cdot 10^{-1}$	$2.1 \cdot 10^{-1}$	$1.9 \cdot 10^{-1}$	$1.7 \cdot 10^{-1}$	$1.7 \cdot 10^{-1}$	$1.8 \cdot 10^{-1}$
	(A)-SW	$4.2 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$4.1 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$4.1 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$4.1 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$

表2：不完整离线计算（“批次”）以及滑动窗口估计量和 A-SW 的联合计算在 100 次独立重复试验中的平均总计算时间，单位为秒。

结论

我们研究了用于单纯形中位数的滑动窗口随机梯度估计量，对应于 $p = 1$ 和 $1 \leq \ell \leq d$。每当有新的观测值可用时，所提出的方法便更新估计量，每次更新需要 $O(d\ell)$次运算。与朴素的不相交分块随机梯度算法和离线估计量相比，该方法在估计精度与计算成本之间实现了较好的权衡，这一点得到了理论结果和数值实验的支持。

仍有若干问题尚待解决。首要方向之一是在更弱的分布假设下，建立 Oja中位数的收敛速率和极限定理。将我们的分析扩展到 $p > 1$ 情形也很有意义。最后，开发平均滑动窗口估计量的渐近协方差矩阵的递归估计量，将有助于构造渐近置信区间和置信椭球（Godichon-Baggioni，2019；Godichon-Baggioni 和 Lu，2024）。

致谢

作者谨向 Davy Paindaveine 致以诚挚的谢意，感谢他富有洞见且富有建设性的讨论，这些讨论启发了本研究。

参考文献

Agarwal, N.、Anil, R.、Koren, T.、Talwar, K. 和 Zhang, C.（2020年）。具有滞后数据管道的随机优化。载于《神经信息处理系统进展》第33卷，10282–10293。Andersen, P. K. 和 Gill, R. D.（1982年）。计数过程的Cox回归模型：大样本研究。《统计年鉴》10，1100–1120。doi:10.1214/aos/1176345976。Benveniste, A.、Métivier, M. 和 Priouret, P.（1990年）。《自适应算法与随机逼近》。数学应用22。Springer-Verlag，柏林。doi:10.1007/978-3-642-75894-2。Brown, B. M.（1983年）。空间中位数的统计应用。《皇家统计学会杂志》，B辑45，25–30。doi:10.1111/j.2517-6161.1983.tb01226.x。Cardot, H.、Cénac, P. 和 Godichon-Baggioni, A.（2017年）。希尔伯特空间中几何中位数的在线估计：非渐近置信球。《统计年鉴》45，591–614。doi:10.1214/16-AOS1460。Cardot, H.、Cénac, P. 和 Zitt, P.-A.（2013年）。利用平均随机梯度算法在希尔伯特空间中高效、快速地估计几何中位数。《伯努利》19，18–43。doi:10.3150/11-BEJ390。Duflo, M.（1997年）。《随机迭代模型》。数学应用34。Springer-Verlag，柏林。S. S. Wilson译。doi:10.1007/978-3-662-12880-0。Dürre, A. 和 Paindaveine, D.（2022a）。基于 L_p损失函数的多元位置仿射等变推断。《统计年鉴》50，2616–2640。doi:10.1214/22-AOS2199。Dürre, A. 和 Paindaveine, D.（2022b）。《基于 L_p 损失函数的多元位置仿射等变推断》的补充材料。doi:10.1214/22-AOS2199SUPP。Fischer, D.、Mosler, K.、Möttönen, J.、Nordhausen, K.、Pokotylo, O. 和 Vogel, D.（2020年）。在R中计算Oja中位数：OjaNP软件包。《统计软件杂志》92(8)，1–36。doi:10.18637/jss.v092.i08。Godichon-Baggioni, A.（2019年）。平均随机梯度算法中渐近方差的在线估计。《统计规划与推断杂志》203，1–19。doi:10.1016/j.jspi.2019.01.001。Godichon-Baggioni, A. 和 Lu, W.（2024年）。用于估计几何中位数的在线随机牛顿方法及其应用。《多元分析杂志》202，105313。doi:10.1016/j.jmva.2024.105313。Godichon-Baggioni, A. 和 Werge, N.（2025年）。关于流数据的自适应随机优化：一种具有 $O(dN)$ 运算的牛顿方法。《机器学习研究杂志》26(59)，1–49。https://www.jmlr.org/papers/v26/23-1565.html。

Haldane，J. B. S.（1948年）。关于多元分布的中位数的注记。Biometrika 35，414–417。doi:10.1093/biomet/35.3-4.414。Hoeffding，W.（1948年）。具有渐近正态分布的一类统计量。《数学统计年鉴》19，293–325。doi:10.1214/aoms/1177730196。Hoeffding，W.（1961年）。U统计量的大数强律。统计学研究所Mimeo系列第302号。北卡罗来纳大学，教堂山。https://repository.lib.ncsu.edu/handle/1840.4/2128。Kemperman，J. H. B.（1987年）。Banach空间上有限测度的中位数。载于《基于 L_1-范数及相关方法的统计数据分析》（Y. Dodge编），217–230。North-Holland，阿姆斯特丹。Kushner，H. J. 和 Yang，J.（1993年）。通过迭代平均进行随机逼近：一般过程的最优渐近收敛速率。SIAM控制与优化杂志31，1045–1062。doi:10.1137/0331047。Kushner，H. J. 和 Yin，G. G.（2003年）。《随机逼近与递归算法及其应用》第二版。数学应用35。Springer-Verlag，纽约。doi:10.1007/b97441。Miles，R. E.（1971年）。各向同性随机单纯形。《应用概率学进展》3，353–382。doi:10.2307/1426176。Oja，H.（1983年）。多元分布的描述性统计。《统计学与概率通讯》1，327–332。doi:10.1016/0167-7152(83)90054-8。Oja，H. 和 Niinimaa，A.（1985年）。多元正态性情形下广义中位数的渐近性质。《皇家统计学会杂志》，B辑47，372–377。doi:10.1111/j.2517-6161.1985.tb01366.x。Pelletier，M.（2000年）。平均随机算法的渐近几乎处处有效性。SIAM控制与优化杂志39，49–72。doi:10.1137/S0363012998308169。Polyak，B. T. 和 Juditsky，A. B.（1992年）。通过平均化加速随机逼近。SIAM控制与优化杂志30，838–855。doi:10.1137/0330046。Robbins，H. 和 Monro，S.（1951年）。一种随机逼近方法。《数学统计年鉴》22，400–407。doi:10.1214/aoms/1177729586。Robbins，H. 和 Siegmund，D.（1971年）。关于非负几乎超鞅的收敛定理及其若干应用。载于《统计学中的优化方法》（J. S. Rustagi编），233–257。Academic Press，纽约。doi:10.1016/B978-0-12-604550-5.50015-8。Rockafellar，R. T. 和 Wets，R. J.-B.（1998年）。《变分分析》。数学科学基础丛书317。Springer-Verlag，柏林。doi:10.1007/978-3-642-02431-3。Rosén，B.（1967年）。关于相依随机变量和的中心极限定理。《概率论与相关领域》7，48–82。doi:10.1007/BF00532097。Ruppert，D.（1988年）。从慢收敛Robbins–Monro过程中得到的有效估计量。技术报告781，运筹学与工业工程学院，康奈尔大学，伊萨卡，纽约州。https://hdl.handle.net/1813/8664。Schneider，R. 和 Weil，W.（2008年）。《随机几何与积分几何》。概率及其应用。Springer-Verlag，柏林。doi:10.1007/978-3-540-78859-1。Sherman，J. 和 Morrison，W. J.（1950年）。给定矩阵的一个元素发生变化时相应逆矩阵的调整。《数学统计年鉴》21，124–127。doi:10.1214/aoms/1177729893。

滑动窗口估计量的证明

在整个附录 A 和 B.3 节中，对于 $k \geq 1$，我们记 $\mathcal{F}_0 := \sigma(\mu_0,\ c_\gamma)$ 和 $\mathcal{F}_k := \sigma(\mu_0,\ c_\gamma,\ X_1,\ \ldots\ldots,\ X_k)$。对于滑动窗口递推和 $k \geq \ell$，我们记 $Z_k := (X_{k-\ell+1},\ \ldots\ldots,\ X_k)$。在递归估计量的证明中，我们首先在给定$\mathcal{F}_0$的条件下进行讨论。由于二元组 $(\mu_0,\ c_\gamma)$ 与序列 $(X_i)_{i\geq 1}$ 相互独立，给定$\mathcal{F}_0$ 时，$\left(\mu_0, c_\gamma, (X_i)_{i\geq 1}\right)$ 的条件分布的一个版本为 $\delta_{(\mu_0,c_\gamma)} \otimes P^{\otimes\infty}$，其中 δ_z 表示在 z 处的狄拉克概率测度。

A.1 定理 3.1 的证明

对于任意 $\mathbf{x} \in \left(\mathbb{R}^d\right)^\ell$ 和 $u,\ v \in \mathbb{R}^d$，由凸性、(2.1) 和 (2.2) 可得

$$g(\mathbf{x}; v) \geq g(\mathbf{x}; u) + \langle \nabla_\mu g(\mathbf{x}; u), v - u \rangle, \qquad \|\nabla_\mu g(\mathbf{x}; u)\| \leq \frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell},$$
$$|g(\mathbf{x}; u) - g(\mathbf{x}; v)| \leq \frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell} \|u - v\|. \tag{A.1}$$

对于 $k \geq \ell$ 和 $u \in \mathbb{R}^d$，定义

$$\xi_k(u) \coloneqq g(Z_k; u) - G(u).$$

根据 (A.1)，对于每个 $k \geq \ell$ 和 $u,\ v \in \mathbb{R}^d$，

$$|\xi_k(u) - \xi_k(v)| \leq \frac{1}{\ell}\left(m_{\ell-1}(\mathrm{Simpl}(Z_k)) + \mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))\right]\right)\|u - v\| \quad \text{a.s.} \tag{A.2}$$

证明。令 $V_n \coloneqq \|\mu_n - \mu^*\|^2$。根据 (3.1) 和 (A.1)，通过归纳法可得，对于每个 $n \geq 0$、$\mathbb{E}[V_n \mid \mathcal{F}_0] < \infty$ 几乎处处成立。对于 $n \geq \ell - 1$，对连续更新求和可得

$$\|\mu_n - \mu_{n+1-\ell}\| \leq \sum_{j=\max(\ell, n+2-\ell)}^{n} \gamma_j \frac{m_{\ell-1}(\mathrm{Simpl}(Z_j))}{\ell}. \tag{A.3}$$

求和的下限是 $\max(\ell,\ n+2-\ell)$，因为前 ℓ 个迭代值等于 μ_0。对于 $\ell = 1$，求和为空，因此不存在重叠误差。由于序列 $(X_i)_{i \geq 1}$ 独立于 $\mathcal{F}_0$，且所有窗口具有相同的分布，条件柯西–施瓦茨不等式给出，对于每个 $j \geq \ell$，

$$\mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(Z_{n+1}))\, m_{\ell-1}(\mathrm{Simpl}(Z_j)) \mid \mathcal{F}_0\right] \leq \mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^2\right] \quad \text{a.s.},$$
$$\mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(Z_j)) \mid \mathcal{F}_0\right] = \mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))\right] \quad \text{a.s.}$$

(A.3) 中的和至多包含 $\ell - 1$ 项，并且对于所有充分大的 n，该和中的每个步长 γ_j 至多为 $2^\gamma \gamma_n$。这意味着存在一个有限常数 $C > 0$，使得对于所有充分大的 n，

$$\begin{aligned}
&\frac{1}{\ell}\mathbb{E}\left[\left(m_{\ell-1}(\mathrm{Simpl}(Z_{n+1})) + \mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))\right]\right)\|\mu_n - \mu_{n+1-\ell}\| \mid \mathcal{F}_0\right] \\
&\quad \leq \frac{\mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^2\right] + \left(\mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))\right]\right)^2}{\ell^2} \sum_{j=\max(\ell, n+2-\ell)}^{n} \gamma_j \\
&\quad \leq C\gamma_n \quad \text{a.s.}
\end{aligned} \tag{A.4}$$

在 $\mathbb{E}[V_n \mid \mathcal{F}_0]$ 上的一致界。（A.1）中的次梯度不等式给出

$$V_{n+1} \leq V_n - 2\gamma_{n+1}\left(g(Z_{n+1}; \mu_n) - g(Z_{n+1}; \mu^*)\right) + \gamma_{n+1}^2 \frac{m_{\ell-1}(\mathrm{Simpl}(Z_{n+1}))^2}{\ell^2}. \tag{A.5}$$

窗口 $Z_{n+1} = (X_{n+2-\ell},\ \ldots\ldots,\ X_{n+1})$ 与$\mathcal{F}_{n+1-\ell}$独立。因此

$$\mathbb{E}\left[g(Z_{n+1}; \mu_{n+1-\ell}) - g(Z_{n+1}; \mu^*) \mid \mathcal{F}_0\right] = \mathbb{E}\left[G(\mu_{n+1-\ell}) - G(\mu^*) \mid \mathcal{F}_0\right] \geq 0 \quad \text{a.s.}$$

由 (A.1) 及其非负性，对每个 $n \geq \ell - 1$，

$$\mathbb{E}\left[g(Z_{n+1}; \mu_n) - g(Z_{n+1}; \mu^*) \mid \mathcal{F}_0\right] \geq -\frac{1}{\ell}\mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(Z_{n+1}))\|\mu_n - \mu_{n+1-\ell}\| \mid \mathcal{F}_0\right] \quad \text{a.s.}$$

在 (A.5) 中取给定$\mathcal{F}_0$ 的条件期望，并利用此界、(A.3) 和前述柯西-施瓦茨界，可得

$$\mathbb{E}\left[V_{n+1} \mid \mathcal{F}_0\right] \leq \mathbb{E}\left[V_n \mid \mathcal{F}_0\right] + \frac{\mathbb{E}\left[m_{\ell-1}\left(\operatorname{Simpl}\left(\mathbf{X}\right)\right)^2\right]}{\ell^2}\left(\gamma_{n+1}^2 + 2\gamma_{n+1}\sum_{j=\max(\ell,n+2-\ell)}^{n}\gamma_j\right) \quad \text{a.s.}$$

利用 $2\gamma_{n+1}\gamma_j \leq \gamma_{n+1}^2 + \gamma_j^2$并对 n 求和可知，$\sup_n \mathbb{E}\left[V_n \mid \mathcal{F}_0\right] \leq V_0 + Kc_\gamma^2$ 几乎处处成立，其中常数 K 定义为

$$K \coloneqq \frac{2\ell-1}{\ell^2}\mathbb{E}\left[m_{\ell-1}\left(\operatorname{Simpl}\left(\mathbf{X}\right)\right)^2\right]\sum_{k=\ell}^{\infty}k^{-2\gamma} < \infty.$$

事实上，由二重和得到的各项中，每个平方步长至多出现 2 $(\ell-1)$ 次。

$g\ (Z_{n+1};\ \mu_n)\ -g\ (Z_{n+1};\ \mu^*)$ 的分解。对于 $n \geq \ell-1$，有

$$g\left(Z_{n+1};\mu_n\right) - g\left(Z_{n+1};\mu^*\right) = G\left(\mu_n\right) - G\left(\mu^*\right) + \varepsilon_{n+1} + \rho_{n+1},$$

其中

$$\begin{aligned}\varepsilon_{n+1} &\coloneqq \xi_{n+1}\left(\mu_{n+1-\ell}\right) - \xi_{n+1}\left(\mu^*\right),\\ \rho_{n+1} &\coloneqq \xi_{n+1}\left(\mu_n\right) - \xi_{n+1}\left(\mu_{n+1-\ell}\right).\end{aligned}$$

于是 $\mathbb{E}\left[\varepsilon_{n+1} \mid \mathcal{F}_{n+1-\ell}\right] = 0$ 几乎处处成立。根据（A.2）以及 Z_{n+1} 与 $\mathcal{F}_{n+1-\ell}$的独立性，存在有限常数 $C_\varepsilon > 0$ 使得

$$\mathbb{E}\left[\varepsilon_{n+1}^2 \mid \mathcal{F}_{n+1-\ell}\right] \leq C_\varepsilon V_{n+1-\ell} \quad \text{a.s.}$$

对 $\mathcal{F}_0$ 取条件期望，并利用 $\mathbb{E}\left[V_n \mid \mathcal{F}_0\right]$ 上的一致界，得到

$$\sup_{n\geq\ell-1}\mathbb{E}\left[\varepsilon_{n+1}^2 \mid \mathcal{F}_0\right] \leq C_\varepsilon\left(V_0 + Kc_\gamma^2\right) < \infty \quad \text{a.s.}$$

因此,

$$\sum_{n\geq\ell-1}\mathbb{E}\left[\left(\gamma_{n+1}\varepsilon_{n+1}\right)^2 \mid \mathcal{F}_0\right] \leq C_\varepsilon c_\gamma^2\left(V_0 + Kc_\gamma^2\right)\sum_{k\geq\ell}k^{-2\gamma} < \infty \quad \text{a.s.}$$

在以 $\mathcal{F}_0$,为条件并带有延迟 ℓ的情况下应用引理 D.1(i)，得到

$$\mathbb{P}\left[\sum_{n\geq\ell-1}\gamma_{n+1}\varepsilon_{n+1}\ \text{converges} \mid \mathcal{F}_0\right] = 1 \quad \text{a.s.}$$

由公式 (A.2) 和 (A.4)、假设 A1 以及 $\sum_{k\geq\ell}\gamma_k^2 < \infty$ 可得

$$\sum_{n\geq\ell-1}\mathbb{E}\left[\gamma_{n+1}|\rho_{n+1}| \mid \mathcal{F}_0\right] < \infty \quad \text{and} \quad \sum_{n\geq\ell-1}\mathbb{E}\left[\gamma_{n+1}^2\frac{m_{\ell-1}\left(\operatorname{Simpl}\left(Z_{n+1}\right)\right)^2}{\ell^2} \mid \mathcal{F}_0\right] < \infty \quad \text{a.s.}$$

由于求和项均非负，因此可得

$$\sum_{n\geq\ell-1}\gamma_{n+1}|\rho_{n+1}| < \infty \quad \text{a.s.}, \qquad \text{and} \qquad \sum_{n\geq\ell-1}\gamma_{n+1}^2\frac{m_{\ell-1}\left(\operatorname{Simpl}\left(Z_{n+1}\right)\right)^2}{\ell^2} < \infty \quad \text{a.s.}$$

$(\mu_n)_n$ 的几乎必然收敛。对于 $n \geq \ell$，我们设

$$a_n \coloneqq -2\gamma_n \left(\varepsilon_n + \rho_n\right) + \gamma_n^2 \frac{m_{\ell-1}\left(\mathrm{Simpl}\left(Z_n\right)\right)^2}{\ell^2}.$$

将上述分解代入 (A.5) 得到

$$V_n \leq V_{n-1} - 2\gamma_n \left(G\left(\mu_{n-1}\right) - G\left(\mu^*\right)\right) + a_n.$$

上述三个级数的收敛性给出

$$\mathbb{P}\left[\sum_{n\geq\ell} a_n \text{ converges} \mid \mathcal{F}_0\right] = 1 \quad \text{a.s.}$$

由于 $G(\mu_{n-1}) \geq G(\mu^*)$，关于 V_n 的不等式表明 $(V_n - \sum_{k=\ell}^n a_k)_{n\geq\ell}$ 是非增的。$\sum_{k\geq\ell} a_k$ 和 $V_n \geq 0$ 的收敛性表明该序列有下界。因此，$(V_n)_n$ 几乎处处收敛到一个有限极限。对关于 V_n 的不等式求和并利用 $V_n \geq 0$，得到 $\sum_{n\geq\ell} \gamma_n \left(G(\mu_{n-1}) - G(\mu^*)\right) < \infty$ 几乎处处。由于 $\sum_{n\geq\ell} \gamma_n = \infty$ 几乎处处，我们得到

$$\liminf_{n\to\infty} \left(G\left(\mu_n\right) - G\left(\mu^*\right)\right) = 0 \quad \text{a.s.}$$

$(V_n)_n$ 的收敛性意味着 $(\mu_n)_n$ 几乎处处有界。由 G 的连续性及其极小化点的唯一性，存在一个子序列几乎处处收敛到 μ^*。因此，$\liminf_{n\to\infty} V_n = 0$几乎处处成立，而 $(V_n)_n$ 的收敛性给出 $V_n \to 0$ 几乎处处。我们已经证明

$$\sup_n \mathbb{E}\left[V_n \mid \mathcal{F}_0\right] \leq V_0 + Kc_\gamma^2 \quad \text{and} \quad \mathbb{P}\left[\mu_n \xrightarrow[n\to\infty]{} \mu^* \mid \mathcal{F}_0\right] = 1 \quad \text{a.s.}$$

取期望可得

$$\sup_n \mathbb{E}\left[V_n\right] \leq \mathbb{E}\left[V_0\right] + K\mathbb{E}\left[c_\gamma^2\right] < \infty \quad \text{and} \quad \mathbb{P}\left[\mu_n \xrightarrow[n\to\infty]{} \mu^*\right] = 1.$$

这就证明了定理 3.1。 □

A.2 定理 3.2 的证明

定理 3.2 的证明分为三部分。对于附录 A.1 中引入的函数 ξ_k，以及 $k \geq \ell$ 和 $u \in \mathbb{R}^d$，我们采用如下记号

$$\nabla\xi_k\left(u\right) \coloneqq \nabla_\mu g\left(Z_k; u\right) - \nabla G\left(u\right),$$

按照约定，在 $g(Z_k;\cdot)$不可微的点处，$\nabla_\mu g(Z_k; u) = 0$，如式 (2.2) 所示。由 Z_k 与$\mathcal{F}_{k-\ell}$ 的独立性可得，$\mathbb{E}\left[\nabla\xi_k\left(\mu_{k-\ell}\right) \mid \mathcal{F}_{k-\ell}\right] = 0$ 几乎处处成立。对于 $k \geq 2\ell - 1$，递推式 (3.1) 可写为

$$\mu_k = \mu_{k-1} - \gamma_k \left(\nabla G\left(\mu_{k-1}\right) + \left(\nabla\xi_k\left(\mu_{k-1}\right) - R_k\right) + R_k\right),$$

其中

$$R_k \coloneqq \mathbb{E}\left[\nabla\xi_k\left(\mu_{k-1}\right) \mid \mathcal{F}_{k-2\ell+1}\right].$$

向量 R_k 是$\mathcal{F}_{k-2\ell+1}$-可测的，且

$$\mathbb{E}\left[\nabla\xi_k\left(\mu_{k-1}\right) - R_k \mid \mathcal{F}_{k-2\ell+1}\right] = 0 \quad \text{a.s.}$$

由于 $\nabla G(\mu^*) = 0$，对于每个 $k \geq \ell$，$\nabla\xi_k(\mu^*) = \nabla_\mu g(Z_k; \mu^*)$。对于 $k \geq 0$，我们将泰勒余项定义为

$$\delta_k \coloneqq \nabla G\left(\mu_k\right) - H\left(\mu_k - \mu^*\right).$$

对于 $k \geq 2\ell - 1$，递推变为

$$\mu_k - \mu^* = (I_d - \gamma_k H)(\mu_{k-1} - \mu^*) - \gamma_k ((\nabla \xi_k (\mu_{k-1}) - R_k) + R_k + \delta_{k-1}).$$

令 n_H 为满足 $\gamma_{n_H} \|H\|_{\mathrm{op}} \leq 1/2$的最小正整数。该整数是$\mathcal{F}_0$-可测的。由于$(\gamma_k)_k$ 是递减的，对于每个 $k \geq n_H$，都有 $\gamma_k \|H\|_{\mathrm{op}} \leq 1/2$ 。对于 $n \geq k \geq n_H - 1$，我们定义

$$\Pi_{n,k} \coloneqq \prod_{j=k+1}^{n} (I_d - \gamma_j H), \quad \text{and} \quad \Pi_{n,n} = I_d.$$

我们定义 $K_0 \coloneqq \max\{n_H, 2\ell - 1\}$。随机整数 K_0和矩阵$\Pi_{n,k}$是$\mathcal{F}_0$-可测的。对于每个 $n \geq K_0$，从 K_0 开始迭代此递推可得

$$\mu_n - \mu^* = \Pi_{n,K_0}(\mu_{K_0} - \mu^*) - \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} ((\nabla \xi_k (\mu_{k-1}) - R_k) + R_k + \delta_{k-1}). \tag{A.6}$$

这些因子可交换，因为它们都是 H的多项式。令 $\lambda_{\min}(H) > 0$ 表示
H $j \geq n_H$ H $\|I_d - \gamma_j H\|_{\mathrm{op}} =$的最小特征值。对于每个，取
$1 - \gamma_j \lambda_{\min}$ H $n \geq k \geq n_H - 1$
标准正交基对角化可得（）。因此，对于，

$$\|\Pi_{n,k}\|_{\mathrm{op}} \leq \prod_{j=k+1}^{n} (1 - \gamma_j \lambda_{\min}(H)) \leq \exp\left(-\lambda_{\min}(H) \sum_{j=k+1}^{n} \gamma_j\right) \quad \text{a.s.} \tag{A.7}$$

