arXiv:2304.00770v1 [stat.ML] 3 Apr 2023

用于估计几何中位数及其应用的在线随机牛顿方法

Antoine Godichon-Baggioni(*) 和 Wei Lu(**)(*)索邦大学，概率、统计与建模实验室(**) 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
antoine.godichon_baggioni@upmc.fr, wei.lu@insa-rouen.fr

摘要

在大样本情境下，少数个体就可能使均值等基本统计指标失真。自动检测这些异常个体并非易事，稳健方法则是一种替代策略。本文聚焦于随机变量几何中位数的估计；几何中位数是一种稳健的集中趋势指标。为处理按顺序到达的大量数据，本文提出了用于估计几何中位数的在线随机牛顿算法，并给出了其收敛速率。由于中位数及海森矩阵的估计值可以递归更新，我们还确定了中位数在任意指定方向上的置信区间，并进行了在线统计检验。

关键词： 几何中位数；随机牛顿算法；在线估计；随机优化

1 引言

随着测量技术的发展和计算机存储能力的提升，大样本观测如今已十分常见。在这样的大样本情境下，即使只有少数个体，也可能使均值等基本统计指标失真。自动检测这些异常个体并不容易，因此采用稳健方法是一种颇具吸引力的替代方案。众所周知，中位数是衡量集中趋势的稳健指标；本文关注 $\mathbb{R}^p$ 中随机变量的几何中位数。几何中位数也称为空间中位数或多变量 L_1 中位数，最早由 [9] 和 [13]提出。它定义为随机变量观测值的 L_1 距离之和的最小化点。它具有良好的稳健性，例如其崩溃点为 0.5 [16, 14]。

近年来，几何中位数在机器学习领域受到越来越多的关注。例如，在 [17], 作者提出了一种 L_1 中位数滤波器，作为其网格去噪方法的工具，有助于模型保留几何特征；基于几何中位数建立了一种截线提取优化算法 [20], 作者考虑采用几何中位数，是因为它具有抗噪性；一种图像滤波算法 [7] 已经

提出了一种基于空间中位数滤波器的算法，其性能优于均值滤波器，因为当图像中存在噪声像素时，中位数比均值更加稳健。

本文聚焦于几何中位数的估计。人们提出了一种名为 Weiszfeld 算法的迭代算法 [19, 12, 18], , 并在 [1]中对该方法进行了改进。该算法简单且快速，但当数据按顺序从大到无法载入内存的文件中读取时，这一过程并不适用。为解决这一问题，鉴于几何中位数被定义为某个凸函数的极小化点，[6]提出了一种用于估计几何中位数的平均随机梯度算法。然而，作为一种一阶算法，在实际应用中，它可能对待最小化函数的海森矩阵结构非常敏感[2, 3]。

为解决这一问题，我们在此提出用于估计几何中位数的新型随机牛顿型算法。随机牛顿算法面临的一项困难是更新海森矩阵估计的逆。我们对海森矩阵的逆进行递归估计，该方法基于 Sherman-Morrison 公式 [8], , 从而避免了代价高昂的矩阵求逆计算。为克服可能出现的初始化问题，我们还提出了加权平均版本[3]。得益于这些算法的渐近效率，并且由于可以递归估计协方差矩阵，我们引入了针对所选方向上几何中位数的在线置信区间，并开展在线统计假设检验。

本文组织如下：第2节介绍一般框架，并说明递归估计海森矩阵的逆的方法。第3节介绍随机牛顿算法，并给出其收敛速率。此外，还通过模拟研究比较不同算法的性能。第4节旨在为几何中位数建立递归置信区间并开展在线统计检验。证明汇集于第5节。

2 框架

2.1 一般框架

随机变量 X 在 $\mathbb{R}^p$ 中取值，其几何中位数 m 是凸函数 $G:\ \mathbb{R}^p \longrightarrow \mathbb{R}$ 的极小值点，该函数对所有 $h \in \mathbb{R}^p$ 均有定义，其表达式为 [13]

$$G(h) =: \mathbb{E}\left[g(X,h)\right] = \mathbb{E}\left[\|X-h\| - \|X\|\right].$$

请注意，该定义并未假设中位数唯一，也未要求 $\|X\|$的一阶矩存在。从现在起，我们假设以下条件成立。

- **假设1**。 随机变量 X 不会集中在单个点附近：存在$C_6 > 0$ ，使得对所有 $h \in \mathbb{R}^p$,

$$\mathbb{E}\left[\frac{1}{\|X-h\|^6}\right] \leq C_6.$$

- **假设 2.** 随机变量 X 未集中在一条直线上：对于所有 $h \in \mathbb{R}^p$，存在 $h' \in \mathbb{R}^p$ 使得 $\langle h, h' \rangle \neq 0$ 且

$$\mathbb{V}\mathrm{ar}\left[\langle X, h' \rangle\right] > 0.$$

请注意，在文献 [6], 中，假设1的阶通常取为2；出于技术原因，我们将其提高到6，也就是说，这样做是为了获得后文将介绍的海森矩阵估计的收敛速率。根据 [14], ，假设 2 确保函数G 严格凸，因此中位数 m 是唯一确定的。如 [6], 所示，函数 G 处处可微，并且可以验证，对所有 $h \in \mathbb{R}^p$ ，其梯度定义为:

$$\nabla G(h) = -\mathbb{E}\left[\frac{X-h}{\|X-h\|}\right].$$

此外，函数 G 处处二阶可微，其海森矩阵为 [15]

$$\nabla^2 G(h) = \mathbb{E}\left[\frac{1}{\|X-h\|}\left(I_p - \frac{(X-h)(X-h)^T}{\|X-h\|^2}\right)\right].$$

根据 [6], $\nabla^2 G(h)$ 在假设 1 和 2 下是正定的。这对于用于估计几何中位数的随机牛顿型算法尤为重要，因为这类算法会利用函数 G 的海森矩阵所提供的信息。

2.2 平均随机梯度算法的一些回顾

文献 [6] 提出了一种用于估计几何中位数的平均随机梯度算法。给定 X_1、 X_2，…， X_n，X_{n+1}，…， 个 X的独立同分布副本，随机梯度算法定义如下:

$$m_{n+1}^{(SG)} = m_n^{(SG)} + \gamma_n \frac{X_{n+1} - m_n^{(SG)}}{\left\|X_{n+1} - m_n^{(SG)}\right\|}, \tag{1}$$

其中， γ_n 是一个下降步长序列。其平均版本通过对所有已估计的历史值取平均来构造，递归定义如下:

$$\overline{m}_{n+1} = \overline{m}_n + \frac{1}{n+1}\left(m_{n+2}^{(SG)} - \overline{m}_n\right) \tag{2}$$

其中， $m_0^{(SG)}$ 有界且 $m_0 = m_0^{(SG)}$ 。因此，可以轻松更新估计值。该算法已得到深入研究：其渐近效率见 [6], ，而 L^p 收敛速率则在 [10]中推导得出。此外， [4] 还研究了该算法的非渐近行为，并基于改进的 L^2 收敛速率推导出非渐近置信球。然而，它是一种一阶算法，因此可能对我们试图最小化的函数 [3], 的海森矩阵结构相当敏感；例如，这意味着随机变量 X 相当集中于一条直线附近。

2.3 如何估计海森矩阵的逆矩阵

随后我们将重点讨论牛顿型方法，它更适合处理病态问题。随机牛顿算法面临的一个主要难点是，我们必须能够更新海森矩阵估计的逆，并尽可能降低计算时间成本。现在我们说明如何递归地估计海森矩阵的逆 $\nabla^2 G(m)$

在已知 m 时。其思路是得到一个形式为 $\frac{1}{n}\sum_{k=1}^{n} a_k\phi_k\phi_k^T$ 的估计，以应用 Riccati公式 [8]。我们有

$$\nabla^2 g(X,h) = \frac{1}{\|X-h\|}\left(I_p - \frac{(X-h)(X-h)^T}{\|X-h\|^2}\right).$$

注意到 $\left(I_p - \frac{(X-h)(X-h)^T}{\|X-h\|^2}\right)^2 = I_p - \frac{(X-h)(X-h)^T}{\|X-h\|^2}$，于是有

$$\nabla^2 g(X,h) = \|X-h\|\left(\nabla^2 g(X,h)\right)^2. \tag{3}$$

此外，根据泰勒定理，

$$\nabla g(X,h+\alpha Z) - \nabla g(X,h) = \int_0^1 \nabla^2 g(X,h+t\alpha Z)dt\alpha Z,$$

