arXiv:2603.06176v1 [math.ST] 6 Mar 2026

基于离散观测的高维 Lévy 驱动 Ornstein–Uhlenbeck 过程稀疏估计

Niklas Dexheimer* Natalia Jeszka*

2026年3月9日

摘要

我们研究基于离散观测的 Lévy 驱动 Ornstein–Uhlenbeck 过程的高维漂移估计。假设漂移矩阵具有稀疏性，我们分析基于近似似然构造的 Lasso 和 Slope 估计量，并推导出精确的非渐近预言不等式。我们的界将离散化误差与随机波动的影响区分开来，并证明在高频制下，选择合适的调节参数即可达到极小极大最优收敛速率。我们还量化了根据 Lévy 噪声达到这些速率所需的样本复杂度。这些结果将随机过程高维统计的理论拓展到范围显著更广的噪声机制，尤其是纯跳过程。研究还表明，Lasso 和 Slope 在跳驱动系统中依然具有竞争力，并为 Lévy 过程作为自然建模选择的应用场景提供了实用的推断指导。

1. 引言

高维统计是现代统计研究的核心领域，尤其关注参数数量显著超过样本量的情形。在这类情形中，一个常见的假设是底层参数具有稀疏性。传统估计技术通常不足以应对这一问题，因此人们引入了惩罚估计量来克服这些局限。Lasso 估计量 [24] 是研究最广泛的例子之一。它专为稀疏设置设计，通过在估计目标中加入 L_1 惩罚项来实现这一点。Lasso 估计量的一种推广是 Slope 估计量[4]，，它在 L_1 惩罚项中引入了额外的权重参数。如今，这些估计量在经典情形中的性质已得到充分理解，例如在稀疏线性回归中，其极小极大最优性已得到验证 [3]。近年来，人们日益关注将高维统计理论拓展到连续时间情形，尤其是扩散过程 [1, 7, 8, 11, 17]。在这一背景下，一类重要的模型是多元 Ornstein– Uhlenbeck（OU）过程。这些过程是 d维随机过程 $\mathbf{X} = (X_t)_{t\geq 0}$，满足如下形式的随机微分方程（SDE）

$$\mathrm{d}X_t = -\boldsymbol{A}_0 X_t \,\mathrm{d}t + \mathrm{d}W_t, \quad t \geq 0, \tag{1.1}$$

其中 $\boldsymbol{A}_0 \in \mathrm{R}^{d\times d}$ 称为漂移矩阵，且$(W_t)_{t\geq 0}$ 是标准 d维布朗运动。标量 OU 过程常用于利率建模，即 Vasicek 模型。因此，多元 OU 过程是用于建模的自然候选

*特温特大学应用数学系，Drienerlolaan 5，7522 NB
恩斯赫德，荷兰。电子邮箱：
n.dexheimer@utwente.nl/n.jeszka@utwente.nl

银行间借贷 [5, 16]。银行之间的相互作用编码于漂移矩阵 A_0之中；在此设定下，当银行数量非常多时，自然可以假设 A_0的稀疏性。推广该模型的一种直接方法是纳入随机冲击，例如关键利率的变化，即将布朗运动替换为一般 d维 Lévy 过程。Lévy 驱动的 OU 过程的其他应用例如见于计算神经科学，其中它们被用于对生物神经网络中的突触后膜电位进行建模，参见第 2.1节（见 [10]）。该模型一个特别有趣之处在于，驱动 Lévy 过程是纯跳过程。从理论角度看，将（1.1）中的噪声扩展到一般Lévy过程，也会将 $\mathbf{X}$ 的可能极限分布类从正态分布拓宽到自可分解分布类，详情参见[20, 22]。

主要贡献 本文研究漂移矩阵 A_0 的估计，该矩阵属于一个 d维 Lévy 驱动的 OU 过程 $\mathbf{X}$，基于 $n \in \mathbb{N}$ 个等距离散观测，观测对象为$\mathbf{X}$，相邻间隔为$\Delta_n > 0$。我们尤其关注 A_0 稀疏的情形，因此采用 Lasso 与 Slope 估计量。为此，我们分别定义局部化截断似然与对比函数，见第 2.2 节。总之，我们的主要贡献如下。

1. 我们给出估计量的 L_2误差的尖锐oracle不等式，见定理 3.1。这些不等式将离散化、截断与随机波动对总误差的贡献分离开来。

2. 两种估计量的随机误差均被证明具有如下阶

$$\frac{s \log\left(\frac{\mathrm{e}d^2}{s}\right)}{T},$$

其中 s 表示 A_0 的稀疏性，$T = n\Delta_n$ 表示观测时长。若可获得直至时刻 T 的$\mathbf{X}$ 的连续观测记录，则这是稀疏性约束下的极小极大最优收敛速率（见 [11]中的定理 2.7）。此外，离散化误差以 $d^2\Delta_n^2$为上界，从而改进了文献中的已有结果。文中给出了依赖于背景驱动 Lévy 过程的截断水平的最优选择。据我们所知，这是首个针对扩散过程的高频离散观测、为稀疏回归保证极小极大最优性的结果。

3. 这些结果在对背景驱动 Lévy 过程的最小假设下成立，即只要 BDLP 具有 p 阶矩并满足 $p > 2$。我们的分析尤其适用于纯跳过程和各向异性噪声；据我们所知，这些情形此前尚未在随机过程的高维统计中得到研究。

4. 我们还针对 $\mathbf{X}$ 的局部化经验协方差矩阵证明了一个新颖的矩阵Bernstein型集中不等式（见命题 4.3），从而验证了理论分析 Lasso 估计量所需的受限特征值性质的一个类似性质（见 [7, 11] 的第 3.1 节）。由此得到的样本复杂度至多关于稀疏性 s 呈多项式增长，其具体形式取决于 BDLP 的性质。

相关工作 本文所研究的相同设定已在 [11] 中进行过探讨，但该研究采用了高度理想化的假设，即可以获得关于 $\mathbf{X}$ 的连续观测记录。作者证明，对于恰当选择的调节参数，所提出的估计量能够达到极小极大最优的收敛速率。尽管如此，除不现实的连续观测方案外，该论文还作出了多项限制性假设。首先， [11]假设

已知所观测 Lévy 驱动的 OU 过程的连续鞅部分，因为计算似然函数需要这一信息（见 [23]）。然而，即使给定 $\mathbf{X}$ 的连续观测，只要背景驱动 Lévy 过程（BDLP）具有无限跳跃活性，该连续鞅部分就难以辨识。此外，如果 BDLP 是纯跳过程，这些估计量便无法使用，因为它们基于似然函数，因而也依赖于连续鞅部分，而纯跳过程中并不存在连续鞅部分。对 [11] 的分析还依赖于 $\mathbf{X}$ 的经验协方差矩阵具有足够强的集中性（见 $(\mathcal{H})$ （位于 [11]中）），这与 Lasso 估计量分析中的通常做法一致（更多详情见第 4.2节）。然而，在存在跳跃的情况下，只能验证非常弱的集中性（见 [11, 命题 4。2]）。若 BDLP 的 Lévy 测度具有重尾，这会导致样本复杂度随维数 d呈指数级增长，这是非常不理想的。

将 [11] 的结果转化为离散观测方案的传统方法，是采用跳跃过滤方法来近似似然函数中的连续鞅部分。 [19] 使用了这一方法，对离散观测的标量 Lévy 驱动的 OU 过程的漂移参数进行最大似然估计。然而，正如 [19], 中注 3.4 所述，只要 BDLP 的漂移非零，就“没有希望”恢复连续鞅部分，这排除了许多类型的各向异性鞅噪声。此外，如上所述，纯跳过程也无法采用跳跃过滤，因此我们使用了另一种方法。

首篇研究离散观测扩散过程高维估计的论文是 [1]。与我们的工作相比，作者研究了更一般的一类参数扩散过程，其噪声为标准布朗运动。此外，作者还研究了经典 OU 过程，因为这类过程不属于所考虑的一般框架。我们在注3.2中详细比较了本文与该论文的结果；简言之，通过对 OU 情形进行详细分析，我们同时改进了随机误差和离散化误差的上界。 [8] 研究了基于离散观测值、适用于非常一般的扩散过程的自适应弹性网络估计量。该论文不仅关注漂移的参数估计，也关注波动率的估计。该估计量定义为给定原始估计量的惩罚版本，而在大多数情况下，原始估计量为最大似然估计。由于论文的研究范围非常广泛，作者主要关注证明弹性网络估计量的表现与原始估计量相当。另有一篇相关论文 [28] 研究了重尾 β-混合时间序列的稀疏回归。事实上，根据所施加的假设（见假设$(\mathcal{A}_0)$），由$\mathbf{X}$表示的 Lévy 驱动的 OU 过程，根据 β的结果是[20], -混合的；而在（1.1）中允许噪声为一般 Lévy 过程，相当于允许 [28]所研究的时间序列动态中存在重尾噪声。与 [28] 相比，我们提高了 Lasso 估计量的收敛速率，同时也允许重尾噪声；这类噪声仅具有多项式矩，而非次 Weibull 噪声。

结构 本文结构安排如下。第 2 节介绍数学框架、记号和估计量。随后，我们在第 3 节给出主要结果，并在第 4 节证明这些结果。接着，我们在第 5 节使用合成数据进行模拟研究。最后，本文在第 6 节给出结论和未来展望。附录 A 包含一些有用的辅助结果。

2. 预备知识与记号

在严格介绍我们的数学模型之前，我们先固定一些基本记号。对于$x \in \mathrm{R}^d$， $a_0 > 0$，我们用R^d 中以 x 为中心、半径为 a_0 的开球表示为 $B(x, a_0)$，对于 $n \in \mathrm{N}$，令 $[n] := \{1, \ldots\ldots, n\}$，对于集合 S，我们用 S^{c}表示其补集。对于对称矩阵$\mathbf{M} \in \mathrm{R}^{d\times d}$，我们用 $\lambda_{\max}(\mathbf{M})$和 $\lambda_{\min}(\mathbf{M})$分别表示其最大和最小特征值

$\mathbf{M}$，对于任意正定矩阵$\mathbf{M} \in \mathrm{R}^{d\times d}$，我们用 $\kappa(\mathbf{M}) \coloneqq \lambda_{\max}(\mathbf{M})/\lambda_{\min}(\mathbf{M})$表示$\mathbf{M}$的条件数。对于一般的$\mathbf{M} \in \mathrm{R}^{d_1\times d_2}$，我们定义

$$\|\mathbf{M}\|_p \coloneqq \left(\sum_{1\le i\le d_1, 1\le j\le d_2} |M_{ij}|^p\right)^{1/p}, \quad p \ge 1,$$

以及 $\|\mathbf{M}\|_\infty \coloneqq \max_{1\le i\le d_1,\ 1\le j\le d_2} |\mathbf{M}_{ij}|$。稍微滥用一下记号，我们用 $\|\mathbf{M}\|_0$，表示矩阵 $\mathbf{M}$ 的非零元素个数，即 $\|\mathbf{M}\|_0 \coloneqq \sum_{1\le i\le d_1,\ 1\le j\le d_2} \mathbb{1}(M_{ij} \ne 0)$。此外，我们令 $\|\mathbf{M}\|$ 表示谱范数，即 $\|\mathbf{M}\|^2 = \lambda_{\max}(\mathbf{M}^\top\mathbf{M})$。

所有实部均为正的实 $d \times d$ 矩阵构成的集合记为$M_+(\mathrm{R}^d)$，即当且仅当 $\|\mathrm{e}^{-t\mathbf{M}}\| \to 0$ as $t \to \infty$时，$\mathbf{M} \in M_+(\mathrm{R}^d)$。

给定 $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d) \in \mathrm{R}^d$，我们用 $(\beta_1^\#, \ldots, \beta_d^\#)$ 表示$|\beta_1|, \ldots, |\beta_d|$的非递增重排，并令

$$\|\boldsymbol{\beta}\|_\star \coloneqq \sum_{j=1}^{d} \sqrt{\log\left(\frac{2d}{j}\right)}\beta_j^\#.$$

由 [4] $\|\cdot\|_\star$ 可知，它在 R^d 上定义了一个范数。对于 $\mathbf{M} \in \mathrm{R}^{d_1\times d_2}$，我们设（稍微滥用一下记号） $\|\mathbf{M}\|_\star \coloneqq \|\mathrm{vec}(\mathbf{M})\|_\star$，其中 vec 表示通常的向量化算子。于是有

$$\|\mathbf{M}\|_\star = \sum_{j=1}^{d_1 d_2} \mathrm{vec}(\mathbf{M})_j^\# \sqrt{\log\left(\frac{2d_1 d_2}{j}\right)}. \tag{2.1}$$

2.1. 数学模型

全文中，我们令 $\mathbf{Z} = (Z_t)_{t\ge 0}$ 表示定义在滤过概率空间 $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathrm{P})$ 上并适应于滤过 $(\mathcal{F}_t)_{t\ge 0}$的 d 维 Lévy 过程。我们将考虑 d 维 Lévy 驱动的 Ornstein–Uhlenbeck（OU）过程，该过程是随机微分方程的强解$\mathbf{X} = (X_t)_{t\ge 0}$

$$\mathrm{d}X_t = -\boldsymbol{A}_0 X_t\,\mathrm{d}t + \mathrm{d}Z_t, \quad t > 0, \qquad X_0 \sim \pi, \tag{2.2}$$

其中 $\boldsymbol{A}_0 \in \mathrm{R}^{d\times d}$ 是漂移参数，π 是 $\mathbf{X}$ 的初始分布。过程 $\mathbf{Z}$ 称为背景驱动 Lévy过程（BDLP）。我们假设初始条件X_0 与 $\mathbf{Z}$ 相互独立。

随机微分方程（2.2）的解可表示为

$$X_t = \mathrm{e}^{-t\boldsymbol{A}_0}X_0 + \int_0^t \mathrm{e}^{-(t-s)\boldsymbol{A}_0}\,\mathrm{d}Z_s, \quad t > 0, \tag{2.3}$$

这由 Itô 公式推出（例如参见 [20] 中的公式（1.1）和（1.2））。

记（b, $\mathbf{C} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$, ν）为$\mathbf{Z}$的生成三元组，即，$b \in \mathrm{R}^d$，$\mathbf{C} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \in \mathrm{R}^{d\times d}$ 是一个对称的半正定矩阵，而 ν 是一个 Lévy 测度，即R^d上的 σ-有限测度，并满足

$$\nu(\{\mathbf{0}\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} \min\{1, \|z\|^2\}\,\nu(\mathrm{d}z) < \infty.$$

Lévy–Itô 分解（例如参见 [2] 中的定理 2.4.16）随后推出

$$Z_t = bt + \Sigma W_t + \int_0^t\int_{\|z\|\ge 1} zN(\mathrm{d}s, \mathrm{d}z) + \int_0^t\int_{\|z\|<1} z\widetilde{N}(\mathrm{d}s, \mathrm{d}z), \quad t \ge 0,$$

其中 $(W_s)_{s\ge 0}$ 是一个 d维布朗运动，N 是定义在 $[0,\infty) \times \mathrm{R}^d$ 上的泊松随机测度，其强度测度为 $\lambda \otimes \nu$，而 $\widetilde{N}$ 表示其补偿测度。

对于 BDLP $\mathbf{Z}$，或等价地，对于其 Lévy 测度 ν，我们区分以下几种情况。

- BDLP **Z** 是连续的，当且仅当 $\nu(\mathrm{R}^d) = 0$，也就是说，**Z** 是布朗运动。

- 过程**Z** 具有有界跳跃，即存在 $a_0 > 0$, 使得

$$\nu(B(0, a_0)^{\mathrm{c}}) = 0.$$

- 对于 $p \geq 2$, 过程**Z**, 及其 Lévy 测度 ν分别具有 p 阶矩，即

$$\nu_p := \int \|z\|^p \, \nu(\mathrm{d}z) < \infty.$$

- 此外，若**Z**, 及 ν, 具有二阶矩，且存在常数 $c_\alpha > 0$, 使得**Z**满足以下条件，则称其分别为参数为 $\alpha > 0$, 的次 Weibull：

$$\int \|x\|^2 \exp(c_\alpha \|x\|^\alpha) \, \nu(\mathrm{d}x) \leq 2\nu_2.$$

  由连续性可知，如果 $\|x\|^2 \exp(c\|x\|^\alpha)$对某个$c > 0$可积 ν-，则这样的常数总是存在。与 $\alpha = 2$ 通常的次高斯范数(参见 [27]中的定义 2.5.6)， $\|x\|^2$ 这一项是为确保原点附近的可积性而添加的。

2.2. Lasso 和 Slope 估计量

下文的目标是基于该过程的离散时间观测，估计在（2.2）中定义的 Lévy 驱动 OU 过程 **X** 的漂移参数 $A_0 \in \mathrm{R}^{d\times d}$ 。更具体地，对于 $n \in \mathrm{N}$, $T > 0$,，假设可以观测到**X** 的观测值 $(X_{t_i})_{i=0,\ldots,n}$，其中$0 = t_0 < t_1 < \ldots < t_n = T$. 为简单起见，假设观测时刻 $(t_i)_{i\in\{0,\ldots,n\}}$ 等间距，即满足 $t_i = iT/n, i = 0, \ldots, n$，并将观测间距记为$\Delta_n = T/n$。