证明定理 *3.2* 中$(\mu_n)_n$ 的几乎必然收敛速率。根据命题 F.1，假设 A2 蕴含假设 A1。因此，定理 3.1 给出 $\mu_n \to \mu^*$ 几乎处处。由于 $\gamma\nu > 2$，我们取 q 使得

$$\frac{2}{\gamma} < q < \nu.$$

由于 $\sum_{j=K_0+1}^{n} \gamma_j \sim c_\gamma n^{1-\gamma}/(1-\gamma)$，(A.7) 给出

$$\|\Pi_{n,K_0}(\mu_{K_0} - \mu^*)\| \leq \|\mu_{K_0} - \mu^*\| \exp\left(-\lambda_{\min}(H) \sum_{j=K_0+1}^{n} \gamma_j\right) = O\left(e^{-cn^{1-\gamma}}\right) \quad \text{a.s.}, \tag{A.8}$$

其中 $c : = \lambda_{\min}(H) c_\gamma / (2(1-\gamma)) > 0$。

首先假设 $\ell \geq 2$。令 $C_1 > 0$ 为使得 (C.1) 中关于单纯形体积的乘积界成立的常数，并设 $C : = 2^\gamma (\ell - 1) C_1$。对于每个 $k \geq 2\ell - 1$，(A.3) 给出

$$\begin{aligned}
\|\mu_{k-1} - \mu_{k-\ell}\| &\leq \sum_{j=k-\ell+1}^{k-1} \gamma_j \frac{m_{\ell-1}(\mathrm{Simpl}(Z_j))}{\ell} \\
&\leq C_1 \sum_{j=k-\ell+1}^{k-1} \gamma_j \prod_{i=j-\ell+1}^{j} (1 + \|X_i\|) \\
&\leq 2^\gamma C_1 \gamma_k \sum_{j=k-\ell+1}^{k-1} \prod_{i=k-2\ell+2}^{k} (1 + \|X_i\|) \\
&= C \gamma_k \prod_{i=k-2\ell+2}^{k} (1 + \|X_i\|) \quad \text{a.s.}
\end{aligned} \tag{A.9}$$

向量 $\mu_{k-\ell}$、步长 γ_k和乘积 $\prod_{i=k-2\ell+2}^{k-\ell}(1 + \|X_i\|)$ 是 $\mathcal{F}_{k-\ell}$-可测的，并且 Z_k与 $\mathcal{F}_{k-\ell}$独立。令 $t : = C\gamma_k \prod_{i=k-2\ell+2}^{k-\ell}(1 + \|X_i\|)$，其中 C是 (A.9) 中的常数。根据 (A.9)， μ_{k-1} 几乎处处属于以 $\mu_{k-\ell}$ 为中心、半径为

$t\prod_{i=k-\ell+1}^{k}(1+\|X_i\|)$ 的闭球。我们将引理 C.2 的梯度界 (C.5) 应用于 Z_k，取 $u=\mu_{k-\ell}$ 和 $v=\mu_{k-1}$，得到

$$\begin{aligned}&\|\nabla_\mu g(Z_k;\mu_{k-1})-\nabla_\mu g(Z_k;\mu_{k-\ell})\|\\&\quad\le\begin{cases}\dfrac{2m_{\ell-1}(\mathrm{Simpl}(Z_k))}{\ell}\min\left(1,\dfrac{t\prod_{i=k-\ell+1}^{k}(1+\|X_i\|)}{\mathrm{dist}(\mu_{k-\ell},\mathrm{aff}(Z_k))}\right), & \text{if } \ell<d,\\ \dfrac{2m_{d-1}(\mathrm{Simpl}(Z_k))}{d}\mathbf{1}_{\left\{\mathrm{dist}(\mu_{k-\ell},\mathrm{aff}(Z_k))\le t\prod_{i=k-\ell+1}^{k}(1+\|X_i\|)\right\}}, & \text{if } \ell=d,\end{cases}\quad\text{a.s.}\end{aligned}$$

期望在以$\mathcal{F}_{k-\ell}$为条件后，引理 C.2 中的界 (C.6) 给出一个常数 $C_2\ge 0$ 使得

$$\mathbb{E}\left[\|\nabla_\mu g(Z_k;\mu_{k-1})-\nabla_\mu g(Z_k;\mu_{k-\ell})\|\mid\mathcal{F}_{k-\ell}\right]\le C_2 t\quad\text{a.s.}$$

对 (A.9) 取条件期望，得到

$$\mathbb{E}\left[\|\mu_{k-1}-\mu_{k-\ell}\|\mid\mathcal{F}_{k-\ell}\right]\le t\mathbb{E}\left[\prod_{i=k-\ell+1}^{k}(1+\|X_i\|)\mid\mathcal{F}_{k-\ell}\right]=t\left(\mathbb{E}\left[1+\|X_1\|\right]\right)^{\ell}\quad\text{a.s.}$$

由三角不等式以及命题 F.1 中 ∇G 的利普希茨连续性，现在可得一个常数 $C\ge 0$，使得对每个 $k\ge 2\ell-1$,

$$\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1})-\nabla\xi_k(\mu_{k-\ell})\|\mid\mathcal{F}_{k-\ell}\right]\le C\gamma_k\prod_{i=k-2\ell+2}^{k-\ell}(1+\|X_i\|)\quad\text{a.s.}$$

右侧的乘积与$\mathcal{F}_{k-2\ell+1}$ 独立，且其期望为 $\left(\mathbb{E}\left[1+\|X_1\|\right]\right)^{\ell-1}<\infty$。因此，给定$\mathcal{F}_{k-2\ell+1}$ 取条件期望，可得一个常数 $C\ge 0$，使得对每个 $k\ge 2\ell-1$,

$$\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1})-\nabla\xi_k(\mu_{k-\ell})\|\mid\mathcal{F}_{k-2\ell+1}\right]\le C\gamma_k\quad\text{a.s.}$$

$\nabla\xi_k(\mu_{k-\ell})$ 的条件中心化和塔式性质给出

$$\mathbb{E}\left[\nabla\xi_k(\mu_{k-\ell})\mid\mathcal{F}_{k-2\ell+1}\right]=0\quad\text{a.s.}$$

因此，根据 R_k 的定义和条件詹森不等式，对于每个 $k\ge 2\ell-1$,

$$\begin{aligned}\|R_k\|&=\left\|\mathbb{E}\left[\nabla\xi_k(\mu_{k-1})-\nabla\xi_k(\mu_{k-\ell})\mid\mathcal{F}_{k-2\ell+1}\right]\right\|\\&\le\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1})-\nabla\xi_k(\mu_{k-\ell})\|\mid\mathcal{F}_{k-2\ell+1}\right]\le C\gamma_k\quad\text{a.s.}\end{aligned}\tag{A.10}$$

对于 $\ell=1$，这两个界成立，因为 $\mu_{k-\ell}=\mu_{k-1}$ 和 $R_k=0$。根据条件詹森不等式、(A.1) 中的梯度界以及 (C.2) 中的矩界，存在常数$C\ge 0$，使得对于每个 $k\ge K_0+1$,

$$\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1})-R_k\|^q\mid\mathcal{F}_{k-2\ell+1}\right]\le C\quad\text{a.s.}$$

由于 $q>2/\gamma$，引理 D.4 在条件于 $\mathcal{F}_0$ 且延迟为 $2\ell-1$ 的情况下应用，给出

$$\left\|\sum_{k=K_0+1}^{n}\gamma_k\Pi_{n,k}\left(\nabla\xi_k(\mu_{k-1})-R_k\right)\right\|^2=O(\gamma_n\log n)\quad\text{a.s.}$$

此外，引理 D.2 中关于加权和的界（D.1）结合 $a=\gamma$ 和 $b=0$ 可得

$$\left\|\sum_{k=K_0+1}^{n}\gamma_k\Pi_{n,k}R_k\right\|\le C\sum_{k=K_0+1}^{n}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}=O(\gamma_n)\quad\text{a.s.}\tag{A.11}$$

这些界和（A.6）表明，存在一个有限随机变量 $C_0 \geq 0$，使得对于每个 $n > K_0$,

$$\|\mu_n - \mu^*\| \leq C_0\sqrt{\gamma_n \log n} + \sum_{k=K_0+1}^{n} \gamma_k \|\Pi_{n,k}\|_{\mathrm{op}} \|\delta_{k-1}\| \quad \text{a.s.}$$

我们定义 κ_n：当 $\mu_n \neq \mu^*$ 时为 $= \|\delta_n\| / \|\mu_n - \mu^*\|$，否则为 $\kappa_n := 0$。由于 $\nabla G(\mu^*) = 0$，两种情况下均有 $\|\delta_n\| = \kappa_n \|\mu_n - \mu^*\|$。根据假设 A3，$\nabla G$ 在 μ^* 处可微，其导数为 H，因此当 $u \to 0$ 时，$\nabla G(\mu^* + u) - Hu = o(\|u\|)$。由于 $\mu_n \to \mu^*$ 几乎处处成立，$\kappa_n \to 0$ 几乎处处成立。在上述界成立且 $\kappa_n \to 0$ 的概率为 1 的事件上，引理 D.3 适用，其中

$$u_n = \|\mu_n - \mu^*\|, \qquad v_n = (C_0 + 1)\sqrt{\gamma_n \log n}, \qquad n_0 = K_0.$$

由此可得

$$\|\mu_n - \mu^*\| = O\left(\sqrt{\gamma_n \log n}\right) \quad \text{a.s.},$$

这就证明了所述的几乎处处收敛速率。 □

关于平均值的断言的证明 $\bar{\mu}_n$ 定理 *3.2* 中的断言。我们采用 Polyak 和 Juditsky（1992 年，定理 2）证明中的线性化和平均化论证，并针对重叠窗口之间的依赖关系给出额外的估计。对于上文选取的指数 q，我们令 $\beta := (q-2)/(q-1)$。

的分解 $H(\bar{\mu}_n - \mu^*)$。重新整理并对（3.1）求和可得

$$(n - \ell + 1) H(\bar{\mu}_n - \mu^*) = T_{1,n} - T_{2,n} - T_{3,n}, \tag{A.12}$$

其中

$$T_{1,n} := \sum_{k=\ell}^{n} \frac{\mu_k - \mu_{k+1}}{\gamma_{k+1}}, \qquad T_{2,n} := \sum_{k=\ell}^{n} \nabla\xi_{k+1}(\mu_{k+1-\ell}),$$
$$T_{3,n} := \sum_{k=\ell}^{n} (\delta_k + \nabla\xi_{k+1}(\mu_k) - \nabla\xi_{k+1}(\mu_{k+1-\ell})).$$

应用分部求和可得

$$T_{1,n} = \frac{\mu_\ell - \mu^*}{\gamma_{\ell+1}} - \frac{\mu_{n+1} - \mu^*}{\gamma_{n+1}} + \sum_{k=\ell+1}^{n} (\mu_k - \mu^*)\left(\frac{1}{\gamma_{k+1}} - \frac{1}{\gamma_k}\right).$$

根据中值定理，$0 \leq \gamma_{k+1}^{-1} - \gamma_k^{-1} \leq c_\gamma^{-1}\gamma k^{\gamma-1}$。由于 $\gamma < 1$，由该界以及 $(\mu_n)_n$ 的几乎处处收敛速率可得

$$\|T_{1,n}\|^2 = O(n^\gamma \log n) = o(n) \quad \text{a.s.}$$

（A.10）中关于 R_k 的界以及 $1/2 < \gamma < 1$ 可得

$$\sum_{k=2\ell-1}^{n} \|R_k\| = O\left(n^{1-\gamma}\right) = o(\sqrt{n}) \quad \text{a.s.}$$

假设 A3 以及 $(\mu_k)_k$ 的几乎处处收敛速率给出

$$\|\delta_k\|^2 = O\left(k^{-\gamma(1+\alpha)} (\log k)^{1+\alpha}\right) \quad \text{a.s.}$$

由于 $1/2 < \gamma(1+\alpha)/2 < 1$，因此

$$\left(\sum_{k=\ell}^{n} \|\delta_k\|\right)^2 = O\left(n^{2-\gamma(1+\alpha)} (\log n)^{1+\alpha}\right) = o(n) \quad \text{a.s.}$$

$T_{2,\ n}$ 和 $T_{3,\ n}$的近似。与$F_{k-\ell}$ 的独立性给出

$$\mathbb{E}\left[\nabla \xi_k(\mu_{k-\ell}) - \nabla \xi_k(\mu^*) \mid \mathcal{F}_{k-\ell}\right] = 0 \quad \text{a.s.}$$

由命题 F.1 中的界 (F.1)，存在常数 $C \geq 0$，使得对每个 $k \geq \ell$,

$$\begin{aligned} &\mathbb{E}\left[\|\nabla \xi_k(\mu_{k-\ell}) - \nabla \xi_k(\mu^*)\|^2 \mid \mathcal{F}_{k-\ell}\right] \\ &\quad = \mathbb{E}\left[\|\nabla_\mu g(Z_k; \mu_{k-\ell}) - \nabla_\mu g(Z_k; \mu^*)\|^2 \mid \mathcal{F}_{k-\ell}\right] - \|\nabla G(\mu_{k-\ell}) - \nabla G(\mu^*)\|^2 \\ &\quad \leq C \min\left(\|\mu_{k-\ell} - \mu^*\|^\beta, 1\right) \quad \text{a.s.} \end{aligned} \tag{A.13}$$

（μ_k）$_k$的几乎处处收敛速率表明，右端几乎处处为 $O\left(k^{-\gamma\beta/2} (\log k)^{\beta/2}\right)$。由于 $\gamma\beta > 0$，因此

$$\sum_{k=\ell}^{\infty} \frac{1}{k} \mathbb{E}\left[\|\nabla \xi_k(\mu_{k-\ell}) - \nabla \xi_k(\mu^*)\|^2 \mid \mathcal{F}_{k-\ell}\right] < \infty \quad \text{a.s.}$$

对于 $k \geq 2\ell - 1$，我们定义

$$\eta_k \coloneqq \nabla \xi_k(\mu_{k-1}) - \nabla \xi_k(\mu_{k-\ell}) - R_k.$$

R_k 的定义和塔式性质给出 $\mathbb{E}[\eta_k \mid \mathcal{F}_{k-2\ell+1}] = 0$ 几乎处处成立。由 (A.1)，对每个 $k \geq 2\ell - 1$,

$$\|\nabla \xi_k(\mu_{k-1}) - \nabla \xi_k(\mu_{k-\ell})\| \leq \frac{2 m_{\ell-1}(\mathrm{Simpl}(Z_k))}{\ell} + 2\mathbb{E}\left[\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\ell}\right] \quad \text{a.s.}$$

由于 Z_k 与 $\mathcal{F}_{k-2\ell+1}$独立，(C.2) 中的矩界给出一个常数 $C_q \geq 0$，使得对每个 $k \geq 2\ell - 1$,

$$\mathbb{E}\left[\|\nabla \xi_k(\mu_{k-1}) - \nabla \xi_k(\mu_{k-\ell})\|^q \mid \mathcal{F}_{k-2\ell+1}\right] \leq C_q \quad \text{a.s.}$$

因此，条件方差恒等式、(A.10) 以及命题 F.1 的证明中以 (F.2) 形式给出的条件 Hölder 不等式可得一个常数 $C \geq 0$，使得对于每个$k \geq 2\ell - 1$,

$$\mathbb{E}\left[\|\eta_k\|^2 \mid \mathcal{F}_{k-2\ell+1}\right] = \mathbb{E}\left[\|\nabla \xi_k(\mu_{k-1}) - \nabla \xi_k(\mu_{k-\ell})\|^2 \mid \mathcal{F}_{k-2\ell+1}\right] - \|R_k\|^2 \leq C\gamma_k^\beta \quad \text{a.s.} \tag{A.14}$$

由于 $\gamma_k = c_\gamma k^{-\gamma}$ 和 $\gamma\beta > 0$，(A.14) 给出

$$\sum_{k=2\ell-1}^{\infty} \frac{1}{k} \mathbb{E}\left[\|\eta_k\|^2 \mid \mathcal{F}_{k-2\ell+1}\right] < \infty \quad \text{a.s.}$$

分别在给定$\mathcal{F}_0$, 且延迟为 ℓ 和 $2\ell-1$ 的条件分布下应用引理 D.1(iii)，可得

$$\sum_{k=\ell}^{n} \left(\nabla \xi_k(\mu_{k-\ell}) - \nabla \xi_k(\mu^*)\right) = o\left(\sqrt{n}\right) \quad \text{a.s.}, \quad \text{and} \quad \sum_{k=2\ell-1}^{n} \eta_k = o\left(\sqrt{n}\right) \quad \text{a.s.}$$

结合上述两个估计量以及 $\sum_{k=\ell}^{n}\|\delta_k\|$ 和 $\sum_{k=2\ell-1}^{n}\|R_k\|$ 的界，可得

$$T_{3,n}=o\left(\sqrt{n}\right)\quad\text{a.s.},\quad\text{and}\quad T_{2,n}=\sum_{k=\ell}^{n}\nabla\xi_{k+1}\left(\mu^*\right)+o\left(\sqrt{n}\right)\quad\text{a.s.}$$

代入 (A.12) 并乘以 $H^{-1}/\sqrt{n}$ 可得

$$\frac{n-\ell+1}{\sqrt{n}}\left(\bar{\mu}_n-\mu^*\right)=-H^{-1}\frac{1}{\sqrt{n}}\sum_{k=\ell}^{n}\nabla\xi_{k+1}\left(\mu^*\right)+o\left(1\right)\quad\text{a.s.}\tag{A.15}$$

的渐近正态性$(\bar{\mu}_n)_n$。序列 $(\nabla\xi_k(\mu^*))_{k\geq\ell}$ 是严格平稳的，因为对具有共同分布 P 的独立观测的每个窗口都应用了相同的可测函数，并且它与$\mathcal{F}_0$独立。根据 $\nabla\xi_k$ 的定义、命题 F.1 中的梯度公式以及界 (A.1)，对于每个 $k\geq\ell$,

$$\mathbb{E}\left[\nabla\xi_k\left(\mu^*\right)\right]=0,\quad\text{and}\quad\mathbb{E}\left[\left\|\nabla\xi_k\left(\mu^*\right)\right\|^2\right]\leq\frac{1}{\ell^2}\mathbb{E}\left[m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{X}\right)\right)^2\right]<\infty.$$

序列 $(\nabla\xi_k(\mu^*))_{k\geq\ell}$ 是 $(\ell-1)$-相关的：对于每个整数 $a\geq\ell$，序列 $(\nabla\xi_k(\mu^*))_{\ell\leq k\leq a}$ 和 $(\nabla\xi_k(\mu^*))_{k\geq a+\ell}$ 分别依赖于 $(X_i)_{1\leq i\leq a}$ 和 $(X_i)_{i\geq a+1}$，因此彼此独立。对于 $\ell\geq2$，Duflo（1997）中的中心极限定理（在命题 D.5 中重述）适用于 $m=\ell-1$。对于 $\ell=1$，我们使用 $m=1$；连续项之间的协方差因独立性而为零。对于每个 $k\geq\ell$和 $j\in\{0,\ldots,\ell-1\}$，定理 3.2 中 Γ_j 的定义给出 $\mathbb{E}\left[\nabla\xi_k(\mu^*)\ \nabla\xi_{k+j}\left(\mu^*\right)^\top\right]=\Gamma_j$。将命题 D.5 条件应用于$\mathcal{F}_0$，并将其应用于序列 $(\nabla\xi_{k+1}(\mu^*))_{k\geq\ell}$，得到弱收敛：

$$\mathcal{L}\left(\frac{1}{\sqrt{n-\ell+1}}\sum_{k=\ell}^{n}\nabla\xi_{k+1}\left(\mu^*\right)\mid\mathcal{F}_0\right)\xrightarrow[n\to\infty]{}\mathcal{N}_d\left(0,\Sigma\right)\quad\text{a.s.}$$

由于 $(n-\ell+1)/n\to1$，由 (A.15) 和 Slutsky 定理可得 $(\bar{\mu}_n)_n$ 的条件渐近正态性，其协方差为 $H^{-1}\Sigma H^{-1}$。

($(\bar{\mu}_n)_n$ 的几乎必然收敛速率。对于 $a\in\{0,\ldots,\ell-1\}$，按剩余类 $a+1$ 模 ℓ 编号的序列 $(\nabla\xi_{a+j\ell+1}(\mu^*))_{j\geq1}$ 是独立同分布的，因为相应的窗口彼此不相交。将 Duflo（1997，第 6.4.23 定理，第 209 页）中的重对数律应用于每个坐标，并对 d 个坐标和 ℓ个剩余类求和，得到

$$\left\|\sum_{k=\ell}^{n}\nabla\xi_{k+1}\left(\mu^*\right)\right\|=O\left(\sqrt{n\log\log n}\right)\quad\text{a.s.}$$

由公式 (A.15) 和前述界可得

$$\frac{n-\ell+1}{\sqrt{n}}\left\|\bar{\mu}_n-\mu^*\right\|=O\left(\sqrt{\log\log n}\right)+o\left(1\right)\quad\text{a.s.}$$

由于 $(n-\ell+1)/n\to1$，因此

$$\left\|\bar{\mu}_n-\mu^*\right\|^2=O\left(\frac{\log\log n}{n}\right)\quad\text{a.s.}$$

这证明了所述 $(\bar{\mu}_n)_n$ 的几乎处处收敛速率。

□

定理 *3.2* 中 $(\mu_n)_n$ 的渐近正态性证明。

对 $\frac{1}{\sqrt{\gamma_n}}(\mu_n - \mu^*)$ 的近似。在 (A.6) 中加上并减去 $\nabla\xi_k(\mu^*)$，可得，对任意$n \geq K_0$,

$$\begin{aligned}\mu_n - \mu^* = \Pi_{n,K_0}(\mu_{K_0} - \mu^*) - \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} \nabla\xi_k(\mu^*) - \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} R_k \\ - \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k}(\nabla\xi_k(\mu_{k-1}) - R_k - \nabla\xi_k(\mu^*)) - \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k}\delta_{k-1}.\end{aligned} \tag{A.16}$$

由界 (A.8) 和 (A.11) 可得

$$\|\Pi_{n,K_0}(\mu_{K_0} - \mu^*)\| = o(\sqrt{\gamma_n}) \quad \text{and} \quad \left\| \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} R_k \right\| = o(\sqrt{\gamma_n}) \quad \text{a.s.}$$

根据假设 A3，$(\mu_n)_n$ 的几乎处处收敛速率、引理 D.2 中的界 (D.1)，并且由于 $\alpha > 0$,

$$\left\| \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} \delta_{k-1} \right\| = O\left(\gamma_n^{(1+\alpha)/2} (\log n)^{(1+\alpha)/2}\right) = o(\sqrt{\gamma_n}) \quad \text{a.s.} \tag{A.17}$$

对于 $k \geq 2\ell - 1$，我们令

$$\widetilde{\eta}_k \coloneqq \nabla\xi_k(\mu_{k-1}) - R_k - \nabla\xi_k(\mu^*).$$

根据 R_k 的定义以及 $\nabla\xi_k(\mu^*)$ 的中心化，可得

$$\mathbb{E}[\widetilde{\eta}_k \mid \mathcal{F}_{k-2\ell+1}] = 0 \quad \text{a.s.}$$

根据 η_k 的定义，

$$\widetilde{\eta}_k = \eta_k + \nabla\xi_k(\mu_{k-\ell}) - \nabla\xi_k(\mu^*).$$

由于它被 1 和 $\mu_k \to \mu^*$ 几乎处处所界，条件支配收敛适用于 $\min\left(\|\mu_{k-\ell} - \mu^*\|^\beta, 1\right)$。利用 $\|x+y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$ 并在给定 $\mathcal{F}_0$, 的条件下对 (A.13) 和 (A.14) 取条件期望，我们得到一个常数 $C \geq 0$ ，使得

$$\begin{aligned}\mathbb{E}\left[\|\widetilde{\eta}_k\|^2 \mid \mathcal{F}_0\right] &\leq 2\mathbb{E}\left[\|\eta_k\|^2 \mid \mathcal{F}_0\right] + 2\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-\ell}) - \nabla\xi_k(\mu^*)\|^2 \mid \mathcal{F}_0\right] \\ &\leq C\gamma_k^\beta + C\mathbb{E}\left[\min\left(\|\mu_{k-\ell} - \mu^*\|^\beta, 1\right) \mid \mathcal{F}_0\right] \xrightarrow[k\to\infty]{} 0 \quad \text{a.s.}\end{aligned} \tag{A.18}$$

设 i, j 为满足 $2\ell - 1 \leq i < j$ 且 $j - i$ 是 $2\ell - 1$的倍数的整数。于是 $i \leq j - 2\ell + 1$，因此 $\widetilde{\eta}_i$ 是 $\mathcal{F}_{j-2\ell+1}$- 可测的。先在 $\mathcal{F}_{j-2\ell+1}$ 下、再在 $\mathcal{F}_0$ 下取条件期望，得到

$$\mathbb{E}\left[\widetilde{\eta}_i \widetilde{\eta}_j^\top \mid \mathcal{F}_0\right] = 0 \quad \text{a.s.}$$