其中 $Z\sim\mathcal{N}(0,I_p)$ 且 $\alpha>0$。因此， $\nabla^2 G(m)$ 的一个估计量为

$$\hat{H}_n = \frac{1}{n+1}\left(\sum_{k=1}^{n}\frac{\|X_k-m\|}{\alpha_k^2}\Phi_k\Phi_k^T + H_0\right), \tag{4}$$

其中 $\alpha_k = k^{-\frac{1}{\ln(k+1)}}$， $H_0 = I_p$ 且 Φ_k 定义为

$$\Phi_k := \nabla g(X_k, m+\alpha_k Z_k) - \nabla g(X_k,m) = \int_0^1 \nabla^2 g(X_k, m+t\alpha_k Z_k)dt\alpha_k Z_k,$$

其中，对于任意 $k\geq 1$，$(Z_k)_k$ 均为标准独立高斯向量。事实上，可以验证 $\mathbb{E}\left[\frac{\|X_k-m\|}{\alpha_k^2}\Phi_k\Phi_k^T\right] \to \nabla^2 G(m)$。此外，借助 Riccati公式 [8]， $H_{n+1}^{-1} = (n+1)^{-1}\hat{}\, H_n^{-1}$可以轻松更新为

$$H_{n+1}^{-1} = H_n^{-1} - \frac{\|X_{n+1}-m\|}{\alpha_{n+1}^2}\left(1+\frac{\|X_{n+1}-m\|}{\alpha_{n+1}^2}\phi_{n+1}^T H_n^{-1}\phi_{n+1}\right)^{-1} H_n^{-1}\phi_{n+1}\phi_{n+1}^T H_n^{-1}.$$

因此，已知 m 后，我们便能够以复杂度$\mathcal{O}\left(p^2\right)$（而非 $\mathcal{O}\left(p^3\right)$）在每次迭代中递归估计海森矩阵的逆。

3 随机牛顿方法

本节介绍两种用于估计中位数 m 的随机牛顿方法：一种随机牛顿算法及其加权平均版本。我们还给出它们收敛性的理论保证。回顾一下，$(X_n)_{n\geq 1}$ 是一列相互独立且同分布的随机向量，与向量 X 同分布；$(Z_n)_{n\geq 1}$ 是一列相互独立的标准高斯向量。

3.1 随机牛顿算法

3.1.1 算法定义

现在，我们递归地定义随机牛顿估计：对所有 $n \geq 0$，有

$$m_{n+1} = m_n + \frac{1}{n+1}\widetilde{H}_n^{-1}\frac{X_{n+1} - m_n}{\|X_{n+1} - m_n\|}, \tag{5}$$

其中，m_0 有界。令$(\beta_n)^{n\geq 1}$为对任意$n \geq 1$定义的严格正实数序列，其定义为 $\tilde{\beta}_n = \frac{c_\beta}{n^\beta}$，其中 $0 < \beta < \frac{1}{2}$ 且$c_\beta > 0$。矩阵 $\widetilde{H}_n$ 对任意 $n \geq 0$ 定义为

$$\widetilde{H}_n = \overline{H}_n + \frac{1}{n+1}\sum_{k=1}^{n}\tilde{\beta}_k Z_k Z_k^T, \tag{6}$$

其中，$\overline{H}_n$ 是海森矩阵的递归估计 $\nabla^2 G(m)$，定义如下:

$$\overline{H}_n = \frac{1}{n+1}\left(\sum_{k=1}^{n}\frac{\|X_k - m_{k-1}\|}{\alpha_k^2}\phi_k\phi_k^T + H_0\right),$$

其中，对任意 $k \geq 1$，$\phi_k = \nabla g(X_k, m_{k-1} + \alpha_k Z_k) - \nabla g(X_k, m_{k-1})$，$\alpha_k = \frac{1}{k\ln(k+1)}$，且 H_0 为对称正定矩阵。我们加入项 $\sum_{k=1}^{n}\tilde{\beta}_k Z_k Z_k^T$，以控制海森矩阵估计的特征值（参见第 5 节）；这对于保证算法 [3]的收敛性是必要的。借助Riccati公式 [8]，$H_n^{-1} = (n+1)^{-1}H_n^{-1}$ 可分两步更新，得到

$$H_{n+1/2}^{-1} = H_n^{-1} - \frac{\|X_{n+1} - m_n\|}{\alpha_{n+1}^2}\left(1 + \frac{\|X_{n+1} - m_n\|}{\alpha_{n+1}^2}\phi_{n+1}^T H_n^{-1}\phi_{n+1}\right)^{-1} H_n^{-1}\phi_{n+1}\phi_{n+1}^T H_n^{-1},$$

$$H_{n+1}^{-1} = H_{n+1/2}^{-1} - \tilde{\beta}_k\left(1 + \tilde{\beta}_k Z_{n+1}^T H_{n+1/2}^{-1} Z_{n+1}\right)^{-1} H_{n+1/2}^{-1} Z_{n+1} Z_{n+1}^T H_{n+1/2}^{-1}.$$

因此，该算法使我们能够递归地更新海森矩阵的估计和几何中位数的估计。

3.1.2 收敛性结果

下面的定理给出了随机牛顿估计的几乎必然收敛速率以及渐近效率。请注意，其渐近效率使我们能够构造置信区间并进行检验（见第4节）。

定理1 假设假设*1*和假设*2*成立，则由(5) 定义的随机牛顿估计 m_n 几乎必然收敛至 m，且

$$\|m_n - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.$$

此外，(6) 中定义的海森矩阵估计对所有 $\delta > 0$均满足

$$\left\|\widetilde{H}_n - H\right\|^2 = \mathcal{O}\left(\max\left\{\frac{(\ln n)^{1+\delta}}{n}, \frac{c_\beta}{n^{2\beta}}\right\}\right) a.s.$$

最后，

$$\sqrt{n}\left(m_n - m\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, H^{-1}\Sigma H^{-1}\right),$$

其中 $\Sigma = \mathbb{E}\left[\nabla g(X, m)\nabla g(X, m)^T\right]$.

证明见第5节。值得注意的是，为了控制海森矩阵估计的特征值，需要付出的代价是估计的收敛速率有所降低。更具体地说，这会出现一项，其收敛速率为 $n^{-2\beta}$，而不是 n^{-1}。

3.2 加权平均随机牛顿算法

3.2.1 算法定义

为了改善初始值不佳时估计量的实际表现，我们现在引入加权平均随机牛顿算法（WASN） [3], 其递归定义如下：$n \geq 0$

$$\hat{m}_{n+1} = \hat{m}_n + \frac{c_\gamma}{\left(n + 1 + c'_\gamma\right)^\gamma} \widetilde{H}_{n,\tau}^{-1} \frac{X_{n+1} - \hat{m}_n}{\|X_{n+1} - \hat{m}_n\|} \tag{7}$$

$$m_{n+1,\tau} = (1 - \tau_{n+1})m_{n,\tau} + \tau_{n+1}\hat{m}_{n+1}, \tag{8}$$

其中 $c_\gamma > 0$、 $c'_\gamma \geq 0$ 和 $\gamma \in \left(\frac{1}{2}, 1\right)$。加权平均序列 $(\tau_n)_{n\geq 1}$ 按如下方式选取： $\tau_n = \ln($ $_{n+1}{}^{\omega})\ln(k+1)^\omega$ ，对任意 $n \geq \bar{0}$ 和 $\omega \geq 0$ 均成立。注意，当 $\omega = 0$ 时，对应于平均随机牛顿算法（ASN）。海森矩阵的递归估计定义如下：

$$\widetilde{H}_{n,\tau} = \frac{1}{n+1}\left(\sum_{k=1}^{n} \frac{\|X_k - m_{k-1,\tau}\|}{\alpha_k^2}\phi_{k,\tau}\phi_{k,\tau}^T + H_0\right) + \frac{1}{n+1}\sum_{k=1}^{n} \tilde{\beta}_k Z_k Z_k^T. \tag{9}$$

其中，对任意 $k \geq 1$, $\phi_{k,\,\tau} = \nabla g\left(X_k,\ m_{k-1,\,\tau} + \alpha_k Z_k\right) - \nabla g\left(X_k,\ m_{k-1,\,\tau}\right)$。为了控制 $\widetilde{H}_{n,\,\tau}$ 的特征值， $(\tilde{\beta}_n)_{\,n\geq 1}$ 应为由 $\tilde{\beta}_n = \frac{c_\beta}{n^{1-\beta}}$ 定义的实数序列，其中 $0 < \beta < \gamma - \frac{1}{2}$ 且 $c_\beta > 0$。按照随机牛顿算法的相同步骤，我们始终可以利用 Riccati公式 [8] 更新 $H_{n,\,\tau}^{-1} = (n+1)^{-1}H_{n\,\tau}^{-1}$。

3.2.2 收敛性结果

下述定理表明，在相同假设下，WASN 估计量仍然是渐近有效的。

定理2 若假设*1*和假设*2*成立，则加权平均随机牛顿估计 $\hat{}m_n$ 和 m_n，τ 几乎必然收敛至 m。此外，

$$\|\hat{m}_n - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n^\gamma}\right) a.s. \qquad \textit{and} \qquad \|m_{n,\tau} - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.$$

此外，由(9) 定义的*Hessian*估计对所有 $\delta > 0$均满足

$$\left\|\widetilde{H}_{n,\tau} - H\right\|^2 = \mathcal{O}\left(\max\left\{\frac{(\ln n)^{1+\delta}}{n}, \frac{c_\beta}{n^{2\beta}}\right\}\right) a.s.$$

最后，

$$\sqrt{n}\left(m_{n,\tau} - m\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(0, H^{-1}\Sigma H^{-1}\right),$$

其中 $\Sigma = \mathbb{E}\left[\nabla g\left(X,\ m\right) \nabla g\left(X,\ m\right)^T\right]$.