参数统计中的一种经典方法是采用最大似然估计量（MLE）。给定截至时间 $T > 0$ 的<研究 id='1'>X 连续观测记录，负对数似然函数 $\mathcal{L}_T$: $\mathrm{R}^{d\times d} \to \mathrm{R}$ 在对 BDLP **Z**作出适当假设的条件下，可通过应用 Girsanov 定理（见 [23]）显式计算

$$\mathcal{L}_T(\boldsymbol{A}) = \frac{1}{T}\int_0^T (\boldsymbol{A}X_{s-})^\top \, \mathrm{d}X_s^c + \frac{1}{2T}\int_0^T \|\boldsymbol{A}X_{s-}\|^2 \, \mathrm{d}s, \quad \boldsymbol{A} \in \mathbb{R}^{d\times d},$$

其中，$(X_t^c)_{t\in[0,T]}$ 表示**X** 的连续鞅部分， X_{s-} 表示**X** 在时间$s \in [0, T]$处的左极限。基于离散观测的一种自然的似然函数形式为

$$\frac{1}{T}\sum_{i=1}^n (\boldsymbol{A}X_{t_{i-1}})^\top \Delta X_i + \frac{\Delta_n}{2T}\sum_{i=1}^n \|\boldsymbol{A}X_{t_{i-1}}\|^2 \tag{2.4}$$

其中

$$\Delta X_i = X_{t_i} - X_{t_{i-1}}.$$

与真实似然函数的一个关键区别在于，连续鞅部分被原过程的增量所替代**X**。直接计算表明，关于 $\boldsymbol{A}$ 最小化（2.4）等价于最小化对比函数

$$\frac{1}{T}\sum_{i=1}^n \|\Delta X_i - \Delta_n \boldsymbol{A}X_{t_{i-1}}\|^2, \tag{2.5}$$

即二者均由参数 $\boldsymbol{A}$ 的同一取值实现最小化。

由于我们关注的是漂移矩阵可能具有稀疏性的高维情形 A_0，因此我们打算将似然函数的惩罚版本用作估计过程的最小化目标，例如经典的 Lasso 估计量 [24]。给定 $\mathbf{X}$ 的连续观测记录以及对连续鞅部分的了解，[11] 采用了 $\mathcal{L}_T$ 的惩罚版本来定义漂移矩阵的 Lasso（带有 $\|\cdot\|_1$ 惩罚）和 Slope（带有 $\|\cdot\|_\star$ 惩罚）估计量，其中 $\mathbf{X}$。基于连续OU过程的离散观测，即 BDLP $\mathbf{Z}$ 为标准布朗运动时，[1] 将 Lasso 估计量定义为对比函数（2.5）的最小化解，并采用 $\|\cdot\|_1$ 惩罚。然而，这种方法并不适用于一般的莱维驱动情形，下面将对此进行说明。因此，我们提出以下伪似然函数

$$\mathcal{L}_n^D(A) := \frac{1}{T}\sum_{i=1}^{n}(AX_{t_{i-1}})^\top \Delta X_i \mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}} + \frac{\Delta_n}{2T}\sum_{i=1}^{n}\|AX_{t_{i-1}}\|^2\mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}, \quad (2.6)$$

其中 $B \subseteq \mathrm{R}^d$ 是半径为 $b := \mathrm{rad}(B)$ 的 $< \infty$有界集，且 $\eta > 0$ 称为截断水平。关于 B 的典型选取，将在假设($\mathcal{A}_0$)之后讨论；依赖于 BDLP η Z 的**具体取值见表**1。如上所述，最小化伪似然 $\mathcal{L}_n^D(A)$ 等价于最小化局部化截断对比函数

$$R_T(A) := \frac{1}{T}\sum_{i=1}^{n}\|\Delta X_i - \Delta_n AX_{t_{i-1}}\|^2\mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}.$$

定义伪似然 $\mathcal{L}_n^D$的首要原因是，在离散化负对数似然函数$\mathbf{X}$ 中，连续鞅部分被原过程的增量替代（2.4）。只有当 BDLP $\mathbf{Z}$ 为布朗运动时，这样做才有充分依据，因为此时$\mathbf{X}$ 的连续鞅部分就是$\mathbf{X}$ 本身。解决这一问题的经典方法是通过跳跃过滤来恢复连续鞅部分；这种方法已用于文献 [19]中的标量 Lévy 驱动的 OU 过程的最大似然估计。更具体地说，如果相应增量的范数超过阈值2.5，就从对比函数（ $\eta > 0$）中剔除相应的加项；其中 η 当观测值之间的间距0 趋于 Δ_n 时，趋于0。然而，正如引言中所解释的，这会排除许多类型的 Lévy 过程，尤其是纯跳过程。为了滤除$\mathbf{Z}$ 中的过大跳跃，我们仍然忽略过大的增量，但不再令截断水平 η 随0 一起 Δ_n 趋于零，而是让它随观测时长 T一同增大。这也会减少被忽略的观测值数量，在实际应用中这样做更为合理。

我们对比函数（2.5）所做的另一项修改是忽略$\mathbf{X}$的过大观测值。这是因为对比函数的形式源于连续鞅部分的次高斯性。具体而言，它类似于正态分布均值的经典负对数似然函数。众所周知，在重尾情形下（例如一般的莱维驱动情形），此类 L_2准则并不适用，因为它们过于强调极端离群值。

基于伪似然 $\mathcal{L}_n^D$，我们将调节参数为$\lambda_{\mathrm{L}} > 0$的 Lasso 估计量 $\hat{}\, A_{\mathrm{L}}$ 定义为

$$\hat{A}_{\mathrm{L}} \in \mathrm{argmin}_{A\in\mathbb{R}^{d\times d}}(\mathcal{L}_n^D(A) + \lambda_{\mathrm{L}}\|A\|_1), \quad (2.7)$$

类似地，我们将调节参数为 $\lambda_{\mathrm{S}} > 0$ 的 Slope 估计量 $\hat{}\, A_{\mathrm{S}}$ 定义为

$$\hat{A}_{\mathrm{S}} \in \mathrm{argmin}_{A\in\mathbb{R}^{d\times d}}(\mathcal{L}_n^D(A) + \lambda_{\mathrm{S}}\|A\|_\star), \quad (2.8)$$

其中 $\|\cdot\|_\star$ 定义见（2.1）。我们的分析要求 $\mathbf{X}$ 具有适当的稳定性，特别是要求其存在不变分布 μ，见 [22] 中的定理4.1和4.2，或 [20]

证明，只要满足 $A_0 \in M_+(\mathbb{R}^d)$ 和 $\mathbb{E}[(1 \vee \log(\|Z_1\|))] < \infty$，$\mathbf{X}$就存在不变分布 μ。若不变分布 μ 存在，则对于集合$B \subseteq \mathbb{R}^d$ ，将其截断协方差矩阵记为

$$\mathbf{C}_\infty(B) := \mathbb{E}[YY^\top \mathbb{1}_B(Y)], \quad \text{where} \quad Y \sim \mu,$$

并将其真实协方差矩阵记为 $\mathbf{C}_\infty := \mathbf{C}_\infty(\mathbb{R}^d)$。 在本文余下部分，我们对所考虑的 OU 过程作出以下假设，该假设意味着$\mathbf{X}$ 具有足够的稳定性以及 A_0 的可识别性。

$(\mathcal{A}_0)$ $A_0 \in M_+(\mathbb{R}^d)$ 中，BDLP $\mathbf{Z}$ 是一个平方可积鞅，并且 $\lambda_{\min}(\mathbf{C}_\infty(B)) > 0$。此外，初始分布与不变分布 μ 重合，即 $\pi = \mu$， 因此$\mathbf{X}$是平稳的。

假设 $(\mathcal{A}_0)$ 由 [20, 22]此前的结果表明是适定的。事实上， [20] 甚至表明，只要 $\mathbf{X}$ 满足 $A_0 \in M_+(\mathbb{R}^d)$ 且 $\mathbf{Z}$ 具有某个 p 阶矩，其中 $p > 0$，则 $(\mathcal{A}_0)$成立。平稳性假设因此仅为方便起见；在实践中，可以通过设置足够长的预烧期来规避该假设，更多详情参见 [6]中的引理 4.2。在$(\mathcal{A}_0)$ 下，矩阵也

$$\boldsymbol{\nu}_2 := \int zz^\top \nu(\mathrm{d}z) \tag{2.9}$$

这一点在我们的陈述中将很重要，并且是良定义的。我们想强调，我们假设的是$\mathbf{C}_\infty(B)$的正定性，而非$\mathbf{C}$的正定性。该假设弱于要求$\lambda_{\min}(\mathbf{C}) > 0$, 因为根据 [20, 22]的结果， $\lambda_{\min}(\mathbf{C}) > 0$, 对任意勒贝格测度大于0的 B 都蕴含 $\lambda_{\min}(\mathbf{C}_\infty(B)) > 0$ 。该条件还允许$\mathbf{Z}$ 为纯跳过程，这与 [11]的假设形成对照。上述假设充当 A_0的可识别性条件，因为$\mathbf{C}_\infty(B)$可解释为给定模型中的费希尔信息。

除调节参数和截断水平 η, 之外，我们的估计量还要求我们选择集合 $B \subset \mathbb{R}^d$。为此，我们回顾高维随机变量的所谓薄壳现象，参见例如 [27]中的第 3.1 节。该现象表明，高维随机向量的平方范数强烈集中在其协方差矩阵的迹附近。因此，为了在集合 B, 中有足够数量的观测值，其半径b 应随维数像$\sqrt{d}$. 那样增长。对此的标准选择是 $B = B(0, c\sqrt{d})$, 即以$b = c\sqrt{d}$, 为半径的开球，其中 $c > 0$是与维数无关的常数。

为介绍我们的主要结果，我们还引入以下量。对于任意随机过程 $(Y_t)_{t\in[0,T]}$， 以及 $B \subseteq \mathbb{R}^d$, $\eta > 0$， $T > 0$，我们定义经验范数

$$\|Y\|^2_{l^2_n(B,X)} := \frac{\Delta_n}{T} \sum_{i=1}^n \|Y_{t_{i-1}}\|^2 \mathbb{1}_B(X_{t_{i-1}}),$$

$$\|Y\|^2_{l^2_n(B,X,\eta)} := \frac{\Delta_n}{T} \sum_{i=1}^n \|Y_{t_{i-1}}\|^2 \mathbb{1}_B(X_{t_{i-1}}) \mathbb{1}_{\{\|\Delta X_i\| < \eta\}}.$$

Mor此外，我们将经验协方差矩阵的截断类比量记为$\mathbf{X}$ b y

$$\hat{\mathbf{C}}_n(B) := \frac{\Delta_n}{T} \sum_{i=1}^n X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}_B(X_{t_{i-1}}),$$

$$\hat{\mathbf{C}}^\eta_n(B) := \frac{\Delta_n}{T} \sum_{i=1}^n X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}_B(X_{t_{i-1}}) \mathbb{1}_{\{\|\Delta X_i\| < \eta\}}.$$

最后，对于 $i \in [n]$，我们还引入

$$\Delta \widetilde{Z}_i := \int_{t_{i-1}}^{t_i} \exp(-(t_i - u)A_0)\, \mathrm{d}Z_u. \tag{2.10}$$

3. 主要结果

本文的主要结果由以下针对 Lasso 和 Slope 估计量的 oracle 不等式给出。其证明及所有必要的辅助结果见第4节。相关结论要求观测时域长度为 $T_\star(\varepsilon_0,\eta)$，它决定了给定问题的样本复杂度。其详细定义见第 4 节（见命题 4.7）。显然，样本复杂度取决于置信水平 $\varepsilon_0 \in (0,1)$，也取决于截断参数 $\eta > 0$。由于根据 Lévy 测度的尾部 ν选择不同的 η 更有利 ν，因此我们先给出一个一般性表述，并省略对 η的依赖。第 2.1 节所介绍情形的具体表述见推论 3.3。

定理 3.1。假设 $(\mathcal{A}_0)$。设 $\varepsilon_0 \in (0,1)$ 给定，假设 $b \geq 1, \eta \geq 2b(\exp(\Delta_n\|\boldsymbol{A}_0\|) - 1)$ 并定义

$$\gamma(\Delta_n) := \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)(\lambda_{\max}(\mathbf{C}_\infty(B)) \vee 1)\exp(\Delta_n\|\boldsymbol{A}_0\|),$$

其中 $\boldsymbol{\nu}_2$ 定义见 (2.9)。那么，存在一个普适常数 $c_\star > 0$ 使得对于任意 $T \geq T_\star(\varepsilon_0,\eta)$以下结论成立。

a) 令 $\hat{}\,\boldsymbol{A}_\mathsf{L}$ 为 *Lasso* 估计量，其调节参数 λ_L 满足

$$\lambda_\mathsf{L} \geq 2c_\star\left(\frac{\log(2ed^2/s)\gamma(\Delta_n)}{T}\right)^{1/2}. \tag{3.1}$$

那么，对于任意 $\varepsilon \in (0,1)$ 以及任意 s-稀疏矩阵 $\boldsymbol{A} \in \mathrm{R}^{d\times d}$，界

$$\begin{aligned}
&\|(\hat{\boldsymbol{A}}_\mathsf{L} - \boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)} + \lambda_\mathsf{L}\|\hat{\boldsymbol{A}}_\mathsf{L} - \boldsymbol{A}\|_1 \\
&\leq \|(\boldsymbol{A} - \boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)} + 3\Delta_n^2\lambda_{\max}(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\left(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\right)^2 \\
&\quad + \frac{3}{\Delta_n}\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2) \\
&\quad + \frac{48\lambda_\mathsf{L}^2}{\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta/2)}\left(s \vee \frac{\log(2/\varepsilon)}{\log(2ed^2/s)}\right),
\end{aligned}$$

以大于 $1 - \varepsilon_0 - \varepsilon$ 的概率成立。

b) 令$\hat{}\,\boldsymbol{A}_\mathsf{S}$ 为 *Slope* 估计量，其调节参数 λ_S 满足

$$\lambda_\mathsf{S} \geq 2c_\star\left(\frac{\gamma(\Delta_n)}{T}\right)^{1/2}. \tag{3.2}$$

那么，对于任意 $\varepsilon \in (0,1)$ 以及任意 s-稀疏矩阵 $\boldsymbol{A} \in \mathrm{R}^{d\times d}$，以下界

$$\begin{aligned}
&\|(\hat{\boldsymbol{A}}_\mathsf{S} - \boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)} + \lambda_\mathsf{S}\|\hat{\boldsymbol{A}}_\mathsf{S} - \boldsymbol{A}\|_\star \\
&\leq \|(\boldsymbol{A} - \boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)} + 3\Delta_n^2\lambda_{\max}(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\left(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\right)^2 \\
&\quad + \frac{3}{\Delta_n}\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2) \\
&\quad + \frac{48\lambda_\mathsf{S}^2}{\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta/2)}\left(s\log\left(\frac{2ed^2}{s}\right) \vee \log(2/\varepsilon)\right),
\end{aligned}$$

以大于 $1 - \varepsilon_0 - \varepsilon$ 的概率成立。

在继续陈述关于 ν 尾部的不同假设下的上界之前，我们先讨论上述结果的几个方面。

备注3.2. a) 上述两个结论都是关于经验范数的尖锐oracle不等式 $\|\cdot\|_{l_n^2(B,X,\eta)}$。这与 [3], 中的定理 4.2 和 6.1 相呼应；该文研究了线性回归框架下的 Lasso 和 Slope 估计量。

b) 我们看到，误差分解为四个部分。

- 偏差项，用于衡量真实漂移矩阵 A_0 与一个 s-稀疏矩阵A的接近程度。自然地，如果 A_0 本身是 s-稀疏的，这一项就会消失。
- 离散化误差由观测间距的大小决定 Δ_n。当 $\Delta_n \lesssim \|A_0\|^{-1}/20$ 时，该误差的阶为 $\Delta_n^2\|A_0\|_2^4 \lesssim \Delta_n^2\|A_0\|_0^2$。相比之下， [1], 中的推论 4 讨论了连续 OU 过程的 Lasso 估计量，其离散化误差阶为 $\Delta_n s d^4 \log(d)$。特别是，这一结果隐含地假设 $\Delta_n \lesssim s^{-2}$（参见命题 6.9 的证明，尤其是其中对 [1]中公式 (49) 推导的引用）。造成这种差异的原因在于估计问题所采用的方法不同。我们研究的是 Lévy 驱动的 OU 过程，而 [1] 关注的是漂移函数具有一般参数结构的连续扩散过程。为了界定这一情形下的离散化误差，作者通过 Talagrand 通用链方法得到了一个高概率界，并将这一结果应用于连续 OU 过程。另一方面，我们直接利用 Lévy 驱动的 OU 过程具有显式解这一事实，通过泰勒近似给出离散化误差的确定性界（见命题4.2）。