由于 K_0 且 $\gamma_k \Pi_{n,\,k}$ 是 $\mathcal{F}_0$-可测的，通过划分为剩余类并应用施瓦茨 y–

不等式可得

$$
\begin{aligned}
&\frac{1}{\gamma_n}\mathbb{E}\left[\left\|\sum_{k=K_0+1}^{n}\gamma_k\Pi_{n,k}\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \\
&\quad\le \frac{2\ell-1}{\gamma_n}\sum_{a=0}^{2\ell-2}\mathbb{E}\left[\left\|\sum_{\substack{K_0<k\le n\\ k\equiv a \pmod{2\ell-1}}}\gamma_k\Pi_{n,k}\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \\
&\quad= \frac{2\ell-1}{\gamma_n}\sum_{k=K_0+1}^{n}\mathbb{E}\left[\left\|\gamma_k\Pi_{n,k}\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \\
&\quad\le \frac{2\ell-1}{\gamma_n}\sum_{k=K_0+1}^{n}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}^2\mathbb{E}\left[\left\|\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \quad \text{a.s.}
\end{aligned}
$$

由引理 D.2 中的界 (D.2)，存在一个有限的 $\mathcal{F}_0$-可测随机变量 $C_\Pi \ge 0$，使得对于每个 $n \ge K_0+1$,

$$
\frac{1}{\gamma_n}\sum_{k=K_0+1}^{n}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}^2 \le C_\Pi \quad \text{a.s.}
$$

我们固定一个整数 $K \ge 2\ell-1$。当 K 固定且 n 趋于无穷时，指数界 (A.7) 给出

$$
\begin{aligned}
\frac{1}{\gamma_n}\sum_{k=K_0+1}^{K-1}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}^2\mathbb{E}\left[\left\|\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] &= O\left(\gamma_n^{-1}e^{-\lambda_{\min}(H)c_\gamma n^{1-\gamma}}\right) \\
&= o(1) \quad \text{a.s.}
\end{aligned}
$$

对于这个固定的 K，将前述二阶矩界中非负项之和在 K 处分开，可得

$$
\begin{aligned}
&\frac{1}{\gamma_n}\mathbb{E}\left[\left\|\sum_{k=K_0+1}^{n}\gamma_k\Pi_{n,k}\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \\
&\quad\le \frac{2\ell-1}{\gamma_n}\sum_{k=K_0+1}^{K-1}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}^2\mathbb{E}\left[\left\|\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \\
&\qquad+\frac{2\ell-1}{\gamma_n}\sum_{k=\max\{K,K_0+1\}}^{n}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}^2\mathbb{E}\left[\left\|\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \\
&\quad= o(1)+\frac{2\ell-1}{\gamma_n}\sum_{k=\max\{K,K_0+1\}}^{n}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}^2\mathbb{E}\left[\left\|\widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \\
&\quad\le o(1)+\frac{2\ell-1}{\gamma_n}\left(\sum_{k=K_0+1}^{n}\gamma_k^2\left\|\Pi_{n,k}\right\|_{\mathrm{op}}^2\right)\sup_{j\ge K}\mathbb{E}\left[\left\|\widetilde{\eta}_j\right\|^2 \mid \mathcal{F}_0\right] \\
&\quad\le o(1)+(2\ell-1)C_\Pi\sup_{j\ge K}\mathbb{E}\left[\left\|\widetilde{\eta}_j\right\|^2 \mid \mathcal{F}_0\right] \quad \text{a.s.}
\end{aligned}
$$

由 (A.18) ，当 $K\to\infty$ 时，上确界几乎处处趋于零。我们首先取上极限 or

在 n 中，然后令 $K \to \infty$。由条件马尔可夫不等式可得，对每个 $\varepsilon > 0$,

$$\begin{aligned} &\mathbb{P}\left[\left\|\sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} \widetilde{\eta}_k\right\| > \varepsilon\sqrt{\gamma_n} \mid \mathcal{F}_0\right] \\ &\quad \leq \frac{1}{\varepsilon^2 \gamma_n} \mathbb{E}\left[\left\|\sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} \widetilde{\eta}_k\right\|^2 \mid \mathcal{F}_0\right] \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.} \end{aligned} \tag{A.19}$$

将 (A.8)、(A.11)、(A.17) 和 (A.19) 代入 (A.16) 得到

$$\frac{1}{\sqrt{\gamma_n}}\left(\mu_n - \mu^*\right) = -\frac{1}{\sqrt{\gamma_n}} \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} \nabla \xi_k\left(\mu^*\right) + o_{\mathbb{P}(\cdot|\mathcal{F}_0)}(1), \tag{A.20}$$

其中，$o_{\mathbb{P}(\cdot|\mathcal{F}_0)}(1)$ 表示在给定 $\mathcal{F}_0$ 的几乎处处条件分布下依概率收敛于零。

极限分布。</style>序列 $(())$ 是严格平稳、中心化、平方可积且 $()$-依赖的，并且与 F_0 独立。令 (3.3) 中定义的 V_H，在条件于 $\mathcal{F}_0$, 的情况下应用引理 D.6，可得

$$\mathcal{L}\left(\frac{1}{\sqrt{\gamma_n}} \sum_{k=K_0+1}^{n} \gamma_k \Pi_{n,k} \nabla \xi_k\left(\mu^*\right) \mid \mathcal{F}_0\right) \xrightarrow[n\to\infty]{} \mathcal{N}_d\left(0, V_H\right) \quad \text{a.s.}$$

条件分布下的公式 (A.20) 和 Slutsky 定理证明了定理 3.2 中所述的 $(\mu_n)_n$ 的条件渐近正态性。

由于 c_γ 是 $\mathcal{F}_0$-可测的，由 $\gamma_n^{-1/2} = c_\gamma^{-1/2}\sqrt{n^\gamma}$ 可得以 $\sqrt{n^\gamma}$ 归一化后的条件协方差。

字符　　化 V_H。为证明定理 3.2 后所述的刻画，我们对

$$\frac{d}{dt}\left(e^{-tH}\Sigma e^{-tH}\right) = -He^{-tH}\Sigma e^{-tH} - e^{-tH}\Sigma e^{-tH}H$$

关于 $[0,\infty)$ 积分并得到 $HV_H + V_H H = \Sigma$。在使 H 对角化的基下，其对应的特征值为 λ_1、……、λ_d，由（$\lambda_i + \lambda_j$）（V_H）可得唯一性$_{ij}$ $= \Sigma_{ij}$，因为所有 $\lambda_i + \lambda_j$ 均为正数。

□

定理 *3.2* 的证明完毕。我们从给定 $\mathcal{F}_0$ 的条件分布转到原始分布。对于表示定理 3.2 中一个几乎处处速率断言的每个事件 E，前面的证明给出 $\mathbb{P}[E \mid \mathcal{F}_0] = 1$ 几乎处处成立；因此 $\mathbb{P}[E] = \mathbb{E}[\mathbb{P}[E \mid \mathcal{F}_0]] = 1$。

对于迭代序列，取 $Y_n := \gamma_n^{-1/2}(\mu_n - \mu^*)$ 和 $Q := \mathcal{N}_d(0,\ V_H)$；对于平均值，在其附加假设下，取 $Y_n := \sqrt{n}(\bar{\mu}_n - \mu^*)$ 和 $Q := \mathcal{N}_d(0,\ H^{-1}\Sigma H^{-1})$。在这两种情况下，$Q$ 均为确定性的，并且对于每个有界连续函数 φ: $\mathbb{R}^d \to \mathbb{R}$，条件极限给出

$$\mathbb{E}[\varphi(Y_n) \mid \mathcal{F}_0] \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} \varphi(x)\, Q(dx) \quad \text{a.s.}$$

由于 $|\mathbb{E}[\varphi(Y_n) \mid \mathcal{F}_0]| \leq \sup_{x\in\mathbb{R}^d} |\varphi(x)|$，利用塔式性质和控制收敛，并在给定 c_γ 的条件下也成立，可得

$$\begin{aligned} \mathbb{E}[\varphi(Y_n)] = \mathbb{E}[\mathbb{E}[\varphi(Y_n) \mid \mathcal{F}_0]] &\xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} \varphi(x)\, Q(dx), \\ \mathbb{E}[\varphi(Y_n) \mid c_\gamma] = \mathbb{E}[\mathbb{E}[\varphi(Y_n) \mid \mathcal{F}_0] \mid c_\gamma] &\xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} \varphi(x)\, Q(dx) \quad \text{a.s.} \end{aligned}$$

因此，$(Y_n)_n$ 在原始分布下以及在几乎所有给定 c_γ 的条件分布下，都依分布收敛于 Q 。

对于定理 3.2 后讨论的归一化 $\sqrt{n^\gamma}$ ，给定$\mathcal{F}_0$ 的条件高斯极限的协方差为 $c_\gamma V_H$。由于 $\sigma(c_\gamma) \subset \mathcal{F}_0,$ ，由条件支配收敛可得

$$\mathbb{E}\left[\varphi\left(\sqrt{n^\gamma}(\mu_n-\mu^*)\right) \mid c_\gamma\right] \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^d} \varphi(x)\,\mathcal{N}_d(0,c_\gamma V_H)(dx) \quad \text{a.s.}$$

取期望可得

$$\mathbb{E}\left[\varphi\left(\sqrt{n^\gamma}(\mu_n-\mu^*)\right)\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[\int_{\mathbb{R}^d} \varphi(x)\,\mathcal{N}_d(0,c_\gamma V_H)(dx)\right],$$

这表明，无条件极限是这些高斯分布在c_γ的分布下的混合。

□

A.3 命题 3.3 的证明

证明。 沿用定理 3.2 证明中的记号。设 q 为假设 A4 中的指数，并令 $\beta := (q-2)/(q-1)$。由假设 A1 和定理 3.1， $\mu_n \to \mu^*$ 几乎处处成立。

的可微性。 假设 A4 中的矩条件、 $X_1, \dots, X_\ell$的独立性以及乘积界 (C.1) 给出 $\mathbb{E}[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^q] < \infty$ 和 $G(0) < \infty$。界 (A.1) 随即表明， G 在 $\mathbb{R}^d$上有限。对于每个 $u \in \mathbb{R}^d$, Dürre 和 Paindaveine（2022b）的引理 S.1.3 给出 $\mathbb{P}[u \in \mathrm{aff}(\mathbf{X})] = 0$，因此 $g(\mathbf{X};\cdot)$ 在 u 处几乎处处可微。对于每个非零$v \in \mathbb{R}^d$,

$$\frac{|g(\mathbf{X};u+v) - g(\mathbf{X};u) - \langle \nabla_\mu g(\mathbf{X};u), v\rangle|}{\|v\|} \le \frac{2m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\ell} \quad \text{a.s.}$$

控制收敛定理适用：右端是可积的，并且 $g(\mathbf{X};\cdot)$ 在 u 处的可微性使得当 $v \to 0$ 时，左端几乎处处收敛到零。因此

$$\frac{G(u+v) - G(u) - \langle \mathbb{E}[\nabla_\mu g(\mathbf{X};u)], v\rangle}{\|v\|} \xrightarrow[v\to 0]{} 0.$$

因此， G 可微，并且对于每个 $u \in \mathbb{R}^d$， $\nabla G(u) = \mathbb{E}[\nabla_\mu g(\mathbf{X};u)]$。
情形 $\ell = 1$。这里 $g(x;u) = \|x-u\|$，因此 $\|\nabla_\mu g(x;u)\| \le 1$。由此，对于每个 $k \ge 1$, $R_k = \eta_k = 0$，并且 $(\nabla\xi_k(\mu_{k-1}))_{k\ge1}$ 是关于 $(\mathcal{F}_k)_{k\ge0}$ 的鞅差序列。对于每个 $k \ge 1$， $\|\nabla\xi_k(\mu_{k-1})\| \le 2$ 几乎处处成立。结合假设 A3 和引理 D.4，$(\mu_n)_n$ 的几乎处处收敛速率的证明仍然适用。

令 e_1、.....、 e_d表示 $\mathbb{R}^d$ 的标准基。对于 $t \ne 0$，由凸性可得

$$0 \le \sum_{j=1}^d \frac{\|X_1-\mu^*+te_j\| + \|X_1-\mu^*-te_j\| - 2\|X_1-\mu^*\|}{t^2} \xrightarrow[t\to0]{} \frac{d-1}{\|X_1-\mu^*\|} \quad \text{a.s.},$$

因为 P 没有原子。Fatou 引理和假设 A3 得出

$$(d-1)\mathbb{E}\left[\frac{1}{\|X_1-\mu^*\|}\right] \le \liminf_{t\to0} \sum_{j=1}^d \frac{G(\mu^*+te_j)+G(\mu^*-te_j)-2G(\mu^*)}{t^2} = \mathrm{tr}(H) < \infty.$$

由于 $d-1>0$, $\mathbb{E}[1/\|X_1-\mu^*\|] \le \mathrm{tr}(H)/(d-1) < \infty$。对于非零向量 a, b，三角不等式给出 $\left\|\frac{a}{\|a\|} - \frac{b}{\|b\|}\right\| \le 2\min\left(1, \frac{\|a-b\|}{\|a\|}\right)$。对于每个 $u \in \mathbb{R}^d$，向量 μ^*-X_1和 $u-X_1$几乎处处非零，因为 P没有原子。应用不等式，其中 $a := \mu^*-X_1$和 $b := u-X_1$，并取期望可得

$$\mathbb{E}\left[\|\nabla_\mu g(X_1;u) - \nabla_\mu g(X_1;\mu^*)\|^2\right] \le 4\min\left(\|u-\mu^*\|\,\mathbb{E}\left[\frac{1}{\|X_1-\mu^*\|}\right], 1\right).$$

对于每个 $k \geq 1$，向量 μ_{k-1} 是$\mathcal{F}_{k-1}$-可测的，而 X_k 与$\mathcal{F}_{k-1}$独立。因此，根据条件方差恒等式，存在一个常数 $C \geq 0$ 使得

$$\begin{aligned}
&\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1}) - \nabla\xi_k(\mu^*)\|^2 \mid \mathcal{F}_{k-1}\right] \\
&\quad = \mathbb{E}\left[\|\nabla_\mu g(X_k;\mu_{k-1}) - \nabla_\mu g(X_k;\mu^*)\|^2 \mid \mathcal{F}_{k-1}\right] - \|\nabla G(\mu_{k-1}) - \nabla G(\mu^*)\|^2 \\
&\quad \leq C\min\left(\|\mu_{k-1} - \mu^*\|^\beta, 1\right) \quad \text{a.s.}
\end{aligned}$$

这就证明了 $\ell = 1$ 的 (A.13)。对于 $R_k = \eta_k = 0$，关于平均值以及 $(\mu_n)_n$ 渐近正态性的断言，其证明同样适用。

情形 $2 \leq \ell < d$。引理 C.2 在假设 A4 和 A5 下适用。由于 $\prod_{i=1}^{\ell}(1+\|X_i\|) \geq 1$，其期望界 (C.6) 给出一个常数 $C \geq 0$，使得对每个 $u,\ v \in \mathbb{R}^d$,

$$\|\nabla G(u) - \nabla G(v)\| \leq \mathbb{E}\left[\|\nabla_\mu g(\mathbf{X};u) - \nabla_\mu g(\mathbf{X};v)\|\right] \leq C\|u - v\|.$$

因此 ∇G 是 Lipschitz 的。(A.1) 中的界还给出

$$\mathbb{E}\left[\|\nabla_\mu g(\mathbf{X};u) - \nabla_\mu g(\mathbf{X};v)\|^q\right] \leq \left(\frac{2}{\ell}\right)^q \mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^q\right] < \infty.$$

由 Hölder 不等式可得

$$\begin{aligned}
&\mathbb{E}\left[\|\nabla_\mu g(\mathbf{X};u) - \nabla_\mu g(\mathbf{X};v)\|^2\right] \\
&\quad \leq \left(\mathbb{E}\left[\|\nabla_\mu g(\mathbf{X};u) - \nabla_\mu g(\mathbf{X};v)\|\right]\right)^\beta \left(\mathbb{E}\left[\|\nabla_\mu g(\mathbf{X};u) - \nabla_\mu g(\mathbf{X};v)\|^q\right]\right)^{1/(q-1)} \\
&\quad \leq C\|u - v\|^\beta.
\end{aligned}$$

将此结果与 (A.1) 中的二阶矩界结合，即得 (F.1)。以 $\mathcal{F}_{k-\ell}$ 为条件，并利用 Z_k 与 $\mathcal{F}_{k-\ell}$ 的独立性，再由条件方差恒等式可得 (A.13)。

对每个 $k \geq 2\ell - 1$，将引理 C.2 中的增量界 (A.9)、逐点界 (C.5) 和期望界 (C.6) 以 $\mathcal{F}_{k-\ell}$ 为条件应用，可得

$$\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1}) - \nabla\xi_k(\mu_{k-\ell})\| \mid \mathcal{F}_{k-\ell}\right] \leq C\gamma_k \prod_{i=k-2\ell+2}^{k-\ell}(1 + \|X_i\|) \quad \text{a.s.}$$

右侧的乘积与$\mathcal{F}_{k-2\ell+1}$独立，且其期望为 $(\mathbb{E}[1+\|X_1\|])^{\ell-1}$。对给定的$\mathcal{F}_{k-2\ell+1}$ 取条件期望，并利用 R_k 的定义，可得

$$\|R_k\| \leq \mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1}) - \nabla\xi_k(\mu_{k-\ell})\| \mid \mathcal{F}_{k-2\ell+1}\right] \leq C\gamma_k \quad \text{a.s.}$$

由界 (A.1) 和 q 阶矩可得

$$\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1}) - \nabla\xi_k(\mu_{k-\ell})\|^q \mid \mathcal{F}_{k-2\ell+1}\right] \leq C \quad \text{a.s.}$$

因此，由条件方差恒等式和条件 Hölder 不等式可得 e

$$\mathbb{E}\left[\|\eta_k\|^2 \mid \mathcal{F}_{k-2\ell+1}\right] \leq C\gamma_k^\beta \quad \text{a.s.}$$

由相同的包络和条件詹森不等式可得

$$\mathbb{E}\left[\|\nabla\xi_k(\mu_{k-1}) - R_k\|^q \mid \mathcal{F}_{k-2\ell+1}\right] \leq C \quad \text{a.s.}$$

因此，我们得到了 (A.10) 和 (A.14)，以及几乎处处速率证明中所使用的条件矩界。由于 $\gamma q > 2$，定理 3.2 证明的三个部分给出了该命题的条件性结论。附录 A.2 末尾的论证则在原始分布下给出了这些结论。

□

B 离线估计与协方差比较

在假设 A1 和 A3 下，(A.1) 中的界给出 $\mathbb{E}[\|\nabla_\mu g(\mathbf{X};\mu^*)\|^2] < \infty$。由于 G 在其极小化点处可微，对 (A.1) 中的次梯度不等式取期望可得$\mathbb{E}[\nabla_\mu g(\mathbf{X};\mu^*)] = 0$。

根据 g 的对称性以及 (2.2) 中的约定，函数 $\mathbf{x} \mapsto \nabla_\mu g(\mathbf{x};\mu^*)$ 关于 $x_1,\dots,x_\ell$ 是对称的。对于 $1 \le j \le \ell$，将引理 E.1 应用于 $h = \nabla_\mu g(\cdot;\,\mu^*)$，得到一个可测的对称 Hoeffding 投影 $\psi_j : (\mathbb{R}^d)^j \to \mathbb{R}^d$（Hoeffding，1961），满足

$$\psi_j\left(X_1,\dots,X_j\right) = \sum_{A\subseteq\{1,\dots,j\}} (-1)^{j-|A|}\,\mathbb{E}\left[\nabla_\mu g\left(\mathbf{X};\mu^*\right) \mid X_i,\ i\in A\right] \quad \text{a.s.}$$

特别地，

$$\psi_1(X_1) = \mathbb{E}[\nabla_\mu g(\mathbf{X};\mu^*) \mid X_1] \quad \text{a.s.}, \qquad \text{and} \qquad \mathbb{E}[\psi_1(X_1)\psi_1(X_1)^\top] = \zeta_1.$$

对于由正整数构成的有限集 $A = \{i_1 < \cdots < i_j\}$，记作 $X_A \coloneqq \left(X_{i_1},\dots,X_{i_j}\right)$。

B.1 定理 3.4 的证明

证明。 在假设 A2 下，命题 F.1 给出假设 A1；在另一种情况下，直接施加假设 A1。在这两种情况下，根据假设 A2 或 A4， P 都是 $(\ell-1)$-光滑的，并且具有有限二阶矩。

对于可测对称函数 $J : (\mathbb{R}^d)^\ell \to \mathbb{R}$，我们定义 $n \ge \ell$ 的 U 统计量（Hoeffding，1948年）

$$U_n\left(J\right) \coloneqq \binom{n}{\ell}^{-1} \sum_{1\le i_1<\cdots<i_\ell\le n} J\left(X_{i_1},\dots,X_{i_\ell}\right).$$

按照 U 统计量的术语，这样的可测对称函数称为核函数。因此，式 (2.3) 中的经验目标函数满足 $G_n(u) = U_n\left(g\left(\cdot;u\right)\right)$。

对于任意整数 $1 \le i_1 < \cdots < i_\ell$，Dürre 和 Paindaveine（2022b）的引理 S.1.3(ii) 表明，μ^* 几乎处处不属于 $\mathrm{aff}(X_{i_1},\dots,X_{i_\ell})$。由于指标的取值方式有可数种，故以概率 1，式 (2.3) 中的每一项对任意 $n \ge \ell$ 均在 μ^* 处可微，且

$$\nabla\widehat{G}_n\left(\mu^*\right) = \binom{n}{\ell}^{-1} \sum_{1\le i_1<\cdots<i_\ell\le n} \nabla_\mu g\left(X_{i_1},\dots,X_{i_\ell};\mu^*\right).$$

存在性与可测选择。 令 $n \ge \ell+1$。根据 Dürre 和 Paindaveine（2022b）的引理 S.1.3(i)，点 $X_1,\dots,X_n$ 几乎处处不包含于维数为 $\le \ell-1$ 的仿射子空间中。因此，Dürre 和 Paindaveine（2022a）的定理 2.2–2.3 表明， G_n 是强制的，且其极小化点构成非空紧凸集。对于每个 $u \in \mathbb{R}^d$， $\widehat{G}_n(u)$ 是 $\sigma(X_1,\dots,X_n)$-可测的，而 G_n 连续。因此，Rockafellar 和 Wets（1998）的定理 14.37 给出一个 $\sigma(X_1,\dots,X_n)$-可测随机向量 $\widehat{\ }$，它几乎处处最小化 G_n。令 $\widehat{\mu}_n$ 为任意一个可测极小化点。

目标函数的局部展开。对于 $u \in \mathbb{R}^d$，我们考虑 $\widehat{G}_n$ 从 μ^*到 $\mu^* + u/\sqrt{n}$的增量。与 Dürre 和 Paindaveine（2022b，公式（S.2.14））一样，我们从每个求和项中减去线性项，并将余项乘以 n：

$$J_{n,u}\left(\mathbf{x}\right) \coloneqq n\left(g\left(\mathbf{x};\mu^* + u/\sqrt{n}\right) - g\left(\mathbf{x};\mu^*\right) - \frac{u^\top \nabla_\mu g\left(\mathbf{x};\mu^*\right)}{\sqrt{n}}\right).$$

平均化得到

$$n\left(\widehat{G}_n\left(\mu^* + u/\sqrt{n}\right) - \widehat{G}_n\left(\mu^*\right)\right) = u^\top\sqrt{n}\,\nabla\widehat{G}_n\left(\mu^*\right) + U_n\left(J_{n,u}\right).$$

我们证明，序列 $(U_n(J_{n,\,u}))_n$ 依概率收敛到 $\frac{1}{2}u^\top Hu$，且对 u 有界集上的情形一致成立。

令 $u \in \mathbb{R}^d$。由凸性可得 $J_{n,\,u} \geq 0$。$g(\mathbf{X};\cdot)$ 在 μ^* 处的可微性及包络界 (A.1) 给出

$$\frac{J_{n,u}(\mathbf{X})}{\sqrt{n}} \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.,} \quad \text{and} \quad 0 \leq \frac{J_{n,u}(\mathbf{X})}{\sqrt{n}} \leq \frac{2m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\ell} \|u\|.$$

根据假设 A1，该主控函数具有有限二阶矩。因此，由控制收敛定理可得

$$\mathbb{E}[J_{n,u}(\mathbf{X})^2] = o(n).$$

沿着从 μ^* 到 $\mu^* + u/\sqrt{n}$ 的线段对假设 A3 中的界积分，可得

$$\mathbb{E}\left[J_{n,u}\left(\mathbf{X}\right)\right] = n\left(G\left(\mu^* + u/\sqrt{n}\right) - G\left(\mu^*\right)\right) = \tfrac{1}{2}u^\top Hu + o\left(1\right).$$