证明见第5节。

3.3 方法比较

我们进行了一次数值实验，以比较随机牛顿算法（SN）、平均随机牛顿算法（ASN）、加权平均随机牛顿算法（WASN）以及文献 [6] 中提出的平均随机梯度下降法（ASGD）的性能。对于WASN，我们选择 $\tau_n = \ln(\sum_{k=0}^{n} n+1\ ^2)\ln(k+1)^2$。在该实验中，我们生成了来自高斯随机向量 $X \sim \mathcal{N}\left(0_p,\ \Sigma\right)$ 的样本，其中样本量为 $p \equiv 10$，并考虑了协方差矩阵 Σ 的两种结构，定义如下

(i) $\Sigma_{ij} = 0.5^{|i-j|}$;

(ii) Σ 为对角矩阵，其中 $\Sigma_{1,\ 1} = 1000$和 $\Sigma_{i,\ i} = 1$ 适用于 $i \neq 1$.

为了评估 算法的性能，我们计算以下均方误差 ：

$$MSE(\hat{m}) = \mathbb{E}\left[\|m - \hat{m}\|^2\right],$$

其中$\hat{}m$ 是中位数的估计值。我们通过蒙特卡洛实验来估计这一误差，使用$N = 400$ 个样本；对于每个样本，我们生成 $n = 15000$ 个 X的副本。为了观察初始化$\hat{}m$的影响，我们考虑四种不同的初始化： $m_0 = rU$ 取 $U \sim \mathcal{N}_p\left(0, I_p\right)$，以及$r = 1$、5、10 或15。

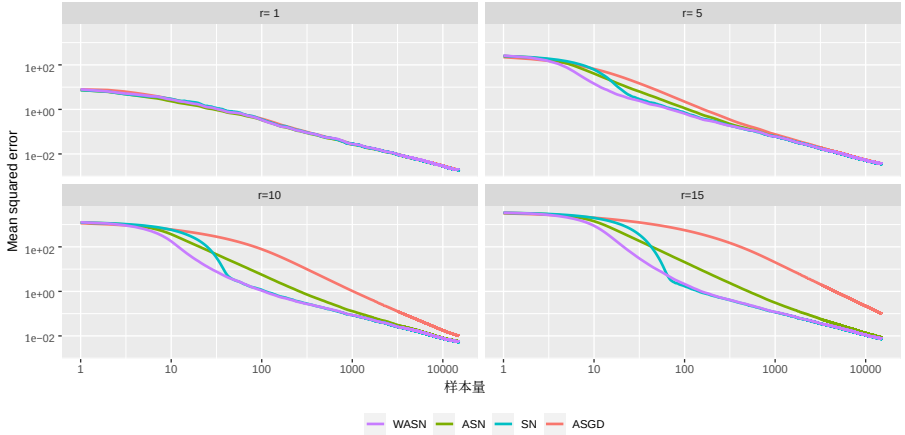

图1：均方误差随 样本量的变化（结构 (i)） .

对于结构 (i)，在初始化较好时，四种算法的性能相同。然而，随着初始化变差，我们可以看到，二阶方法比 ASGD 收敛得更快。

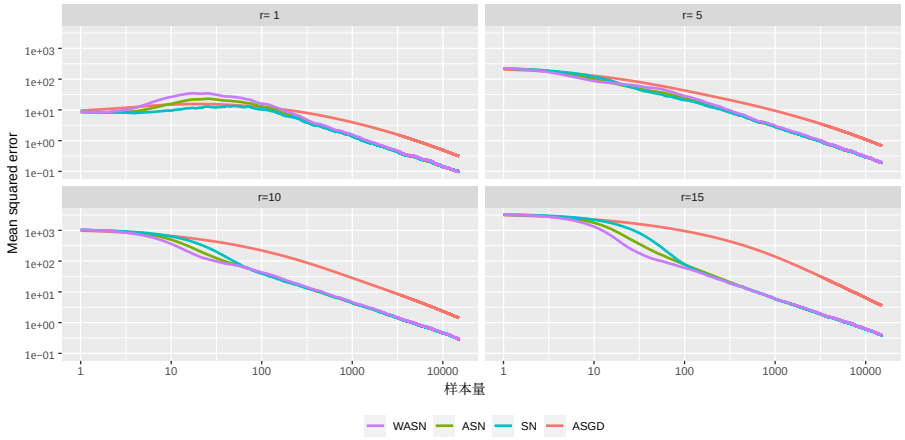

图 2：E　结构 (ii) 下均方误差随样本量的变化　.

对于结构 (ii)，我们发现牛顿型算法的表现远优于 ASGD。即使采用良好的初始化，ASGD 的收敛速度也明显慢于 WASN、ASN 和 SN。因此，ASGD 对海森矩阵的结构更为敏感。值得注意的是，对于较差的初始化，WASN 估计量似乎收敛得更快；在条件更差的问题中，也就是海森矩阵结构更差时，这一现象可能更加明显 [3]。

4 置信区间与检验

本节将提出关于中位数的置信区间和统计检验。这些结果由定理 1 和定理 2 得出，因此需要递归估计协方差矩阵 Σ，其定义如下：

$$\Sigma = \mathbb{E}\left[\nabla g(X,m)\nabla g(X,m)^T\right]$$

此处假设其为正。下文中，$\widetilde{m}_n$ 表示几何中位数的任一渐近有效估计量。例如，$\widetilde{m}_n$ 可以是公式 (2) 中定义的 ASGD 估计量、公式 (5) 中定义的 SN 估计量，或公式 (8) 中定义的 WASN 估计量。

4.1 协方差估计

一种自然的递归估计方法为 Σ

$$\overline{\Sigma}_n = \frac{1}{n+1}\left(\sum_{k=1}^{n}\frac{(X_k-\widetilde{m}_{k-1})}{\|X_k-\widetilde{m}_{k-1}\|}\frac{(X_k-\widetilde{m}_{k-1})^T}{\|X_k-\widetilde{m}_{k-1}\|}+\Sigma_0\right),$$

其中 Σ_0 是对称正定的。与 H_n 一样，Riccati公式（[8]）使我们能够递归更新矩阵 $W_{n+1}^{-1}=(n+1)^{-1}\Sigma_n^{-1}$：

$$W_{n+1}^{-1}=W_n^{-1}-\left(1+\frac{\left(\widetilde{X}_{n+1}\right)^T}{\left\|\widetilde{X}_{n+1}\right\|}W_n^{-1}\frac{\left(\widetilde{X}_{n+1}\right)}{\left\|\widetilde{X}_{n+1}\right\|}\right)^{-1}W_n^{-1}\frac{\left(\widetilde{X}_{n+1}\right)}{\left\|\widetilde{X}_{n+1}\right\|}\frac{\left(\widetilde{X}_{n+1}\right)^T}{\left\|\widetilde{X}_{n+1}\right\|}W_n^{-1}, \tag{10}$$

其中 $\widetilde{X}_{n+1}:=X_{n+1}-\widetilde{m}_n$。这一性质对于构建在线检验尤其重要（见第4.2节）。下一个定理给出了 Σ_n的收敛速率。

定理 3 设 $\widetilde{m}_n$ 是由 (2)、(5) 或 (8)定义的估计量。假设假设*1*和假设*2*成立，则对任意 $\delta>0$,

$$\left\|\overline{\Sigma}_n-\Sigma\right\|^2=o\left(\frac{(\ln n)^{1+\delta}}{n}\right)a.s.$$

证明见第5节。

4.2 置信区间与统计假设检验

回顾一下，在假设1和假设2成立时，

$$\sqrt{n}\left(\widetilde{m}_n-m\right)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}\left(0,H^{-1}\Sigma H^{-1}\right).$$