- 截断误差直接取决于 Lévy 测度的尾部 ν 以及截断水平 η。在下文中，我们将选择 η ，使其在 $\tilde{T}$趋于无穷时也趋于无穷。由于 ΔZ_1 是长度为 Δ_n的随机积分，见（2.10），因此误差的这一部分会足够快地减小。

- 随机误差的形式为 $s\log(2ed^2/s)/T$，其中调节参数选择得当。对于具有 d^2 个参数的稀疏线性回归（见 [3]中的定理 7.1）以及可获得连续观测的 Lévy 驱动的 OU 过程（见 [11] 中的定理 2.7 ），已知这一收敛速率是极小极大最优的。与通常得到的收敛速率 $s\log(d^2)/T$ （例如见 [1, 28]）相比，差异在于对数项。尽管这一项通常影响很小，但更精细的界表明，Lasso 和 Slope 估计量也能在非稀疏估计问题中达到极小极大最优速率，即当 $s = d^2$ 时。

c) 估计量的调节参数选择只通过项 $\exp(\Delta_n\|A_0\|)$ 依赖于 Δ_n ，而该项在高频框架下可以忽略。这与 [1], 显著不同，在后者中， λ 必须选得大于 $d^2\Delta_n^{3/2}\sqrt{\log(d^2) + \log(1/\varepsilon)}$，参见 [1] 中 λ_2^{OU} 的定义。其原因在于界定离散化误差的方法不同，这一点已在 b) 中讨论。此外，Slope 估计量的最优 λ 完全不依赖于维数和稀疏性，而 Lasso 估计量的最优调节参数则依赖于先验未知的稀疏性 s。在 [3] 和 [11] 中也是如此。不过，由于遍历性假设（$\mathcal{A}_0$）意味着 $s \geq d$，总可以将对数项选为 $\log(2ed)$，从而使 Lasso 估计量达到通常得到的收敛速率 $s\log(d)/T$（见 b)）。因此， λ 的最优选择本质上只依赖于观测时域 T ，尤其完全不依赖于置信水平 $\varepsilon_0 + \varepsilon$；例如， [1, 28] 中就并非如此。与 [28], 不同，调节参数也不依赖于未知的混合系数。

为便于阐述下一个结果，我们对截断水平 η引入以下条件。

$(\mathcal{N})$ 存在一个常数 $c_\eta > 0$，使得

$$\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2) \leq c_\eta \frac{d\Delta_n}{T}.$$

定理 3.1 中的截断误差阶为 d/T，只要 η 满足条件（$\mathcal{N}$）。在这种情况下，该误差相较于阶为 $s\log(2ed^2/s)/T$ 的随机误差可忽略，因为遍历性假设（$\mathcal{A}_0$）意味着 $s \geq d$。命题 4.11 针对 Z 为连续过程、具有有界跳跃、为次 Weibull 分布或具有多项式矩的情形，给出了确保假设（$\mathcal{N}$）成立所需的 η 大小的充分条件。我们在表 1 中汇总了这些条件及其对样本复杂度 $T_\star$的影响。为避免记号过于繁杂，我们对 $\varepsilon \in (0,1)$ 引入如下记号，

$$T_{\star,0} = T_{\star,0}(\varepsilon) := d\log(d/\varepsilon)^2 \vee \frac{d}{c_{\mathrm{mix},2}} \log\left(\frac{c_{\mathrm{mix},1} d}{c_{\mathrm{mix},2}\varepsilon}\right)^2,$$

其中 $c_{\mathrm{mix}},1, c_{\mathrm{mix}},2 > 0$ 是 β-混合系数中的常数，见（4.3）。这些常数通常难以显式求得。不过，对于 Z 为连续过程的情形，我们在命题 A.4 中证明，$c_{\mathrm{mix}},1$ 的阶为 d，而 $c_{\mathrm{mix}},2$ 仅取决于 C 和 $\mathbf{C}_\infty$ 的特征值。因此，在这种情况下，$T_\star,0$ 的阶为 $d\log(d/\varepsilon)^2$。此外，我们只关注较小置信水平 ε 的情形，并在表 1 中省略不直接依赖维数的项；显式表达式可见第 4 节和命题 4.11。我们在注 3.4 中详细讨论这些结果。

	η的最小阶	$T_\star(\varepsilon,\ \eta)$的最小阶
Z 是连续的	$\sqrt{\text{对数}(T)d\Delta_n}$	$d^2\Delta_n$ 对数(T)对数$(1/\varepsilon) \vee T_\star,\ 0$
Z 有有界跳跃	$\sqrt{d}\log(T)$	d^2 对数(T)对数$(1/\varepsilon) \vee T_{\star,0}$
Z是 α 次 Weibull	$\sqrt{d}\log(T)^{1+1/\alpha}$	d^2 对数$(T)^{2+2/\alpha}$ 对数$(1/\varepsilon) \vee T_{\star,0}$
Z 具有 p 阶矩	$T^{1/p}d^{1/2-1/p}$	$d^{2-2/p}T^{2/p}\log(1/\varepsilon) \vee T_{\star,0}$

表 1：满足条件 η 所需的截断水平 $(\mathcal{N})$ 的最小阶数（基于 BDLP Z的不同假设）。此外，还给出了对样本复杂度 $T_\star$的影响。

现在我们可以陈述下一个结果，特别是它给出了所研究估计量在弗罗贝尼乌斯范数下的误差界。

推论 3.3。设所有条件均与定理 3.1中相同。此外，假设 η 满足条件$(\mathcal{N})$，且 $T \geq 2c_\eta d\Delta_n$ 并且 A_0 是 s-稀疏的。那么存在一个普适常数 $c > 0$，使得以下陈述成立。

a) 对任意 $\varepsilon \in (0,1)$，界

$$\begin{aligned}
&\|(\hat{A}_{\mathsf{L}} - A_0)X\|_2^2 + \frac{\lambda_{\mathsf{L}}}{\lambda_{\min}(\mathbf{C}_\infty(B))}\|\hat{A}_{\mathsf{L}} - A\|_1 \\
&\leq c\Delta_n^2 \kappa(\mathbf{C}_\infty(B))\|A_0\|_2^4\Big(1 + \exp(\Delta_n\|A_0\|_2)\Delta_n\|A_0\|_2\Big)^2 \\
&\quad + c\frac{\lambda_{\mathsf{L}}^2}{(\lambda_{\min}(\mathbf{C}_\infty(B)) \wedge 1)^2}\left(s \vee \frac{\log(2/\varepsilon)}{\log(2ed^2/s)}\right),
\end{aligned}$$

以大于 $1 - \varepsilon_0 - \varepsilon$ 的概率成立。

b) 对任意 $\varepsilon \in (0,1)$ 该界

$$\|(\hat{A}_{\mathsf{S}} - A_0)X\|_2^2 + \frac{\lambda_{\mathsf{S}}}{\lambda_{\min}(\mathbf{C}_\infty(B))}\|\hat{A}_{\mathsf{S}} - A_0\|_1$$

$$\le c\Delta_n^2\kappa(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\Big(1+\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2$$
$$+c\frac{\lambda_{\mathsf{S}}^2}{(\lambda_{\min}(\mathbf{C}_\infty(B))\wedge 1)^2}\left(s\log\left(\frac{2ed^2}{s}\right)\vee\log(2/\varepsilon)\right),$$

以大于 $1-\varepsilon_0-\varepsilon$ 的概率成立。

推论 3.3 的证明见第 4 节。本节最后简要讨论上述结果以及表 1 中所列的样本复杂度。

备注 3.4. a) 若调节参数 λ_{L} 和 λ_{S} 分别取值使其满足不等式（3.1）和（3.2），且等号成立，并且 $\Delta_n\|\boldsymbol{A}_0\|_2\in O(1)$，则推论 3.3的上界基本上为

$$\Delta_n^2 d^2\|\boldsymbol{A}_0\|_2^4+\frac{s\log(2ed^2/s)}{T}\le\Delta_n^2(d^2\|\boldsymbol{A}_0\|^4\wedge s^2\|\boldsymbol{A}_0\|_\infty^4)+\frac{s\log(2ed^2/s)}{T}.$$

我们看到有两种不同的情形。忽略对数项，可得：当 $s\Delta_n^2T\to\infty$ 时，离散化误差在渐近意义下占主导；而当 $s\Delta_n^2T\to 0$ 时，随机误差占主导。这表明，只要满足 $\Delta_n\in o((sT)^{-1/2})$，就能达到与连续观测相同的极小极大最优收敛速率，见 [11],。关于连续 OU 过程， [1] 中推论 4 得到的类似充分条件为 $\Delta_n\in o((d^4T)^{-1})$，该条件更为严格。

b）推论 3.3 所要求的样本复杂度由两项中的较大者给出：一项类似于定理 3.1 的假设，即$T_\star$，0；另一项则是满足条件（$\mathcal{N}$）所必需的复杂度。后者自然取决于 Lévy-测度 ν 的尾部，见表 1。当 Z 为连续过程时，只要 $\Delta_n\in O(1/d)$，样本复杂度 $T_\star$ 在相差对数因子的意义下等于$T_\star$，0。项 $T_\star$，0 在相差对数因子的意义下，与连续观测下连续 BDLP Z 所需的样本复杂度相同（见 [11]中的第 3.1 节）。因此，我们将 [11] 的结果恢复为 $\Delta_n\to 0$。然而，如果 BDLP Z 不连续，则根据置信水平的大小， $T_\star$ 可能会被条件（$\mathcal{N}$）所要求的样本复杂度所支配。在次 Weibull 情形下，我们因此需要长度为 $\gtrsim d^2$ 的观测时域，忽略对数项。相比之下，[28], 研究了带有参数为 $\alpha\in(0,2)$的次 Weibull 噪声的指数 β-混合时间序列的 Lasso 估计量，并推导出阶为 $s^{1+2/\alpha}$ 的样本复杂度（见 [28] 中的推论 9 以及其中对 γ 的定义）；对于给定问题，由于假设（$\mathcal{A}_0$），当 $\alpha\in(0,2)$ 时，该复杂度严格支配 d^2 。

4. 主要结果的证明

在本节中，我们始终假设($\mathcal{A}_0$) 成立，无需另行说明。证明本文主要结果的第一步是为惩罚估计量推导以下不等式。该不等式基于凸分析中的确定性论证，并改编自 [3]中的引理 A.2。在高维统计中，这类不等式通常称为“基本不等式”。

引理 4.1. 设 $h\colon\mathrm{R}^{d\times d}\to\mathrm{R}$ 是一个凸函数，且 $\hat{}\boldsymbol{A}$ 是最小化问题$\min_{A\in\mathbb{R}^{d\times d}}\big(\mathcal{L}_n^D(A)+h(A)\big)$的一个解。那么 $\hat{}\boldsymbol{A}$ 对所有 $\boldsymbol{A}\in\mathrm{R}^{d\times d}$均满足，

$$\|(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)X\|_{l_n^2(B,X,\eta)}^2$$
$$\le\|(\boldsymbol{A}-\boldsymbol{A}_0)X\|_{l_n^2(B,X,\eta)}^2-\|(\hat{\boldsymbol{A}}-\boldsymbol{A})X\|_{l_n^2(B,X,\eta)}^2+2\Big(h(\boldsymbol{A})-h(\hat{\boldsymbol{A}})\Big)$$

$$+\frac{2}{T}\Bigg(\sum_{i=1}^{n}\Big((\boldsymbol{A}-\hat{\boldsymbol{A}})X_{t_{i-1}}\Big)^{\top}(\exp(-\Delta_n\boldsymbol{A}_0)-\mathbb{I}_d+\Delta_n\boldsymbol{A}_0)X_{t_{i-1}}\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})$$
$$+\sum_{i=1}^{n}\Big((\boldsymbol{A}-\hat{\boldsymbol{A}})X_{t_{i-1}}\Big)^{\top}\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})\Bigg),$$

其中 ΔZ_i 定义于 (2.10)。

引理的证明 4.1。 我们首先计算 $\mathcal{L}_n^D$ 的梯度。为此，我们首先注意到，引理 A.1 意味着 $\Delta X_i = (\exp(-\Delta_n\boldsymbol{A}_0)-\mathbb{I}_d)X_{t_{i-1}}+\Delta Z_i$。因此，我们可以改写为

$$\begin{aligned}\mathcal{L}_n^D(\boldsymbol{A}) =&\frac{1}{T}\sum_{i=1}^{n}(\boldsymbol{A}X_{t_{i-1}})^{\top}(\exp(-\Delta_n\boldsymbol{A}_0)-\mathbb{I}_d+\Delta_n\boldsymbol{A}_0)X_{t_{i-1}}\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\\&+\frac{1}{T}\sum_{i=1}^{n}(\boldsymbol{A}X_{t_{i-1}})^{\top}\Delta\widetilde{Z}_i\mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\\&-\frac{\Delta_n}{T}\sum_{i=1}^{n}(\boldsymbol{A}X_{t_{i-1}})^{\top}\boldsymbol{A}_0X_{t_{i-1}}\mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}+\frac{\Delta_n}{2T}\sum_{i=1}^{n}\|\boldsymbol{A}X_{t_{i-1}}\|^2\mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\\=&\operatorname{tr}\Big(\boldsymbol{A}\varepsilon_n(\boldsymbol{A}_0,B,\eta)^{\top}-\boldsymbol{A}\hat{\mathbf{C}}_n^{\eta}(B)\boldsymbol{A}_0^{\top}+\frac{1}{2}\boldsymbol{A}\hat{\mathbf{C}}_n^{\eta}(B)\boldsymbol{A}^{\top}\Big),\end{aligned}$$

其中

$$\begin{aligned}\varepsilon_n(\boldsymbol{A}_0,B,\eta)^{\top}\coloneqq\frac{1}{T}\Bigg(&\sum_{i=1}^{n}X_{t_{i-1}}((\exp(-\Delta_n\boldsymbol{A}_0)-\mathbb{I}_d+\Delta_n\boldsymbol{A}_0)X_{t_{i-1}})^{\top}\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})\\&+\sum_{i=1}^{n}X_{t_{i-1}}\Big(\Delta\widetilde{Z}_i\Big)^{\top}\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})\Bigg).\end{aligned}$$

通过上述表达式，直接计算可得

$$\nabla\mathcal{L}_n^D(\boldsymbol{A})=(\boldsymbol{A}-\boldsymbol{A}_0)\hat{\mathbf{C}}_n^{\eta}(B)+\varepsilon_n(\boldsymbol{A}_0,B,\eta).$$

为简化记号，我们现在定义函数 $f\equiv\mathcal{L}_n^D+h$。由于 f 是凸函数，因此 $\mathbf{0}$ 属于 f 在 $\hat{}\boldsymbol{A}$处的次微分。根据 Moreau–Rockafellar定理，存在矩阵$\mathbf{B}$ 属于 h 在 $\hat{}\boldsymbol{A}$ 处的次微分，使得

$$\mathbf{0}=(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)\hat{\mathbf{C}}_n^{\eta}+\varepsilon_n(\boldsymbol{A}_0,B,\eta)+\mathbf{B}.$$

于是有 $\operatorname{tr}(\mathbf{B}^{\top}(\boldsymbol{A}-\hat{}\boldsymbol{A}))\le h(\boldsymbol{A})-h(\hat{}\boldsymbol{A})$， 因为迹是与弗罗贝尼乌斯范数 $\|\cdot\|_2$相关联的内积。因此，

$$\begin{aligned}&\|(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)X\|_{l_n^2(B,X,\eta)}^2-\|(\boldsymbol{A}-\boldsymbol{A}_0)X\|_{l_n^2(B,X,\eta)}^2+\|(\hat{\boldsymbol{A}}-\boldsymbol{A})X\|_{l_n^2(B,X,\eta)}^2\\&=\operatorname{tr}\Big(\big[(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)^{\top}(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)-(\boldsymbol{A}-\boldsymbol{A}_0)^{\top}(\boldsymbol{A}-\boldsymbol{A}_0)+(\hat{\boldsymbol{A}}-\boldsymbol{A})^{\top}(\hat{\boldsymbol{A}}-\boldsymbol{A})\big]\hat{\mathbf{C}}_n^{\eta}(B)\Big)\\&=2\operatorname{tr}\Big((\hat{\boldsymbol{A}}-\boldsymbol{A}_0)^{\top}(\hat{\boldsymbol{A}}-\boldsymbol{A})\hat{\mathbf{C}}_n^{\eta}(B)\Big)\\&=2\operatorname{tr}\Big((\hat{\boldsymbol{A}}-\boldsymbol{A})^{\top}(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)\hat{\mathbf{C}}_n^{\eta}(B)\Big)\\&=2\operatorname{tr}\Big((\boldsymbol{A}-\hat{\boldsymbol{A}})^{\top}(\varepsilon_n(\boldsymbol{A}_0,B,\eta)+\mathbf{B})\Big)\\&\le2\Big(\operatorname{tr}\Big((\boldsymbol{A}-\hat{\boldsymbol{A}})^{\top}\varepsilon_n(\boldsymbol{A}_0,B,\eta)\Big)+h(\boldsymbol{A})-h(\hat{\boldsymbol{A}})\Big),\end{aligned}$$