对于可测对称函数 $J : (\mathbb{R}^d)^\ell \to \mathbb{R}$，满足 $\mathbb{E}[J(\mathbf{X})^2] < \infty$，令 I 和 B 遍历 $\{1,\ldots,n\}$ 中大小为 ℓ 的子集。若 $I \cap B = \varnothing$，则随机变量 $J(X_I)$ 和 $J(X_B)$ 相互独立。若 $I \cap B \neq \varnothing$，由柯西–施瓦茨不等式可得 $|\mathrm{Cov}(J(X_I)\ J(X_B))| \leq \mathbb{E}[J(\mathbf{X})^2]$。对于每个 I，满足 $I \cap B \neq \varnothing$ 的集合 B 的数量至多为 $\ell\binom{n-1}{\ell-1}$。因此，

$$\mathrm{Var}\left(U_n\left(J\right)\right) = \binom{n}{\ell}^{-2} \sum_{|I|=|B|=\ell} \mathrm{Cov}(J(X_I), J(X_B)) \leq \frac{\ell\binom{n-1}{\ell-1}}{\binom{n}{\ell}} \mathbb{E}[J(\mathbf{X})^2] = \frac{\ell^2}{n}\mathbb{E}[J(\mathbf{X})^2]. \quad \text{(B.1)}$$

对 $J = J_{n,\,u}$ 应用 (B.1)，并使用两个矩估计可得，对每个固定的$u \in \mathbb{R}^d$，

$$U_n\left(J_{n,u}\right) \xrightarrow[n\to\infty]{\mathbb{P}} \tfrac{1}{2}u^\top Hu.$$

随机函数 $u \mapsto U_n(J_{n,\,u})$ 是有限的凸函数。根据 Andersen and Gill（1982）的定理 II.1，将其应用于凹函数 $u \mapsto -U_n(J_{n,\,u})$，可知逐点依概率收敛在 $\mathbb{R}^d$ 的紧子集上一致。因此，对任意 $R > 0$，

$$\Delta_n\left(R\right) \coloneqq \sup_{\|u\|\leq R} \left|U_n\left(J_{n,u}\right) - \tfrac{1}{2}u^\top Hu\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0. \quad \text{(B.2)}$$

极小化点的局部化。由于 $\mathbb{E}[\nabla_\mu g(\mathbf{X};\mu^*)] = 0$，将 (B.1) 应用于标量函数$\mathbf{x} \mapsto (\nabla_\mu g(\mathbf{x};\mu^*))_r$，$1 \leq r \leq d$，并对 r 求和可得

$$\sup_{n\geq\ell+1} \mathbb{E}\left[\left\|\sqrt{n}\,\nabla\widehat{G}_n(\mu^*)\right\|^2\right] \leq \ell^2\mathbb{E}\left[\|\nabla_\mu g(\mathbf{X};\mu^*)\|^2\right] < \infty.$$

我们将局部目标函数的极小化点与其二次近似的极小化点进行比较。令

$$\widehat{u}_n \coloneqq \sqrt{n}\left(\widehat{\mu}_n - \mu^*\right) \quad \text{and} \quad v_n \coloneqq -H^{-1}\sqrt{n}\,\nabla\widehat{G}_n\left(\mu^*\right).$$

对于固定的 $M > 0$，令 $R > 0$ 足够大，使得

$$\tfrac{1}{2}\lambda_{\min}(H)R^2 - MR - 1 > 0, \quad \text{and} \quad R > \left\|H^{-1}\right\|_{\mathrm{op}} M.$$

在 $\left\|\sqrt{n}\,\nabla\widehat{G}_n(\mu^*)\right\| \leq M$ 且 $\Delta_n(R) \leq 1$ 的事件上，函数 $u \mapsto U_n(J_{n,u}) + u^\top\sqrt{n}\,\nabla\widehat{G}_n(\mu^*)$ 在零点处为零，并且在半径为 R 的球面上严格为正。因此，根据凸性，该函数对于 $\|u\| \geq R$ 严格为正。最小值至多为零，因此 $\|\widehat{u}_n\| < R$。此外，$\|v_n\| \leq \left\|H^{-1}\right\|_{\mathrm{op}} M < R$。

二次函数 $\frac{1}{2}u^\top Hu + u^\top\sqrt{n}\,\nabla\widehat{G}_n(\mu^*)$ 在 $\widehat{u}_n$ 与 v_n 处的差等于 $\frac{1}{2}(\widehat{u}_n - v_n)^\top H(\widehat{u}_n - v_n)$。由于 $\widehat{u}_n$ 使 $u \mapsto U_n(J_{n,u}) + u^\top\sqrt{n}\,\nabla\widehat{G}_n(\mu^*)$ 取得最小值，由（B.2）可得

$$\tfrac{1}{2}\lambda_{\min}(H)\,\|\widehat{u}_n - v_n\|^2 \le 2\Delta_n(R).$$

对于每个 $\varepsilon > 0$，前述比较给出

$$\mathbb{P}[\|\widehat{u}_n - v_n\| > \varepsilon] \le \mathbb{P}[\left\|\sqrt{n}\,\nabla\widehat{G}_n(\mu^*)\right\| > M] + \mathbb{P}\left[\Delta_n(R) > \min\left(1, \frac{\lambda_{\min}(H)\varepsilon^2}{4}\right)\right].$$

对于上文选定的固定 M 和 R，由 (B.2)，第二个概率在 $n \to \infty$ 时趋于零。由马尔可夫不等式和 L^2 界可得

$$\limsup_{n\to\infty}\mathbb{P}[\|\widehat{u}_n - v_n\| > \varepsilon] \le \frac{\ell^2}{M^2}\mathbb{E}[\|\nabla_\mu g(\mathbf{X};\mu^*)\|^2].$$

令 M 趋于无穷，并将 $o_{\mathbb{P}}(1)$记为一个随机向量序列，该序列在 $n \to \infty$ 时依概率收敛到零，由此得到

$$\sqrt{n}\,(\widehat{\mu}_n - \mu^*) = -H^{-1}\sqrt{n}\,\nabla\widehat{G}_n(\mu^*) + o_{\mathbb{P}}(1). \tag{B.3}$$

Hoeffding 分解与渐近正态性。我们对 $h = \nabla_\mu g(\cdot;\,\mu^*)$ 应用引理 E.1。对于 $1 \le j \le \ell$，集合$\{1,\dots,n\}$ 中每个大小为 j 的子集，都包含在 $\binom{n-j}{\ell-j}$ 个大小为 ℓ 的子集中。因此，对$\{1,\dots,n\}$ 中所有大小为 ℓ 的子集上的 (E.2) 取平均，得到

$$\nabla\widehat{G}_n(\mu^*) = \sum_{j=1}^{\ell}\frac{\binom{\ell}{j}}{\binom{n}{j}}\sum_{\substack{A\subseteq\{1,\dots,n\}\\|A|=j}}\psi_j(X_A).$$

减去含有 $j = 1$ 的项，并利用引理 E.1 中的正交性，得到

$$\mathbb{E}\left[\left\|\nabla\widehat{G}_n(\mu^*) - \frac{\ell}{n}\sum_{i=1}^{n}\psi_1(X_i)\right\|^2\right] = \sum_{j=2}^{\ell}\frac{\binom{\ell}{j}^2}{\binom{n}{j}}\mathbb{E}\left[\|\psi_j(X_1,\dots,X_j)\|^2\right] = O(n^{-2}).$$

因此，

$$\sqrt{n}\,\nabla\widehat{G}_n(\mu^*) = \frac{\ell}{\sqrt{n}}\sum_{i=1}^{n}\psi_1(X_i) + o_{\mathbb{P}}(1).$$

随机向量 $\psi_1(X_i)$、$i \ge 1$ 相互独立同分布，均值为零，具有协方差 ζ_1 和有限二阶矩。Duflo（1997）的中心极限定理（见命题 D.5）适用于 $m = 1$。结合 Slutsky 定理，可得

$$\sqrt{n}\,\nabla\widehat{G}_n(\mu^*) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d(0, \ell^2\zeta_1).$$

方程（B.3）和 Slutsky 定理随后给出（3.4）。 □

B.2 引理 3.5 的证明

证明。对于彼此不同的非空正整数集合 A，B，且每个集合的大小至多为 ℓ，由引理 E.1 可得

$$\operatorname{Cov}\left(\psi_{|A|}(X_A), \psi_{|B|}(X_B)\right) = 0.$$

对于 $k \in \{0,\dots,\ell-1\}$，将 $\nabla_\mu g(X_1,\dots,X_\ell;\,\mu^*)$ 和 $\nabla_\mu g(X_{1+k},\dots,X_{\ell+k};\,\mu^*)$ 分解可得

$$\Gamma_k = \sum_{\varnothing\neq A\subseteq\{k+1,\dots,\ell\}}\operatorname{Var}\left(\psi_{|A|}(X_A)\right) = \sum_{j=1}^{\ell-k}\binom{\ell-k}{j}\operatorname{Var}\left(\psi_j(X_1,\dots,X_j)\right). \tag{B.4}$$

为对 k 求和，我们对 $1 \le j \le \ell$ 使用二项式恒等式 $\sum_{k=1}^{\ell-j} \binom{\ell-k}{j} = \binom{\ell}{j+1}$，并采用约定 $\binom{\ell}{\ell+1} = 0$。由于每个 Γ_k 都是对称的，由 (3.3) 可得

$$\begin{aligned}\Sigma = \Gamma_0 + 2\sum_{k=1}^{\ell-1}\Gamma_k &= \sum_{j=1}^{\ell}\left(\binom{\ell}{j} + 2\sum_{k=1}^{\ell-j}\binom{\ell-k}{j}\right)\mathrm{Var}\left(\psi_j(X_1,\ldots,X_j)\right) \\ &= \sum_{j=1}^{\ell}\left(\binom{\ell}{j} + 2\binom{\ell}{j+1}\right)\mathrm{Var}\left(\psi_j(X_1,\ldots,X_j)\right). \end{aligned} \tag{B.5}$$

对于 $1 \le j \le \ell$，我们有

$$\binom{\ell}{j} + 2\binom{\ell}{j+1} = \binom{\ell}{j}\frac{2\ell - j + 1}{j+1}.$$

由于 $\mathrm{Var}(\psi_1(X_1)) = \zeta_1$，由 (B.4) 和 (B.5) 可得

$$\Sigma - \ell^2\zeta_1 = \sum_{j=2}^{\ell}\binom{\ell}{j}\frac{2\ell - j + 1}{j+1}\mathrm{Var}\left(\psi_j(X_1,\ldots,X_j)\right) \succeq 0$$

and

$$\ell\Gamma_0 - \Sigma = \sum_{j=2}^{\ell}\binom{\ell}{j}\frac{(\ell+1)(j-1)}{j+1}\mathrm{Var}\left(\psi_j(X_1,\ldots,X_j)\right) \succeq 0.$$

Thus

$$\ell^2\zeta_1 \preceq \Sigma \preceq \ell\Gamma_0.$$

假设　设 $\ell \ge 2$。我们将 $\Sigma - \ell^2\zeta_1$ 与 $\ell\Gamma_0 - \ell^2\zeta_1$进行比较。令 $k = 0$，由公式 (B.4) 可得 s

$$\ell\Gamma_0 - \ell^2\zeta_1 = \ell\sum_{j=2}^{\ell}\binom{\ell}{j}\mathrm{Var}\left(\psi_j(X_1,\ldots,X_j)\right).$$

对于 $2 \le j \le \ell$，我们有 $3 \le j+1 \le \ell+1$，因此

$$1 \le \frac{2(\ell+1)}{j+1} - 1 = \frac{2\ell - j + 1}{j+1} \le \frac{2\ell - 1}{3}.$$

矩阵 $\mathrm{Var}(\psi_j(X_1,\ldots,X_j))$ 是半正定的，因此

$$\frac{1}{\ell}\left(\ell\Gamma_0 - \ell^2\zeta_1\right) \preceq \Sigma - \ell^2\zeta_1 \preceq \frac{2\ell-1}{3\ell}\left(\ell\Gamma_0 - \ell^2\zeta_1\right).$$

在左侧和右侧分别乘以 H^{-1}，即可得到引理后备注中所述的界：

$$\frac{1}{\ell}\left(V_{\mathrm{ASGD}} - V_{\mathrm{offline}}\right) \preceq V_{\mathrm{SW}} - V_{\mathrm{offline}} \preceq \frac{2\ell-1}{3\ell}\left(V_{\mathrm{ASGD}} - V_{\mathrm{offline}}\right).$$

对于 $\ell = 2$，两个系数均为 $1/2$，因此 $V_{\mathrm{SW}} = \frac{1}{2}(V_{\mathrm{offline}} + V_{\mathrm{ASGD}})$。

等式　y 种情况。假设 $\ell \ge 2$。令 $w \in \mathbb{R}^d$ 为一个固定的单位向量。根据 (B.4) 和 (B.5)　，

$$w^\top\left(\ell\Gamma_0 - \Sigma\right)w = \sum_{j=2}^{\ell}\binom{\ell}{j}\frac{(\ell+1)(j-1)}{j+1}\mathbb{E}\left[\left(w^\top\psi_j\left(X_1,\ldots,X_j\right)\right)^2\right].$$

$w^\top\left(\Sigma - \ell^2\zeta_1\right)w$ 的展开包含相同的项，且系数也都严格为正。因此，任一二次型为零，当且仅当对于每个$2 \le j \le \ell$，$w^\top\psi_j(X_1,\ldots,X_j) = 0$ 几乎处处成立。在这种情况下，我们令 $a(x) := w^\top\psi_1(x)$。Hoeffding 分解（E.2）给出

$$w^\top\nabla_\mu g\left(\mathbf{X};\mu^*\right) = \sum_{i=1}^{\ell} a\left(X_i\right) \qquad \text{a.s.}$$

我们证明该恒等式蕴含 $a(X_1) = 0$ 几乎处处成立。若 x_1，……， x_ℓ 属于半径为 r 的球，则向量 $x_i - x_\ell$、 $1 \le i \le \ell - 1$ 的范数至多为 $2r$。因此，单纯形 Simpl（$\mathbf{x}$）的体积的 Gram 行列式公式、Hadamard 不等式以及（A.1）给出

$$\left|w^\top \nabla_\mu g\left(\mathbf{x}; \mu^*\right)\right| \le \frac{m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{x}\right)\right)}{\ell} \le \frac{1}{\ell!} \prod_{i=1}^{\ell-1} \|x_i - x_\ell\| \le \frac{(2r)^{\ell-1}}{\ell!}.$$

令 $\varepsilon > 0$，并选择 $r > 0$ 使得 $(2r)^{\ell-1}/\ell! < \ell\varepsilon$。对于半径为 r 的每个球 B，在 $X_1, \ldots, X_\ell$ 属于 B 的事件上，我们有 $\sum_{i=1}^{\ell} a(X_i) = w^\top \nabla_\mu g(\mathbf{X}; \mu^*) < \ell\varepsilon$ 几乎处处成立。由独立性可得，

$$\mathbb{P}\left[X_1 \in B,\ a(X_1) > \varepsilon\right]^\ell = \mathbb{P}\left[X_i \in B,\ a(X_i) > \varepsilon \text{ for every } 1 \le i \le \ell\right] = 0.$$

由于 $\mathbb{R}^d$ 是半径为 r 的球的可数并，$\mathbb{P}[a(X_1) > \varepsilon] = 0$。由于 $\varepsilon > 0$ 是任意的，$a(X_1) \le 0$ 几乎处处成立。由于 $\mathbb{E}[a(X_1)] = 0$，我们有 $a(X_1) = 0$ 几乎处处成立。因此，$w^\top \nabla_\mu g(\mathbf{X}; \mu^*) = 0$ 几乎处处成立，或等价地，$w^\top \Gamma_0 w = 0$。反之，若 $w^\top \Gamma_0 w = 0$，则 $w^\top \nabla_\mu g(\mathbf{X}; \mu^*) = 0$ 几乎处处成立。因此，对于每个 $A \subseteq \{1, \ldots, \ell\}$，我们有 $\mathbb{E}[w^\top \nabla_\mu g(\mathbf{X}; \mu^*) \mid X_i,\ i \in A] = 0$ 几乎处处成立。根据 ψ_j 的定义，对于每个 $1 \le j \le \ell$，$w^\top \psi_j(X_1, \ldots, X_j) = 0$ 几乎处处成立。因此我们证明了

$$\ker\left(\Sigma - \ell^2 \zeta_1\right) = \ker\left(\ell\Gamma_0 - \Sigma\right) = \ker\left(\Gamma_0\right). \tag{B.6}$$

对于 $v \in \mathbb{R}^d$ 使得 $v^\top V_{\mathrm{ASGD}} v > 0$，我们有 $H^{-1}v \notin \ker(\Gamma_0)$。因此，由公式（B.6）可得

$$v^\top V_{\mathrm{offline}} v < v^\top V_{\mathrm{SW}} v < v^\top V_{\mathrm{ASGD}} v.$$

不同的协方差矩阵。我们通过反证法证明 $\Gamma_0 \neq 0$。假设 $\Gamma_0 = 0$。根据 Γ_0 的定义，$\nabla_\mu g(\mathbf{X}; \mu^*) = 0$ 几乎处处成立。公式（2.2）给出 $g(\mathbf{X}; \mu^*) = 0$ 几乎处处成立，因此在这种情况下 $G(\mu^*) = 0$。令 $u \in \mathbb{R}^d \setminus \{0\}$。由单纯形体积公式（2.1），对于每个 $t \in \mathbb{R}$，

$$G(\mu^* + tu) = |t| G(\mu^* + u).$$

由于 $G(\mu^*) = 0$，除以 $t > 0$ 并令 t 趋于零，得到 $G(\mu^* + u) = u^\top \nabla G(\mu^*) = 0$，这与假设 A1 中极小化点的唯一性相矛盾。因此 $\Gamma_0 \neq 0$。根据（B.6），两个半正定矩阵 $\Sigma - \ell^2\zeta_1$ 和 $\ell\Gamma_0 - \Sigma$ 均非零。因此，矩阵 V_{offline}、V_{SW} 和 V_{ASGD} 两两不同。

密度假设下的正定性。假设 A2 成立。我们证明 Γ_0 是正定的。单纯形体积的 Gram 行列式公式表明，对于任意 $\mathbf{x}$ 和 μ，

$$g(\mathbf{x}; \mu)^2 = \frac{1}{(\ell!)^2} \det\left(\left((x_i - \mu)^\top (x_j - \mu)\right)_{i,j=1}^{\ell}\right). \tag{B.7}$$

格拉姆矩阵的每个元素都是 $\mathbf{x}$ 和 μ 的坐标的多项式，因此 $g(\mathbf{x}; \mu)^2$ 也是如此。对于单位向量 w，方向导数

$$Q_w(\mathbf{x}) := \frac{d}{dt} g(\mathbf{x}; \mu^* + tw)^2 \Big|_{t=0}$$

因此，它是 $\mathbf{x}$ 的坐标的多项式。对单纯形体积公式 (2.1) 的平方求导，并利用 (2.2)，可得

$$Q_w(\mathbf{x}) = 2\left(\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell}\right)^2 w^\top \Gamma(\mathbf{x})(\mu^* - x_\ell) = 2g(\mathbf{x}; \mu^*)\, w^\top \nabla_\mu g(\mathbf{x}; \mu^*).$$

我们在超平面 $\{x \in \mathbb{R}^d\colon\ w^\top(x-\mu^*) = 1\}$ 中选取 ℓ 个仿射无关点 $x_1, \dots, x_\ell$。对于这个元组，$\Gamma(\mathbf{x})w = w$ 以及 $w^\top\Gamma(\mathbf{x})(\mu^* - x_\ell) = -1$，因此

$$Q_w(\mathbf{x}) = -2\left(\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell}\right)^2 < 0.$$

因此，Q_w 是一个非零多项式，所以其零点集的勒贝格测度为零。由于$\mathbf{X}$ 的联合分布具有勒贝格密度，$Q_w(\mathbf{X}) \neq 0$ 几乎处处成立，因此

$$w^\top\Gamma_0 w = \mathbb{E}\left[\left(w^\top\nabla_\mu g(\mathbf{X};\mu^*)\right)^2\right] > 0.$$

因此，Γ_0 是正定的。公式（B.6）给出 $\Sigma - \ell^2\zeta_1 \succ 0$ 和 $\ell\Gamma_0 - \Sigma \succ 0$。 □

B.3 不相交分块上的重复更新

对于固定整数 $r \geq 1$，一旦有新的由 ℓ个连续随机向量组成的区块可用，我们就执行 r 次连续的梯度更新。对于 $m \geq 1$，记 $B_m := (X_{(m-1)\ell+1}, \dots, X_{m\ell})$。从 $\mu_0^{(r)} := \mu_0$开始，对于 $j \geq 1$，定义

$$\mu_j^{(r)} := \mu_{j-1}^{(r)} - \gamma_{\lceil j/r\rceil}\nabla_\mu g\left(B_{\lceil j/r\rceil}; \mu_{j-1}^{(r)}\right),$$

其中 $\gamma_m := c_\gamma m^{-\gamma}$，$c_\gamma$ 以及固定指数 γ 均如第3节所述。每次更新后都会重新计算梯度。经过 $n \geq \ell$ 个观测值后，该递推已执行 $r\lfloor n/\ell\rfloor$ 次更新，其平均估计量为

$$\bar{\mu}_n^{(r)} := \frac{1}{r\lfloor n/\ell\rfloor}\sum_{j=1}^{r\lfloor n/\ell\rfloor}\mu_j^{(r)}.$$

命题 B.1。 在假设 *A1* 下，序列 $\left(\mu_j^{(r)}\right)_j$ 几乎处处收敛到 μ^*。若假设 *A2* 和 *A3* 成立，且 $\gamma\nu > 2$ 和 $\gamma(1+\alpha) > 1$，则

$$\sqrt{n}\left(\bar{\mu}_n^{(r)} - \mu^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d\left(0, \ell H^{-1}\Gamma_0 H^{-1}\right).$$

如果将步长 $\gamma_{\lceil j/r\rceil}$ 替换为 γ_j，则相同的结论仍然成立。

证明。 在整个证明过程中，我们省略上标 (r)。如附录 A 开头所述，我们在给定 $\mathcal{F}_0$，的条件下进行推导。为证明渐近正态性，我们采用 Polyak 和 Juditsky（1992年，定理 2）的平均化方法，与定理 3.2 中关于 $\bar{\mu}_n$ 的断言的证明相同。这里我们还控制了重复使用每个分块所造成的误差。我们首先考虑更新 j 时的步长 $\gamma_{\lceil j/r\rceil}$。对于每个分块索引 $m \geq 1$ 和每个 $1 \leq s \leq r$，有

$$\mu_{r(m-1)+s} = \mu_{r(m-1)+s-1} - \gamma_m\nabla_\mu g(B_m; \mu_{r(m-1)+s-1}).$$

对于任意 $m \geq 1$ 和 $0 \leq s \leq r$，包络函数 (A.1) 给出

$$\left\|\mu_{r(m-1)+s} - \mu_{r(m-1)}\right\| \leq \frac{s\gamma_m}{\ell}m_{\ell-1}(\mathrm{Simpl}(B_m)). \tag{B.8}$$

几乎必然收敛。对 r 次更新的次梯度不等式求和，并使用 (A.1) 中的 Lipschitz 界，得到

$$\begin{aligned}\|\mu_{rm} - \mu^*\|^2 \leq\ & \left\|\mu_{r(m-1)} - \mu^*\right\|^2 - 2r\gamma_m\left(g(B_m;\mu_{r(m-1)}) - g(B_m;\mu^*)\right)\\ &+ \frac{r^2\gamma_m^2}{\ell^2}m_{\ell-1}(\mathrm{Simpl}(B_m))^2.\end{aligned}$$

由归纳法可知，在条件分布下，每次迭代的结果均为平方可积。由于 B_m 与$\mathcal{F}_{(m-1)\ell}$独立，取条件期望可得

$$\begin{aligned}\mathbb{E}[\|\mu_{rm}-\mu^*\|^2 \mid \mathcal{F}_{(m-1)\ell}] \le \left\|\mu_{r(m-1)}-\mu^*\right\|^2 - 2r\gamma_m\big(G(\mu_{r(m-1)})-G(\mu^*)\big) \\ + \frac{r^2\gamma_m^2}{\ell^2}\mathbb{E}[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^2].\end{aligned} \tag{B.9}$$

假设 A1 给出 $\mathbb{E}[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^2] < \infty$。由于 $\sum_m \gamma_m = \infty$ and $\sum_m \gamma_m^2 < \infty$，Robbins–Siegmund 定理（Robbins 和 Siegmund，1971年，定理 1，第 233 页）表明，$\left(\|\mu_{rm}-\mu^*\|^2\right)_m$ 几乎处处收敛，并且 $\sum_m \gamma_m(G(\mu_{r(m-1)}) - G(\mu^*)) < \infty$ 几乎处处成立。因此，序列 $(\mu_{rm})_m$ 有界，且 $\liminf_m G(\mu_{rm}) = G(\mu^*)$ 几乎处处成立。由 G 的连续性及其极小化点的唯一性，某个子序列收敛到 μ^*。因此，平方距离的收敛性意味着 $(\mu_{rm})_m$ 几乎处处收敛到 μ^*。假设 A1 中的矩条件和 $\sum_m \gamma_m^2 < \infty$ 通过给定$\mathcal{F}_0$的条件期望得到，$\sum_m \gamma_m^2 m_{\ell-1}(\mathrm{Simpl}(B_m))^2 < \infty$ 几乎处处成立。因此，公式 (B.8) 意味着 $(\mu_j)_j$几乎处处收敛到 μ^*。