因此，对于任意 $x_0\in\mathbb{R}^p\backslash\{0\}$

$$\frac{\sqrt{n}}{\sqrt{x_0^T\overline{S}_n^{-1}\overline{\Sigma}_n\overline{S}_n^{-1}x_0}}\left(x_0^T\widetilde{m}_n-x_0^Tm\right)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,1),$$

其中

$$\overline{S}_n=\frac{1}{n+1}\left(\sum_{k=1}^{n}\frac{\|X_k-\widetilde{m}_{k-1}\|}{\alpha_k^2}\widetilde{\phi}_k\widetilde{\phi}_k^T+S_0\right),$$

其中 S_0 为对称正定，$(Z_k)_k$ 为标准独立高斯向量，且 ϕ_k 由 $\phi_k=\nabla g(X_k,\widetilde{m}_{k-1}+\alpha_kZ_k)-\nabla g(X_k,\widetilde{m}_{k-1})$定义。由于 S_n^{-1}和 Σ_n可以递归计算（ S_n^{-1} 的更新见第 3.1.1 节），因此我们可以计算 x_0^Tm的在线置信区间，也就是说，可以确定任意指定方向上的中位数置信区间。此外，由于 $\widetilde{m}_n$ 是渐近有效的，因此有

$$n\left(\widetilde{m}_n-m\right)^T\overline{H}_n^*\overline{\Sigma}_n^{-1}\overline{H}_n^*\left(\widetilde{m}_n-m\right)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{X}_p^2,$$

其中

$$\overline{H}_n^*=\frac{1}{n+1}\left(\sum_{k=1}^{n}\frac{1}{\|X_k-\widetilde{m}_{k-1}\|}\left(I_p-\frac{(X_k-\widetilde{m}_{k-1})(X_k-\widetilde{m}_{k-1})^T}{\|X_k-\widetilde{m}_{k-1}\|^2}\right)+\overline{H}_0^*\right)$$

其中 H_0^* 为对称正定。因此， H_n^* 可以通过递归方式计算。回想一下， Σ_n^{-1} 也可以利用（10）递归更新，从而我们能够以显著性水平 $\alpha \in (0,1)$ 进行在线统计假设检验： H_0 : "$m = m_{test}$" 对比 H_1 : "$m \neq m_{test}$"。我们通过以下方式计算检验统计量 Z_n

$$Z_n = n\left(\tilde{m}_n - m_{test}\right)^T \overline{H}_n^* \overline{\Sigma}_n^{-1} \overline{H}_n^* \left(\tilde{m}_n - m_{test}\right),$$

并且，当 $Z_n > \zeta_{1-\alpha,p}$ 时，我们拒绝原假设，其中 $\zeta_{1-\alpha,p}$ 是自由度为 p 的卡方分布的 $1-\alpha$ 阶分位数。

4.3 模拟

现在，我们通过研究H_0下的经验水平来评估不同算法的性能。为此，我们生成大小为 $n = 3000$ 的高斯随机向量 $X \sim \mathcal{N}(0_p, \Sigma)$样本，其中 $p = 10$，并考察第3.3节中定义的协方差矩阵 Σ 的两种结构。我们通过对 $N = 1000$ 个样本进行实验来计算经验水平。我们考虑两种不同的初始化方式： $m_0 = rU$，其中 $U \sim \mathcal{N}_p(0, I_p)$，以及 $r = 1$ 或5。

Σ的结构	m_0	算法	经验水平（%）
(i)	U	WASN	6.1
		ASN	5.2
		SN	5.8
		ASGD	6.3
	$5U$	WASN	5.9
		ASN	10.7
		SN	5.4
		ASGD	23.0
(ii)	U	WASN	18.9
		ASN	16.8
		SN	44.3
		ASGD	56.8
	$5U$	WASN	19.1
		ASN	22.0
		SN	50.4
		ASGD	97.2

表 1：在 H_0下不同算法达到的经验水平

WASN 的表现尤为突出：即使 m_0 的初始化效果不佳，它所达到的经验水平仍然较低，并且更接近 5% 的理论水平。与之前实验中的观察一致，算法 ASN 对初始化很敏感。我们还可以看到，在所有考察的情形中，ASGD 达到的经验水平都是最高的。总体而言，所提出的二阶方法优于 ASGD；在考虑结构 (ii) 的情形下，改进更为显著。

此外，由于统计检验在原假设下具有卡方极限分布，我们现在关注检验统计量的模拟分布与

理论分布之间的接近程度。为此，我们绘制了通过不同算法获得的估计概率密度以及卡方概率密度。可以看出，采用二阶估计时，估计分布更接近理论分布，尤其是采用 WASN 估计量时。

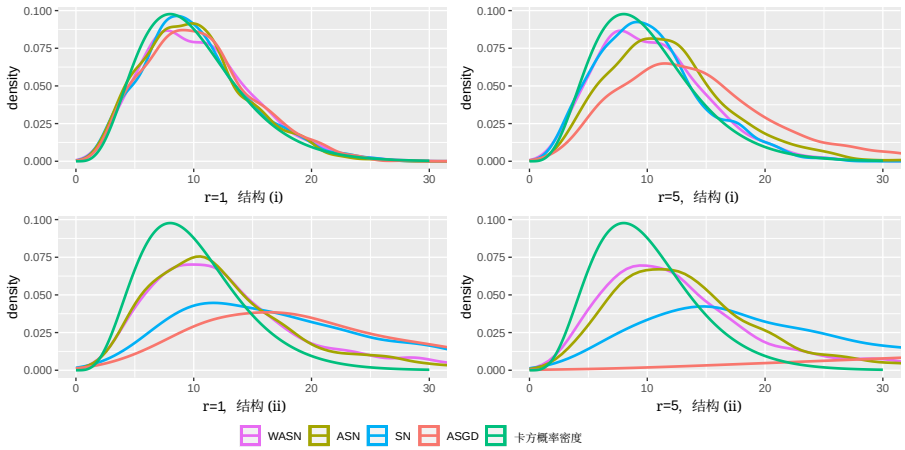

图 3: H_0 下检验统计量的模拟分布和理论分布

5 证明

在下文　中，$\|\cdot\|$ 表示向量的欧几里得范数或矩阵的谱范数　.

5.1 定理 1 和定理 2 的证明

注意，定理 1 和定理 2 的证明非常相似。因此，我们给出定理 1 的证明，并通过评注指出其中的差异。我们的目标是在 [3] 中应用定理 3.3（或定理 4.3）。为此，我们将验证 [3]，中称为**(A1b)**、**(A1c)**、**(A2a)**、**(A2b)**、**(A2c)**、**(H1)**、**(H2a)** 和**(H2b)**的假设均得到满足。

5.1.1 待最小化函数条件的验证

首先，我们将验证该函数应满足的假设。

(A1a) 的验证。假设1确保中位数 m 是方程的唯一解（参见 [14]和 [6]）

$$\nabla G(h) = 0,$$

因此，假设**(A1a)**成立。

(A1b) 的验证。回忆一下，$\nabla g(X,h) = -\frac{X-h}{\|X-h\|}$，因此对所有 $h \in \mathbb{R}^p$，

$$\|\nabla g(X,h)\| \leq 1,$$

假设**(A1b)** 因此成立。

对 (A1c) 的验证。对于所有 $h \in \mathbb{R}^p$

$$\Sigma(h) = \mathbb{E}\left[\nabla g(X,h)\nabla g(X,h)^T\right] = \mathbb{E}\left[\frac{(X-h)(X-h)^T)}{\|X-h\|^2}\right].$$

函数 Σ 在 $\mathbb{R}^p$上连续，因此满足假设**(A1c)** 。

验证 (A2a)。对所有 $h \in \mathbb{R}^p$,

$$\left\|\nabla^2 G(h)\right\| \leq \mathbb{E}\left[\frac{1}{\|X-h\|}\left\|I_p - \frac{(X-h)(X-h)^T)}{\|X-h\|^2}\right\|\right] \leq \mathbb{E}\left[\frac{1}{\|X-h\|}\right],$$

且假设1保证

$$\mathbb{E}\left[\frac{1}{\|X-h\|}\right] \leq C_6^{\frac{1}{6}},$$

因此满足假设**(A2a)**。

验证 (A2b)。在假设1和假设 2下，对于所有 $h \in \mathbb{R}^p$ ，海森矩阵 $\nabla^2 G(h)$正定（见 [6]第2.2节），因此满足假设**(A2b)** 。

[1]**验证 (A2c)**。在假设1下，海森矩阵 $\nabla^2 G(h)$ 是$6C_6^3$-利普希茨的（见引理 1），因此满足假设**(A2c)** 。