至此，证明完成。 ■

引理 4.1 给出的基本界表明，为了控制估计量的误差，我们必须界定以下误差项

$$\begin{aligned}\mathcal{D}(\mathbf{M}) &:= \sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^{\top}(\exp(-\Delta_n \boldsymbol{A}_0) - \mathbb{I}_d + \Delta_n \boldsymbol{A}_0)X_{t_{i-1}}\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})\\ \mathcal{S}(\mathbf{M}) &:= \sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^{\top}\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}}),\end{aligned} \tag{4.1}$$

其中我们回顾一下

$$\Delta\widetilde{Z}_i = \int_{t_{i-1}}^{t_i} \mathrm{e}^{-(t_i-u)\boldsymbol{A}_0}\,\mathrm{d}Z_u.$$

下文中，我们将 $\mathcal{D}$ 称为“离散化误差”，因为当观测间隔 Δ_n 趋于 0 时，它也趋于 0。我们将 $\mathcal{S}$ 称为“随机误差”，它可视为连续观测下所观察到的随机误差在离散时间情形下的对应物，见 [11]。具体而言，随机误差本质上是一个鞅变换，是连续观测下在 [11]的引理 2.1 中所观察到的伊藤积分 ε_T 的离散对应物。

4.1. 离散化误差

利用泰勒展开和经典不等式，我们得到以下关于离散化误差的确定性界。

命题 4.2. 对于任意 $\mathbf{M} \in \mathrm{R}^{d\times d}$, $c > 0$, 离散化误差满足以下界 *e*

$$\frac{2}{T}\mathcal{D}(\mathbf{M}) \le \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X,\eta)} + c\Delta_n^2\|\hat{\mathbf{C}}_n^{\eta}(B)\|\|\boldsymbol{A}_0\|_2^4\Big(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2.$$

证明。 在证明的余下部分，我们记

$$\mathbf{R}_n := \Delta_n\boldsymbol{A}_0 - \mathbb{I}_d + \exp(-\Delta_n\boldsymbol{A}_0).$$

结合柯西–施瓦茨不等式和杨氏不等式，可得对于任意 $c > 0$,

$$\begin{aligned}\frac{2}{T}\mathcal{D}(\mathbf{M}) &= \frac{2}{T}\sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^{\top}\mathbf{R}_n X_{t_{i-1}}\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})\\ &\le \frac{2}{T}\sum_{i=1}^{n}\|\mathbf{M}X_{t_{i-1}}\|\|\mathbf{R}_n X_{t_{i-1}}\|\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})\\ &\le \frac{\Delta_n}{Tc}\sum_{i=1}^{n}\|\mathbf{M}X_{t_{i-1}}\|^2\mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}} + \frac{c}{T\Delta_n}\sum_{i=1}^{n}\|\mathbf{R}_n X_{t_{i-1}}\|^2\mathbb{1}_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\\ &= \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X,\eta)} + \frac{c}{T\Delta_n}\sum_{i=1}^{n}\mathrm{tr}(X_{t_{i-1}}^{\top}\mathbf{R}_n^{\top}\mathbf{R}_n X_{t_{i-1}})1_B(X_{t_{i-1}})\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\\ &= \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X,\eta)} + \frac{c}{\Delta_n^2}\mathrm{tr}(\mathbf{R}_n\hat{\mathbf{C}}_n^{\eta}(B)\mathbf{R}_n^{\top})\\ &\le \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X,\eta)} + \frac{c}{\Delta_n^2}\|\hat{\mathbf{C}}_n^{\eta}(B)\|\|\mathbf{R}_n\|_2^2\end{aligned} \tag{4.2}$$

这里还用到了迹算子的循环性质以及以下众所周知的不等式：$\mathrm{tr}(\boldsymbol{A}\mathbf{C}\boldsymbol{A}^\top) \le \|\mathbf{C}\|\|\boldsymbol{A}\|_2^2$，对于任意 $\boldsymbol{A}$，$\mathbf{C} \in \mathrm{R}^{d\times d}$，其中$\mathbf{C}$为半正定矩阵。现在结合泰勒近似和三角不等式可得

$$\begin{aligned}
\|\mathbf{R}_n\|_2 &= \|\Delta_n \boldsymbol{A}_0 - \mathbb{I}_d + \exp(-\Delta_n \boldsymbol{A}_0)\|_2 \\
&= \Big\| \sum_{k=2}^{\infty} \frac{(-\Delta_n \boldsymbol{A}_0)^k}{k!} \Big\|_2 \\
&\le \sum_{k=2}^{\infty} \frac{\|\Delta_n \boldsymbol{A}_0\|_2^k}{k!} \\
&= \exp(\Delta_n \|\boldsymbol{A}_0\|_2) - 1 - \Delta_n \|\boldsymbol{A}_0\|_2 \\
&\le \frac{1}{2}(\Delta_n \|\boldsymbol{A}_0\|_2)^2 + \frac{1}{6}\exp(\Delta_n \|\boldsymbol{A}_0\|_2)(\Delta_n \|\boldsymbol{A}_0\|_2)^3 \\
&= (\Delta_n \|\boldsymbol{A}_0\|_2)^2 \Big(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n \|\boldsymbol{A}_0\|_2)\Delta_n \|\boldsymbol{A}_0\|_2 \Big).
\end{aligned}$$

将此界代入（4.2）即可完成证明。 ■

我们看到，命题4.2 中的上界取决于截断经验协方差矩阵 $\hat{}\mathbf{C}_n^\eta(B)$ 的算子范数。 为了控制这一项，我们需要 $\hat{}\mathbf{C}_n^\eta(B)$ 具有足够强的集中性。 下一节将讨论这一问题。

4.2. 经验协方差矩阵的集中性

命题 4.2 表明，只要经验协方差矩阵 $\hat{}\mathbf{C}_n^\eta(B)$ 具有足够强的集中性，就足以控制离散化误差。这种集中性通常是 Lasso 和 Slope 估计量理论分析所需的，形式为所谓的受限特征值条件（例如，参见 [7]中的第 3.1 节和 [11]）。对于连续 OU 过程 [7]，通过 Malliavin 微积分获得了经验协方差矩阵的集中性（参见 [7]中的命题 3.2）； [1] 则通过相同的技术得到了类似结果（参见 [1]中的命题 3.3）。

然而，对于不连续的 OU 过程，这种方法并不适用。在 [11]中，这个问题通过假设经验协方差矩阵具有适当的行为得到了处理（见 [11]中的假设（$\mathcal{H}$））。这一假设也通过应用马尔可夫不等式得到了验证。不幸的是，这一结果导致样本复杂度极大，且随维数呈指数增长。

因此，我们采用另一种方法，该方法利用了这样一个事实：只要 Lévy 测度 ν 具有多项式矩，Lévy 驱动的 OU 过程$\mathbf{X}$ 便是指数 β-混合的（见[20]）。由于我们假设 ν 具有二阶矩（见 $(\mathcal{A}_0)$），因此存在常数$c_{\mathrm{mix},1}$,$c_{\mathrm{mix},2} > 0$,，使得$\mathbf{X}$ 的 β-混合系数（记为 $\beta_{\mathbf{X}}$）有如下上界：

$$\beta_{\mathbf{X}}(t) \le c_{\mathrm{mix},1} \exp(-c_{\mathrm{mix},2} t), \tag{4.3}$$

对所有 $t \ge 0$均成立。常数 $c_{\mathrm{mix},1}$,$c_{\mathrm{mix},2}$ 的具体取值通常难以获得。我们在命题 A.4中给出了连续情形的上界。

如 [12], 定理 3.1 的证明中那样，利用 Berbee 耦合，混合性质使我们能够用一系列独立随机变量之和来近似经验协方差矩阵。由于我们采用了似然函数的截断版本，这些随机变量也有界。因此，我们可以对经验协方差矩阵的独立近似应用经典的伯恩斯坦不等式，从而得到以下结果。

命题 4.3. 对于任意 $u > 0, m_T \in (\Delta_n, T/6)$, 成立

$$\mathbb{P}(\|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\| > u) \le 4d \exp\Big(- \frac{Tu^2}{4b^2 m_T (6\lambda_{\max}(\mathbf{C}_\infty(B)) + u)} \Big) + 2c_{\mathrm{mix},1} \frac{T}{m_T} \exp(-c_{\mathrm{mix},2} m_T).$$

证明。我们首先假设 m_T 满足 $n_T := (T+\Delta_n)/(2m_T) \in \mathbb{N}$. 对于 $j \in [n_T]$，我们定义 e

$$\begin{aligned} X^{j,1} &:= (X_{t_i})_{t_i \in [2(j-1)m_T,(2j-1)m_T)\cap\mathbb{T}}, \quad X^{j,2} := (X_{t_i})_{t_i \in [(2j-1)m_T,2jm_T)\cap\mathbb{T}}, \\ n^{j,1} &:= |[2(j-1)m_T,(2j-1)m_T)\cap\mathbb{T}|, \quad n^{j,2} := |[(2j-1)m_T,2jm_T)\cap\mathbb{T}|, \end{aligned} \tag{4.4}$$

其中 $\mathbb{T} := \{i\Delta_n : \ i \in [n]\}$ 表示观测时刻。于是 $X^{j,1}$ 是 $X_{[2(j-1)m_T,(2j-1)m_T)}$ 的可测函数，对所有 $j \in [n_T]$，并且对 $X^{j,2}$也有类似结果。因此，我们可以仿照 [12], 中定理 3.1 的证明进行论证，从而构造随机变量 $\hat{X}^{j,k}$， $k \in [2]$， $j \in [n_T]$，使得

$$\begin{aligned} &(1) \quad X^{j,k} \overset{(d)}{=} \hat{X}^{j,k}, \quad \text{for all } j \in [n_T], \\ &(2) \quad \mathbb{P}(X^{j,k} \neq \hat{X}^{j,k}) \le \beta_{\mathbf{X}}(m_T), \quad \text{for all } j \in [n_T], \\ &(3) \quad \hat{X}^{1,k}, \dots, \hat{X}^{n_T,k} \text{ are independent.} \end{aligned} \tag{4.5}$$

对于 $k \in [2]$。

接下来，我们将经验协方差矩阵拆分为由 $X^{j,1}$和 $X^{j,2}$贡献的部分。更确切地说，对于 $j \in [n_T]$，我们定义

$$S(X^{j,1}) := \sum_{t_i \in [2(j-1)m_T,(2j-1)m_T)\cap\mathbb{T}} X_{t_i}X_{t_i}^\top \mathbb{1}_B(X_{t_i}), \quad S(X^{j,2}) := \sum_{t_i \in [(2j-1)m_T,2jm_T)\cap\mathbb{T}} X_{t_i}X_{t_i}^\top \mathbb{1}_B(X_{t_i}),$$

并类似地定义 $S(\hat{X}^{j,k})$对于 $j \in [n_T]$， $k \in [2]$。根据构造可知，$\hat{\mathbf{C}}_n(B) = \frac{1}{n}\sum_{j=1}^{n_T}(S(X^{j,1}) + S(X^{j,2}))$。此外，式（4.5）中的性质意味着

$$\begin{aligned} &\mathbb{P}(\|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\| > u) \\ &\le \mathbb{P}\left(\|\frac{1}{n}\sum_{j=1}^{n_T}(S(X^{j,1}) - n^{j,1}\mathbf{C}_\infty(B)\|) + \|\frac{1}{n}\sum_{j=1}^{n_T}(S(X^{j,2}) - n^{j,2}\mathbf{C}_\infty(B)\|) > u\right) \\ &\le \mathbb{P}\left(\|\frac{1}{n}\sum_{j=1}^{n_T}(S(X^{j,1}) - n^{j,1}\mathbf{C}_\infty(B))\| > \frac{u}{2}\right) + \mathbb{P}\left(\|\frac{1}{n}\sum_{j=1}^{n_T}(S(X^{j,2}) - n^{j,2}\mathbf{C}_\infty(B))\| > \frac{u}{2}\right) \\ &\le \mathbb{P}\left(\|\frac{1}{n}\sum_{j=1}^{n_T}(S(\hat{X}^{j,1}) - n^{j,1}\mathbf{C}_\infty(B))\| > \frac{u}{2}\right) + \mathbb{P}\left(\|\frac{1}{n}\sum_{j=1}^{n_T}(S(\hat{X}^{j,2}) - n^{j,2}\mathbf{C}_\infty(B))\| > \frac{u}{2}\right) \\ &\quad + \mathbb{P}(\exists j \in [n_T] : X^{j,1} \neq \hat{X}^{j,1}) + \mathbb{P}(\exists j \in [n_T] : X^{j,2} \neq \hat{X}^{j,2}) \\ &\le \mathbb{P}\left(\|\sum_{j=1}^{n_T}\left(S(\hat{X}^{j,1}) - n^{j,1}\mathbf{C}_\infty(B)\right)\| > \frac{nu}{2}\right) + \mathbb{P}\left(\|\sum_{j=1}^{n_T}\left(S(\hat{X}^{j,2}) - n^{j,2}\mathbf{C}_\infty(B)\right)\| > \frac{nu}{2}\right) \\ &\quad + 2\frac{T}{m_T}\beta_X(m_T), \end{aligned} \tag{4.6}$$

其中，最后一步中我们结合了事实

$n_T = \ T+\Delta_n \ \ 2m_T \ \le(\ )/(\ )/$，并应用了并集界。现 $T \ m_T$ $\mathbb{E}[S \ X^{j\,k}\ ] = n^{j\,k}{}_\infty \ B$ $j \in [v_n] \ k \in [2]$ 在，根据定义，对所有 ,., 均有 (ˆ),C( ),。此外，我们注意到 $n^{j,k} \le \lceil m_T/\Delta_n \rceil \le 2m_T/\Delta_n$,, 因为 $m_T \ge \Delta_n$,, 因此

$$\begin{aligned} \|S(\hat{X}^{j,k}) - n^{j,k}\mathbf{C}_\infty(B)\| &\le 2n^{j,k}b^2 \\ &\le 4\frac{m_T}{\Delta_n}b^2 \end{aligned}$$

并且由于 $\left(S(\hat{X}^{j,k}) - n^{j,k}\mathbf{C}_\infty(B)\right)^2$是一个半正定矩阵，Courant–Fischer 定理给出

$$\|\sum_{j=1}^{n_T}\mathbb{E}\left[\left(S(\hat{X}^{j,1}) - n^{j,1}\mathbf{C}_\infty(B)\right)^2\right]\|$$

$$
\begin{aligned}
&\le \sum_{j=1}^{n_T} \max_{u\in\mathbb{R}^d:\|u\|=1} u^\top \mathbb{E}\Big[\Big(S(\hat{X}^{j,1}) - n^{j,1}\mathbf{C}_\infty(B)\Big)^2\Big]u \\
&\le \sum_{j=1}^{n_T} \max_{u\in\mathbb{R}^d:\|u\|=1} u^\top \mathbb{E}\Big[(S(\hat{X}^{j,1}))^2\Big]u \\
&= \sum_{j=1}^{n_T} \max_{u\in\mathbb{R}^d:\|u\|=1} \sum_{t_i,t_j\in[2(j-1)m_T,(2j-1)m_T)\cap\mathbb{T}} \mathbb{E}\Big[|u^\top X_{t_i} X_{t_i}^\top X_{t_j} X_{t_j}^\top u \mathbb{1}_B(X_{t_i})\mathbb{1}_B(X_{t_j})|\Big] \\
&\le b^2 \sum_{j=1}^{n_T} \max_{u\in\mathbb{R}^d:\|u\|=1} \sum_{t_i,t_j\in[2(j-1)m_T,(2j-1)m_T)\cap\mathbb{T}} \mathbb{E}\Big[|u^\top X_{t_i}||X_{t_j}^\top u|\mathbb{1}_B(X_{t_i})\mathbb{1}_B(X_{t_j})\Big] \\
&\le b^2 n_T (n^{j,k})^2 \lambda_{\max}(\mathbf{C}_\infty(B)) \\
&\le 4b^2 n_T \frac{m_T^2}{\Delta_n^2}\lambda_{\max}(\mathbf{C}_\infty(B)),
\end{aligned}
$$