重复更新误差与收敛速率。 现在假设假设 A2 和 A3 成立，且 $\gamma\nu > 2$ 和 $\gamma(1+\alpha) > 1$。我们将 r 次更新之和与在 $\mu_{r(m-1)}$ 处进行的一次步长为 $r\gamma_m$ 的更新进行比较。定义

$$\varepsilon_m := \frac{1}{r}\sum_{s=1}^{r}\big(\nabla_\mu g(B_m;\mu_{r(m-1)+s-1}) - \nabla_\mu g(B_m;\mu_{r(m-1)})\big).$$

Then

$$\mu_{rm} = \mu_{r(m-1)} - r\gamma_m\big(\nabla_\mu g(B_m;\mu_{r(m-1)}) + \varepsilon_m\big).$$

公式 (B.8) 和 (C.1) 给出

$$\max_{0\le s\le r}\left\|\mu_{r(m-1)+s}-\mu_{r(m-1)}\right\| \le C\gamma_m \prod_{i=(m-1)\ell+1}^{m\ell}(1+\|X_i\|).$$

给定 $\mathcal{F}_{(m-1)\ell}$，初始点 $\mu_{r(m-1)}$ 是固定的，且 B_m 服从分布 $P^{\otimes\ell}$。因此，逐点界 (C.5) 和引理 C.2 中的期望界（取 $t = C\gamma_m$）给出

$$\mathbb{E}[\|\varepsilon_m\| \mid \mathcal{F}_{(m-1)\ell}] \le C\gamma_m. \tag{B.10}$$

令 $q \in (2/\gamma, \nu)$。我们定义

$$Y_m := \nabla_\mu g(B_m;\mu_{r(m-1)}) - \nabla G(\mu_{r(m-1)}) + \varepsilon_m - \mathbb{E}[\varepsilon_m \mid \mathcal{F}_{(m-1)\ell}].$$

序列 $(Y_m)_m$ 是关于滤过 $(\mathcal{F}_{m\ell})_m$ 的鞅差序列。根据 ε_m 的定义和包络函数 (A.1)，

$$\left\|\nabla_\mu g(B_m;\mu_{r(m-1)}) + \varepsilon_m\right\| = \left\|\frac{1}{r}\sum_{s=1}^{r}\nabla_\mu g(B_m;\mu_{r(m-1)+s-1})\right\| \le \frac{m_{\ell-1}(\mathrm{Simpl}(B_m))}{\ell}.$$

向量 Y_m 等于 $\nabla_\mu g(B_m;\mu_{r(m-1)}) + \varepsilon_m$ 减去其在给定 $\mathcal{F}_{(m-1)\ell}$ 时的条件期望。因此，由三角不等式、条件詹森不等式和 (C.2)，对每个 $m \ge 1$，

$$\mathbb{E}[\|Y_m\|^q \mid \mathcal{F}_{(m-1)\ell}] \le \frac{2^q}{\ell^q}\mathbb{E}[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^q] < \infty \quad \text{a.s.}$$

对于 $k \ge 0$，我们将 $\delta_k := \nabla G(\mu_k) - H(\mu_k - \mu^*)$ 记为泰勒余项。线性化递推为

$$\mu_{rm}-\mu^* = (I_d - r\gamma_m H)(\mu_{r(m-1)}-\mu^*) - r\gamma_m\big(Y_m + \mathbb{E}[\varepsilon_m \mid \mathcal{F}_{(m-1)\ell}] + \delta_{r(m-1)}\big).$$

令 m_0 为满足 $r\gamma_{m_0}\|H\|_{\mathrm{op}} \le 1/2$ 的最小正整数。该整数是 $\mathcal{F}_0$-可测的。由于 $(\gamma_m)_m$ 递减，对于每个 $m \ge m_0$，都有 $r\gamma_m\|H\|_{\mathrm{op}} \le 1/2$。对于 $m \ge k \ge m_0$，令

$$\Pi_{m,k} \coloneqq \prod_{j=k+1}^{m} (I_d - r\gamma_j H).$$

对于每个 $m \ge m_0$，从 m_0 开始迭代递推可得

$$\mu_{rm} - \mu^* = \Pi_{m,m_0}(\mu_{rm_0} - \mu^*) - \sum_{k=m_0+1}^{m} r\gamma_k \Pi_{m,k}\big(Y_k + \mathbb{E}[\varepsilon_k \mid \mathcal{F}_{(k-1)\ell}] + \delta_{r(k-1)}\big).$$

引理 D.4（延迟一阶，步长为 $r\gamma_k$）和引理 D.2（$\left\|\mathbb{E}[\varepsilon_k \mid \mathcal{F}_{(k-1)\ell}]\right\| \le C\gamma_k$）给出：对每个 $m > m_0$，

$$\|\mu_{rm} - \mu^*\| \le C_0\sqrt{\gamma_m \log m} + \sum_{k=m_0+1}^{m} r\gamma_k \|\Pi_{m,k}\|_{\mathrm{op}} \left\|\delta_{r(k-1)}\right\| \quad \text{a.s.}$$

对于有限随机变量 $C_0 \ge 0$。假设 A3 和几乎必然收敛给出 $\|\delta_{rm}\| = o(\|\mu_{rm} - \mu^*\|)$ 几乎处处成立。因此，将引理 D.3 应用于步长 $r\gamma_m$、$n_0 = m_0$ 以及 $v_m = (C_0+1)\sqrt{\gamma_m}\log m$，可得

$$\|\mu_{rm} - \mu^*\|^2 = O(m^{-\gamma}\log m) \quad \text{a.s.}$$

渐近正态性。令 $\xi_m^* \coloneqq \nabla_\mu g(B_m;\ \mu^*)$ 以及

$$D_m \coloneqq \nabla_\mu g(B_m; \mu_{r(m-1)}) - \nabla G(\mu_{r(m-1)}) - \xi_m^*.$$

序列 $(D_m)_m$ 是关于 $(\mathcal{F}_{m\ell})_m$ 的鞅差序列。根据 (F.1)，令 $\beta = (q-2)/(q-1)$，$(\mu_{rm})_m$ 的几乎必然收敛和条件支配收敛给出 $\mathbb{E}[\|D_m\|^2 \mid \mathcal{F}_0] \to 0$ 几乎处处成立。对于每个 $m \ge 1$，在条件分布下的正交性给出

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{m}}\sum_{k=1}^{m} D_k\right\|^2 \mid \mathcal{F}_0\right] = \frac{1}{m}\sum_{k=1}^{m}\mathbb{E}[\|D_k\|^2 \mid \mathcal{F}_0] \xrightarrow[m\to\infty]{} 0 \quad \text{a.s.}$$

界 (B.10) 和塔式性质给出

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\varepsilon_k\right\| \mid \mathcal{F}_0\right] \le \frac{C}{\sqrt{m}}\sum_{k=1}^{m}\gamma_k = O(m^{1/2-\gamma}) \xrightarrow[m\to\infty]{} 0 \quad \text{a.s.}$$

收敛速率、假设 A3 和 $\gamma(1+\alpha) > 1$ 蕴含

$$\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\left\|\delta_{r(k-1)}\right\| = O\Big(m^{(1-\gamma(1+\alpha))/2}(\log m)^{(1+\alpha)/2}\Big) \xrightarrow[m\to\infty]{} 0 \quad \text{a.s.}$$

分部求和给出

$$\sum_{k=1}^{m}\frac{\mu_{r(k-1)} - \mu_{rk}}{r\gamma_k} = \frac{\mu_0 - \mu^*}{r\gamma_1} - \frac{\mu_{rm} - \mu^*}{r\gamma_m} + \sum_{k=1}^{m-1}(\mu_{rk} - \mu^*)\left(\frac{1}{r\gamma_{k+1}} - \frac{1}{r\gamma_k}\right).$$

由于 $\gamma_{k+1}^{-1} - \gamma_k^{-1} = O(k^{\gamma-1})$，

$$\left\|\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\frac{\mu_{r(k-1)} - \mu_{rk}}{r\gamma_k}\right\| = O\Big(m^{(\gamma-1)/2}\sqrt{\log m}\Big) \xrightarrow[m\to\infty]{} 0 \quad \text{a.s.}$$

将递推式除以 $r\gamma_k$ 后求和，得到

$$\frac{H}{\sqrt{m}}\sum_{k=1}^{m}(\mu_{r(k-1)}-\mu^*)=\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\frac{\mu_{r(k-1)}-\mu_{rk}}{r\gamma_k}-\frac{1}{\sqrt{m}}\sum_{k=1}^{m}\xi_k^*-\frac{1}{\sqrt{m}}\sum_{k=1}^{m}(D_k+\varepsilon_k+\delta_{r(k-1)}).$$

此外，(B.8) 给出

$$\sqrt{m}\,\mathbb{E}\left[\left\|\bar{\mu}_{\ell m}-\frac{1}{m}\sum_{k=1}^{m}\mu_{r(k-1)}\right\|\mid\mathcal{F}_0\right]\le\frac{C\mathbb{E}[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))]}{\ell\sqrt{m}}\sum_{k=1}^{m}\gamma_k\xrightarrow[m\to\infty]{}0\quad\text{a.s.}$$

因此，

$$\sqrt{m}(\bar{\mu}_{\ell m}-\mu^*)=-\frac{H^{-1}}{\sqrt{m}}\sum_{k=1}^{m}\xi_k^*+o_{\mathbb{P}(\cdot|\mathcal{F}_0)}(1).$$

随机向量 ξ_k^* 独立同分布，均值为零，协方差为 Γ_0，且序列 $(\xi_k^*)_k$ 独立于 $\mathcal{F}_0$。Duflo（1997）的中心极限定理（在命题 D.5 中重述）在依赖参数为 1 时适用。结合 Slutsky 定理，可得弱收敛：

$$\mathcal{L}\left(\sqrt{m}(\bar{\mu}_{\ell m}-\mu^*)\mid\mathcal{F}_0\right)\xrightarrow[m\to\infty]{}\mathcal{N}_d(0,H^{-1}\Gamma_0H^{-1})\quad\text{a.s.}$$

由于极限分布是确定性的，附录 A.2 末尾使用有界连续函数的论证表明，在原始分布下也有相同的收敛性。

对于 $n\ge\ell$，令 $m=\lfloor n/\ell\rfloor$。根据定义，$\bar{\mu}_n=\bar{\mu}_{\ell m}$。由于 $n/m\to\ell$，Slutsky 定理给出所述收敛性。

按更新编序的步长。现在我们在第 j 次更新时使用步长 γ_j，并再次将迭代值记为 μ_j。对于每个 $m\ge1$ 和 $1\le s\le r$，

$$0\le\frac{\gamma_{r(m-1)+s}}{\gamma_{rm}}-1=\left(1-\frac{r-s}{rm}\right)^{-\gamma}-1\le\frac{C}{m}.$$

包络函数 (A.1) 给出将 γ_m 替换为 $C\gamma_{rm}$ 后的 (B.8)。由于 $\sum_{s=1}^{r}\gamma_{r(m-1)+s}\ge r\gamma_{rm}$ 和 $G(\mu_{r(m-1)})\ge G(\mu^*)$，次梯度计算同样给出将 γ_m 替换为 γ_{rm}、将 r^2 替换为 C 后的 (B.9)。因此，几乎必然收敛论证在假设 A1 下仍然适用。

如果我们设定下式，则将 γ_m 替换为 γ_{rm} 后，分块递推仍然成立：

$$\varepsilon_m\coloneqq\frac{1}{r}\sum_{s=1}^{r}\left(\frac{\gamma_{r(m-1)+s}}{\gamma_{rm}}\nabla_\mu g(B_m;\mu_{r(m-1)+s-1})-\nabla_\mu g(B_m;\mu_{r(m-1)})\right).$$

将系数 $\gamma_{r(m-1)+s}/\gamma_{rm}$ 替换为 1，会使该余项改变一个范数至多为 $Cm_{\ell-1}(\mathrm{Simpl}(B_m))/(m\ell)$ 的向量。在渐近正态性的假设下，引理 C.2 和包络函数因此给出

$$\mathbb{E}[\|\varepsilon_m\|\mid\mathcal{F}_{(m-1)\ell}]\le C(\gamma_{rm}+m^{-1})=O(\gamma_{rm}),\quad\text{and}\quad\|\varepsilon_m\|\le\frac{C}{\ell}m_{\ell-1}(\mathrm{Simpl}(B_m)).$$

包络函数和 (C.2) 给出了对于 $(Y_m)_m$，阶数为 q 的有界条件矩。由于 $\gamma_{rm}=r^{-\gamma}\gamma_m$ 是 γ_m 的固定正倍数，收敛速率和平均化论证可将 γ_m 替换为 γ_{rm} 后应用。

□

C 几何估计

根据 Dürre 和 Paindaveine（2022b）的引理 S.1.2，存在常数 $C>0$，使得对于每个 $\mathbf{x}\in(\mathbb{R}^d)^\ell$，

$$\frac{1}{\ell}m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{x}\right)\right)+g\left(\mathbf{x};0\right)\le C\prod_{i=1}^{\ell}\left(1+\|x_i\|\right).\tag{C.1}$$

在假设 A2 下，由独立性可得，对于 $0 < q < \nu$,

$$\mathbb{E}\left[m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{X}\right)\right)^{q}\right] < \infty, \quad \text{and} \quad \mathbb{E}\left[\prod_{i=1}^{\ell}\left(1+\|X_i\|\right)^{q}\right] = \left(\mathbb{E}\left[\left(1+\|X_1\|\right)^{q}\right]\right)^{\ell} < \infty. \tag{C.2}$$

在假设 A2 下， P 对维数至多为$\ell-1$的每个仿射子空间赋予零质量。因此，Dürre 和 Paindaveine（2022b）的引理 S.1.3 给出，对于每个 $u \in \mathbb{R}^d$,

$$\mathrm{dist}(u, \mathrm{aff}(\mathbf{X})) > 0 \quad \text{a.s.}$$

接下来，我们对 dist(u,aff(**X**)) 较小的加权概率给出一个界，该界对 u 一致成立。

令 Gr($\ell-1, d$) 表示 $\mathbb{R}^d$ 中所有 ($\ell-1$) 维线性子空间构成的集合，称为格拉斯曼流形。我们用 ϑ 表示其唯一的旋转不变的博雷尔概率测度（Schneider 和 Weil，2008年，定理 13.2.11）。每个维度为 $\ell-1$ 的仿射子空间都有唯一表示形式 $y+V$，其中 $V \in$ Gr($\ell-1, d$) 且 $y \in V^{\perp}$。根据仿射 Blaschke–Petkantschin 公式（ Miles，1971 年，式 (27)，第 361 页），存在常数 $c > 0$ ，使得对于定义在 $(\mathbb{R}^d)^{\ell}$ 上的任意非负可测函数 F ,

$$\begin{aligned}&\int_{(\mathbb{R}^d)^{\ell}} F(\mathbf{x})\, d\mathbf{x} \\ &\quad = c\int_{\mathrm{Gr}(\ell-1,d)}\int_{V^{\perp}}\int_{V^{\ell}} F(y+z_1,\ldots,y+z_{\ell})\, m_{\ell-1}(\mathrm{Simpl}(\mathbf{z}))^{d-\ell+1}\, d\mathbf{z}\, dy\, d\vartheta(V),\end{aligned} \tag{C.3}$$

其中 $\mathbf{z} \coloneqq (z_1, \ldots, z_{\ell})$ 且 $d\mathbf{z} = dz_1 \cdots dz_{\ell}$。测度 dy和 dz_i 分别是 $V^{\perp}$ 和 V上的欧几里得勒贝格测度，其维度分别为 $d-\ell+1$ 和 $\ell-1$。

对于 V 的任意线性子空间 $\mathbb{R}^d$，我们将 Γ_V 记为到 V上的正交投影。对于 $V \in$ Gr($\ell-1$, d)、$y \in V^{\perp}$ 和 $\mathbf{z} \in V^{\ell}$，我们设 $\mathbf{x} \coloneqq (y+z_1, \ldots, y+z_{\ell})$。对于几乎处处的 $\mathbf{z} \in V^{\ell}$，有 aff($\mathbf{z}$) $= V$，因此 aff($\mathbf{x}$) $= y+V$。于是，对于每个 $u \in \mathbb{R}^d$,

$$\Gamma(\mathbf{x}) = \Gamma_{V^{\perp}}, \quad \text{and} \quad \mathrm{dist}(u, \mathrm{aff}(\mathbf{x})) = \left\|\Gamma_{V^{\perp}} u - y\right\|.$$

引理 C.1。 在假设 *A2* 下，对于每个 $0 \le a < \nu - 1$ ，存在一个有限常数$C_a > 0$ ，使得对于每个 $t > 0$,

$$\sup_{u \in \mathbb{R}^d} \mathbb{E}\left[\frac{m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{X}\right)\right)}{\ell}\prod_{i=1}^{\ell}\left(1+\|X_i\|\right)^{a}\mathbf{1}_{\{\mathrm{dist}(u,\mathrm{aff}(\mathbf{X}))\le t\}}\right] \le C_a t^{d-\ell+1}.$$

证明。 对于 $V \in$ Gr($\ell-1$, d) 和 $y \in V^{\perp}$，我们设

$$b_a\left(V, y\right) \coloneqq \frac{c}{\ell}\int_{V^{\ell}} m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{z}\right)\right)^{d-\ell+2}\prod_{i=1}^{\ell}\left(\left(1+\|y+z_i\|\right)^{a} f\left(y+z_i\right)\right)\, d\mathbf{z}. \tag{C.4}$$

Since $y \perp V$，我们对每个 $z_i \in V$都有 $\|y+z_i\| \ge \|z_i\|$。因此，假设 A2 给出 s

$$(1+\|y+z_i\|)^{a} f(y+z_i) \le C_f(1+\|y+z_i\|)^{a-d-\nu} \le C_f(1+\|z_i\|)^{a-d-\nu}.$$

我们利用标准正交基将 V 与 $\mathbb{R}^{\ell-1}$ 等距同构。将此界与应用于$\mathbf{z}$的 (C.1) 结合，可得到一个常数 $C > 0$，使得对于每个 $V \in$ Gr($\ell-1, d$)和 $y \in V^{\perp}$,

$$\begin{aligned}b_a(V,y) &\le C\int_{V^{\ell}}\prod_{i=1}^{\ell}(1+\|z_i\|)^{-(\ell+\nu-2-a)}\, d\mathbf{z} \\ &= C\left(\int_{\mathbb{R}^{\ell-1}}(1+\|w\|)^{-(\ell+\nu-2-a)}\, dw\right)^{\ell}.\end{aligned}$$

由于 $a < \nu - 1$，我们有 $\ell + \nu - 2 - a > \ell - 1$，因此在 $\mathbb{R}^{\ell-1}$ 上的积分有限。所得的界与 V 和 y 无关，因此

$$\sup_{V \in \mathrm{Gr}(\ell-1,d)} \sup_{y \in V^\perp} b_a(V, y) < \infty.$$

对于每个 $u \in \mathbb{R}^d$ 和 $t > 0$，根据单纯形体积的平移不变性和仿射积分公式 (C.3)，可得

$$\begin{aligned}\mathbb{E}\left[\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\ell} \prod_{i=1}^{\ell} (1 + \|X_i\|)^a \mathbf{1}_{\{\mathrm{dist}(u, \mathrm{aff}(\mathbf{X})) \le t\}}\right] \\ = \int_{\mathrm{Gr}(\ell-1,d)} \int_{V^\perp} \mathbf{1}_{\{\|\Gamma_{V^\perp} u - y\| \le t\}} b_a(V, y)\, dy\, d\vartheta(V).\end{aligned}$$

集合 $\{y \in V^\perp \colon \|\Gamma_{V^\perp} u - y\| \le t\}$ 是 $V^\perp$ 中半径为 t 的球，其体积与 $t^{d-\ell+1}$ 成正比。将一致界在此球上积分，再关于概率测度 ϑ 积分，可得

$$\int_{\mathrm{Gr}(\ell-1,d)} \int_{V^\perp} \mathbf{1}_{\{\|\Gamma_{V^\perp} u - y\| \le t\}} b_a(V, y)\, dy\, d\vartheta(V) \le C t^{d-\ell+1} \int_{\mathrm{Gr}(\ell-1,d)} d\vartheta(V) = C t^{d-\ell+1}.$$

这就证明了该引理。 □

引理 C.2。 设 $\mathbf{x}$ 是一个仿射无关组。对于每个 $u \in \mathbb{R}^d$ |aff($\mathbf{x}$)、 $v \in \mathbb{R}^d$ 和$t > 0$，使得 $\|v - u\| \le t \prod_{i=1}^{\ell} (1 + \|x_i\|)$，有

$$\|\nabla_\mu g(\mathbf{x}; v) - \nabla_\mu g(\mathbf{x}; u)\| \le \begin{cases} \dfrac{2 m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell} \min\left(1, \dfrac{t \prod_{i=1}^{\ell} (1 + \|x_i\|)}{\mathrm{dist}(u, \mathrm{aff}(\mathbf{x}))}\right), & \text{if } \ell < d, \\ \dfrac{2 m_{d-1}(\mathrm{Simpl}(\mathbf{x}))}{d} \mathbf{1}_{\{\mathrm{dist}(u, \mathrm{aff}(\mathbf{x})) \le t \prod_{i=1}^{d} (1 + \|x_i\|)\}}, & \text{if } \ell = d. \end{cases} \tag{C.5}$$

在假设 *A2* 下，或者当 $2 \le \ell < d$ 时在假设 *A4* 和 *A5* 下，存在一个常数 $C \ge 0$，使得对于每个 $t > 0$,

$$\begin{aligned}\sup_{u \in \mathbb{R}^d} \mathbb{E}\left[\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\ell} \min\left(1, \frac{t \prod_{i=1}^{\ell} (1 + \|X_i\|)}{\mathrm{dist}(u, \mathrm{aff}(\mathbf{X}))}\right)\right] \le Ct \quad \text{if } \ell < d, \\ \sup_{u \in \mathbb{R}^d} \mathbb{E}\left[\frac{m_{d-1}(\mathrm{Simpl}(\mathbf{X}))}{d} \mathbf{1}_{\{\mathrm{dist}(u, \mathrm{aff}(\mathbf{X})) \le t \prod_{i=1}^{d} (1 + \|X_i\|)\}}\right] \le Ct \quad \text{if } \ell = d.\end{aligned} \tag{C.6}$$

对于 $\ell < d$，我们还有

$$\sup_{u \in \mathbb{R}^d} \mathbb{E}\left[\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{X})) \prod_{i=1}^{\ell} (1 + \|X_i\|)}{\ell\, \mathrm{dist}(u, \mathrm{aff}(\mathbf{X}))}\right] < \infty. \tag{C.7}$$

证明。 我们首先考虑 $\ell < d$ 的情形。对于非零向量 a, b，通过加上并减去 $b/\|a\|$，利用 $|\,\|a\| - \|b\|\,| \le \|a - b\|$ 和二的界，可得 $\left\| \frac{a}{\|a\|} - \frac{b}{\|b\|} \right\| \le 2 \min\left(1, \frac{\|a-b\|}{\|a\|}\right)$。我们设~~$a$~~ := Γ(**x**)($u - x_\ell$) 和 $b :=$ Γ(**x**)($v - x_\ell$)。由于 $u \notin$ **aff**(**x**)，我们有 $\|a\| =$ **dist**(u,**aff**(**x**)) > 0。投影 Γ(**x**) 是正交的，因此 $\|a - b\| \le \|u - v\|$。对于 $b \ne 0$，归一化向量的不等式和 (2.2) 给出了 (C.5) 的第一行。如果 $b = 0$，所选的次梯度满足 $\nabla_\mu g(\mathbf{x}; v) = 0$。在这种情况下，

$$\mathrm{dist}(u, \mathrm{aff}(\mathbf{x})) = \|a\| = \|\Gamma(\mathbf{x})(u - v)\| \le \|u - v\| \le t \prod_{i=1}^{\ell} (1 + \|x_i\|).$$

因此，(C.5) 第一行中的最小值为一，且该不等式的左侧等于 $m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))/\ell$。这也证明了 $b=0$ 时的第一行。

对于 $\ell=d$，存在一个垂直于超平面 $\mathrm{aff}(\mathbf{x})$ 的单位向量 $N\in\mathbb{R}^d$，且 $\Gamma(\mathbf{x})=NN^\top$。令 $\mathrm{sgn}(0):=0$，则由 (2.2) 可得，对任意 $\mu\in\mathbb{R}^d$，

$$\nabla_\mu g(\mathbf{x};\mu)=\frac{m_{d-1}(\mathrm{Simpl}(\mathbf{x}))}{d}N\,\mathrm{sgn}(N^\top(\mu-x_d)).$$