5.1.2 海森估计量特征值的可控性与一致性 cy

验证 (H1)。我们将验证海森估计量的特征值是否受到良好控制。回顾一下

$$\widetilde{H}_n = \overline{H}_n + \frac{1}{n+1}\sum_{k=1}^{n}\tilde{\beta}_k Z_k Z_k^T,$$

其中 $\tilde{\beta}_k = \frac{c_\beta}{k^\beta}$ 其中 $0 < \beta < \frac{1}{2}$且 $c_\beta > 0$。

注 5.1 对于 *WASN*，关于 β 的条件应为 $\beta < \gamma - \frac{1}{2}$，而不是 $\beta < \frac{1}{2}$。 –

因此我们有

$$\lambda_{min}(\tilde{H}_n) \geq \frac{\lambda_{min}(H_0)}{n+1} + \frac{1}{n+1}\lambda_{min}\left(\sum_{k=1}^{n}\tilde{\beta}_k Z_k Z_k^T\right).$$

此外,

$$\left(\frac{c_\beta}{1-\beta}n^{1-\beta}\right)^{-1}\sum_{k=1}^{n}\tilde{\beta}_k Z_k Z_k^T \xrightarrow[n\to+\infty]{a.s.} I_p,$$

因此 $\lambda_{max}\left(\widetilde{H}_n^{-1}\right)=\mathcal{O}\left(n^{\beta}\right)$。对于 H_n 的最大特征值，我们有

$$\begin{aligned}\left\|\overline{H}_n\right\| &\leq \frac{1}{n+1}\sum_{k=1}^{n}\frac{\|X_k-m_{k-1}\|}{\alpha_k^2}\|\phi_k\|^2+\frac{1}{n+1}\|H_0\| \\ &= \frac{1}{n+1}\sum_{k=1}^{n}\frac{\|X_k-m_{k-1}\|}{\alpha_k^2}\|\nabla g(X_k,m_{k-1})-\nabla g(X_k,m_{k-1}+\alpha_k Z_k)\|^2+\frac{1}{n+1}\|H_0\| \\ &= \frac{1}{n+1}\sum_{k=1}^{n}\frac{\|X_k-m_{k-1}\|}{\alpha_k^2}\left\|\frac{X_k-m_{k-1}}{\|X_k-m_{k-1}\|}-\frac{X_k-(m_{k-1}+\alpha_k Z_k)}{\|X_k-(m_{k-1}+\alpha_k Z_k)\|}\right\|^2+\frac{1}{n+1}\|H_0\|.\end{aligned}$$

由于 $\left\|\frac{C-A}{\|C-A\|}-\frac{B-A}{\|B-A\|}\right\|\leq 2\frac{\|C-B\|}{\|B-A\|}$（见 [5]第25页），我们有

$$\begin{aligned}\left\|\overline{H}_n\right\| &\leq \frac{4}{n+1}\sum_{k=1}^{n}\frac{\|X_k-m_{k-1}\|}{\alpha_k^2}\left\|\frac{\alpha_k Z_k}{\|X_k-m_{k-1}\|}\right\|^2+\frac{1}{n+1}\|H_0\| \\ &= \frac{1}{n+1}\sum_{k=1}^{n}\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}+\frac{1}{n+1}\|H_0\|.\end{aligned}$$

由假设1、 Z_k 与 X_k 的独立性以及霍尔德不等式可得

$$\mathbb{E}\left[\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}\,|\,\mathcal{F}_{k-1}\right]=4\mathbb{E}\left[\|Z_k\|^2\right]\mathbb{E}\left[\|X_k-m_{k-1}\|^{-1}\right]\leq 4pC_6^{\frac{1}{6}},$$

因此

$$\frac{1}{n+1}\sum_{k=1}^{n}\mathbb{E}\left[\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}\,|\,\mathcal{F}_{k-1}\right]\leq 4pC_6^{\frac{1}{6}}.$$

此外，采用类似的计算可得

$$\mathbb{E}\left[\left(\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}\right)^2\,|\,\mathcal{F}_{k-1}\right]=16\mathbb{E}\left[\|Z_k\|^4\right]\mathbb{E}\left[\|X_k-m_{k-1}\|^{-2}\right]\leq 16p(p+2)C_6^{\frac{1}{3}},$$

借助鞅大数定律，对所有 $\delta>0$,

$$\left(\frac{1}{n}\sum_{k=1}^{n}\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}-\mathbb{E}\left[\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}\,|\,\mathcal{F}_{k-1}\right]\right)^2=o\left(\frac{(\ln n)^{1+\delta}}{n}\right)a.s.$$

因此，

$$\begin{aligned}\left\|\overline{H}_n\right\| \leq& \left\|\frac{1}{n+1}\sum_{k=1}^{n}\mathbb{E}\left[\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}\,|\,\mathcal{F}_{k-1}\right]\right\|+\left\|\frac{1}{n+1}\sum_{k=1}^{n}\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}-\mathbb{E}\left[\frac{4\|Z_k\|^2}{\|X_k-m_{k-1}\|}\,|\,\mathcal{F}_{k-1}\right]\right\| \\ &+\frac{1}{n+1}\|H_0\| \\ =& \mathcal{O}(1)\ a.s.\end{aligned}$$

然后，

$$\left\|\widetilde{H}_n\right\| \leq \left\|\overline{H}_n\right\| + \frac{1}{n+1}\sum_{k=1}^{n}\tilde{\beta}_k = \mathcal{O}(1)\ a.s.$$

根据定理 3.1，在 [3]，中， $\widetilde{H}_n^{-1}$ 和 $\widetilde{H}_n$ 的最大特征值可以得到控制，随机牛顿估计量满足

$$m_n \xrightarrow[n\to+\infty]{a.s.} m.$$

注 5.2 对于 *WASN*，根据定理 *4.1*，在 *[3]*，中，估计量 $\hat{}m_n$ 几乎处处收敛到 m，这意味着 $m_{n,\tau}$ 几乎处处收敛。

5.1.3 海森估计量的收敛性与收敛速率

(H2a) 的验证。 现在我们验证海森估计量是否收敛到 $\nabla^2 G(m)$。定义

$$X_{k,t} := X_k - (m_{k-1} + t\alpha_k Z_k),$$

and

$$w_{k,t} := \frac{1}{\|X_{k,t}\|}\left(I_p - \frac{X_{k,t}X_{k,t}^T}{\|X_{k,t}\|^2}\right).$$

于是有

$$\begin{aligned}
\overline{H}_n &= \frac{1}{n+1}\sum_{k=1}^{n}\frac{\|X_k - m_{k-1}\|}{\alpha_k^2}\int_0^1 w_{k,t}dt\alpha_k Z_k\alpha_k Z_k^T\int_0^1 w_{k,t}dt + \frac{1}{n+1}H_0 \\
&= \frac{1}{n+1}\sum_{k=1}^{n}\|X_k - m_{k-1}\|\int_0^1 w_{k,t}dtZ_kZ_k^T\int_0^1 w_{k,t}dt + \frac{1}{n+1}H_0 \\
&= \overbrace{\frac{1}{n+1}\sum_{k=1}^{n}\|X_k - m_{k-1}\|\int_0^1 w_{k,t} - w_{k,0}dtZ_kZ_k^T\int_0^1 w_{k,t}dt}^{\mathcal{M}_{1,n}} \\
&+ \overbrace{\frac{1}{n+1}\sum_{k=1}^{n}\|X_k - m_{k-1}\| w_{k,0}Z_kZ_k^T\int_0^1 w_{k,t} - w_{k,0}dt}^{\mathcal{M}_{2,n}} \\
&+ \overbrace{\frac{1}{n+1}\sum_{k=1}^{n}\|X_k - m_{k-1}\| w_{k,0}Z_kZ_k^T w_{k,0}}^{\mathcal{M}_{3,n}} + \frac{1}{n+1}H_0
\end{aligned}$$

收敛性为$\mathcal{M}_{3,\ n}$。我们定义

$$\mathcal{Y}_k := \|X_k - m_{k-1}\| w_{k,0}Z_kZ_k^T w_{k,0},$$

注意到有

$$\mathcal{M}_{3,n} = \frac{1}{n+1}\sum_{k=1}^{n}\mathcal{Y}_k = \frac{1}{n+1}\sum_{k=1}^{n}\mathbb{E}\left[\mathcal{Y}_k \,|\, \mathcal{F}_{k-1}\right] + \frac{1}{n+1}\sum_{k=1}^{n}\mathcal{Y}_k - \mathbb{E}\left[\mathcal{Y}_k \,|\, \mathcal{F}_{k-1}\right].$$