这里用到了柯西–施瓦茨不等式和Hölder不等式。类似的论证可得 $k=2$的相同界。由于 $\hat{X}^{j,k}$ 在固定 $k=1$ 时相互独立，2，因此我们可以应用矩阵伯恩斯坦不等式（ [26]中的定理 1.4），从而得到

$$
\begin{aligned}
&\mathbb{P}\Big(\|\sum_{j=1}^{n_T}\Big(S(\hat{X}^{j,1}) - n^{j,1}\mathbf{C}_\infty(B)\Big)\| > \frac{nu}{2}\Big) + \mathbb{P}\Big(\|\sum_{j=1}^{n_T}\Big(S(\hat{X}^{j,2}) - n^{j,2}\mathbf{C}_\infty(B)\Big)\| > \frac{nu}{2}\Big) \\
&\le \mathbb{P}\Big(\lambda_{\max}\Big(\sum_{j=1}^{n_T}\Big(S(\hat{X}^{j,1}) - n^{j,1}\mathbf{C}_\infty(B)\Big)\Big) > \frac{nu}{2}\Big) + \mathbb{P}\Big(\lambda_{\max}\Big(\sum_{j=1}^{n_T}\Big(n^{j,1}\mathbf{C}_\infty(B) - S(\hat{X}^{j,1})\Big)\Big) > \frac{nu}{2}\Big) \\
&\quad + \mathbb{P}\Big(\lambda_{\max}\Big(\sum_{j=1}^{n_T}\Big(S(\hat{X}^{j,2}) - n^{j,2}\mathbf{C}_\infty(B)\Big)\Big) > \frac{nu}{2}\Big) + \mathbb{P}\Big(\lambda_{\max}\Big(\sum_{j=1}^{n_T}\Big(n^{j,2}\mathbf{C}_\infty(B) - S(\hat{X}^{j,2})\Big)\Big) > \frac{nu}{2}\Big) \\
&\le 4d\exp\Big(-\frac{(nu)^2}{8b^2 n_T \frac{m_T^2}{\Delta_n^2}\lambda_{\max}(\mathbf{C}_\infty(B)) + \frac{4num_T}{3\Delta_n}b^2}\Big) \\
&= 4d\exp\Big(-\frac{(nu)^2\Delta_n^2}{8b^2 n_T m_T^2\lambda_{\max}(\mathbf{C}_\infty(B)) + \frac{4n\Delta_n u m_T}{3}b^2}\Big) \\
&= 4d\exp\Big(-\frac{T^2u^2}{4b^2 m_T(T+\Delta_n)\lambda_{\max}(\mathbf{C}_\infty(B)) + \frac{4Tum_T}{3}b^2}\Big) \\
&\le 4d\exp\Big(-\frac{3Tu^2}{4b^2 m_T(6\lambda_{\max}(\mathbf{C}_\infty(B)) + u)}\Big),
\end{aligned}
$$

其中我们再次使用了 $n_T m_T = (T+\Delta_n)/2 \le T$。结合（4.6），可得

$$
\mathbb{P}(\|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\| > u) \le 4d\exp\Big(-\frac{3Tu^2}{4b^2 m_T(6\lambda_{\max}(\mathbf{C}_\infty(B)) + u)}\Big) + 2\frac{T}{m_T}\beta_X(m_T). \tag{4.7}
$$

现在令 $\Delta_n < m_T \le T/6$ 给定，并设 $\widetilde{n}_T := \lfloor (T+\Delta_n)/(2m_T)\rfloor$，$\widetilde{m}_T := (T+\Delta_n)/(2\widetilde{n}_T)$。根据定义 $\widetilde{m}_T \ge m_T$，且

$$
\begin{aligned}
\widetilde{m}_T &= (T+\Delta_n)/(2\widetilde{n}_T) \\
&\le \frac{T+\Delta_n}{(T+\Delta_n)/(2m_T) - 1} \\
&= 2m_T\frac{T+\Delta_n}{T+\Delta_n - 2m_T}
\end{aligned}
$$

$$\leq 3m_T,$$

其中我们使用了 $m_T \leq T/6 < (T+\Delta_n)/6$。因此， （4.7）成立，其中 $\widetilde{m}_T$。结合 $\widetilde{m}_T$ 在（4.7）中的界与混合性质（4.3），即可完成证明。 ■

为推导我们的结果，我们需要下列事件发生

$$\mathfrak{C}_0 := \{\|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\| \leq \frac{\lambda_{\min}(\mathbf{C}_\infty(B))}{2}\}.$$

直接计算表明 $\lambda_{\min}(\hat{\mathbf{C}}_n(B)) \geq \lambda_{\min}(\mathbf{C}_\infty(B))/2$ 在$\mathfrak{C}_0$, 上成立，这也是我们关注该事件的原因之一。$\mathfrak{C}_0$ 的概率可简单地由命题4.3 在 T足够大时给出上界。 T 的下界依赖于 b，即局部化集合 B的半径。如第2.2节所述， b 应选取为 $\sqrt{d}$的量级。 –

推论 4.4. 假设 $b \geq 1$，并给定 $\varepsilon \in (0,\ 1)$ 。则对任意

$$T \geq T_0(\varepsilon) := \left(\frac{52b\lambda_{\max}(\mathbf{C}_\infty(B))}{3\lambda_{\min}(\mathbf{C}_\infty(B))^2}\log\left(\frac{8d}{\varepsilon}\right) \vee \frac{12b}{c_{\mathrm{mix},2}}\log\left(\frac{144b^2c_{\mathrm{mix},1}}{c_{\mathrm{mix},2}\varepsilon}\right)\right)^2 \vee 36\Delta_n^2 b^2,$$

如下界成立

$$\mathbb{P}\left(\mathfrak{C}_0\right) \geq 1-\varepsilon.$$

证明。 由于 $b \geq 1$， $T \geq 36\Delta_n^2 b^2$，我们可以应用命题4.3 并取 $m_T = \sqrt{T}/(6b)$，只要$T \gtrsim 1$。于是我们得到

$$\begin{aligned}
&\mathbb{P}\left(\|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\| > \frac{\lambda_{\min}(\mathbf{C}_\infty(B))}{2}\right)\\
&\leq 4d\exp\left(-\frac{T\lambda_{\min}(\mathbf{C}_\infty(B))^2}{8b^2m_T(12\lambda_{\max}(\mathbf{C}_\infty(B)) + \lambda_{\min}(\mathbf{C}_\infty(B)))}\right) + 2c_{\mathrm{mix},1}\frac{T}{m_T}\exp(-c_{\mathrm{mix},2}m_T)\\
&\leq 4d\exp\left(-\frac{T\lambda_{\min}(\mathbf{C}_\infty(B))^2}{104b^2m_T\lambda_{\max}(\mathbf{C}_\infty(B))}\right) + 2c_{\mathrm{mix},1}\frac{T}{m_T}\exp(-c_{\mathrm{mix},2}m_T)\\
&= 4d\exp\left(-\frac{3\sqrt{T}\lambda_{\min}(\mathbf{C}_\infty(B))^2}{52b\lambda_{\max}(\mathbf{C}_\infty(B))}\right) + 12bc_{\mathrm{mix},1}\sqrt{T}\exp\left(-c_{\mathrm{mix},2}\frac{\sqrt{T}}{6b}\right)\\
&\leq 4d\exp\left(-\frac{3\sqrt{T}\lambda_{\min}(\mathbf{C}_\infty(B))^2}{52b\lambda_{\max}(\mathbf{C}_\infty(B))}\right) + 144b^2\frac{c_{\mathrm{mix},1}}{c_{\mathrm{mix},2}}\exp\left(-c_{\mathrm{mix},2}\frac{\sqrt{T}}{12b}\right)
\end{aligned}$$

其中，我们在最后一步使用了初等不等式 $x \leq \exp(x)$，该式对一切 $x \in \mathbf{R}$ 成立。因此，若 $T \geq T_0(\varepsilon)$，则成立

$$\mathbb{P}(\mathfrak{C}_0) = \mathbb{P}\left(\|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\| \leq \frac{\lambda_{\min}(\mathbf{C}_\infty(B))}{2}\right) \geq 1-\varepsilon,$$

由此完成证明。 ■

推论 4.4 讨论了未经增量截断的经验协方差矩阵 $\hat{\mathbf{C}}_n(B)$。然而，我们的结果依赖于截断经验协方差矩阵 $\hat{\mathbf{C}}_n^\eta(B)$。下面的结果使用矩阵 Freedman 不等式（见 [25]）证明，在温和的假设下，若$\hat{C}_n^\eta(B)$ 成立，则可得到充分的集中性。$\mathfrak{C}_0$ 。

命题 4.5。 假设 $b \geq 1$ 且给定 $\varepsilon \in (0,\ 1)$ 。定义事件

$$\mathfrak{C}_\eta^+ := \left\{\lambda_{\max}(\hat{\mathbf{C}}_n^\eta(B)) \leq 2\lambda_{\max}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_+))\right\} \cap \mathfrak{C}_0,,$$

$$\mathfrak{C}_\eta^- := \left\{\lambda_{\min}(\hat{\mathbf{C}}_n^\eta(B)) \geq \frac{1}{4}\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)\right\} \cap \mathfrak{C}_0,$$
$$\mathfrak{C} := \mathfrak{C}_\eta^+ \cap \mathfrak{C}_\eta^-,$$

其中 η_+ 和 η_- 定义于引理 A.2中。则对于任意

$$T \geq T_1(\varepsilon,\eta) := T_0(\varepsilon/2) \vee 48\Delta_n b^2 \log(2d/\varepsilon)\frac{\mathbb{P}(\|\Delta\widetilde{Z}_1\| > \eta_-)\kappa(\mathbf{C}_\infty(B)) + \mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)}{\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-))}$$

成立

$$\mathbb{P}(\mathfrak{C}) \geq 1 - \varepsilon.$$

证明。 为便于表述，我们对 $a \in \mathbb{R}$引入以下记号

$$\mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|>a\}} := \mathbb{1}_{\{\|\Delta\widetilde{Z}_i\|>a\}} - \mathbb{P}(\|\Delta\widetilde{Z}_i\| > a), \quad i \in [n].$$

~由于 ΔZ_i 是独立同分布随机变量，结合引理A.2，由三角不等式可得

$$\begin{aligned}
&\lambda_{\max}(\hat{\mathbf{C}}_n^\eta(B)) \\
&\leq \lambda_{\max}\Big(\frac{1}{n}\sum_{i=1}^n X_{t_{i-1}}X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|\leq\eta_+\}}\mathbb{1}_B(X_{t_{i-1}})\Big) + \mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_+))\lambda_{\max}(\hat{\mathbf{C}}_n(B)) \\
&\leq \lambda_{\max}\Big(\frac{1}{n}\sum_{i=1}^n X_{t_{i-1}}X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|\leq\eta_+\}}\mathbb{1}_B(X_{t_{i-1}})\Big) \\
&\quad + \mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_+))(\|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\| + \lambda_{\max}(\mathbf{C}_\infty(B)))
\end{aligned}$$

类似地

$$\begin{aligned}
&\lambda_{\min}(\hat{\mathbf{C}}_n^\eta(B)) \\
&\geq \lambda_{\min}\Big(\frac{1}{n}\sum_{i=1}^n X_{t_{i-1}}X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|\leq\eta_-\}}\mathbb{1}_B(X_{t_{i-1}})\Big) + \mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)\lambda_{\min}(\hat{\mathbf{C}}_n(B)) \\
&\geq \lambda_{\min}\Big(\frac{1}{n}\sum_{i=1}^n X_{t_{i-1}}X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|\leq\eta_-\}}\mathbb{1}_B(X_{t_{i-1}})\Big) \\
&\quad + \mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)(\lambda_{\min}(\mathbf{C}_\infty(B)) - \|\hat{\mathbf{C}}_n(B) - \mathbf{C}_\infty(B)\|).
\end{aligned}$$

根据上述界，对于任意 $u > 0$，我们得到

$$\begin{aligned}
&\mathbb{P}\Big(\lambda_{\max}(\hat{\mathbf{C}}_n^\eta(B)) > u + \frac{3}{2}\lambda_{\max}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_+)), \mathfrak{C}_0\Big) \\
&\leq \mathbb{P}\left(\lambda_{\max}\Big(\sum_{i=1}^n X_{t_{i-1}}X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|\leq\eta_+\}}\mathbb{1}_B(X_{t_{i-1}})\Big) > nu, \mathfrak{C}_0\right),
\end{aligned} \tag{4.8}$$

and

$$\begin{aligned}
&\mathbb{P}\Big(\lambda_{\min}(\hat{\mathbf{C}}_n^\eta(B)) < -u + \frac{1}{2}\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)), \mathfrak{C}_0\Big) \\
&\leq \mathbb{P}\left(\lambda_{\min}\Big(\sum_{i=1}^n X_{t_{i-1}}X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|\leq\eta_-\}}\mathbb{1}_B(X_{t_{i-1}})\Big) < -nu, \mathfrak{C}_0\right) \\
&= \mathbb{P}\left(\lambda_{\max}\Big(-\sum_{i=1}^n X_{t_{i-1}}X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta\widetilde{Z}_i\|\leq\eta_-\}}\mathbb{1}_B(X_{t_{i-1}})\Big) > nu, \mathfrak{C}_0\right).
\end{aligned} \tag{4.9}$$

~ ~
对任意 $i \in [n]$，且因为 ΔZ_i 是独立同分布的，因此$\sum_{i=1}^n X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta \widetilde{Z}_i\| > a\}} \mathbb{1}_B(X_{t_{i-1}})$ 对任意 $a \in$R的取值都是一个矩阵鞅。此外，对于任意 $i \in [n]$，几乎必然有

$$\lambda_{\max}(X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta \widetilde{Z}_i\| \le \eta_+\}} \mathbb{1}_B(X_{t_{i-1}})) \le b^2(1 - \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le \eta_+)) = b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| > \eta_+),$$

类似地

$$\lambda_{\max}(-X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta \widetilde{Z}_i\| \le \eta_-\}} \mathbb{1}_B(X_{t_{i-1}})) \le b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le \eta_-).$$

此外，对于任意 $a \in \mathrm{R}$，我们得到

$$\begin{aligned}
&\| \sum_{i=1}^n \mathbb{E}[(X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta \widetilde{Z}_i\| \le a\}} \mathbb{1}_B(X_{t_{i-1}}))^2 | \mathcal{F}_{t_{i-1}}] \| \\
&\le b^2 \mathbb{E}[(\mathbb{1}^c_{\{\|\Delta \widetilde{Z}_1\| \le a\}})^2] \| \sum_{i=1}^n X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}_B(X_{t_{i-1}}) \| \\
&= n b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| > a) \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le a) \| \hat{\mathbf{C}}_n(B) \|,
\end{aligned}$$

因此，集合包含关系

$$\begin{aligned}
\mathfrak{C}_0 \subseteq \Big\{ &\| \sum_{i=1}^n \mathbb{E}[(X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta \widetilde{Z}_i\| \le a\}} \mathbb{1}_B(X_{t_{i-1}}))^2 | \mathcal{F}_{t_{i-1}}] \| \\
&\le \frac{3}{2} n b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| > a) \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le a) \lambda_{\max}(\mathbf{C}_\infty(B)) \Big\},
\end{aligned}$$

对任意 $a \in \mathrm{R}$均成立。上述结果使我们能够应用矩阵鞅的Freedman不等式（参见 [25]），由此可得，对于任意 $u > 0$

$$\begin{aligned}
&\mathbb{P}\left(\lambda_{\max}\Big(\sum_{i=1}^n X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta \widetilde{Z}_i\| \le \eta_+\}} \mathbb{1}_B(X_{t_{i-1}}) \Big) > nu, \mathfrak{C}_0 \right) \\
&\le d \exp\left(- \frac{nu^2}{b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| > \eta_+)(3\mathbb{P}(\|\Delta \widetilde{Z}_1\| \le \eta_+) \lambda_{\max}(\mathbf{C}_\infty(B)) + \frac{2}{3}u)} \right),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\left(\lambda_{\max}\Big(- \sum_{i=1}^n X_{t_{i-1}} X_{t_{i-1}}^\top \mathbb{1}^c_{\{\|\Delta \widetilde{Z}_i\| \le \eta_-\}} \mathbb{1}_B(X_{t_{i-1}}) \Big) > nu, \mathfrak{C}_0 \right) \\
&\le d \exp\left(- \frac{nu^2}{b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le \eta_-)(3\mathbb{P}(\|\Delta \widetilde{Z}_1\| > \eta_-) \lambda_{\max}(\mathbf{C}_\infty(B)) + \frac{2}{3}u)} \right).
\end{aligned}$$

因此结合上述不等式与（4.8）和（4.9），可得

$$\begin{aligned}
&\mathbb{P}\Big(\lambda_{\max}(\hat{\mathbf{C}}_n^\eta(B)) > 2\lambda_{\max}(\mathbf{C}_\infty(B)) \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le \eta_+)), \mathfrak{C}_0 \Big) \\
&\le d \exp\left(- \frac{3n \lambda_{\max}(\mathbf{C}_\infty(B)) \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le \eta_+))}{40 b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| > \eta_+)} \right) \\
&= d \exp\left(- \frac{3T \lambda_{\max}(\mathbf{C}_\infty(B)) \mathbb{P}(\|\Delta \widetilde{Z}_1\| \le \eta_+))}{40 \Delta_n b^2 \mathbb{P}(\|\Delta \widetilde{Z}_1\| > \eta_+)} \right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\Big(\lambda_{\min}(\hat{\mathbf{C}}_n^{\eta}(B)) < \frac{1}{4}\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-)), \mathfrak{C}_0\Big) \\
&\le d\exp\left(-\frac{n\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-))}{16b^2(3\mathbb{P}(\|\Delta\widetilde{Z}_1\| > \eta_-)\kappa(\mathbf{C}_\infty(B)) + \frac{1}{6}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-))}\right) \\
&= d\exp\left(-\frac{T\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-))}{16\Delta_n b^2(3\mathbb{P}(\|\Delta\widetilde{Z}_1\| > \eta_-)\kappa(\mathbf{C}_\infty(B)) + \frac{1}{6}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-))}\right).
\end{aligned}$$