如果 $\mathrm{dist}(u,\mathrm{aff}(\mathbf{x}))>t\prod_{i=1}^d(1+\|x_i\|)$，则对 $\|v-u\|$ 所作的有界性假设意味着，u 和 v 位于由 $\mathrm{aff}(\mathbf{x})$ 界定的同一个开半空间内，因此 $\nabla_\mu g(\mathbf{x};u)=\nabla_\mu g(\mathbf{x};v)$。差的范数始终至多为 $2m_{d-1}(\mathrm{Simpl}(\mathbf{x}))/d$。这证明了 (C.5) 的第二行。

对于 $\ell<d$，我们先证明 (C.7)，然后推出 (C.6) 中的第一个不等式。界 (C.7) 是用到 $\mathrm{aff}(\mathbf{X})$ 的距离表示的假设 A5。在假设 A2 下，(C.2) 对 $\{\mathrm{dist}(u,\mathrm{aff}(\mathbf{X}))\ge 1\}$ 对 (C.7) 中期望的贡献给出了界，该界关于 u 是一致的。由引理 C.1，结合 $a=1$，存在一个常数 $C\ge 0$，使得对于任意 $u\in\mathbb{R}^d$ 和 $j\ge 0$，满足 $2^{-j-1}<\mathrm{dist}(u,\mathrm{aff}(\mathbf{X}))\le 2^{-j}$ 的部分被界定为

$$2^{j+1}\mathbb{E}\left[\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\ell}\prod_{i=1}^{\ell}(1+\|X_i\|)\mathbf{1}_{\{\mathrm{dist}(u,\mathrm{aff}(\mathbf{X}))\le 2^{-j}\}}\right]\le C2^{-j(d-\ell)}.$$

由于 $d-\ell\ge 1$，对 $j\ge 0$ 求和即可在假设 A2 下证明 (C.7)。在这两种情况下，用其第二个参数对最小值进行界定，即可得到 (C.6) 中的第一个不等式。

对于 $\ell=d$，假设 A2 成立。我们固定 $a\in(1,\ \nu-1)$，并根据 $2^j\le\prod_{i=1}^d(1+\|X_i\|)<2^{j+1}$（$j\ge 0$）进行划分。在这些事件中的每一个上，半径 $t\prod_{i=1}^d(1+\|X_i\|)$ 至多为 $t2^{j+1}$。将引理 C.1 中的 t 替换为 $t2^{j+1}$ 后应用该引理，可得到常数 $C,\ C'\ge 0$，使得

$$\begin{aligned}
&\mathbb{E}\left[\frac{m_{d-1}(\mathrm{Simpl}(\mathbf{X}))}{d}\mathbf{1}_{\{\mathrm{dist}(u,\mathrm{aff}(\mathbf{X}))\le t\prod_{i=1}^d(1+\|X_i\|)\}}\right]\\
&\quad\le\sum_{j=0}^{\infty}2^{-ja}\mathbb{E}\left[\frac{m_{d-1}(\mathrm{Simpl}(\mathbf{X}))}{d}\prod_{i=1}^{d}(1+\|X_i\|)^a\mathbf{1}_{\{\mathrm{dist}(u,\mathrm{aff}(\mathbf{X}))\le t2^{j+1}\}}\right]\\
&\quad\le Ct\sum_{j=0}^{\infty}2^{-j(a-1)}\le C't.
\end{aligned}$$

这就证明了 (C.6) 中的第二个不等式。 □

D 概率与矩阵估计

在本节中，$(\mathcal{F}_k)_{k\ge 0}$ 是一个任意的滤过。

D.1 具有固定延迟的鞅差

引理 D.1. 设 d 和 m 为正整数，且 $(U_k)_{k\ge 1}$ 是取值于 $\mathbb{R}^d$ 的平方可积随机向量序列，关于滤过 $(F_k)_{k\ge 0}$ 适应。假设对每个 $k\ge m$，

$$\mathbb{E}\left[U_k\mid\mathcal{F}_{k-m}\right]=0\quad a.s.$$

则以下结论成立。

(i) 若 $\sum_{k=1}^{\infty}\mathbb{E}\left[\|U_k\|^2\right]<\infty$，则 $\sum_{k=1}^{\infty}U_k$ 几乎处处收敛，且在 L^2 中收敛。

(ii) 若 $\sum_{k=m}^{\infty}\mathbb{E}\left[\|U_k\|^2\mid F_{k-m}\right]<\infty$ 几乎处处成立，则 $\sum_{k=1}^{\infty}U_k$ 几乎处处收敛。

(iii) 若

$$\sum_{k=m}^{\infty} \frac{1}{k} \mathbb{E}\left[\|U_k\|^2 \mid \mathcal{F}_{k-m}\right] < \infty \quad a.s.,$$

于是 $\sum_{k=1}^{n} U_k = o(\sqrt{n})$ 几乎处处成立。

证明。 对于 (i) 和 (ii)，固定 $a \in \{0, \ldots, m-1\}$，并令 $k_j := a + jm$ （ $j \geq 0$）。由于对于 $j \geq 1$ 有 $k_j - m = k_{j-1}$， $\sum_{j\geq 1} U_{k_j}$ 的部分和构成相对于 $\left(\mathcal{F}_{k_j}\right)_{j\geq 0}$ 的平方可积鞅。

在 (i) 的假设下，由正交性可知，对每个 $N \geq 1$,

$$\mathbb{E}\left[\left\|\sum_{j=1}^{N} U_{k_j}\right\|^2\right] = \sum_{j=1}^{N} \mathbb{E}\left[\left\|U_{k_j}\right\|^2\right] \leq \sum_{k=1}^{\infty} \mathbb{E}\left[\|U_k\|^2\right] < \infty.$$

因此，部分和在 L^2中有界。对每个坐标应用 Duflo（1997，命题1.3.7(2)，第16页），可得其几乎必然收敛且在 L^2中收敛。

对于 (ii)，该假设给出

$$\sum_{j=1}^{\infty} \mathbb{E}\left[\left\|U_{k_j}\right\|^2 \mid \mathcal{F}_{k_{j-1}}\right] \leq \sum_{k=m}^{\infty} \mathbb{E}\left[\|U_k\|^2 \mid \mathcal{F}_{k-m}\right] < \infty \quad \text{a.s.}$$

对该鞅的每个坐标应用 Duflo（1997，定理1.3.15(1)，第20页），可得 $\sum_{j\geq 1} U_{k_j}$ 几乎必然收敛。

对 m 剩余类求和可得 $\sum_{k=m}^{\infty} U_k$几乎处处收敛，并且在情形 (i) 下也在L^2 收敛。加上 $\sum_{k=1}^{m-1} U_k$ 即可证明 (i) 和 (ii)。

在 (iii) 的假设下，将断言 (ii) 应用于 $\left(U_k/\sqrt{k}\right)_{k\geq 1}$ 可知， $\sum_{k=1}^{\infty} U_k/\sqrt{k}$几乎处处收敛。将克罗内克引理（Duflo，1997 年，引理 1.3.14，第 19 页）应用于每个坐标，即可得到 (iii)。

□

D.2 矩阵乘积

在本小节中， H 是一个对称正定的 $d \times d$ 矩阵，c_γ 是一个正实数，且 $0 < \gamma < 1$。对于 $k \geq 1$，我们将 $\gamma_k := c_\gamma k^{-\gamma}$ 设为 $\lambda := \lambda_{\min}(H)$。与附录 A.2 中一样， n_H 是满足 $\gamma_{n_H} \|H\|_{\text{op}} \leq 1/2$的最小正整数。对于 $n \geq k \geq n_H - 1$，我们设

$$\Pi_{n,k} := \prod_{j=k+1}^{n} (I_d - \gamma_j H) \qquad \text{and} \qquad p_{n,k} := \prod_{j=k+1}^{n} (1 - \lambda\gamma_j).$$

引理 D.2. 对于每个 $n \geq n_H$,

$$\sum_{k=n_H}^{n} \gamma_k \|\Pi_{n,k}\|_{\text{op}} \leq \frac{1}{\lambda}.$$

对于每个 $a \geq 0$ 和 $b \in \mathbb{R}$,

$$\sum_{k=\max\{n_H,2\}}^{n} \gamma_k \|\Pi_{n,k}\|_{\text{op}} \, k^{-a} \left(\log k\right)^b = O\left(n^{-a} \left(\log n\right)^b\right), \tag{D.1}$$

$$\sum_{k=n_H}^{n} \gamma_k^2 \|\Pi_{n,k}\|_{\text{op}}^2 = O\left(\gamma_n\right). \tag{D.2}$$

对于每个固定的整数 $K \geq n_H - 1$，都存在 $c > 0$ 使得

$$\max_{n_H - 1 \leq k \leq K} \|\Pi_{n,k}\|_{\text{op}} = O\left(e^{-cn^{1-\gamma}}\right) \quad as\ n \to \infty.$$

证明。对角化 H，对于 $n \geq k \geq n_H - 1$，可得

$$\|\Pi_{n,k}\|_{\mathrm{op}} \leq p_{n,k} \leq \exp\left(-\lambda \sum_{j=k+1}^{n} \gamma_j\right) \leq 1. \tag{D.3}$$

对恒等式 $\lambda\gamma_k p_{n,k} = p_{n,k} - p_{n,k-1}$ 求和，利用望远镜求和可得，

$$\lambda \sum_{k=n_H}^{n} \gamma_k p_{n,k} = 1 - p_{n,n_H-1} \leq 1,$$

从而证明第一个不等式。对于 $k \leq n/2$，我们有 $\sum_{j=k+1}^{n} \gamma_j \geq (n-k)c_\gamma n^{-\gamma} \geq (c_\gamma/2)n^{1-\gamma}$。因此 $\|\Pi_{n,k}\|_{\mathrm{op}} \leq \exp\left(-\lambda c_\gamma n^{1-\gamma}/2\right)$。式 (D.1) 中每个求和项的其他因子至多呈多项式增长，因此对 $k \leq n/2$ 的求和为 $o\left(n^{-a}(\log\ n)^b\right)$。对于式 (D.1} 求和中的 $k > n/2$，我们有如下关于 $n \geq 4$ 的一致界：

$$2^{-|b|} \leq \frac{k^{-a}(\log k)^b}{n^{-a}(\log n)^b} = \left(\frac{n}{k}\right)^a \left(\frac{\log k}{\log n}\right)^b \leq 2^{a+|b|}.$$

引理的第一个不等式给出

$$\begin{aligned}\sum_{\substack{n_H \leq k \leq n \\ k > n/2}} \gamma_k \|\Pi_{n,k}\|_{\mathrm{op}}\, k^{-a}(\log k)^b &\leq 2^{a+|b|} n^{-a}(\log n)^b \sum_{k=n_H}^{n} \gamma_k \|\Pi_{n,k}\|_{\mathrm{op}} \\ &\leq \frac{2^{a+|b|}}{\lambda}\, n^{-a}(\log n)^b.\end{aligned}$$

结合 $k \leq n/2$ 的界，即可证明 (D.1)。

为证明 (D.2)，对 $k \leq n/2$ 的求和再次指数级小。对于 $k > n/2$，$\gamma_k \leq 2^\gamma \gamma_n$ 和 $\|\Pi_{n,k}\|_{\mathrm{op}} \leq 1$，因此

$$\sum_{k>n/2} \gamma_k^2 \|\Pi_{n,k}\|_{\mathrm{op}}^2 \leq 2^\gamma \gamma_n \sum_{k>n/2} \gamma_k \|\Pi_{n,k}\|_{\mathrm{op}} \leq \frac{2^\gamma}{\lambda}\gamma_n.$$

最后，对于固定的整数 $K \geq n_H - 1$，每个指标 $n_H - 1 \leq k \leq K$ 对于所有充分大的 n 都满足 $k \leq n/2$。因此，$k \leq n/2$ 的指数界证明了最后一个断言。 □

引理 D.3。令 $n_0 \geq n_H$ 为整数，令 $(u_n)_n$ 和 $(\kappa_n)_n$ 为非负实数序列，并令 $(v_n)_{n>n_0}$ 为正实数序列。假设对于每个 $n > n_0$，

$$u_n \leq v_n + \sum_{k=n_0+1}^{n} \gamma_k \|\Pi_{n,k}\|_{\mathrm{op}} \kappa_{k-1} u_{k-1}.$$

若 $(\kappa_n)_n$ 收敛到零，并且

$$\frac{v_{n-1}}{v_n} = 1 + o(\gamma_n),$$

则 $u_n = O(v_n)$。比值条件特别适用于 $v_n = cn^{-a}(\log n)^b$，其中 $c > 0$、$a \geq 0$ 和 $b \in \mathbb{R}$。

证明。设

$$s_n := \sum_{k=n_0+1}^{n} \gamma_k p_{n,k} \kappa_{k-1} u_{k-1}.$$

对于 $n > n_0$，由 (D.3) 得到 $u_n \le v_n + s_n$，因此

$$\begin{aligned} s_{n+1} &= (1 - \lambda\gamma_{n+1})s_n + \gamma_{n+1}\kappa_n u_n \\ &\le \big(1 - (\lambda - \kappa_n)\gamma_{n+1}\big)s_n + \gamma_{n+1}\kappa_n v_n. \end{aligned}$$

由于 $v_n/v_{n+1} = 1 + o(\gamma_{n+1})$ 和 $\kappa_n \to 0$，存在一个整数 $N > n_0$ 使得，对每个$n \ge N$，

$$\begin{aligned} \frac{s_{n+1}}{v_{n+1}} &\le \frac{v_n}{v_{n+1}}\left(\big(1 - (\lambda - \kappa_n)\gamma_{n+1}\big)\frac{s_n}{v_n} + \gamma_{n+1}\kappa_n\right) \\ &\le \left(1 - \frac{\lambda}{2}\gamma_{n+1}\right)\frac{s_n}{v_n} + \frac{\lambda}{2}\gamma_{n+1}. \end{aligned}$$

由归纳法，对每个 $n \ge N$，

$$\frac{s_n}{v_n} \le \max\left\{\frac{s_N}{v_N}, 1\right\}.$$

结论由 $u_n \le v_n + s_n$ 得出。对于 $v_n = cn^{-a}(\log n)^b$，我们有 $v_{n-1}/v_n = 1 + O(n^{-1}) = 1 + o(\gamma_n)$，因为 $\gamma < 1$。 □

以下引理将 Duflo（1997 年，《推论 6.4.25》，第 211 页）关于加权鞅和的结果推广到具有固定延迟的条件中心化。

引理 D.4。设 $1/2 < \gamma < 1$。令 $m \ge 1$ 为一个整数，并令 $(U_k)_{k\ge1}$ 为 $\mathbb{R}^d$ 中的 F_k-可测随机向量序列。假设存在 $q > 2/\gamma$ 和一个常数 $C \ge 0$，使得几乎处处对每个 $k \ge m$，

$$\mathbb{E}\left[U_k \mid \mathcal{F}_{k-m}\right] = 0, \quad \textit{and} \quad \mathbb{E}\left[\|U_k\|^q \mid \mathcal{F}_{k-m}\right] \le C.$$

Then

$$\left\|\sum_{k=n_H}^{n} \gamma_k \Pi_{n,k} U_k\right\| = O\left(\sqrt{\gamma_n \log n}\right) \quad a.s.$$

证明。令 $n_0 := \max\{n_H, m\}$。根据引理 D.2，存在 $c > 0$ 使得

$$\left\|\sum_{k=n_H}^{n_0-1} \gamma_k \Pi_{n,k} U_k\right\| = O\left(e^{-cn^{1-\gamma}}\right) \quad \text{a.s.}$$

剩下需要对 $n_0 \le k \le n$上的和给出界。为应用 Duflo 的推论，我们在必要时减小 q 并增大 C，使得 $2/\gamma < q < 4$；条件矩界由詹森不等式保持不变。

对于 $a \in \{0, \dots, m-1\}$ 和 $k \ge n_0$，令 $U_k^{(a)} := U_k \mathbf{1}_{\{k \equiv a \,(\text{模 } m)\}}$。对于 $k \ge n_0 - 1$，记 $\rho_a(k)$ 为满足 $j \equiv a \,(\text{模 } m)$ 的最大正整数 $j \le k$；若不存在，则记为零。对于 $k \ge n_0$，若 $k \equiv a \,(\text{模 } m)$，则有 $\rho_a(k-1) = k - m$，否则为 $U_k^{(a)} = 0$。因此，$\left(U_k^{(a)}\right)_{k\ge n_0}$ 适应于 $\left(\mathcal{F}_{\rho_a(k)}\right)_{k\ge n_0-1}$，并且对于每个 $k \ge n_0$，

$$\mathbb{E}[U_k^{(a)} \mid \mathcal{F}_{\rho_a(k-1)}] = 0, \quad \text{and} \quad \mathbb{E}[\left\|U_k^{(a)}\right\|^q \mid \mathcal{F}_{\rho_a(k-1)}] \le C.$$

条件詹森不等式还给出 $\mathbb{E}[\left\|U_k^{(a)}\right\|^2 \mid \mathcal{F}_{\rho_a(k-1)}] \le C^{2/q}$，正如 Duflo 的推论所要求的那样。

我们固定 H 的一个单位特征向量 e，其特征值为 $h > 0$，并令 $b_n := \prod_{i=n_0}^{n}(1 - h\gamma_i) > 0$ 对 $n \ge n_0 - 1$。对于 $n \ge n_0$，我们定义

$$M_n := \sum_{k=n_0}^{n} \frac{\gamma_k}{b_k} e^\top U_k^{(a)}, \qquad A_n := \sum_{k=n_0}^{n} \frac{\gamma_k^2}{b_k^2}.$$

由于 $\sum_n \gamma_n^2 < \infty$，展开 $\log(1-h\gamma_i)$ 得到 $-\log\, b_n \sim hc_\gamma n^{1-\gamma}/(1-\gamma)$。此外，$\gamma_{n-1}/\gamma_n = 1 + O(n^{-1}) = 1 + o(\gamma_n)$，因此

$$\frac{\gamma_n}{b_n^2} - \frac{\gamma_{n-1}}{b_{n-1}^2} = \frac{\gamma_n - \gamma_{n-1}(1-h\gamma_n)^2}{b_n^2} \sim 2h\frac{\gamma_n^2}{b_n^2}.$$

施托尔茨–切萨罗定理给出 $A_n \sim \gamma_n/(2hb_n^2)$。取对数得到

$$\log A_n = -2\log b_n + \log\gamma_n - \log(2h) + o(1) \sim \frac{2hc_\gamma}{1-\gamma} n^{1-\gamma}.$$

因此 $A_n \to \infty$ 以及 $\log\log A_n = O(\log\, n)$。由于 $\gamma q/2 > 1$，我们有

$$\sum_{n=n_0}^{\infty} \left(\frac{\gamma_n^2}{b_n^2 A_n}\right)^{q/2} < \infty.$$

我们将 Duflo（1997，第 211 页）的推论 6.4.25 应用于 $\left(e^\top U_k^{(a)}\right)_{k\geq n_0}$，其中 $\alpha = q/2 - 1$。该推论中的权重 Φ 及其界 T 均为 γ_k/b_k，且它们的平方和为A_n。因此

$$M_n = O\left(\sqrt{A_n \log\log A_n}\right) \quad \text{a.s.}$$

由于 $e^\top \Pi_{n,\,k} = (b_n/b_k)e^\top$，可得

$$\left| e^\top \sum_{k=n_0}^{n} \gamma_k \Pi_{n,k} U_k^{(a)} \right| = b_n |M_n| = O\left(\sqrt{\gamma_n \log n}\right) \quad \text{a.s.}$$

对 m 个剩余类求和，并使用 H 的标准正交特征基，即可证明该引理。

□

D.3 严格平稳序列与有限依赖

如果对于某个整数 $m \geq 0$，对每个 $a \geq 1$，由 $(Y_i)_{i\leq a}$ 和 $(Y_i)_{i\geq a+m+1}$ 生成的 σ-代数彼此独立，则称序列 $(Y_i)_{i\geq 1}$ 是 m-依赖的。

命题 D.5 (Duflo（1997，命题 2.1.7，第 43 页）)。设 $(Y_i)_{i\geq 1}$ 是 $\mathbb{R}^d$ 中的严格平稳、中心化、平方可积随机向量序列。假设它对于某个整数 $m \geq 1$ 是 m-依赖的。对于 $0 \leq j \leq m$，令

$$\Lambda_j \coloneqq \mathbb{E}\left[Y_1 Y_{1+j}^\top\right] \quad \textit{and} \quad \Sigma_Y \coloneqq \Lambda_0 + \sum_{j=1}^{m} \left(\Lambda_j + \Lambda_j^\top\right).$$

Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d\left(0, \Sigma_Y\right).$$

在独立情形下，该命题对 $m = 1$ 和 $\Sigma_Y = \Lambda_0$ 成立。

引理 D.6（带矩阵系数的和的中心极限定理）。设 d、$n_0 \geq 1$ 和 $m \geq 0$ 为整数，并设 $(Y_k)_{k\geq n_0}$ 为取值于 $\mathbb{R}^d$ 中的严格平稳、中心化、平方可积且 m-依赖的随机向量序列。设 H 为对称正定的 $d \times d$ 矩阵，令 $c_\gamma > 0$ 和 $0 < \gamma < 1$，并设 $\gamma_k \coloneqq c_\gamma k^{-\gamma}$，其中 $k \geq 1$。假设 $\gamma_{n_0} \|H\|_{\mathrm{op}} \leq 1/2$。对于 $n \geq k \geq n_0$，定义 $\Pi_{n,\,k} \coloneqq \prod_{r=k+1}^{n} (I_d - \gamma_r H)$，其中 $\Pi_{n,\,n} = I_d$。对于 $0 \leq j \leq m$，令 $\Lambda_j \coloneqq \mathbb{E}\left[Y_{n_0} Y_{n_0+j}^\top\right]$，并令

$$\Sigma_Y \coloneqq \Lambda_0 + \sum_{j=1}^{m} \left(\Lambda_j + \Lambda_j^\top\right).$$

Then

$$\frac{1}{\sqrt{\gamma_n}}\sum_{k=n_0}^{n}\gamma_k\Pi_{n,k}Y_k \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_d\left(0,\int_0^\infty e^{-tH}\Sigma_Y e^{-tH}\,dt\right).$$

证明。 我们应用 Orey中心极限定理，该定理针对 m-依赖的三角阵，其形式见 Rosén（1967年，定理 13.1，第 67–68 页）。
协方差收敛。对于 $n \geq n_0$，我们设

$$Z_n \coloneqq \sum_{k=n_0}^{n}\gamma_k\Pi_{n,k}Y_k \quad \text{and} \quad Q_n \coloneqq \operatorname{Var}(Z_n).$$

对于 $n > n_0+m$，递推 $Z_n = (I_d-\gamma_n H)Z_{n-1}+\gamma_n Y_n$、严格平稳性和 m-依赖性给出

$$Q_n = (I_d-\gamma_n H)Q_{n-1}(I_d-\gamma_n H)+\gamma_n^2\Lambda_0+\gamma_n\sum_{j=1}^{m}\gamma_{n-j}\left(\Pi_{n,n-j}\Lambda_j+\Lambda_j^\top\Pi_{n,n-j}\right).$$

对于 $1 \leq j \leq m$，我们有 $\gamma_{n-j}/\gamma_n = 1+O(n^{-1})$，并且 $\Pi_{n,\,n-j} = I_d+O(\gamma_n)$ 在算子范数下成立。由于$n^{-1} = o(\gamma_n)$，因此可得

$$Q_n = (I_d-\gamma_n H)Q_{n-1}(I_d-\gamma_n H)+\gamma_n^2\Sigma_Y+O(\gamma_n^3).$$

我们设 $V\coloneqq\int_0^\infty e^{-tH}\Sigma_Y e^{-tH}\,dt$。由于 H 正定，积分是有限的，对 $e^{-tH}\Sigma_Y e^{-tH}$ 的导数进行积分得到 $HV+VH=\Sigma_Y$。结合 $\gamma_{n-1}-\gamma_n = O(\gamma_n/n)$，可得

$$Q_n-\gamma_n V = (I_d-\gamma_n H)(Q_{n-1}-\gamma_{n-1}V)(I_d-\gamma_n H)+O(\gamma_n^3+\gamma_n/n).$$

我们设 $N\coloneqq n_0+m$。由压缩界 (D.3) 和引理 D.2，结合 (D.1) 中的 $a=2\gamma$ 和 $a=1$，通过迭代可得：存在常数 $C\geq 0$，使得对每个 $n>N$,

$$\begin{aligned}\|Q_n-\gamma_n V\|_{\mathrm{op}} &\leq \|\Pi_{n,N}\|_{\mathrm{op}}^2\,\|Q_N-\gamma_N V\|_{\mathrm{op}}+C\sum_{k=N+1}^{n}\gamma_k\,\|\Pi_{n,k}\|_{\mathrm{op}}^2\,(\gamma_k^2+k^{-1})\\ &= O(\gamma_n^2+n^{-1}).\end{aligned}$$

因此,

$$\operatorname{Var}\left(\frac{Z_n}{\sqrt{\gamma_n}}\right) = \frac{Q_n}{\gamma_n} = V+O\left(\gamma_n+\frac{1}{n\gamma_n}\right)\xrightarrow[n\to\infty]{} V.$$

林德伯格条件和极限分布。对于 $n_0 \leq k \leq n$，我们设

$$W_{n,k} \coloneqq \frac{\gamma_k}{\sqrt{\gamma_n}}\Pi_{n,k} \quad \text{and} \quad \delta_n \coloneqq \max_{n_0\leq k\leq n}\|W_{n,k}\|_{\mathrm{op}}.$$

界 (D.2) 给出 $\sum_{k=n_0}^{n}\|W_{n,\,k}\|_{\mathrm{op}}^2 = O(1)$。分别将 (D.3) 应用于 $k \leq n/2$ 和$k > n/2$，得到

$$\delta_n \leq \frac{\gamma_{n_0}}{\sqrt{\gamma_n}}\exp\left(-\frac{c_\gamma\lambda_{\min}(H)}{2}n^{1-\gamma}\right)+2^\gamma\sqrt{\gamma_n}\xrightarrow[n\to\infty]{}0.$$

我们固定一个单位向量 $v\in\mathbb{R}^d$，并设定 $X_{n,\,k}\coloneqq v^\top W_{n,\,k}Y_k$，其中 $n_0\leq k\leq n$。这些变量经过中心化，并构成一个 m-相依三角阵。严格平稳性和平方可积性表明，对每个 $\varepsilon>0$,

$$\sum_{k=n_0}^{n}\mathbb{E}\left[X_{n,k}^2\mathbf{1}_{\{|X_{n,k}|>\varepsilon\}}\right] \leq \left(\sum_{k=n_0}^{n}\|W_{n,k}\|_{\mathrm{op}}^2\right)\mathbb{E}\left[\|Y_{n_0}\|^2\mathbf{1}_{\{\|Y_{n_0}\|>\varepsilon/\delta_n\}}\right]\xrightarrow[n\to\infty]{}0.$$

此外，

$$\sum_{k=n_0}^{n} \operatorname{Var}(X_{n,k}) \le \mathbb{E}\left[\|Y_{n_0}\|^2\right] \sum_{k=n_0}^{n} \|W_{n,k}\|_{\mathrm{op}}^2 = O(1).$$

如果 $v^\top V v > 0$，则行和 σ_n^2 的方差 $= v^\top Q_n v/\gamma_n$： $v^\top V v$ 收敛。对于所有充分大的 n，标准化数组 $(X_{n,\,k}/\sigma_n)_{n_0\le k\le n}$ 满足 Rosén（1967年，定理 13.1）的假设，并且

$$v^\top \frac{Z_n}{\sqrt{\gamma_n}} = \sum_{k=n_0}^{n} X_{n,k} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, v^\top V v\right).$$

如果 $v^\top V v = 0$，则同一投影在 L^2 中收敛于零。由于 v 是任意的，根据 Cramér–Wold 定理可得所述向量收敛。 □

E Hoeffding 分解

我们回顾 经典的 Hoeffding 分解（Hoeffding，1961年），其形式用于附录 B.