首先，我们证明

$$\frac{1}{n+1}\sum_{k=1}^{n}\mathbb{E}\left[\mathcal{Y}_k \,|\, \mathcal{F}_{k-1}\right] \xrightarrow[n\to+\infty]{a.s.} \nabla^2 G(m).$$

我们有

$$w_{k,0} := \frac{1}{\|X_k - m_{k-1}\|}\left(I_p - \left(\frac{X_k - m_{k-1}}{\|X_k - m_{k-1}\|}\right)\left(\frac{X_k - m_{k-1}}{\|X_k - m_{k-1}\|}\right)^T\right) = \nabla^2 g(X_k, m_{k-1}),$$

因此，根据方程（3）

$$\|X_k - m_{k-1}\| w_{k,0}^2 = w_{k,0}.$$

此外，由于中位数估计量满足

$$m_n \xrightarrow[n\to+\infty]{a.s.} m,$$

由连续性可知

$$\mathbb{E}\left[w_{n,0} \,|\, \mathcal{F}_{n-1}\right] = \nabla^2 G(m_n) \xrightarrow[n\to+\infty]{a.s.} \nabla^2 G(m).$$

因此，当 $\{Z_k \quad\}_k$ 是标准独立高斯向量，根据大数定律，我们有 e

$$\begin{aligned}\frac{1}{n+1}\sum_{k=1}^{n}E\left[\mathcal{Y}_k \,|\, \mathcal{F}_{k-1}\right] &= \frac{1}{n+1}\sum_{k=1}^{n}E\left[\|X_k - m_{k-1}\| w_{k,0}^2 \,|\, \mathcal{F}_{k-1}\right] \\ &= \frac{1}{n+1}\sum_{k=1}^{n}E\left[w_{k,0} \,|\, \mathcal{F}_{k-1}\right] \xrightarrow[n\to+\infty]{a.s.} \nabla^2 G(m).\end{aligned}$$

此外，根据假设1以及独立性，

$$\begin{aligned}\mathbb{E}\left[\|\mathcal{Y}_k\|^2 \mid \mathcal{F}_{k-1}\right] &\le \mathbb{E}\left[\|X_k - m_{k-1}\|^2 \|w_{k,0}\|^4 \|Z_k\|^4 \mid \mathcal{F}_{k-1}\right] \\ &\le \mathbb{E}\left[\|\mathcal{Z}_k\|^4 \mid \mathcal{F}_{k-1}\right]\mathbb{E}\left[\left\|\frac{1}{X_k - m_{k-1}}\right\|^2 \mid \mathcal{F}_{k-1}\right] \\ &\le p(p+2)C_6^{\frac{1}{3}},\end{aligned}$$

由此，借助鞅大数定律可得，对所有 $\delta > 0$,

$$\left\|\frac{1}{n}\sum_{k=1}^{n}\mathcal{Y}_k - \mathbb{E}\left[\mathcal{Y}_k \,|\, \mathcal{F}_{k-1}\right]\right\|^2 = o\left(\frac{(\ln n)^{1+\delta}}{n}\right).$$

因此，

$$\mathcal{M}_{3,n} \xrightarrow[n\to+\infty]{a.s.} \nabla^2 G(m).$$

的收敛性$\mathcal{M}_{2,\ n}$：为了得到$\mathcal{M}_{2,\ n}$的收敛速率，我们先介绍 [4]中引理 5.1的一个推广。

引理 1 对于所有 h，$h' \in \mathbb{R}^p$ 和 $0 < q \leq 3$,

$$\left(\mathbb{E}\left[\left\|\nabla^2 g(X,h') - \nabla^2 g(X,h)\right\|^q\right]\right)^{\frac{1}{q}} \leq 6C_6^{\frac{1}{3}} \left\|h' - h\right\|.$$

在我们的情形中，对于所有 $t \in (0,1)$ 和 $q \in (0,3]$,，并考虑滤子 $\mathcal{F}'_k = \sigma\left(X_1, \ldots, X_{k-1}, Z_1, \ldots, Z_k\right)$，我们有

$$\mathbb{E}\left[\|w_{k,t} - w_{k,0}\|^q \,|\mathcal{F}_k\right] \leq 6^q C_6^{\frac{q}{3}} \alpha_k^q \|Z_k\|^q .$$

我们定义

$$\mathcal{W}_k := \|X_k - m_{k-1}\| \, w_{k,0} Z_k Z_k^T \int_0^1 w_{k,t} - w_{k,0} dt.$$

于是,

$$\|\mathcal{M}_{2,n}\| \leq \frac{1}{n+1} \sum_{k=1}^n \mathbb{E}\left[\|\mathcal{W}_k\| \mid \mathcal{F}_{k-1}\right] + \frac{1}{n+1} \sum_{k=1}^n \|\mathcal{W}_k\| - \mathbb{E}\left[\|\mathcal{W}_k\| \mid \mathcal{F}_{k-1}\right].$$

注意到

$$\begin{aligned}
\mathbb{E}\left[\|\mathcal{W}_k\| \mid \mathcal{F}_{k-1}\right] &\leq \mathbb{E}\left[\|X_k - m_{k-1}\| \, \|w_{k,0}\| \, \|Z_k\|^2 \left\|\int_0^1 w_{k,t} - w_{k,0}\right\| dt \,|\, \mathcal{F}_{k-1}\right] \\
&\leq \mathbb{E}\left[\|Z_k\|^2 \int_0^1 \|w_{k,t} - w_{k,0}\| \, dt \,|\, \mathcal{F}_{k-1}\right] \\
&= \mathbb{E}\left[\|Z_k\|^2 \int_0^1 \mathbb{E}\left[\|w_{k,t} - w_{k,0}\| \, dt \,|\, \mathcal{F}'_k\right] \mid \mathcal{F}_{k-1}\right],
\end{aligned}$$

其中 $\mathcal{F}_{k'} = \sigma\{X_1, ..., X_{k-1}, Z_1, ..., Z_k\}$。因此，根据引理 1，我们有 ve

$$\mathbb{E}\left[\|\mathcal{W}_k\| \mid \mathcal{F}_{k-1}\right] \leq 6\alpha_k C_6^{\frac{1}{3}} \mathbb{E}\left[\|Z_k\|^3 \mid \mathcal{F}_{k-1}\right].$$

由于 $\alpha_k = \frac{1}{k \ln k+1}$，可得 $\mathbb{E}\left[\|\mathcal{W}_k\| \mid \mathcal{F}_{k-1}\right] = \mathcal{O}\left(\frac{1}{k \ln k}\right)$，因此

$$\frac{1}{n+1} \sum_{k=1}^n \mathbb{E}\left[\|\mathcal{W}_k\| \mid \mathcal{F}_{k-1}\right] = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.$$

此外，根据引理 1，并借助霍尔德不等式,

$$\begin{aligned}
\mathbb{E}\left[\|\mathcal{W}_k\|^2 \mid \mathcal{F}_{k-1}\right] &\leq \mathbb{E}\left[\|Z_k\|^4 \left(\int_0^1 \|w_{k,t} - w_{k,0}\| \, dt\right)^2 \mid \mathcal{F}_{k-1}\right]^2 \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\|Z_k\|^4 \left(\int_0^1 \|w_{k,t} - w_{k,0}\| \, dt\right)^2 \mid \mathcal{F}'_k\right] \mid \mathcal{F}_{k-1}\right] \\
&\leq \mathbb{E}\left[\|Z_k\|^4 \int_0^1 \mathbb{E}\left[\|w_{k,t} - w_{k,0}\|^2 \,|\mathcal{F}'_k\right] dt | \mathcal{F}_{k-1}\right] \\
&\leq 36\alpha_k^2 \mathbb{E}\left[\|Z_k\|^6 \mid \mathcal{F}_{k-1}\right] C_6^{\frac{2}{3}}.
\end{aligned}$$

因此，借助鞅大数定律，

$$\left\| \frac{1}{n} \sum_{k=1}^{n} \mathcal{W}_k - \mathbb{E}\left[\|\mathcal{W}_k\| \mid \mathcal{F}_{k-1} \right] \right\|^2 = o\left(\frac{(\ln n)^{1+\delta}}{n} \right) a.s.$$

因此，我们得到

$$\mathcal{M}_{2,n} = o\left(\frac{(\ln n)^{1+\delta}}{n} \right) a.s.$$

收敛性为$\mathcal{M}_{1,\ n}$。我们定义

$$\mathcal{V}_k := \|X_k - m_{k-1}\| \int_0^1 w_{k,t} - w_{k,0} dt Z_k Z_k^T \int_0^1 w_{k,t} dt.$$