现在只需注意到 $T \ge T_1(\varepsilon)$，由推论 4.4

$$\begin{aligned}
&\mathbb{P}(\mathfrak{C}) \\
&\ge \mathbb{P}(\mathfrak{C}_\eta^+) + \mathbb{P}(\mathfrak{C}_\eta^-) - \mathbb{P}(\mathfrak{C}_0) \\
&\ge \mathbb{P}(\mathfrak{C}_0) - d\exp\left(-\frac{3T\lambda_{\max}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_+))}{40\Delta_n b^2\mathbb{P}(\|\Delta\widetilde{Z}_1\| > \eta_+)}\right) \\
&\quad - d\exp\left(-\frac{T\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-))}{16\Delta_n b^2(3\mathbb{P}(\|\Delta\widetilde{Z}_1\| > \eta_-)\kappa(\mathbf{C}_\infty(B)) + \frac{1}{6}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-))}\right) \\
&\ge 1 - \varepsilon.
\end{aligned}$$

■

4.3. 随机误差

本节推导随机误差的一个一致高概率估计 or

$$\mathcal{S}(\mathbf{M}) := \sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^\top \Delta\widetilde{Z}_i \mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}}), \quad \mathbf{M} \in \mathbb{R}^{d\times d}.$$

该估计的证明依赖于鞅的集中不等式，并结合 Talagrand 通用链方法。然而，由于对增量进行了截断，随机误差 $\mathcal{S}(\cdot)$并非鞅。因此，我们首先推导以下分解，该分解将随机误差界定为一个鞅与另一项之和，我们将后一项称为截断偏差。

引理 4.6. 假设 $\eta_- > 0$。则对于任意 $c > 0$，以下界成立：

$$\begin{aligned}
&\frac{2}{T}|\mathcal{S}(\mathbf{M})| \\
&\le \frac{2}{T}|\sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^\top\Big(\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}} - \mathbb{E}[\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}|\mathcal{F}_{t_{i-1}}]\Big)\mathbb{1}_B(X_{t_{i-1}})| + \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X)} \\
&\quad + \frac{c}{\Delta_n}\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|A_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\| \ge \eta_-),
\end{aligned}$$

其中 η_- 在引理 A.2中定义。

~证明。由于 ΔZ_i 是中心化的，且对任意 $i \in [n]$都与 $\mathcal{F}_{t_{i-1}}$ 独立，因此对任意 $i \in [n]$，有

$$\begin{aligned}
&\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}}) \\
&= \Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}} - \mathbb{E}[\Delta\widetilde{Z}_i|\mathcal{F}_{t_{i-1}}] \\
&= \Big(\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}} - \mathbb{E}[\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}|\mathcal{F}_{t_{i-1}}]\Big) - \mathbb{E}[\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|\ge\eta\}}|\mathcal{F}_{t_{i-1}}],
\end{aligned}$$

从而我们可以将随机误差界定如下

$$\begin{aligned}\frac{2}{T}|\mathcal{S}(\mathbf{M})| &= |\sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^\top \Delta\widetilde{Z}_i \mathbb{1}_{\{\|\Delta X_i\|<\eta\}}\mathbb{1}_B(X_{t_{i-1}})| \\ &\le \frac{2}{T}|\sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^\top\Big(\Delta\widetilde{Z}_i \mathbb{1}_{\{\|\Delta X_i\|<\eta\}} - \mathbb{E}[\Delta\widetilde{Z}_i \mathbb{1}_{\{\|\Delta X_i\|<\eta\}}|\mathcal{F}_{t_{i-1}}]\Big)\mathbb{1}_B(X_{t_{i-1}})| \\ &\quad + \frac{2}{T}|\sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^\top \mathbb{E}[\Delta\widetilde{Z}_i \mathbb{1}_{\{\|\Delta X_i\|\ge\eta\}}|\mathcal{F}_{t_{i-1}}]\mathbb{1}_B(X_{t_{i-1}})|.\end{aligned} \tag{4.10}$$

为了证明该断言，现在只需对第二项求界。结合 Young 不等式和 Hölder 不等式、 $\mathcal{F}_{t_i}$关于 X_{t_i} 的可测性以及 Courant–Fischer 定理，对于任意 $c>0$，我们得到

$$\begin{aligned}&\frac{2}{T}|\sum_{i=1}^{n}(\mathbf{M}X_{t_{i-1}})^\top \mathbb{E}[\Delta\widetilde{Z}_i \mathbb{1}_{\{\|\Delta X_i\|\ge\eta\}}|\mathcal{F}_{t_{i-1}}]\mathbb{1}_B(X_{t_{i-1}})| \\ &\le \frac{\Delta_n}{Tc}\sum_{i=1}^{n}\|\mathbf{M}X_{t_{i-1}}\|^2 + \frac{c}{T\Delta_n}\sum_{i=1}^{n}\mathbb{E}[(\mathbf{M}X_{t_{i-1}}/\|\mathbf{M}X_{t_{i-1}}\|)^\top\Delta\widetilde{Z}_i \mathbb{1}_{\{\|\Delta X_i\|\ge\eta\}}\mathbb{1}_B(X_{t_{i-1}})|\mathcal{F}_{t_{i-1}}]^2 \\ &\le \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X)} + \frac{c}{T\Delta_n}\sum_{i=1}^{n}\mathbb{E}[((\mathbf{M}X_{t_{i-1}}/\|\mathbf{M}X_{t_{i-1}}\|)^\top\Delta\widetilde{Z}_i)^2|\mathcal{F}_{t_{i-1}}]\|\mathbb{P}(\|\Delta X_i\|\ge\eta, X_{t_{i-1}}\in B|\mathcal{F}_{t_{i-1}}) \\ &\le \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X)} + \frac{nc}{T\Delta_n}\lambda_{\max}\Big(\mathbb{E}[\Delta\widetilde{Z}_0(\Delta\widetilde{Z}_0)^\top]\Big)\mathbb{P}(\|\Delta\widetilde{Z}_1\|\ge\eta_-) \\ &\le \frac{1}{c}\|\mathbf{M}X\|^2_{l_n^2(B,X)} + \frac{c}{\Delta_n}\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|A_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\|\ge\eta_-),\end{aligned}$$

其中我们还利用了 $\mathbf{X}$ 的平稳性以及$\mathbf{Z}$ 的增量，并结合 $n/(T\Delta_n)=\Delta_n^{-2}$，马尔可夫不等式以及引理 A.2 和A.3。证明至此完成。 ■

为简化记号，对于 $\mathbf{M}\in\mathrm{R}^{d\times d}, i\in[n]$,

$$\begin{aligned}\Delta\bar{Z}_i &:= \Big(\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}} - \mathbb{E}[\Delta\widetilde{Z}_i\mathbb{1}_{\{\|\Delta X_i\|<\eta\}}|\mathcal{F}_{t_{i-1}}]\Big) \\ \xi_i(\mathbf{M}) &:= (\mathbf{M}X_{t_{i-1}})^\top\bar{Z}_i\mathbb{1}_B(X_{t_{i-1}}) \\ \widetilde{\mathcal{S}}(\mathbf{M}) &:= \sum_{i=1}^{n}\xi_i(\mathbf{M}),\end{aligned}$$

此外，我们引入 $\mathcal{S}^{(d\times d)-1}:=\{\mathrm{M}\in\mathrm{R}^{d\times d}:\ \|\mathrm{M}\|_2=1\}$。 $\widetilde{\mathcal{S}}(\mathrm{M})$ 表示引理 4.6 中所得上界的鞅$\widetilde{\ }$部分。为对 $\mathcal{S}(\mathbf{M})$ 建立一致界，我们利用 [14], 中的结果；这些结果表明，如果鞅的（可预测）二次变差过程性质良好，则其具有次高斯集中性。因此，我们在 (4.11) 中引入事件 Q_T，在该事件上，所述二次变差过程一致地有界于 T，并证明当 T 足够大时，该事件以高概率发生。证明的关键在于结合 Freedman不等式（它表明二次变差过程具有 [13], 意义下的混合尾部），随后应用 [13]中针对具有混合尾部的随机过程的上确界的集中不等式。

命题 4.7. 给定 $\varepsilon\in(0,1)$，并定义事件

$$Q_T := \Big\{\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}}\Big(\sum_{i=1}^{n}\xi_i^2(\mathbf{M}) + \sum_{i=1}^{n}\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}]\Big)\le 4\gamma(\Delta_n)T\Big\}\cap\mathfrak{C}, \tag{4.11}$$

其中

$$\gamma(\Delta_n) \coloneqq \lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)(\lambda_{\max}(\mathbf{C}_\infty(B)) \vee 1)\exp(\Delta_n\|\boldsymbol{A}_0\|).$$

那么存在一个常数 $c_{1,\star}>0$，与 d 无关，使得对所有

$$T \geq T_\star(\varepsilon,\eta) \coloneqq T_1(\varepsilon/2,\eta) \vee c_{1,\star}\gamma^{-1}(\Delta_n)b^2\eta_+^2\Big(d+\sqrt{\log(2/\varepsilon)}\Big)^2,$$

其中 η_+ 由引理 A.2 定义，可得

$$\mathbb{P}(Q_T) \geq 1-\varepsilon.$$

证明。首先，注意到引理 A.3 表明，对于任意 $\mathbf{M} \in \mathrm{R}^{d\times d}$,

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}] \\
&= \sum_{i=1}^n \mathbb{1}_B(X_{t_{i-1}})(\mathbf{M}X_{t_{i-1}})^\top \mathbb{E}[\Delta\bar{Z}_i(\Delta\bar{Z}_i)^\top|\mathcal{F}_{t_{i-1}}]\mathbf{M}X_{t_{i-1}} \\
&\leq \sum_{i=1}^n \mathbb{1}_B(X_{t_{i-1}})(\mathbf{M}X_{t_{i-1}})^\top \mathbb{E}[\Delta\widetilde{Z}_i(\Delta\widetilde{Z}_i)^\top|\mathcal{F}_{t_{i-1}}]\mathbf{M}X_{t_{i-1}} \\
&\leq \Delta_n\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\sum_{i=1}^n \mathbb{1}_B(X_{t_{i-1}})(\mathbf{M}X_{t_{i-1}})^\top\mathbf{M}X_{t_{i-1}} \\
&= T\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\operatorname{tr}(\mathbf{M}\hat{\mathbf{C}}_n(B)\mathbf{M}) \\
&\leq T\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\|\mathbf{M}\|_2^2\lambda_{\max}(\hat{\mathbf{C}}_n(B)),
\end{aligned}$$

这意味着以下集合包含关系

$$\begin{aligned}
&\Big\{\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}}\Big(\sum_{i=1}^n\xi_i^2(\mathbf{M})+\sum_{i=1}^n\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}]\Big) > 4T\gamma(\Delta_n)\Big\}\cap\mathfrak{C} \\
&\subseteq \Big\{\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}}\sum_{i=1}^n(\xi_i^2(\mathbf{M})-\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}])+2\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}}\sum_{i=1}^n\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}] > 4T\gamma(\Delta_n)\Big\}\cap\mathfrak{C} \\
&\subseteq \Big\{\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}}\sum_{i=1}^n(\xi_i^2(\mathbf{M})-\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}]) > T\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\lambda_{\max}(\mathbf{C}_\infty(B))\exp(\Delta_n\|\boldsymbol{A}_0\|)\Big\}\cap\mathfrak{C}.
\end{aligned}$$

因此，为了证明第一个断言，只需验证对于任意 $T \geq T_1(\varepsilon)$

$$\mathbb{P}\Big(\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}}\sum_{i=1}^n(\xi_i^2(\mathbf{M})-\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}]) > \gamma(\Delta_n)T,\mathfrak{C}\Big) \leq \varepsilon.$$

为此，任取$\mathbf{M}_1$,$\mathbf{M}_2 \in \mathcal{S}^{(d\times d)-1}$。则有以下恒等式

$$\xi_i^2(\mathbf{M}_1)-\xi_i^2(\mathbf{M}_2) = ((\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}})^\top\Delta\bar{Z}_i((\mathbf{M}_1+\mathbf{M}_2)X_{t_{i-1}})^\top\Delta\bar{Z}_i\mathbb{1}_B(X_{t_{i-1}}),$$

由引理A.2

$$|\xi_i^2(\mathbf{M}_1)-\mathbb{E}[\xi_i^2(\mathbf{M}_1)|\mathcal{F}_{t_{i-1}}]-\xi_i^2(\mathbf{M}_2)+\mathbb{E}[\xi_i^2(\mathbf{M}_2)|\mathcal{F}_{t_{i-1}}]| \leq 4b^2\eta_+^2\|\mathbf{M}_1-\mathbf{M}_2\|_2.$$

此外，我们得到

$$\sum_{i=1}^n\mathbb{E}[(\xi_i^2(\mathbf{M}_1)-\mathbb{E}[\xi_i^2(\mathbf{M}_1)|\mathcal{F}_{t_{i-1}}]-\xi_i^2(\mathbf{M}_2)+\mathbb{E}[\xi_i^2(\mathbf{M}_2)|\mathcal{F}_{t_{i-1}}])^2|\mathcal{F}_{t_{i-1}}]$$

$$
\begin{aligned}
&\le \sum_{i=1}^{n} \mathbb{E}[(((\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}})^\top \Delta\bar{Z}_i((\mathbf{M}_1+\mathbf{M}_2)X_{t_{i-1}})^\top \Delta\bar{Z}_i)^2 \mathbb{1}_B(X_{t_{i-1}})|\mathcal{F}_{t_{i-1}}] \\
&\le 4b^2\eta_+^2 \sum_{i=1}^{n} \mathbb{E}[(((\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}})^\top \Delta\bar{Z}_i)^2 \mathbb{1}_B(X_{t_{i-1}})|\mathcal{F}_{t_{i-1}}] \\
&= 4b^2\eta_+^2 \sum_{i=1}^{n} ((\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}})^\top \mathbb{E}[\Delta\bar{Z}_i(\Delta\bar{Z}_i)^\top|\mathcal{F}_{t_{i-1}}](\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}}\mathbb{1}_B(X_{t_{i-1}}) \\
&\le 4b^2\eta_+^2 \sum_{i=1}^{n} ((\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}})^\top \mathbb{E}[\Delta\widetilde{Z}_i(\Delta\widetilde{Z}_i)^\top|\mathcal{F}_{t_{i-1}}](\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}}\mathbb{1}_B(X_{t_{i-1}}) \\
&\le 4b^2\eta_+^2\Delta_n\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|A_0\|) \sum_{i=1}^{n} ((\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}})^\top (\mathbf{M}_1-\mathbf{M}_2)X_{t_{i-1}}\mathbb{1}_B(X_{t_{i-1}}) \\
&= 4b^2\eta_+^2 T\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|A_0\|)\,\mathrm{tr}((\mathbf{M}_1-\mathbf{M}_2)\hat{\mathbf{C}}_n(B)(\mathbf{M}_1-\mathbf{M}_2)^\top) \\
&\le 4b^2\eta_+^2 T\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|A_0\|)\|\mathbf{M}_1-\mathbf{M}_2\|_2^2\lambda_{\max}(\hat{\mathbf{C}}_n(B)).
\end{aligned}
$$

由此得到以下集合包含关系

$$
\begin{aligned}
\mathfrak{C} &= \left\{\|\hat{\mathbf{C}}_n(B)-\mathbf{C}_\infty(B)\| \le \frac{\lambda_{\min}(\mathbf{C}_\infty(B))}{2}\right\} \\
&\subseteq \Big\{\forall \mathbf{M}_1,\mathbf{M}_2 \in \mathcal{S}^{(d\times d)-1} : \sum_{i=1}^{n} \mathbb{E}[(\xi_i^2(\mathbf{M}_1)-\mathbb{E}[\xi_i^2(\mathbf{M}_1)|\mathcal{F}_{t_{i-1}}]-\xi_i^2(\mathbf{M}_2)+\mathbb{E}[\xi_i^2(\mathbf{M}_2)|\mathcal{F}_{t_{i-1}}])^2|\mathcal{F}_{t_{i-1}}] \\
&\quad \le 6b^2\eta_+^2 T\gamma(\Delta_n)\|\mathbf{M}_1-\mathbf{M}_2\|_2^2\Big\}
\end{aligned}
$$