引理 *E.1*。设 $(X_i)_{i\ge1}$ 是独立同分布随机向量，取值于 $\mathbb{R}^d$，并令 $\ell \ge 1$ 为整数。设 $\boldsymbol{X}:=(X_1,\ldots,X_\ell)$。令 $h:(\mathbb{R}^d)^\ell \to \mathbb{R}^d$ 为可测且对称的，并满足 $\mathbb{E}[h(\boldsymbol{X})]=0$ 和 $\mathbb{E}[\|h(\boldsymbol{X})\|^2]<\infty$。对于每个 $1\le j\le \ell$，存在可测对称函数 $\psi_j:(\mathbb{R}^d)^j\to\mathbb{R}^d$，满足 $\mathbb{E}[\|\psi_j(X_1,\ldots,X_j)\|^2]<\infty$，并且

$$\psi_j(X_1,\ldots,X_j) = \sum_{A\subseteq\{1,\ldots,j\}} (-1)^{j-|A|}\mathbb{E}[h(\mathbf{X}) \mid X_i,\ i\in A] \quad a.s. \tag{E.1}$$

对于一个有限的正整数集合 $A=\{i_1<\cdots<i_j\}$，我们设 $X_A := (X_{i_1},\ldots,X_{i_j})$。则

$$h(\mathbf{X}) = \sum_{\varnothing\ne A\subseteq\{1,\ldots,\ell\}} \psi_{|A|}(X_A) \quad a.s. \tag{E.2}$$

对于每个 个带有 $|A|\le\ell$ 且满足每个 $k\in A$ 的有限非空正整数集合 A，我们有

$$\mathbb{E}[\psi_{|A|}(X_A) \mid X_i,\ i\in A\setminus\{k\}] = 0 \quad a.s.$$

当 $A=\{k\}$ 时，该恒等式给出 $\mathbb{E}[\psi_1(X_k)]=0$。此外，对于满足 $|A|$、$|B|\le\ell$ 的不同有限非空正整数集合 A、B，我们有

$$\mathbb{E}[\psi_{|A|}(X_A)\psi_{|B|}(X_B)^\top] = 0.$$

证明。 我们将观测值的共同分布记为 P。对于 $1\le r<\ell$ 和 $\mathbf{x}\in(\mathbb{R}^d)^r$，我们定义

$$h_r(\mathbf{x}) := \begin{cases} \displaystyle\int h(\mathbf{x},\mathbf{y})\,P^{\otimes(\ell-r)}(d\mathbf{y}), & \text{if } \displaystyle\int \|h(\mathbf{x},\mathbf{y})\|\,P^{\otimes(\ell-r)}(d\mathbf{y})<\infty,\\ 0, & \text{otherwise.}\end{cases}$$

我们还设定 $h_\ell := h$。根据富比尼定理，函数 h_r 是可测的。 h的对称性意味着每个 h_r 都是对称的。对于每个 $\varnothing\ne A\subseteq\{1,\ldots,\ell\}$， h 的独立性和对称性给出

$$h_{|A|}(X_A) = \mathbb{E}[h(\mathbf{X}) \mid X_i,\ i\in A] \quad \text{a.s.}$$

对于 $1\le j\le\ell$ 和 $(x_1,\ldots,x_j)\in(\mathbb{R}^d)^j$，当 $A=\{i_1<\cdots<i_r\}\subseteq\{1,\ldots,j\}$ 时，我们记 $x_A := (x_{i_1},\ldots,x_{i_r})$，并定义

$$\psi_j(x_1,\ldots,x_j) := \sum_{\varnothing\ne A\subseteq\{1,\ldots,j\}} (-1)^{j-|A|} h_{|A|}(x_A).$$

函数 ψ_j 是可测且对称的。由于 $\mathbb{E}[h(\mathbf{X})]=0$，该定义给出 (E.1)。条件詹森不等式与有限和给出 $\mathbb{E}[\|\psi_j(X_1,\ldots,X_j)\|^2]<\infty$。

对于每个 $\varnothing\neq B\subseteq\{1,\ldots,\ell\}$，由二项式恒等式可得

$$\sum_{B\subseteq A\subseteq\{1,\ldots,\ell\}}(-1)^{|A|-|B|}=\sum_{r=0}^{\ell-|B|}(-1)^r\binom{\ell-|B|}{r}=\begin{cases}1, & \text{if } B=\{1,\ldots,\ell\},\\ 0, & \text{if } B\neq\{1,\ldots,\ell\}.\end{cases}$$

因此，

$$\begin{aligned}\sum_{\varnothing\neq A\subseteq\{1,\ldots,\ell\}}\psi_{|A|}(X_A)&=\sum_{\varnothing\neq A\subseteq\{1,\ldots,\ell\}}\sum_{\varnothing\neq B\subseteq A}(-1)^{|A|-|B|}h_{|B|}(X_B)\\&=\sum_{\varnothing\neq B\subseteq\{1,\ldots,\ell\}}\left(\sum_{B\subseteq A\subseteq\{1,\ldots,\ell\}}(-1)^{|A|-|B|}\right)h_{|B|}(X_B)\\&=h_\ell(\mathbf{X})=h(\mathbf{X}).\end{aligned}$$

这就证明了 (E.2)。

对于 $1\le j\le\ell$ 和 $B\subseteq\{1,\ldots,j-1\}$，$X_j$ 与 $(X_1,\ldots,X_{j-1})$ 的独立性给出

$$\mathbb{E}\Big[\mathbb{E}[h(\mathbf{X})\mid X_i,\ i\in B\cup\{j\}]\mid X_1,\ldots,X_{j-1}\Big]=\mathbb{E}[h(\mathbf{X})\mid X_i,\ i\in B]\quad\text{a.s.}$$

(E.1) 中由 B 和 $B\cup\{j\}$ 索引的项符号相反。因此

$$\mathbb{E}[\psi_j(X_1,\ldots,X_j)\mid X_1,\ldots,X_{j-1}]=0\quad\text{a.s.}$$

由对称性可知，在对任意一个 $j-1$ 个参数进行条件化后，也有相同的恒等式。由同分布性可知，对于每个集合 A 和每个 $k\in A$，所述条件中心化均成立。

对于最后一个断言，在必要时交换 A 和 B 并取转置，我们不妨设 $A\setminus B\neq\varnothing$。固定 $k\in A\setminus B$，并令 $\mathcal{G}:=\sigma(X_i:\ i\in(A\cup B)\setminus\{k\})$。随机向量 X_A 与 $(X_i)_{i\in B\setminus A}$ 独立。因此，刚刚证明的条件中心化给出 $\mathbb{E}[\psi_{|A|}(X_A)\mid\mathcal{G}]=0$ 几乎处处成立。由于 $\psi_{|B|}(X_B)$ 是 $\mathcal{G}$-可测的，

$$\mathbb{E}[\psi_{|A|}(X_A)\psi_{|B|}(X_B)^\top]=\mathbb{E}\left[\mathbb{E}[\psi_{|A|}(X_A)\mid\mathcal{G}]\,\psi_{|B|}(X_B)^\top\right]=0.$$

□

F 目标函数的性质

在假设 A2 下，我们建立 G 的正则性，并验证假设 A1。随后证明命题 2.1，该命题给出了假设 A3 中局部展开的一个充分条件。

F.1 g 和 G的正则性

命题 F.1。在假设 *A2* 下，G 在 $\mathbb{R}^d$ 上有限、严格凸、强制且可微，并且满足

$$\nabla G(u)=\mathbb{E}\left[\nabla_\mu g(\mathbf{X};u)\right],\quad and\quad \|\nabla G(u)\|\le\frac{1}{\ell}\mathbb{E}\left[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))\right].$$

梯度 ∇G 是 *Lipschitz* 的。此外，对于每个 $q\in(2,\nu)$，其中 $\beta:=(q-2)/(q-1)\in(0,1)$，存在常数 $C\ge0$，使得对于每个 u、$v\in\mathbb{R}^d$，

$$\mathbb{E}\left[\|\nabla_\mu g(\mathbf{X};u)-\nabla_\mu g(\mathbf{X};v)\|^2\right]\le C\min\left(\|u-v\|^\beta,1\right).\tag{F.1}$$

特别地，假设 *A1* 成立。

证明。 G 的有限性和凸性由 (C.1) 及 (2.1) 得出。对于任意 $u \in \mathbb{R}^d$，如附录 C 开头所述，我们几乎处处有$u \notin \mathrm{aff}(\mathbf{X})$。因此，底高公式 (2.1) 表明， $g(\mathbf{X};\cdot)$ 几乎处处在 u 处可微。包络函数 (A.1) 给出，对于 $z \neq 0$,

$$\frac{|g(\mathbf{X};u+z) - g(\mathbf{X};u) - \langle \nabla_\mu g(\mathbf{X};u), z\rangle|}{\|z\|} \leq \frac{2m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\ell}.$$

控制收敛定理适用：控制函数可积，且当 $z \to 0$时，商几乎处处趋于零。因此,

$$\frac{G(u+z) - G(u) - \langle \mathbb{E}[\nabla_\mu g(\mathbf{X};u)], z\rangle}{\|z\|} \xrightarrow[z\to 0]{} 0.$$

因此， G 可微，且具有所述梯度。由于 P 具有密度和有限一阶矩，Dürre 和 Paindaveine（2022a）的定理 2.2(ii$_a$) 和 2.3(i$_a$) 可推出 G具有强制性，因而极小化点存在。

接下来证明严格凸性。设 $u, v \in \mathbb{R}^d$互不相同，且令 $t \in$属于 (0,1)。首先假设 $\ell < d$。对于每个 $1 \leq i \leq \ell$，在给定 $X_1, \ldots, X_{i-1}$的条件下，向量 X_i服从分布 P。仿射空间 $\mathrm{aff}(u, v, X_1, \ldots, X_{i-1})$ 的维数至多为 $i < d$，因此 X_i 以条件概率零属于该空间。于是，点 $u, v, X_1, \ldots, X_\ell$几乎处处仿射无关。因此， $\Gamma(\mathbf{X})(u - X_\ell)$ 与 $\Gamma(\mathbf{X})(v - X_\ell)$ 几乎处处不共线。因此，三角不等式不可能取等号。结合 (2.1)，可得

$$g(\mathbf{X}; (1-t)u + tv) < (1-t)\, g(\mathbf{X};u) + t g(\mathbf{X};v) \quad \text{a.s.}$$

取期望即可证明当 $\ell < d$ 时的严格凸性。

现在假设 $\ell = d$。为证明严格凸性，我们首先证明，以正概率，仿射包 $\mathrm{aff}(\mathbf{X})$ 将 u 与v 严格分离。集合 $E := \{x \in \mathbb{R}^d : f(x) > 0\}$ 具有正的勒贝格测度。由于 $d \geq 2$，直线 $\mathrm{aff}(u, v)$ 的勒贝格测度为零。根据勒贝格密度定理，存在集合 E 的一个密度点 $x_0 \in E \setminus \mathrm{aff}(u, v)$，使得

$$m_d(B(x_0, r) \setminus E) = o(r^d) \qquad \text{as } r \downarrow 0.$$

向量 量 $u - x_0$ 和 $v - x_0$ 线性无关。存在一个单位向量 e_0，使得 at

$$e_0^\top (u - x_0) > 0, \quad \text{and} \quad e_0^\top (v - x_0) < 0.$$

我们选取仿射无关点 $z_1, \ldots, z_d \in e_0^\perp \cap B(\mathbf{0}, 1/2)$。格拉姆矩阵的逆的连续性表明， Γ 在仿射无关元组构成的集合上连续。每当$\Gamma(\mathbf{x})e_0 \neq 0$ 时，我们令 $e(\mathbf{x}) := \Gamma(\mathbf{x})e_0 / \|\Gamma(\mathbf{x})e_0\|$。由于 $\Gamma(z_1, \ldots, z_d)e_0 = e_0$，函数 e 在 $(z_1, \ldots, z_d)$ 附近给出了一个连续的单位法向量。因此存在常数 $\varepsilon > 0$ 和 $0 < \delta < 1/2$，使得每个元组 $\mathbf{x} \in \prod_{i=1}^d B(z_i, \delta)$ 都是仿射无关的， $e(\mathbf{x})$ 有定义，并且

$$e(\mathbf{x})^\top (u - x_0) > \varepsilon, \quad \text{and} \quad e(\mathbf{x})^\top (v - x_0) < -\varepsilon.$$

对于 $0 < r < \varepsilon$，我们设

$$B_i(r) := B(x_0 + r z_i, \delta r), \qquad 1 \leq i \leq d.$$

这些球包含于 $B(x_0, r)$ 中。对于每个 $\mathbf{x} \in \prod_{i=1}^d B_i(r)$，对每个 i，我们都有 $(x_i - x_0)/r \in B(z_i, \delta)$。平移和伸缩保持仿射独立性，并且使 Γ，从而也使 e，保持不变。因此，$\mathbf{x}$ 是仿射无关的，且 $e(\mathbf{x})$ 有定义。由于 $\|x_d - x_0\| < r$，我们得到

$$e(\mathbf{x})^\top (u - x_d) > \varepsilon - r > 0, \quad \text{and} \quad e(\mathbf{x})^\top (v - x_d) < -\varepsilon + r < 0.$$

因此， $\mathrm{aff}(\mathbf{x})$ 严格分离 u和 v 。

由于 $m_d(B(x_0,\ r)\setminus E)=o(r^d)$ 当 $r\downarrow 0$ 时，因此我们固定 $r\in(0,\ \varepsilon)$，使得

$$m_d(B(x_0,r)\setminus E)<r^d m_d(B(0,\delta)).$$

对于每个 $1\le i\le d$，我们于是有

$$\begin{aligned} m_d(E\cap B_i(r)) &\ge m_d(B_i(r))-m_d(B(x_0,r)\setminus E)\\ &= r^d m_d(B(0,\delta))-m_d(B(x_0,r)\setminus E)>0. \end{aligned}$$

由于 $f>0$ 在 E上，我们有 $P(B_i(r))>0$，对每个 i均成立。由独立性可知，$X_i\in B_i(r)$ 对所有 i都成立的概率为正。在此事件上，X_1、…、X_d仿射无关，因此$m_{d-1}(\mathrm{Simpl}(\mathbf{X}))>0$。此外，$e(\mathbf{X})^\top(u-X_d)>0$，且 $e(\mathbf{X})^\top(v-X_d)<0$。因此，由底高公式 (2.1) 得

$$\begin{aligned} g(\mathbf{X};(1-t)u+tv) &= \frac{m_{d-1}(\mathrm{Simpl}(\mathbf{X}))}{d}\left|(1-t)e(\mathbf{X})^\top(u-X_d)+te(\mathbf{X})^\top(v-X_d)\right|\\ &< (1-t)g(\mathbf{X};u)+tg(\mathbf{X};v). \end{aligned}$$

对于每个 $\mathbf{x}\in(\mathbb{R}^d)^d$，函数 $g(\mathbf{x};\cdot)$ 都是凸的。由于严格不等式在一个概率为正的事件上成立，取期望即可证明，在 $\ell=d$ 的情形下，G 严格凸。

现在已证明，对每个 $1\le\ell\le d$，严格凸性都成立，因此 G 有唯一极小元。由于$\nu>2$，由 (C.2) 及 $q=2$ 可得假设 A1 中的二阶矩条件。

由于 $\prod_{i=1}^{\ell}(1+\|X_i\|)\ge 1$，引理 C.2 中的界 (C.5) 和 (C.6)，取 $t=\|u-v\|$，可得一个常数 $C\ge 0$，使得对每个 $u,\ v\in\mathbb{R}^d$，

$$\mathbb{E}[\|\nabla_\mu g(\mathbf{X};u)-\nabla_\mu g(\mathbf{X};v)\|]\le C\,\|u-v\|\,.$$

∇G 的公式因此表明 ∇G 是 Lipschitz 的。为证明 (F.1)，我们固定 $q\in(2,\ \nu)$，并令 $U:=\|\nabla_\mu g(\mathbf{X};u)-\nabla_\mu g(\mathbf{X};\ v)\|$。由包络函数 (A.1) 和矩界 (C.2)，

$$\mathbb{E}[U^q]\le\left(\frac{2}{\ell}\right)^q\mathbb{E}[m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))^q]<\infty.$$

由 Hölder 不等式，取 $\beta=(q-2)/(q-1)$，可得

$$\mathbb{E}[U^2]\le(\mathbb{E}[U])^{\beta}\left(\mathbb{E}[U^q]\right)^{1/(q-1)}. \tag{F.2}$$

将这些界与 $\mathbb{E}[U^2]\le 4\mathbb{E}[m_{\ell-1}(\mathbf{Simpl(X)})^2]/\ell^2$ 合并，即可证明 (F.1)。 □

F.2 Hessian 矩阵与局部展开

命题 F.2。 在假设 *A2* 下，如果 $\ell<d$，那么 ∇G 在每个 $u\in\mathbb{R}^d$ 处均可微，并且

$$\nabla^2 G(u)=\frac{1}{\ell}\mathbb{E}\left[\frac{m_{\ell-1}\left(\mathrm{Simpl}\left(\mathbf{X}\right)\right)}{\left\|\Gamma\left(\mathbf{X}\right)\left(u-X_\ell\right)\right\|}\left(\Gamma\left(\mathbf{X}\right)-\frac{\Gamma\left(\mathbf{X}\right)\left(u-X_\ell\right)\left(u-X_\ell\right)^\top\Gamma\left(\mathbf{X}\right)}{\left\|\Gamma\left(\mathbf{X}\right)\left(u-X_\ell\right)\right\|^2}\right)\right].$$

对于每个 $u\in\mathbb{R}^d$，分母几乎处处非零。

证明。 令 $u\in\mathbb{R}^d$。在假设 A2 下，P具有有限一阶矩，并且对每个维数为 $\ell-1$ 的仿射子空间赋予的概率均为零。因此，Dürre 和 Paindaveine（2022b）的引理 S.1.3(ii) 给出 $\|\Gamma(\mathbf{X})(u-X_\ell)\|>0$ 几乎处处成立。由于 $\prod_{i=1}^{\ell}(1+\|X_i\|)\ge 1$，由 (C.7) 得到

$$\mathbb{E}\left[\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\|\Gamma(\mathbf{X})(u-X_\ell)\|}\right]<\infty.$$

Dürre 和 Paindaveine（2022b）的引理 S.2.3 应用于 $p=1$ 和 $\mu_0=u$，现在给出了 ∇G 在 u 处的可微性及所述公式，该公式即为其方程 (S.2.4)。 □

命题 *2.1* 的证明。 根据 (C.4) 中结合 $a=0$ 的定义，对于 $V\in \mathbf{Gr}(\ell-1,\ d)$ 和 $y\in V^{\perp}$，我们有

$$b_0(V,y)=\frac{c}{\ell}\int_{V^{\ell}} m_{\ell-1}(\mathrm{Simpl}(\mathbf{z}))^{d-\ell+2}\prod_{i=1}^{\ell} f(y+z_i)\,d\mathbf{z}.$$

公式 (C.3) 给出

$$G\left(u\right)=\int_{\mathrm{Gr}(\ell-1,d)}\int_{V^{\perp}} b_0\left(V,y\right)\left\|\Gamma_{V^{\perp}}u-y\right\|\,dy\,d\vartheta\left(V\right). \tag{F.3}$$

的界 b_0。 我们首先构造一个有界且可积的函数，以控制b_0 在 y中的变化。对于 $V\in \mathbf{Gr}(\ell-1,\ d)$ 和 $y\in V^{\perp}$，定义

$$\widetilde{b}(V,y):=\frac{c}{\ell}\int_{V^{\ell}} m_{\ell-1}(\mathrm{Simpl}(\mathbf{z}))^{d-\ell+2}\prod_{i=1}^{\ell}(1+\|y+z_i\|)^{-d-\nu}\,d\mathbf{z}.$$

根据假设 A2，$0\le b_0(V,\ y)\ \le C_f^{\ell}\widetilde{b}(V,\ y)$。由于 $y\perp V$，对于每个$z_i\in V$都有 $\|y+z_i\|\ge\|z_i\|$。将 (C.1) 应用于$\mathbf{z}$，可得一个常数 $C>0$，使得

$$m_{\ell-1}(\mathrm{Simpl}(\mathbf{z}))^{d-\ell+2}\prod_{i=1}^{\ell}(1+\|y+z_i\|)^{-d-\nu}\le C\prod_{i=1}^{\ell}(1+\|z_i\|)^{-(\ell-1)-(\nu-1)}.$$

右侧关于 V^{ℓ}的积分不依赖于 V，并且是有限的，因为 $\dim\ V=\ell-1$且 $\nu>1$。仿射积分公式 (C.3) 还给出

$$\int_{\mathrm{Gr}(\ell-1,d)}\int_{V^{\perp}}\widetilde{b}(V,y)\,dy\,d\vartheta(V)=\int_{(\mathbb{R}^d)^{\ell}}\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell}\prod_{i=1}^{\ell}(1+\|x_i\|)^{-d-\nu}\,d\mathbf{x}.$$

应用 (C.1) 并在必要时增大 C 后，我们得到

$$\begin{aligned}&\sup_{V\in\mathrm{Gr}(\ell-1,d)}\ \sup_{y\in V^{\perp}}\widetilde{b}(V,y)<\infty,\\ &\int_{\mathrm{Gr}(\ell-1,d)}\int_{V^{\perp}}\widetilde{b}(V,y)\,dy\,d\vartheta(V)\le C\left(\int_{\mathbb{R}^d}(1+\|x\|)^{-d-\nu+1}\,dx\right)^{\ell}<\infty.\end{aligned} \tag{F.4}$$

特别地，b_0 是一致有界的。

根据假设 A2 以及命题陈述中关于 ∇f 的界，存在常数 $C>0$，使得对于每个 $V\in \mathbf{Gr}(\ell-1,\ d)$，$y,\ v\in V^{\perp}$ 以及 $\mathbf{z}\ \in V^{\ell}$，

$$\begin{aligned}\left|\prod_{i=1}^{\ell}f(y+v+z_i)-\prod_{i=1}^{\ell}f(y+z_i)\right|&=\left|\int_0^1\sum_{j=1}^{\ell}\langle\nabla f(y+sv+z_j),v\rangle\prod_{i\neq j}f(y+sv+z_i)\,ds\right|\\ &\le C\left\|v\right\|\int_0^1\prod_{i=1}^{\ell}(1+\|y+sv+z_i\|)^{-d-\nu}\,ds.\end{aligned}$$