注意到

$$\begin{aligned}
\mathbb{E}\left[\|\mathcal{V}_k\| \mid \mathcal{F}_{k-1} \right] &\leq \mathbb{E}\left[\|X_k - m_{k-1}\| \left\| \int_0^1 w_{k,t} dt \right\| \|Z_k\|^2 \left\| \int_0^1 w_{k,t} - w_{k,0} dt \right\| \mid \mathcal{F}_{k-1} \right] \\
&\leq \mathbb{E}\left[\|Z_k\|^2 \int_0^1 \frac{\|X_{k,t}\| + t\alpha_k \|Z_k\|}{\|X_{k,t}\|} dt \int_0^1 \|w_{k,t} - w_{k,0}\| \, dt \, | \, \mathcal{F}_{k-1} \right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\|Z_k\|^2 \left(1 + \int_0^1 \frac{\alpha_k \|Z_k\|}{\|X_{k,t}\|} dt \right) \int_0^1 \|w_{k,t} - w_{k,0}\| \, dt \, | \, \mathcal{F}'_k \right] \mid \mathcal{F}_{k-1} \right]
\end{aligned}$$

其中 $\mathcal{F}'_k = \sigma\left(X_1, \ \ldots\ldots, \ X_{k-1}, \ Z_1, \ \ldots\ldots, \ Z_k \right)$。因此，根据引理 1 和假设1，利用霍尔德不等式可得

$$\begin{aligned}
\mathbb{E}\left[\|\mathcal{V}_k\| \mid \mathcal{F}_{k-1} \right] &\leq \mathbb{E}\left[\|Z_k\|^2 \int_0^1 \mathbb{E}\left[\|w_{k,t} - w_{k,0}\| \, | \mathcal{F}'_k \right] dt \right] \\
&\quad + \mathbb{E}\left[\alpha_k \|Z_k\|^3 \left(\int_0^1 \mathbb{E}\left[\frac{1}{\|X_{k,t}\|^2} | \mathcal{F}'_k \right] dt \right)^{\frac{1}{2}} \left(\int_0^1 \mathbb{E}\left[\|w_{k,t} - w_{k,0}\|^2 | \mathcal{F}'_k \right] dt \right)^{\frac{1}{2}} | \mathcal{F}_{k-1} \right] \\
&\leq 6\alpha_k C_6^{\frac{1}{3}} \mathbb{E}\left[\|Z_k\|^3 \mid \mathcal{F}_{k-1} \right] + 6\alpha_k^2 \mathbb{E}\left[\|Z_k\|^4 | \mathcal{F}_{k-1} \right] C_6^{\frac{1}{2}}.
\end{aligned}$$

我们有 $\alpha_k = \underline{\frac{1}{k \ln k + 1}}$，因此 $\mathbb{E}\left[\|\mathcal{V}_k\| \mid \mathcal{F}_{k-1} \right] = \mathcal{O}\left(\underline{\frac{1}{k \ln k}} \right)$，这导出

$$\frac{1}{n+1} \sum_{k=1}^{n} \mathbb{E}\left[\|\mathcal{V}_k\| \mid \mathcal{F}_{k-1} \right] = \mathcal{O}\left(\frac{\ln n}{n} \right).$$

此外，根据霍尔德不等式，我们有

$$\begin{aligned}
\mathbb{E}\left[\|\mathcal{V}_k\|^2 \mid \mathcal{F}_{k-1} \right] &\leq \mathbb{E}\left[\|Z_k\|^4 \left(1 + \int_0^1 \frac{\alpha_k \|Z_k\|}{\|X_{k,t}\|} dt \right)^2 \left(\int_0^1 \|w_{k,t} - w_{k,0}\| \, dt \right)^2 \mid \mathcal{F}_{k'} \mid \mathcal{F}_{k-1} \right] \\
&\leq 2\mathbb{E}\left[\|Z_k\|^4 \int_0^1 \|w_{k,t} - w_{k,0}\|^2 \, dt \, | \, \mathcal{F}_{k-1} \right] + 2\alpha_k^2 \mathbb{E}\left[\|Z_k\|^6 \int_0^1 \int_0^1 \frac{1}{\|X_{k,t}\|^2} \left\| w_{k,t'} - w_{k,0} \right\|^2 dt dt' \, | \, \mathcal{F}_{k-1} \right] =: (*)
\end{aligned}$$

因此，应用霍尔德不等式可得，

$$(*) \leq 2\mathbb{E}\left[\|Z_k\|^4 \int_0^1 \mathbb{E}\left[\|w_{k,t}-w_{k,0}\|^2 |\mathcal{F}'_k\right] dt \,|\, \mathcal{F}_{k-1}\right]$$
$$+ 2\alpha_k^2\mathbb{E}\left[\|Z_k\|^6 \int_0^1\int_0^1 \left(\mathbb{E}\left[\frac{1}{\|X_{k,t}\|^6}|\mathcal{F}'_k\right]\right)^{\frac{1}{3}} \left(\mathbb{E}\left[\left\|w_{k,t'}-w_{k,0}\right\|^3 |\mathcal{F}'_k\right]\right)^{\frac{2}{3}} dtdt' \,|\, \mathcal{F}_{k-1}\right].$$

于是，得益于假设1和引理 1，

$$\mathbb{E}\left[\|\mathcal{V}_k\|^2 \mid \mathcal{F}_{k-1}\right] \leq 72\alpha_k^2 C_6^{2/3}\mathbb{E}\left[\|Z_k\|^6\right] + 72\alpha_k^4 C_6\mathbb{E}\left[\|Z_k\|^8\right].$$

借助鞅大数定律，可得

$$\left\|\frac{1}{n}\sum_{k=1}^n \mathcal{V}_k - \mathbb{E}\left[\mathcal{V}_k \,|\, \mathcal{F}_{k-1}\right]\right\|^2 = o\left(\frac{(\ln n)^{1+\delta}}{n}\right) a.s$$

and

$$\mathcal{M}_{1,n} = o\left(\frac{(\ln n)^{1+\delta}}{n}\right) a.s.$$

最后，我们有

$$\overline{H}_n \xrightarrow[n\to+\infty]{a.s.} \nabla^2 G(m).$$

注意到

$$\widetilde{H}_n = \overline{H}_n + \frac{1}{n+1}\sum_{k=1}^n \tilde{\beta}_k Z_k Z_k^T,$$

and

$$\frac{1}{n+1}\sum_{k=1}^n \tilde{\beta}_k Z_k Z_k^T \xrightarrow[n\to+\infty]{a.s.} 0.$$

因此，

$$\widetilde{H}_n \xrightarrow[n\to+\infty]{a.s.} \nabla^2 G(m).$$

根据定理 3.2， [3], 随机牛顿估计量满足

$$\|m_n - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.$$

注 5.3 对于 *WASN*，根据 *[3],* 中的定理 *4.2*，我们有

$$\|\hat{m}_n - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n^\gamma}\right) a.s., \quad \textit{which implies that} \quad \|m_{n,\tau} - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n^\gamma}\right) a.s.$$

5.1.4 海森矩阵估计量的收敛速率与渐近效率

~现在我们给出 H_n的收敛速率。回顾一下,

$$w_{k,0} := \frac{1}{\|X_k - m_{k-1}\|}\left(I_p - \left(\frac{X_k - m_{k-1}}{\|X_k - m_{k-1}\|}\right)\left(\frac{X_k - m_{k-1}}{\|X_k - m_{k-1}\|}\right)^T\right),$$

这意味着

$$\mathbb{E}\left[w_{k,0} \,|\, \mathcal{F}_{k-1}\right] = \nabla^2 G(m_{k-1}).$$

[1]注意， $\nabla^2 G(h)$ 是 $6C_6^{\text{a}}$-利普希茨连续的，因此我们有

$$\mathbb{E}\left[\left\|w_{k,0} - \nabla^2 G(m)\right\|^2 \,|\, \mathcal{F}_{k-1}\right] = \left\|\nabla^2 G(m_{k-1}) - \nabla^2 G(m)\right\|^2 \leq 6C_6^{\frac{1}{3}} \left\|m_{k-1} - m\right\|^2.$$

由于该估计量满足

$$\|m_n - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.,$$

我们有

$$\left\|\mathbb{E}\left[w_{k,0} \,|\, \mathcal{F}_{k-1}\right] - \nabla^2 G(m)\right\|^2 = \mathcal{O}\left(\frac{\ln k}{k}\right) a.s.$$

备注 5.4 对于 *WASN*，这里有 $\|m_{n,\tau} - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n^\gamma}\right) a.s.$，因此 $\left\|\mathbb{E}\left[w_{k,0} \,|\, \mathcal{F}_{k-1}\right] - \nabla^2 G(m)\right\|^2 = \mathcal{O}\left(\frac{\ln k}{k^\gamma}\right) a.s.$