因此，根据Freedman不等式

$$
\begin{aligned}
&\mathbb{P}\Big(|\sum_{i=1}^{n} \xi_i^2(\mathbf{M}_1)-\mathbb{E}[\xi_i^2(\mathbf{M}_1)|\mathcal{F}_{t_{i-1}}]-\xi_i^2(\mathbf{M}_2)+\mathbb{E}[\xi_i^2(\mathbf{M}_2)|\mathcal{F}_{t_{i-1}}]|\mathbb{1}_{\mathfrak{C}} > u\Big) \\
&= \mathbb{P}\Big(|\sum_{i=1}^{n} \xi_i^2(\mathbf{M}_1)-\mathbb{E}[\xi_i^2(\mathbf{M}_1)|\mathcal{F}_{t_{i-1}}]-\xi_i^2(\mathbf{M}_2)+\mathbb{E}[\xi_i^2(\mathbf{M}_2)|\mathcal{F}_{t_{i-1}}]| > u, \mathfrak{C}\Big) \\
&\le 2\exp\left(-\frac{u^2}{b^2\eta_+^2\|\mathbf{M}_1-\mathbf{M}_2\|_2(12\gamma(\Delta_n)T\|\mathbf{M}_1-\mathbf{M}_2\|_2+\frac{8}{3}u)}\right) \\
&\le 2\exp\left(-\frac{u^2}{2b^2\eta_+^2\|\mathbf{M}_1-\mathbf{M}_2\|_2(12\gamma(\Delta_n)T\|\mathbf{M}_1-\mathbf{M}_2\|_2\vee\frac{8}{3}u)}\right) \\
&= 2\exp\left(-\frac{1}{b^2\eta_+^2\|\mathbf{M}_1-\mathbf{M}_2\|_2}\Big(\frac{u^2}{24\gamma(\Delta_n)T\|\mathbf{M}_1-\mathbf{M}_2\|_2}\wedge\frac{3u}{16}\Big)\right).
\end{aligned}
$$

对于任意 $u>0$，，均有

$$
\begin{aligned}
&\mathbb{P}\Big(|\sum_{i=1}^{n} \xi_i^2(\mathbf{M}_1)-\mathbb{E}[\xi_i^2(\mathbf{M}_1)|\mathcal{F}_{t_{i-1}}]-\xi_i^2(\mathbf{M}_2)+\mathbb{E}[\xi_i^2(\mathbf{M}_2)|\mathcal{F}_{t_{i-1}}]|\mathbb{1}_{\mathfrak{C}} > \\
&\qquad \sqrt{24\gamma(\Delta_n)}b\eta_+\|\mathbf{M}_1-\mathbf{M}_2\|_2\sqrt{Tu}+\frac{16}{3}b^2\eta_+^2\|\mathbf{M}_1-\mathbf{M}_2\|_2 u\Big) \le 2\exp(-u),
\end{aligned}
$$

这意味着 $(\sum_{i=1}^{n}(\xi_i^2(\mathbf{M})-\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}])\mathbb{1}_{\mathfrak{C}})_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}}$ 按照 定义（12） 的含义具有混合尾部 [13]。因此，定理 3.5 [13] 蕴含存在某个普适常数

$c_3 > 0$，使得

$$\mathbb{P}\Bigg(\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}} \sum_{i=1}^{n} (\xi_i^2(\mathbf{M}) - \mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}]) \mathbb{1}_{\mathfrak{C}} > c_3 \Big(\gamma_2(\mathcal{S}^{(d\times d)-1}, \sqrt{24\gamma(\Delta_n)T} b\eta_+ \|\cdot\|_2) + \gamma_1(\mathcal{S}^{(d\times d)-1}, \frac{8}{3} b^2\eta_+^2 \|\cdot\|_2) + 2\sqrt{24u\gamma(\Delta_n)T} b\eta_+ + \frac{16}{3} u b^2 \eta_+^2 \Big) \Bigg) \leq \mathrm{e}^{-u}, \tag{4.12}$$

其中 γ_1 和 γ_2 表示 Talagrand γ_α 泛函（见 [13]中的第2节）。为了界定 γ_2泛函，我们采用Talagrand控制测度定理（见 [27]中的定理8.6.1）。为此，令N 为一个 d^2维标准正态随机向量。则对于某个绝对常数 $c_4 > 0$,

$$\begin{aligned} &\gamma_2(\mathcal{S}^{(d\times d)-1}, \sqrt{24\gamma(\Delta_n)T} b\eta_+ \|\cdot\|_2) \\ &\leq c_4 b\eta_+ \sqrt{\gamma(\Delta_n)T}\, \mathbb{E}\Big[\sup_{u\in\mathcal{S}^{d^2-1}} u^\top \mathsf{N} \Big] \\ &\leq c_4 b\eta_+ \sqrt{\gamma(\Delta_n)T}\, d, \end{aligned} \tag{4.13}$$

其中最后一步使用了 [27] 中的例7.5.7。为了界定 γ_1 泛函，我们利用如下事实（例如见 [13]中的公式（4））：对于某个普适常数 $c_5 > 0$

$$\gamma_1(\mathcal{S}^{(d\times d)-1}, \frac{16}{3} b^2\eta_+^2 \|\cdot\|_2) \leq c_5 \int_0^\infty \log(\mathcal{N}(\mathcal{S}^{(d\times d)-1}, \frac{16}{3} b^2\eta_+^2 \|\cdot\|_2, u)\, \mathrm{d}u,$$

其中，对于 $\varepsilon > 0$ 且 $(T,\ d)$ 为一个半度量空间，$\mathcal{N}(T,\ d,\ \varepsilon)$ 表示T的覆盖数，即覆盖 T所需的半径为 d的球的最小数量。于是，我们得到

$$\begin{aligned} &\gamma_1(\mathcal{S}^{(d\times d)-1}, \frac{16}{3} b^2\eta_+^2 \|\cdot\|_2) \\ &\leq c_5 \int_0^\infty \log(\mathcal{N}(\mathcal{S}^{(d\times d)-1}, \frac{16}{3} b^2\eta_+^2 \|\cdot\|_2, u)\, \mathrm{d}u \\ &= c_5 \int_0^\infty \log(\mathcal{N}(\mathcal{S}^{(d\times d)-1}, \|\cdot\|_2, \frac{3}{16 b^2\eta_+^2} u)\, \mathrm{d}u \\ &= c_5 \frac{16}{3} b^2\eta_+^2 \int_0^\infty \log(\mathcal{N}(\mathcal{S}^{(d\times d)-1}, \|\cdot\|_2, u)\, \mathrm{d}u \\ &\leq c_5 \frac{16}{3} b^2\eta_+^2 d^2 \int_0^2 \log\Big(1 + \frac{2}{u}\Big)\, \mathrm{d}u \\ &= c_5 \frac{64}{3} \log(2) b^2\eta_+^2 d^2, \end{aligned} \tag{4.14}$$

其中在倒数第二步中，我们使用了 [27]以及以下事实$\|\mathbf{M}_1 - \mathbf{M}$对于任意 $\mathbf{M}_1, \mathbf{M}_2 \in \mathcal{S}^{(d\times d)-1}$。因此，结合（4.12）、（4.13）和（4.14）可知，存在某个普适常数 $c_6 > 0$ 使得对任意 $\varepsilon \in (0,\ 1)$

$$\mathbb{P}\Bigg(\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}} \sum_{i=1}^{n} (\xi_i^2(\mathbf{M}) - \mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}]) > c_6 \Big(b\eta_+ \sqrt{\gamma(\Delta_n)T} \Big(d + \sqrt{\log(\varepsilon^{-1})} \Big) + b^2\eta_+^2 \Big(d^2 + \log(\varepsilon^{-1}) \Big) \Big), \mathfrak{C} \Bigg)$$

$$= \mathbb{P}\Bigg(\sup_{\mathbf{M}\in\mathcal{S}^{(d\times d)-1}} \sum_{i=1}^{n} (\xi_i^2(\mathbf{M}) - \mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}])\mathbb{1}_{\mathfrak{C}} > c_6\Big(b\eta_+\sqrt{\gamma(\Delta_n)T}\Big(d+\sqrt{\log(\varepsilon^{-1})}\Big) + b^2\eta_+^2\Big(d^2+\log(\varepsilon^{-1})\Big)\Big)\Bigg)$$

$$\leq \varepsilon.$$

注意到 $T \geq c_{1,\star}\gamma^{-1}(\Delta_n)b^2\eta_+^2(d+\sqrt{\log(\varepsilon^{-1})})^2$，其中 $c_{1,\star} = (2c_6 \vee 4c_6^2)$，这意味着

$$c_6\Big(b\eta_+\sqrt{\gamma(\Delta_n)T}\Big(d+\sqrt{\log(\varepsilon^{-1})}\Big) + b^2\eta_+^2\Big(d^2+\log(\varepsilon^{-1})\Big)\Big)$$
$$= \gamma(\Delta_n)Tc_6\Big(b\eta_+(\gamma(\Delta_n)T)^{-1/2}\Big(d+\sqrt{\log(\varepsilon^{-1})}\Big) + (\gamma(\Delta_n)T)^{-1}b^2\eta_+^2\Big(d^2+\log(\varepsilon^{-1})\Big)\Big)$$
$$\leq \gamma(\Delta_n)T,$$

通过应用命题4.5完成证明。 ■

~[14] 的结果现在表明，在事件 Q_T 上，$\mathcal{S}$ 关于弗罗贝尼乌斯范数具有次高斯增量。因此，我们可以应用 Talagrand控制测度定理，对~得到 $\mathcal{S}$ 的统一界，该界取决于其指标集的高斯宽度（参见 [27]中的练习 8.6.5）。回顾一下，集合 $\mathcal{D} \subset \mathbb{R}^{d\times d}$ 的高斯宽度定义为

$$w(\mathcal{D}) := \mathbb{E}\Big[\sup_{\mathbf{M}\in\mathcal{D}} \mathrm{vec}(\mathbf{M})^\top Z\Big], \quad \text{where} \quad Z \sim \mathcal{N}(0, \mathbb{I}_{d^2}).$$

定理 4.8。 给定 $\mathcal{D} \subset \mathbb{R}^{d\times d}$。那么存在一个普适常数 $c_{2,\star} > 0$，使得对任意 $\varepsilon \in (0,1)$，均有

$$\mathbb{P}\Big(\sup_{\mathbf{M}\in\mathcal{D}} |\widetilde{\mathcal{S}}(\mathbf{M})|\mathbb{1}_{Q_T} > c_{2,\star}\sqrt{\gamma(\delta_n)T}\,(w(\mathcal{D}) + \sqrt{\log(2/\varepsilon)}\,\mathrm{rad}(\mathcal{D}))\Big) \leq \varepsilon.$$

证明。 给定 $\mathbf{M}_1,\mathbf{M}_2 \in \mathcal{D}$，并记 $\mathbf{M} := \mathbf{M}_1 - \mathbf{M}_2$。那么， [14] 中的推论 3.4 对任意 $u > 0$给出

$$\mathbb{P}\Big(|\widetilde{\mathcal{S}}(\mathbf{M}_1)\mathbb{1}_{Q_T} - \widetilde{\mathcal{S}}(\mathbf{M}_2)\mathbb{1}_{Q_T}| > u\Big)$$
$$= \mathbb{P}\Big(|\widetilde{\mathcal{S}}(\mathbf{M})| > u, Q_T\Big)$$
$$\leq \mathbb{P}\Big(|\widetilde{\mathcal{S}}(\mathbf{M})| > u, \sum_{i=1}^{n}\xi_i^2(\mathbf{M}) + \sum_{i=1}^{n}\mathbb{E}[\xi_i^2(\mathbf{M})|\mathcal{F}_{t_{i-1}}] \leq 4\gamma(\Delta_n)T\|\mathbf{M}\|_2^2\Big)$$
$$\leq 2\exp\Big(-\frac{u^2}{8\gamma(\Delta_n)T\|\mathbf{M}\|_2^2}\Big).$$

[27] 中的命题 2.5.2 进而表明，$(\widetilde{\mathcal{S}}(\mathbf{M})\mathbb{1}_{Q_T}/(c_7\sqrt{\gamma(\Delta_n)T}))_{\mathbf{M}\in\mathcal{D}}$ 在 [27], 中的练习 8.6.4 所述意义下是次高斯的，其中 $c_7 > 0$ 是一个普适常数。于是，结论由 [27]中的练习 8.6.5 得出。■

根据前面的定理，为了得到随机误差的一致界，现在只需界定某个合适集合的高斯宽度和半径。

推论 4.9。 给定 $\varepsilon_1, \varepsilon_2 \in (0,1)$，并令

$$\mathcal{D}^*(\varepsilon_1) := \{\mathbf{M} \in \mathbb{R}^{d\times d} : (\|\mathbf{M}\|_\star \vee \log(2/\varepsilon_1)^{1/2}\|\mathbf{M}\|_2) = 1\}.$$

那么存在一个普适常数 $c_\star > 0$，使得对于任意 $T \geq T_\star(\varepsilon_2, \eta)$，

$$\mathbb{P}\Big(\sup_{\mathbf{M}\in\mathcal{D}^*(\varepsilon_1)} |\widetilde{\mathcal{S}}(\mathbf{M})| \leq c_\star\sqrt{\gamma(\delta_n)T}, Q_T\Big) \geq 1 - \varepsilon_1 - \varepsilon_2.$$

证明。 为了应用定理 4.8 我们首先证明，高斯宽度 $w(\mathcal{D}^*(\varepsilon_1))$，以及$\sqrt{\text{对数}(2/\varepsilon_1)}$半径$(\mathcal{D}^*(\varepsilon_1))$均由与 d无关的普适常数所上界。由三角不等式可知，对于任意 $\mathbf{M}_1,\mathbf{M}_2 \in \mathcal{D}^*(\varepsilon_1)$，

$$\begin{aligned}\|\mathbf{M}_1 - \mathbf{M}_2\|_2 &\le \|\mathbf{M}_1\|_2 + \|\mathbf{M}_2\|_2 \\ &\le 2\log(2/\varepsilon_1)^{-1/2},\end{aligned}$$

这意味着

$$\sqrt{\log(2/\varepsilon_1)}\,\mathrm{rad}(\mathcal{D}^*(\varepsilon_1)) \le 2.$$

对于高斯宽度，我们可以仿照 [11] 中命题 3.3 的证明进行论证，从而得到

$$w(\mathcal{D}^*(\varepsilon_1)) \le \sqrt{\frac{\pi}{\log(4)}} + 4.$$

因此，命题 4.7 与定理 4.8 表明存在普适常数 $c_\star > 0$，使得对任意 $T \ge T_\star(\varepsilon_2)$ 成立

$$\begin{aligned}&\mathbb{P}\Big(\sup_{\mathbf{M}\in\mathcal{D}^*(\varepsilon_1)} |\widetilde{\mathcal{S}}(\mathbf{M})| \le c_\star\sqrt{\gamma(\delta_n)T}, Q_T\Big) \\ &= \mathbb{P}(Q_T) - \mathbb{P}\Big(\sup_{\mathbf{M}\in\mathcal{D}^*(\varepsilon_1)} |\widetilde{\mathcal{S}}(\mathbf{M})| > c_\star\sqrt{\gamma(\delta_n)T}, Q_T\Big) \\ &\ge 1 - \varepsilon_2 - \mathbb{P}\Big(\sup_{\mathbf{M}\in\mathcal{D}^*(\varepsilon_1)} |\widetilde{\mathcal{S}}(\mathbf{M})|\mathbb{1}_{Q_T} > c_\star\sqrt{\gamma(\delta_n)T}\Big) \\ &\ge 1 - \varepsilon_1 - \varepsilon_2,\end{aligned}$$

由此完成证明。 ■

现在，只需对引理 4.6 中的截断偏差给出界，即可证明我们的主要结果。这是下一节的内容。

4.3.1. 截断偏差

为了界定截断偏差，我们必须界定 BDLP $\mathbf{Z}$ 的增量大于截断水平 η的概率。由于我们假设 BDLP $\mathbf{Z}$ 是一个鞅，因此可以应用 [14],中的定理 3.3，从而得到以下结果。

引理 4.10。 对于任意 $a > 0$、 $u > 0$、 $L \ge 2\Delta_n \exp(2\Delta_n\|A_0\|)\lambda_{\max}(\mathbf{C} + \nu_2)$，均有

$$\mathbb{P}(\|\Delta\widetilde{Z}_1\| \ge u) \le 2d\exp\left(-\left(\frac{u^2}{4dL} \wedge \frac{u}{\frac{4}{3}a\sqrt{d}}\right)\right) + 2\Delta_n L^{-1}\exp(2\Delta_n\|A_0\|)\int_{B(0,\exp(-\Delta_n\|A_0\|)a)^{\mathrm{c}}} \|z\|^2\nu(\mathrm{d}z).$$