公式 (C.4) 和 Tonelli 定理给出

$$\left|b_0(V,y+v)-b_0(V,y)\right|\le C\left\|v\right\|\int_0^1\widetilde{b}(V,y+sv)\,ds. \tag{F.5}$$

由 (F.4) 可知，存在常数 $L_{b_0}\ge 0$，使得对于每个 $V\in \mathbf{Gr}(\ell-1,\ d)$，$b_0(V,\ \cdot)$ 在 $V^{\perp}$ 上是 L_{b_0}-Lipschitz 的。

Hessian *n* 当 $\ell<d$。 </style}对于 $\ell<d$，命题 F.2 和仿射积分公式 (C.3) 给出 ve

$$\nabla^2G\left(u\right)=\iint\frac{b_0\left(V,y\right)}{\left\|\Gamma_{V^{\perp}}u-y\right\|}\left(\Gamma_{V^{\perp}}-\frac{\left(\Gamma_{V^{\perp}}u-y\right)\left(\Gamma_{V^{\perp}}u-y\right)^{\top}}{\left\|\Gamma_{V^{\perp}}u-y\right\|^2}\right)dy\,d\vartheta\left(V\right). \tag{F.6}$$

对于 $y \neq \Gamma_{V^\perp} u$，(F.6) 中括号内的矩阵之差是到$V^\perp \cap (\Gamma_{V^\perp} u - y)^\perp$上的正交投影，因此其算子范数至多为一。对于 $u,\ w \in \mathbb{R}^d$，在 $\nabla^2 G(u+w)$ 中作变量代换$y \mapsto y + \Gamma_{V^\perp} w$，并利用界 (F.5)，可得

$$\begin{aligned}\left\|\nabla^2 G(u+w) - \nabla^2 G(u)\right\|_{\mathrm{op}} &\leq \iint \frac{|b_0(V, y+\Gamma_{V^\perp} w) - b_0(V,y)|}{\|\Gamma_{V^\perp} u - y\|}\, dy\, d\vartheta(V) \\ &\leq C\,\|w\| \int_0^1 \iint \frac{\widetilde{b}(V, y + s\Gamma_{V^\perp} w)}{\|\Gamma_{V^\perp} u - y\|}\, dy\, d\vartheta(V)\, ds \\ &= C\,\|w\| \int_0^1 \iint \frac{\widetilde{b}(V,z)}{\|\Gamma_{V^\perp}(u+sw) - z\|}\, dz\, d\vartheta(V)\, ds,\end{aligned}$$

其中，最后一个等号由变量代换 $z = y + s\Gamma_{V^\perp} w$ 得出。最后一行中关于 V 和 z 的二重积分在 u, w 和 $s \in [0,1]$上一致有界。事实上，在集合$\|\Gamma_{V^\perp}(u+sw) - z\| \leq 1$上，(F.4) 中的一致界适用，且 $\|z\|^{-1}$ 在 $d - \ell + 1 \geq 2$维单位球上可积。在集合 $\|\Gamma_{V^\perp}(u+sw) - z\| > 1$上，被积函数由 $\sim b(V,z)$ 有界，因此 (F.4) 中的第二个界适用。

因此，存在常数 $C > 0$，使得对于每个 $u, w \in \mathbb{R}^d$，

$$\left\|\nabla^2 G(u+w) - \nabla^2 G(u)\right\|_{\mathrm{op}} \leq C\,\|w\|.$$

因此，$\nabla^2 G$ 是 Lipschitz 的，而 G 在 $\mathbb{R}^d$ 上二阶连续可微。

当 $\ell = d$时。设 $n : \mathrm{Gr}(d-1, d) \to \mathbb{R}^d$是一个 Borel可测映射，为每个$V$指定一个单位法向量 $n(V) \perp V$。将 (C.3) 应用于命题 F.1 中的梯度公式，并在关于 $V^\perp$ 的积分中作变量代换 $y = tn(V)$，$t \in \mathbb{R}$，得到

$$\nabla G(u) = \int_{\mathrm{Gr}(d-1,d)} n(V) \int_{\mathbb{R}} \mathrm{sgn}\left(n(V)^\top u - t\right) b_0\left(V, tn(V)\right)\, dt\, d\vartheta(V).$$

对于 $u,\ w \in \mathbb{R}^d$，用在 u 处的内层积分减去在 $u+w$ 处的内层积分，得到

$$2\int_{n(V)^\top u}^{n(V)^\top (u+w)} b_0\left(V, tn(V)\right)\, dt.$$

利用 Lipschitz常数 L_{b_0}，对每个 $u,\ w \in \mathbb{R}^d$，我们得到

$$\begin{aligned}\left\|\nabla G(u+w) - \nabla G(u) - 2\int b_0(V, \Gamma_{V^\perp} u)\Gamma_{V^\perp} w\, d\vartheta(V)\right\| &\leq 2L_{b_0} \int \int_0^{|n(V)^\top w|} s\, ds\, d\vartheta(V) \\ &\leq L_{b_0}\,\|w\|^2.\end{aligned}$$

因此，∇G可微，且

$$\nabla^2 G(u) = 2\int_{\mathrm{Gr}(d-1,d)} b_0(V, \Gamma_{V^\perp} u)\Gamma_{V^\perp}\, d\vartheta(V). \tag{F.7}$$

利用 (F.7)，对每个 $u,\ w \in \mathbb{R}^d$，我们还得到

$$\left\|\nabla^2 G(u+w) - \nabla^2 G(u)\right\|_{\mathrm{op}} \leq 2\int |b_0(V, \Gamma_{V^\perp}(u+w)) - b_0(V, \Gamma_{V^\perp} u)|\, d\vartheta(V) \leq 2L_{b_0}\,\|w\|.$$

因此，$\nabla^2 G$是 Lipschitz 的。

正定性。 令 $u \in \mathbb{R}^d$和 $v \in \mathbb{R}^d \setminus \{0\}$。首先假设 $\ell < d$。正如命题 F.1 的证明中所确立的，点 $u, u+v, X_1, \ldots, X_\ell$几乎处处仿射无关。因此，向量 $a := \Gamma(\mathbf{X})(u - X_\ell)$ 和 $\Gamma(\mathbf{X})v$ 几乎处处线性无关。等号

因此不可能在这两个向量的柯西–施瓦茨不等式中成立。命题 F.2 中的公式给出

$$v^\top \nabla^2 G(u) v = \frac{1}{\ell}\mathbb{E}\left[\frac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{X}))}{\|a\|^3}\left(\|\Gamma(\mathbf{X})v\|^2\,\|a\|^2 - \langle \Gamma(\mathbf{X})v, a\rangle^2\right)\right] > 0.$$

现在假设 $\ell = d$。由于 f 连续且积分为 1，因此存在一个开球 $B(x_0, \varepsilon)$，在该开球上 $f > 0$。令 $V \in \mathrm{Gr}(d-1, d)$ 且 $y \in V^\perp$。如果超平面 $y + V$ 与此球相交，则它们的交集是 $y + V$ 中的一个非空开球，并且包含 d 个仿射无关点 $y + z_1, \ldots, y + z_d$。因此，对于每个 $1 \le i \le d$，都有 $m_{d-1}(\mathrm{Simpl}(\mathbf{z})) > 0$ 且 $f(y + z_i) > 0$。由连续性可知，这些不等式在 V^d 中 $\mathbf{z}$ 的某个开邻域内仍然成立。因此，根据定义 (C.4)，得到

$$b_0(V, y) > 0 \quad \text{whenever } \|y - \Gamma_{V^\perp} x_0\| < \varepsilon.$$

令 σ 为单位球面 $\mathbb{S}^{d-1}$ 上的均匀概率测度 $= \{e \in \mathbb{R}^d : \|e\| = 1\}$。在 $e \mapsto \mathrm{span}(e)^\perp$ 作用下，σ 的像是 $\mathrm{Gr}(d-1, d)$ 上的旋转不变概率测度，因此等于 ϑ。对于 $e \in \mathbb{S}^{d-1}$，到 $\mathrm{span}(e)$ 的正交投影为 $ee^\top$，所以公式 (F.7) 给出

$$v^\top \nabla^2 G(u) v = 2\int_{\mathbb{S}^{d-1}} b_0(\mathrm{span}(e)^\perp, ee^\top u)(e^\top v)^2\, d\sigma(e).$$

集合

$$\mathcal{U} := \{e \in \mathbb{S}^{d-1} : |e^\top (u - x_0)| < \varepsilon\}$$

在 $\mathbb{S}^{d-1}$ 中是开集。由于 $d \ge 2$，存在一个与 $u - x_0$ 正交的单位向量，因此 $\mathcal{U}$ 非空，且 $\sigma(\mathcal{U}) > 0$。由于 $v \ne 0$，我们有 $\sigma(\mathbb{S}^{d-1} \cap \mathrm{span}(v)^\perp) = 0$，因此

$$\sigma(\mathcal{U} \setminus \mathrm{span}(v)^\perp) = \sigma(\mathcal{U}) > 0.$$

对于每个 $e \in \mathcal{U} \setminus \mathrm{span}(v)^\perp$，关系 $\left\|ee^\top (u - x_0)\right\| = |e^\top (u - x_0)| < \varepsilon$ 和 $e^\top v \ne 0$ 表明被积函数严格为正。因此，$v^\top \nabla^2 G(u) v > 0$。

在这两种情况下，对于每个 $u \in \mathbb{R}^d$，$\nabla^2 G(u)$ 都是正定的。

根据命题 F.1，G 存在唯一极小元 μ^*。令 $H := \nabla^2 G(\mu^*)$，则有

$$\nabla G(\mu^* + u) - Hu = \int_0^1 \left(\nabla^2 G(\mu^* + tu) - H\right) u\, dt.$$

根据海森矩阵的 Lipschitz 性质，存在一个有限常数 $C > 0$，且与 u 无关，使得 $\|\nabla G(\mu^* + u) - Hu\| \le (C/2)\,\|u\|^2$。这证明了假设 A3，其中 $\alpha = 1$。 □

备注。当 $\ell = d$ 时，底高公式表明，对于每个 $\mathbf{x}$，$g(\mathbf{x};\cdot)$ 的海森矩阵在 $\mathrm{aff}(\mathbf{x})$ 之外为零。在命题 2.1 的假设下，对于每个 $u \in \mathbb{R}^d$，$g(\mathbf{X};\cdot)$ 在 u 处的海森矩阵因此几乎处处为零，而 (F.7) 中的 $\nabla^2 G(u)$ 是正定的。因此，$\nabla^2 G(u)$ 不可能通过对 $g(\mathbf{X};\cdot)$ 在 u 处的海森矩阵取期望而得到。

命题 *2.1* 后示例的验证。对于非退化高斯密度 f，存在常数 $C, \eta > 0$，使得对每个 $x \in \mathbb{R}^d$，

$$f(x) + \|\nabla f(x)\| \le Ce^{-\eta\|x\|^2}.$$

因此，对于每个 $\nu > 2$，f 满足命题 2.1 的假设。此类密度的有限混合也满足相同的假设。如果 $f \in C^1(\mathbb{R}^d)$ 具有紧支撑，那么 f 和 ∇f 有界，并且在一个紧集之外为零，因此相同的结论仍然成立。

设 f 为如下形式的 Student 密度

$$f(x) \coloneqq c\left(1+\frac{(x-a)^{\top} S^{-1}(x-a)}{\nu}\right)^{-(d+\nu)/2},$$

其中， $a \in \mathbb{R}^d$、 $\nu > 2$、 S是一个对称正定 $d \times d$矩阵， $c > 0$是归一化常数。由于 S 是正定的，存在常数 $C > 0$ ，使得对于每个 $x \in \mathbb{R}^d$,

$$f(x) \leq C\left(1+\|x\|^2\right)^{-(d+\nu)/2} \leq 2^{(d+\nu)/2} C\left(1+\|x\|\right)^{-d-\nu}.$$

微分和这个界给出一个常数 $C' > 0$ ，使得对于每个 $x \in \mathbb{R}^d$,

$$\|\nabla f(x)\| = \frac{(d+\nu)\left\|S^{-1}(x-a)\right\|}{\nu+(x-a)^{\top} S^{-1}(x-a)} f(x) \leq \frac{(d+\nu)\left\|S^{-1}\right\|_{\mathrm{op}}}{2\sqrt{\nu \lambda_{\min}(S^{-1})}} f(x) \leq C'\left(1+\|x\|\right)^{-d-\nu}.$$

这就证明了命题 2.1 的假设成立，其中使用相同的 ν。 □

G 梯度的计算

我们描述滑动窗口递推中梯度的计算，并确立第 3.4 节中所述的运算次数。

我们首先考虑 $\ell \geq 2$的情形，在假设 A2 下；或者当 $\ell < d$时，在假设 A4 下。在这两个假设中的任一个下， P 对维数至多为$\ell - 1$的每个仿射子空间赋予零概率。随后，对每个窗口中的独立观测值依次进行条件化可知，以概率 1，对于每个 $n \geq \ell$，观测值 $X_{n-\ell+1}$、……、 X_n 都是线性无关的，因而也是仿射无关的。

根据格拉姆矩阵求值。 令$\mathbf{x} \coloneqq (x_1, \ldots, x_\ell)$为 $\mathbb{R}^d$中的仿射无关点元组。我们回顾式（2.2）中的次梯度选取：对于每个 $\mu \in \mathbb{R}^d$,

$$\nabla_\mu g(\mathbf{x};\mu) = \begin{cases} \dfrac{m_{\ell-1}(\mathrm{Simpl}(\mathbf{x}))}{\ell} \dfrac{\Gamma(\mathbf{x})(\mu - x_\ell)}{\|\Gamma(\mathbf{x})(\mu - x_\ell)\|}, & \text{if } \Gamma(\mathbf{x})(\mu - x_\ell) \neq 0, \\ 0, & \text{if } \Gamma(\mathbf{x})(\mu - x_\ell) = 0. \end{cases}$$

我们定义由 x_ℓ 得到的边向量 B $d \times(\ell - 1)$矩阵及其格拉姆矩阵 S 为

$$B \coloneqq (x_1 - x_\ell, \ldots, x_{\ell-1} - x_\ell), \quad \text{and} \quad S \coloneqq B^{\top} B.$$

$x_1, \ldots, x_\ell$的仿射独立性意味着 B的列是线性无关的，因此S 是正定的。对于每个 $\mu \in \mathbb{R}^d$，我们有

$$m_{\ell-1}(\mathrm{Simpl}(\mathbf{x})) = \frac{\sqrt{\det S}}{(\ell-1)!}, \quad \text{and} \quad \Gamma(\mathbf{x})(\mu - x_\ell) = \mu - x_\ell - B\left(S^{-1}\left(B^{\top}(\mu - x_\ell)\right)\right).$$

一旦获得 B、 S^{-1} 和 $\det S$ ，通过连续的矩阵—向量乘积计算 $\Gamma(\mathbf{x})(\mu - x_\ell)$，然后求取选定次梯度，需要进行 $O(d\ell + \ell^2)$次 $= O(d\ell)$运算。

更新窗口。 在获得 $n \geq \ell$ 个观测值后，仅存储当前窗口 $X_{n-\ell+1}, \ldots, X_n$。由于置换这些观测值不会改变单纯形体积和选定次梯度，我们将 $X_{n-\ell+1}$ 替换为 X_{n+1} ，而不移动其他观测值。

现在，我们推导替换一个观测值时更新 B、 S^{-1} 和 $\det S$ 的公式。我们用 $j \in \{1, \ldots, \ell\}$ 表示被替换观测值的索引，用 $x'_j \in \mathbb{R}^d$ 表示新的观测值。我们令 $u \coloneqq x'_j - x_j$。我们假设替换后观测值仍然仿射无关

对于连续窗口，替换后的观测值几乎处处仍满足这一条件。更新后的矩阵 B'的定义公式与 B相同，只是将 x_j 替换为 x'_j。因此

$$B' = B + uv^\top, \qquad v \coloneqq \begin{cases} e_j, & \text{if } j < \ell, \\ -\mathbf{1}_{\ell-1}, & \text{if } j = \ell, \end{cases}$$

其中 $e_1, \ldots, e_{\ell-1}$ 是 $\mathbb{R}^{\ell-1}$ 和 $\mathbf{1}_{\ell-1} \coloneqq (1, \ldots, 1)^\top \in \mathbb{R}^{\ell-1}$ 的坐标向量。根据 $z \coloneqq B^\top u + (\|u\|^2/2)v$，更新后的 Gram 矩阵满足

$$\begin{aligned} S' \coloneqq (B')^\top B' &= S + B^\top uv^\top + vu^\top B + \|u\|^2 vv^\top \\ &= S + vz^\top + zv^\top = S + (v, z)(z, v)^\top. \end{aligned}$$

因此，从 S 到 S' 的变化秩至多为二。替换后的观测值仿射独立，因此 S' 是正定的。根据矩阵行列式引理，

$$\det S' = \det S \, \det\left(I_2 + (z, v)^\top S^{-1}(v, z)\right).$$

因此， $I_2 + (z, v)^\top S^{-1}(v, z)$ 非奇异，且根据 Woodbury 公式，

$$(S')^{-1} = S^{-1} - (S^{-1}(v, z))\left(I_2 + (z, v)^\top S^{-1}(v, z)\right)^{-1}((z, v)^\top S^{-1}).$$

运算计数。对于第一个窗口 $X_1, \ldots, X_\ell$，计算 B 和 $S = B^\top B$ 需要 $O(d\ell^2)$ 次运算。通过分解计算 S^{-1} 和 $\det S$ 需要 $O(\ell^3)$ 次运算，因此初始化需要 $O(d\ell^2)$ 次运算，因为 $\ell \le d$。

每次替换时，根据上述公式计算 u、z 和 B' 需要 $O(d\ell)$ 次运算。计算 $S^{-1}(v, z)$ 和 $(z, v)^\top S^{-1}$，然后计算 $(S')^{-1}$ 的公式，需要 $O(\ell^2)$ 次运算。2×2 逆矩阵和行列式更新需要 $O(1)$ 次运算。计入梯度评估后，每次替换的开销为 $O(d\ell)$，因为 $\ell \le d$。对于 $n \ge \ell$，滑动窗口递推对 $X_1, \ldots, X_n$ 执行一次遍历的总开销，包括初始化和 $n - \ell + 1$ 次梯度评估，因此为

$$O\left(d\ell^2 + (n - \ell + 1)\, d\ell\right) = O\left(nd\ell\right).$$

存储观测值 B 和 S^{-1} 需要 $O(d\ell)$ 内存。

对于不相交分块递推，为每个 $\lfloor n/\ell \rfloor$ 完整分块分别计算格拉姆矩阵和梯度，每个分块需要 $O(d\ell^2)$ 次运算。因此，对 $X_1, \ldots, X_n$ 进行一次遍历时，不相交分块递推的总计算量为 $O(\lfloor n/\ell \rfloor d\ell^2) = O(nd\ell)$ 次运算。

情形 $\ell = 1$。对于 $\ell = 1$，约定 $\Gamma(x_1) = I_d$ 和 $m_0(\mathrm{Simpl}(x_1)) = 1$，由 (2.2) 可得，

$$\nabla_\mu g(x_1; \mu) = \begin{cases} \dfrac{\mu - x_1}{\|\mu - x_1\|}, & \text{if } \mu \neq x_1, \\ 0, & \text{if } \mu = x_1. \end{cases}$$

每次梯度计算需要 $O(d)$ 次运算。对 $X_1, \ldots, X_n$ 进行一次遍历时，滑动窗口递推的总计算量为 $O(nd)$ 次运算，内存需求为 $O(d)$。

计算 Oja中位数的另一种方法。因此，对于 $\ell = d$，前述 Gram矩阵计算表明，针对 X_1、.....、X_n 的滑动窗口递推总计需要 $O(nd^2)$ 次运算，内存开销为 $O(d^2)$。第4节中使用的实现基于观测值构成的方阵。替换一个观测值会使该矩阵发生秩至多为一的变化，而下面的行列式公式无需计算正交投影即可给出选定次梯度。

设 $\mathbf{x} \coloneqq (x_1、\ \ldots\ldots、\ x_d)$ 是 $\mathbb{R}^d$ 中线性无关点组成的元组，并设 A 为以 $x_1、\ \ldots\ldots、\ x_d$ 为列的矩阵。令 $\mathbf{1}_d \coloneqq (1\ 、\ \ldots\ldots\ 、\ 1)^\top \in \mathbb{R}^d$。由单纯形体积公式 (B.7) 和矩阵行列式引理可知，对于任意 $\mu \in \mathbb{R}^d$,

$$g(\mathbf{x};\mu) = \frac{\left|\det(A - \mu \mathbf{1}_d^\top)\right|}{d!} = \frac{|\det A|}{d!}\left|1 - \mathbf{1}_d^\top A^{-1}\mu\right|.$$

对于满足 $\mathbf{1}_d^\top A^{-1}\mu \neq 1$ 的 $\mu \in \mathbb{R}^d$，微分得到

$$\nabla_\mu g(\mathbf{x};\mu) = -\frac{|\det A|}{d!}\operatorname{sgn}\left(1 - \mathbf{1}_d^\top A^{-1}\mu\right)(A^{-1})^\top \mathbf{1}_d.$$

由于 $\mathbf{1}_d^\top A^{-1}\mu = 1$ 等价于 $\mu \in \mathsf{aff}(\mathbf{x})$，因此，对于每个 $\mu \in \mathbb{R}^d$，将 $\operatorname{sgn}(0)$ 设为 $= 0$，即可用同一公式得到 (2.2) 中的选定次梯度。

对于上文考虑的将 x_j 替换为 x_j'，其中 $u = x_j' - x_j$，我们将更新后的矩阵记为 A'，并假设 A' 非奇异。设 e_j 为 $\mathbb{R}^d$ 的第 j 个坐标向量。那么 $A' = A + ue_j^\top$，由矩阵行列式引理可得

$$\det A' = \det A\left(1 + e_j^\top A^{-1}u\right).$$

由于 A 和 A' 均非奇异，$1 + e_j^\top A^{-1}u \neq 0$，由秩一更新的 Sherman–Morrison 公式（Sherman 和 Morrison，1950年）可得

$$(A')^{-1} = A^{-1} - \frac{(A^{-1}u)(e_j^\top A^{-1})}{1 + e_j^\top A^{-1}u}.$$

计算 $A^{-1}u$、更新逆矩阵和行列式，以及计算选定次梯度，每次替换需要 $O(d^2)$ 次运算，此外还需进行一次初始 $O(d^3)$ 分解。对于 $n \geq d$，因此，对 $X_1, \ldots, X_n$ 执行一遍滑动窗口递推的总成本（包括初始化）为 $O(nd^2)$ 次运算，内存占用为 $O(d^2)$。与 Gram矩阵计算相比，这种实现更简单，通过秩一逆矩阵更新和更少的用于计算梯度的矩阵—向量乘积，同时保持相同的复杂度阶。

H 步长的高斯校准

我们为第4节中高斯模拟所使用的 $c_\gamma = \sqrt{d!}$ 选择提供依据。

假设 $\ell = d$，并且 $X_1、\ \ldots\ldots、\ X_d$ 独立，且服从分布 $\mathcal{N}_d(0,\ I_d)$。于是$\mu^* = 0$。为计算 Γ_0，对于 $u \in \mathbb{R}^d$，我们定义 $D(u)\colon = \det(X_1 - u、\ \ldots\ldots、\ X_d - u)$。在 $D(u)$ 的多重线性展开中，所有至少在两列中包含 u 的项都为零。因此，存在一个随机向量 a，使得对于每个 $u \in \mathbb{R}^d$，$D(u) = D(0) - a^\top u$。将 Dürre 和 Paindaveine （2022a）的命题 2.1 应用于两个均等于 $x \mapsto x - u$ 的函数，可得

$$\mathbb{E}\left[D(u)^2\right] = d!\det\left(\mathbb{E}\left[(X_1 - u)(X_1 - u)^\top\right]\right) = d!\det\left(I_d + uu^\top\right) = d!\left(1 + \|u\|^2\right).$$

确定 u 中的二次项可得，对于每个 $u \in \mathbb{R}^d$，$u^\top \mathbb{E}\left[aa^\top\right]u = d!\ \|u\|^2$，从而$\mathbb{E}\left[aa^\top\right] = d!I_d$。

由于 $D(0) \neq 0$ 几乎处处成立，在 $u = 0$ 处对 $g(\mathbf{X};u) = |D(u)|/d!$ 求导可得

$$\nabla_\mu g(\mathbf{X};0) = -\frac{\operatorname{sgn}(D(0))}{d!}a.$$

根据对称性，梯度的中心为零，因此

$$\Gamma_0 = \frac{\mathbb{E}\left[aa^\top\right]}{(d!)^2} = \frac{1}{d!}I_d.$$

因此，第4节中使用的选择 $c_\gamma \coloneqq \sqrt{d!}$ 给出 $c_\gamma^2\Gamma_0 = I_d$，所以 $c_\gamma \nabla_\mu g(\mathbf{X};0)$ 的每个坐标的方差均为 1。