因此，由于 $\{Z_k\}_k$ 是标准独立高斯向量，我们有

$$\left\|\frac{1}{n+1}\sum_{k=1}^{n} E\left[\mathcal{Y}_k \,|\, \mathcal{F}_{k-1}\right] - \nabla^2 G(m)\right\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.$$

我们已证明，对所有 $\delta > 0$

$$\left\|\frac{1}{n}\sum_{k=1}^{n} \mathcal{Y}_k - \mathbb{E}\left[\mathcal{Y}_k \,|\, \mathcal{F}_{k-1}\right]\right\|^2 = o\left(\frac{(\ln n)^{1+\delta}}{n}\right) a.s.,$$

因此

$$\left\|\mathcal{M}_{3,n} - \nabla^2 G(m)\right\|^2 = o\left(\frac{(\ln n)^{1+\delta}}{n}\right) a.s.$$

最后，对于所有 $\delta > 0$，海森估计量满足

$$\left\|\widetilde{H}_n - H\right\|^2 = \mathcal{O}\left(\max\left\{\frac{(\ln n)^{1+\delta}}{n}, \frac{c_\beta}{n^{2\beta}}\right\}\right) a.s.$$

根据 [3], 中的定理 3.3，随机牛顿估计量满足

$$\sqrt{n}\left(m_n - m\right) \sim \mathcal{N}\left(0, H^{-1}\Sigma H^{-1}\right),$$

其中 $\Sigma = \mathbb{E}\left[\nabla g(X,\ m)\nabla g(X,\ m)^T\right]$.

注 5.5 对于 *WASN*，我们首先对所有 $\delta > 0$

$$\left\|\widetilde{H}_{n,\tau} - H\right\|^2 = \mathcal{O}\left(\max\left\{\frac{(\ln n)^{1+\delta}}{n^{\gamma}}, \frac{c_\beta}{n^{2\beta}}\right\}\right) a.s.$$

那么根据 *[3],* 定理 *4.3*，我们有

$$\|m_{n,\tau} - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.,$$

由此得到

$$\left\|\widetilde{H}_{n,\tau} - H\right\|^2 = \mathcal{O}\left(\max\left\{\frac{(\ln n)^{1+\delta}}{n}, \frac{c_\beta}{n^{2\beta}}\right\}\right) a.s. \quad \textit{and} \quad \sqrt{n}\left(m_{n,\tau} - m\right) \sim \mathcal{N}\left(0, H^{-1}\Sigma H^{-1}\right).$$

5.2 定理 3 的证明

我们定义

$$\mathcal{T}_k := \frac{(X_k - \widetilde{m}_{k-1})}{\|X_k - \widetilde{m}_{k-1}\|}\frac{(X_k - \widetilde{m}_{k-1})^T}{\|X_k - \widetilde{m}_{k-1}\|},$$

于是有

$$\Sigma_n = \frac{1}{n+1}\sum_{k=1}^{n}\mathcal{T}_k + \frac{1}{n+1}\Sigma_0 = \frac{1}{n+1}\sum_{k=1}^{n}\mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right] + \frac{1}{n+1}\sum_{k=1}^{n}\mathcal{T}_k - \mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right] + \frac{1}{n+1}\Sigma_0.$$

注意，$\mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right] = \Sigma\left(\widetilde{m}_{k-1}\right)$。此外，我们有

$$\Sigma(h) = \mathbb{E}\left[\nabla g(X,h)\nabla g(X,h)^T\right]$$

[1]因此，得益于假设 **(A2c)**，$\Sigma(h)$ 是 $6C_6^6$-利普希茨的（见 [11] 第 6.2 节），这意味着

$$\|\mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right] - \Sigma\|^2 = \|\Sigma(\widetilde{m}_{k-1}) - \Sigma(m)\|^2 \leq 6C_6^{\frac{1}{6}}\|\widetilde{m}_{k-1} - m\|^2.$$

由于估计量 $\widetilde{m}_k$ 满足

$$\|\widetilde{m}_k - m\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.,$$

我们得到

$$\|\mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right] - \Sigma\|^2 = \mathcal{O}\left(\frac{\ln n}{n}\right) a.s.$$

此外，显然

$$\mathbb{E}\left[\|\mathcal{T}_k\|^2 \,|\, \mathcal{F}_{k-1}\right] \leq 1,$$

借助鞅大数定律，从而，

$$\left\|\frac{1}{n}\sum_{k=1}^{n}\mathcal{T}_k - \mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right]\right\|^2 = o\left(\frac{(\ln n)^{1+\delta}}{n}\right) a.s.$$

最后，我们有

$$\begin{aligned}\|\Sigma_n - \Sigma\|^2 &\leq \left\|\frac{1}{n+1}\sum_{k=1}^{n}\mathcal{T}_k - \mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right]\right\|^2 + \left\|\frac{1}{n+1}\sum_{k=1}^{n}\Sigma_n - \mathbb{E}\left[\mathcal{T}_k \,|\, \mathcal{F}_{k-1}\right]\right\|^2 + \left\|\frac{1}{n+1}\Sigma_0\right\|^2 \\ &= o\left(\frac{(\ln n)^{1+\delta}}{n}\right) a.s.\end{aligned}$$

参考文献

[1] Beck, A. 和 Sabach, S.（2015）。Weiszfeld 方法：旧结果与新进展。优化理论与应用期刊，164(1):1–40。

[2] Bercu, B., Godichon, A., 和 Portier, B. (2020). 一种用于逻辑回归参数估计的高效随机牛顿算法。*SIAM* 控制与优化杂志, 58(1):348– 367.

[3] Boyer, C. 和 Godichon-Baggioni, A. (2022). 随机牛顿算法及其加权平均版本的渐近收敛速率。计算优化与应用, pages 1–52.

[4] Cardot, H., Cénac, P., 和 Godichon-Baggioni, A. (2017). 希尔伯特空间中几何中位数的在线估计：非渐近置信球。《统计年刊》, 45(2):591– 614.

[5] Cardot, H., Cénac, P., 和 Monnez, J.-M. (2012). 一种用于大型数据集聚类的快速递归算法：k-中位数。计算统计与数据分析, 56(6):1434–1449.

[6] Cardot, H., Cénac, P. 和 Zitt, P.-A.（2013）。使用平均随机梯度算法，高效快速地估计希尔伯特空间中的几何中位数。伯努利，19(1)：18–43。

[7] Church, J. C., Chen, Y. 和 Rice, S. V.（2008）。用于去除数字图像噪声的空间中值滤波器。载于 *IEEE SoutheastCon2008*，第 618–623 页。IEEE。

[8] Duflo, M.（199 7）。随机迭代模型、随机建模与应用概率，第 34 卷 .

[9] Gini, C. 和 Galvani, L.（1929）。论均值概念在定性特征上的若干扩展。*Metron*，8(1-2)：3–209。

[10] Godichon-Baggioni, A.（2016年）。使用随机梯度算法估计希尔伯特空间中的几何中位数：Lp 收敛速率与几乎必然收敛速率。《多元分析杂志》，146：209–222。

[11] Godichon-Baggioni，A.（2019年）。平均随机梯度算法渐近方差的在线估计。统计规划与推断杂志，203：1–19。

[12] Gower, J.（1974年）。算法 AS 78：中位中心。《皇家统计学会杂志》。*C* 辑（应用统计），23(3)：466–470。

[13] Haldane, J.（1948）。关于多元分布的中位数的注记。*Biometrika*，35(3-4)：414–417。[14] Kemperman, J.（1987）。巴拿赫空间上有限测度的中位数。基于 *L1* 范数及相关方法的统计数据分析（纳沙泰尔，*1987*），第 217–230 页。[15] Koltchinskii, V. I.（1997）。M 估计、凸性与分位数。《统计年刊》，第 435–477 页。[16] Lopuhaa, H. P. 和 Rousseeuw, P. J.（1991）。多元位置和协方差矩阵仿射等变估计量的崩溃点。《统计年刊》，第 229–248 页。[17] Lu, X.、Chen, W. 和 Schaefer, S.（2017）。通过顶点预滤波和 L1 中位数法线滤波实现鲁棒网格去噪。《计算机辅助几何设计》，54：49–60。[18] Vardi, Y. 和 Zhang, C.-H.（2000）。多元 L1 中位数及相关数据深度。《美国国家科学院院刊》，97(4)：1423–1426。[19] Weiszfeld, E.（1937）。关于使给定 n 个点到该点的距离之和最小的点。《东北数学杂志》第一辑，43：355–386。[20] Zhang, Y.、Wang, X.、Li, Q. 和 Yan, F.（2021）。一种基于法向量和 L1 中位数的环形锻件截线提取新方法。《测量》，184：109894。