证明。 该断言的证明依赖于 [14] 中的定理 3.3。为此，我们首先注意到，对任意$i \in [d]$，对任意 $a > 0$ 均成立

$$\begin{aligned}&\int_0^{\Delta_n}\int_{\mathbb{R}^d} (\exp(-(\Delta_n - s)A_0)z)_i^2 \mathbb{1}(|(\exp(-(\Delta_n - s)A_0)z)_i| > a)N(\mathrm{d}s,\mathrm{d}z) \\ &\quad + \int_0^{\Delta_n}\Big(\int_{\mathbb{R}^d} (\exp(-(\Delta_n - s)A_0)z)_i^2\nu(\mathrm{d}z) + (\exp(-(\Delta_n - s)A_0)\mathbf{C}\exp(-(\Delta_n - s)A_0^\top))_{ii}\,\mathrm{d}s \\ &\le \int_0^{\Delta_n}\int_{\mathbb{R}^d} \|\exp(-(\Delta_n - s)A_0)z\|^2\mathbb{1}(\|\exp(-(\Delta_n - s)A_0)z\| > a)N(\mathrm{d}s,\mathrm{d}z)\end{aligned}$$

$$
\begin{aligned}
&+ \lambda_{\max}\Bigg(\int_0^{\Delta_n} \Big(\int_{\mathbb{R}^d} (\exp(-(\Delta_n - s)\boldsymbol{A}_0)z)(\exp(-(\Delta_n - s)\boldsymbol{A}_0)z)^\top \nu(\mathrm{d}z) \\
&+ (\exp(-(\Delta_n - s)\boldsymbol{A}_0)\mathbf{C}\exp(-(\Delta_n - s)\boldsymbol{A}_0^\top))_{ii}\,\mathrm{d}s \Bigg) \\
&\le \exp(2\Delta_n\|\boldsymbol{A}_0\|) \int_0^{\Delta_n} \int_{\mathbb{R}^d} \|z\|^2 \mathbb{1}(\|z\| > \exp(-\Delta_n\|\boldsymbol{A}_0\|)a) N(\mathrm{d}s, \mathrm{d}z) \\
&\quad + \Delta_n \exp(2\Delta_n\|\boldsymbol{A}_0\|)\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2) \\
&=: Q(\Delta_n, a).
\end{aligned}
$$

定理 3.3 在 [14] 中与方程 (1.2) 在 [14] 中相结合，从而对任意 $a,\ u,\ L > 0$

$$
\begin{aligned}
&\mathbb{P}(\|\Delta\widetilde{Z}_1\| \ge u) \\
&\le \mathbb{P}(\sqrt{d}\|\Delta\widetilde{Z}_1\|_\infty \ge u, Q(\Delta_n, a) \le L) + \mathbb{P}(Q(\Delta_n, a) > L) \\
&\le \mathbb{P}(Q(\Delta_n, a) > L) + \sum_{i=1}^{d} \mathbb{P}(|(\Delta\widetilde{Z}_1)_i| \ge u/\sqrt{d}, Q(\Delta_n, a) \le L) \\
&\le \mathbb{P}(Q(\Delta_n, a) > L) + 2d\exp\Big(-\frac{u^2}{2dL}\frac{1}{1 + \frac{au}{3\sqrt{d}L}}) \Big) \\
&= \mathbb{P}(Q(\Delta_n, a) > L) + 2d\exp\Big(-\frac{u^2}{2dL + \frac{2}{3}au\sqrt{d}} \Big) \\
&\le \mathbb{P}(Q(\Delta_n, a) > L) + 2d\exp\Big(-\Big(\frac{u^2}{4dL} \wedge \frac{u}{\frac{4}{3}a\sqrt{d}} \Big) \Big).
\end{aligned}
$$

然后令

$$
\widetilde{L} := L - \Delta_n \exp(2\Delta_n\|\boldsymbol{A}_0\|)\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2), \quad L > 0,
$$

我们得到

$$
\begin{aligned}
\mathbb{P}(Q(\Delta_n, a) > L) &= \mathbb{P}\Big(\exp(2\Delta_n\|\boldsymbol{A}_0\|) \int_0^{\Delta_n} \int_{B(0,\exp(-\Delta_n\|\boldsymbol{A}_0\|)a)^c} \|z\|^2 N(\mathrm{d}s, \mathrm{d}z) > \widetilde{L} \Big) \\
&\le \Delta_n \widetilde{L}^{-1} \exp(2\Delta_n\|\boldsymbol{A}_0\|) \int_{B(0,\exp(-\Delta_n\|\boldsymbol{A}_0\|)a)^c} \|z\|^2 \nu(\mathrm{d}z) \\
&\le 2\Delta_n L^{-1} \exp(2\Delta_n\|\boldsymbol{A}_0\|) \int_{B(0,\exp(-\Delta_n\|\boldsymbol{A}_0\|)a)^c} \|z\|^2 \nu(\mathrm{d}z).
\end{aligned}
$$

至此证明完毕。 ■

前述结果使我们能够在关于 BDLP 的 Lévy 测度尾部的不同假设下给出截断偏差的界。在某些情形中，我们引入观测次数 n 至多随观测时长 T多项式增长这一技术性假设。这大大简化了证明，并且在实践中总可以通过舍弃部分观测值来满足该假设。

命题 4.11. 假设 $n \le c_\delta T^\delta$。则下列结论成立。

1. 假设**Z** 是连续的，即 $\nu(\mathrm{R}^d) = 0$。则对任意

$$
\eta \ge \eta_0^{\mathrm{cont}} := \sqrt{32\delta\log(T)d\Delta_n \exp(2\Delta_n\|\boldsymbol{A}_0\|)\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)}
$$

成立

$$
\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \ge \eta/2) \le \frac{2\mathrm{e}c_\delta d}{T}.
$$

2. 假设 $\mathbf{Z}$ 具有有界跳跃，即存在一个 $a_0 > 0$，使得 $\nu(B(0,a_0)^{\mathsf{c}}) = 0$。那么对于任意

$$\eta \geq \eta_0^{\text{bounded}} := \sqrt{d\delta\log(T)}\exp(\Delta_n\|\boldsymbol{A}_0\|)\Big(\sqrt{32\Delta_n\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)} \vee \frac{8}{3}a_0\sqrt{\delta\log(T)}\Big),$$

成立

$$\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2) \leq \frac{2\mathrm{e}c_\delta d}{T}.$$

3. 假设 ν 是参数为 $\alpha > 0$的次 *Weibull* 分布。那么对于任意

$$\eta \geq \eta_0^{\text{Wei},\alpha} := \sqrt{d}\exp(\Delta_n\|\boldsymbol{A}_0\|)\Big(\sqrt{32\delta\log(T)\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)} \vee \frac{8\delta\log(T)^{1+1/\alpha}}{3c_\alpha^{1/\alpha}}\Big),$$

成立

$$\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2) \leq \frac{2\mathrm{e}c_\delta d}{T}.$$

4. 假设 ν 具有某个 $p \geq 2$阶矩，其中 $p \geq 2$。那么对于任意

$$\eta \geq \eta_0^{\text{poly},p} := T^{1/p}d^{1/2-1/p},$$

成立

$$\begin{aligned}&\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2)\\ &\leq \frac{2^{2p-1}d}{T}\mathrm{e}^{p\Delta_n\|\boldsymbol{A}_0\|_{\text{op}}}\Big(\Delta_n^{p/2-1}\big(\mathrm{e}\lambda_{\max}(\mathbf{C})^{p/2}(p/d+1)^{p/2} + c_1(\nu_2/d)^{p/2}\big) + c_1\nu_p d^{-p/2}\Big).\end{aligned}$$

证明。 我们首先研究连续情形，即 $\nu(\mathbf{R}^d) = 0$。 然后，我们可以将 a 在引理4.10 取得任意小，从而得到

$$\begin{aligned}&\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2)\\ &\leq \frac{2d}{\Delta_n}\exp\Big(-\frac{\eta^2}{32d\Delta_n\exp(2\Delta_n\|\boldsymbol{A}_0\|)\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)}\Big)\\ &\leq \frac{2c_\delta d}{T}\exp\Big(\delta\log(T) - \frac{\eta^2}{32d\Delta_n\exp(2\Delta_n\|\boldsymbol{A}_0\|)\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)}\Big),\end{aligned}$$

因此，选择 $\eta \geq \eta_0^{\text{cont}} := \sqrt{32\delta\log(T)d\Delta_n\exp(2\Delta_n\|\boldsymbol{A}_0\|)\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)}$ 可得

$$\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2) \leq \frac{2\mathrm{e}c_\delta d}{T}.$$

现在，假设$\mathbf{Z}$ 具有有界跳跃，即存在 $a_0 > 0$, 使得

$$\nu(B(0,a_0)^{\mathsf{c}}) = 0.$$

由引理4.10可得

$$\begin{aligned}&\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \geq \eta/2)\\ &\leq \frac{2c_\delta d}{T}\exp\Big(\delta\log(T) - \Big(\frac{\eta^2}{32d\Delta_n\exp(2\Delta_n\|\boldsymbol{A}_0\|)\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)} \wedge \frac{\eta}{\frac{8}{3}\exp(\Delta_n\|\boldsymbol{A}_0\|)a_0\sqrt{d}}\Big)\Big).\end{aligned}$$

因此，选择

$$\eta \geq \eta_0^{\text{bounded}} := \sqrt{d\delta\log(T)}\exp(\Delta_n\|\boldsymbol{A}_0\|)\Big(\sqrt{32\Delta_n\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)} \vee \frac{8}{3}a_0\sqrt{\delta\log(T)}\Big),$$

即完成证明。现在假设 ν 是参数为 $\alpha > 0$ 的次 *Weibull* 随机变量，也就是说，存在一个常数 $c_\alpha > 0$，使得

$$\int \|z\|^2 \exp(c_\alpha \|z\|^\alpha) \nu(\mathrm{d}z) \le 2\nu_2.$$

由引理 4.10 可得，对任意 $u,\ a > 0,\ L > 2\Delta_n \exp(2\Delta_n \|\boldsymbol{A}_0\|) \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)$

$$\begin{aligned}
&\Delta_n^{-1} \mathbb{P}(\|\Delta \widetilde{Z}_1\| \ge \eta/2) \\
&\le 2\Delta_n^{-1} d \exp\Big(- \Big(\frac{\eta^2}{16 d L} \wedge \frac{\eta}{\frac{8}{3} a \sqrt{d}} \Big) \Big) + 2L^{-1} \exp(2\Delta_n \|\boldsymbol{A}_0\|) \int_{B(0, \exp(-\Delta_n \|\boldsymbol{A}_0\|) a)^{\mathrm{c}}} \|z\|^2 \nu(\mathrm{d}z) \\
&\le 2\Delta_n^{-1} d \exp\Big(- \Big(\frac{\eta^2}{16 d L} \wedge \frac{\eta}{\frac{8}{3} a \sqrt{d}} \Big) \Big) + 4\nu_2 L^{-1} \exp(2\Delta_n \|\boldsymbol{A}_0\|) \exp(-c_\alpha (\exp(-\Delta_n \|\boldsymbol{A}_0\|) a)^\alpha).
\end{aligned}$$

取 $a = \exp(\Delta_n \|\boldsymbol{A}_0\|)(c_\alpha^{-1} \log(T))^{1/\alpha},\ L = 2\exp(2\Delta_n \|\boldsymbol{A}_0\|) \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)$，于是可推出，若 $\Delta_n \le 1$

$$\begin{aligned}
&\Delta_n^{-1} \mathbb{P}(\|\Delta \widetilde{Z}_1\| \ge \eta/2) \\
&\le 2\Delta_n^{-1} d \exp\Big(- \Big(\frac{\eta^2}{32 \exp(2\Delta_n \|\boldsymbol{A}_0\|) \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2) d} \wedge \frac{\eta}{\frac{8}{3} \exp(\Delta_n \|\boldsymbol{A}_0\|)(c_\alpha^{-1} \log(T))^{1/\alpha} \sqrt{d}} \Big) \Big) \\
&\quad + 2 \frac{\nu_2}{\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2) T} \\
&\le 2c_\delta \frac{d}{T} \exp\Big(\delta \log(T) - \Big(\frac{\eta^2}{32 \exp(2\Delta_n \|\boldsymbol{A}_0\|) \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2) d} \wedge \frac{\eta}{\frac{8}{3} \exp(\Delta_n \|\boldsymbol{A}_0\|)(c_\alpha^{-1} \log(T))^{1/\alpha} \sqrt{d}} \Big) \Big) \\
&\quad + 2 \frac{d}{T}
\end{aligned}$$

因此，取

$$\eta \ge \eta_0^{\mathrm{Wei},\alpha} := \sqrt{d} \exp(\Delta_n \|\boldsymbol{A}_0\|) \Big(\sqrt{32\delta \log(T) \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)} \vee \frac{8\delta \log(T)^{1+1/\alpha}}{3 c_\alpha^{1/\alpha}} \Big)$$

即可完成证明。最后，我们考察多项式情形，即假设存在$p \ge 2$，使得

$$\nu_p = \int \|z\|^p \nu(\mathrm{d}z) < \infty.$$

对于这种情形，我们直接应用马尔可夫不等式，可得

$$\begin{aligned}
&\Delta_n^{-1} \mathbb{P}(\|\Delta \widetilde{Z}_1\| \ge \eta/2) \\
&\le 2^p \Delta_n^{-1} \eta^{-p} \mathbb{E}[\|\Delta \widetilde{Z}_1\|^p] \\
&\le \frac{2^{2p-1}}{\Delta_n \eta^p} \Big(\mathbb{E}\Big[\Big\| \int_0^{\Delta_n} \exp(-(\Delta_n - s)\boldsymbol{A}_0) \boldsymbol{\Sigma} \, \mathrm{d}W_s \Big\|^p \Big] + \mathbb{E}\Big[\Big\| \int_0^{\Delta_n} \int_{\mathbb{R}^d} \exp(-(\Delta_n - s)\boldsymbol{A}_0) z \widetilde{N}(\mathrm{d}s, \mathrm{d}z) \Big\|^p \Big] \Big).
\end{aligned} \tag{4.15}$$

由于 $\int_0^{\Delta_n} \mathrm{e}^{-(\Delta_n - s)\boldsymbol{A}_0} \boldsymbol{\Sigma} \, \mathrm{d}W_s$ 服从协方差矩阵为 $\int_0^{\Delta_n} \mathrm{e}^{-(\Delta_n - u)\boldsymbol{A}_0} \mathbf{C} \mathrm{e}^{-(\Delta_n - u)\boldsymbol{A}_0^\top} \mathrm{d}s$, 的中心多元正态分布，我们得到，其中 N 表示一个 d 维标准正态随机向量

$$\mathbb{E}[\| \int_0^{\Delta_n} \mathrm{e}^{-(\Delta_n - s)\boldsymbol{A}_0} \boldsymbol{\Sigma} \, \mathrm{d}W_s \|^p]$$

$$
\begin{aligned}
&\le \lambda_{\max}\Big(\int_0^{\Delta_n} \mathrm{e}^{-(\Delta_n-u)\boldsymbol{A}_0}\mathbf{C}\mathrm{e}^{-(\Delta_n-u)\boldsymbol{A}_0^\top}\,\mathrm{d}s\Big)^{p/2}\mathbb{E}[\|N\|^p]\\
&= \lambda_{\max}\Big(\int_0^{\Delta_n} \mathrm{e}^{-(\Delta_n-u)\boldsymbol{A}_0}\mathbf{C}\mathrm{e}^{-(\Delta_n-u)\boldsymbol{A}_0^\top}\,\mathrm{d}s\Big)^{p/2}2^{p/2}\frac{\Gamma\left(\frac{p+d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\\
&\le (2\Delta_n\lambda_{\max}(\mathbf{C})\exp(2\Delta_n\|\boldsymbol{A}_0\|))^{p/2}\frac{\Gamma\left(\frac{p+d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}\\
&\le \mathrm{e}(2\Delta_n\lambda_{\max}(\mathbf{C})\exp(2\Delta_n\|\boldsymbol{A}_0\|))^{p/2}\Big(\frac{p+d}{2\mathrm{e}}\Big)^{p/2}\Big(\frac{p+d}{d}\Big)^{d/2}\\
&\le \mathrm{e}(\Delta_n\lambda_{\max}(\mathbf{C})\exp(2\Delta_n\|\boldsymbol{A}_0\|))^{p/2}(p+d)^{p/2},
\end{aligned}
$$

其中最后几步用到了斯特林公式以及不等式 $1+x\le\exp(x)$, 该不等式对任意 $x\in\mathbf{R}$, 均成立。此外，考察 [2] 中定理 4.4.23 的证明可知，存在某个常数 $c_1>0$, 它仅依赖于 p 且与维数无关，

$$
\begin{aligned}
&\mathbb{E}[\|\int_0^{\Delta_n}\int_{\mathbb{R}^d}\mathrm{e}^{-(\Delta_n-u)\boldsymbol{A}_0}z\widetilde{N}(\mathrm{d}s,\mathrm{d}z)\|^p]\\
&\le c_1\Big(\int_0^{\Delta_n}\int_{\mathbb{R}^d}\|\mathrm{e}^{-(\Delta_n-u)\boldsymbol{A}_0}z\|^p\nu(\mathrm{d}z)\,\mathrm{d}s+\Big(\int_0^{\Delta_n}\int_{\mathbb{R}^d}\|\mathrm{e}^{-(\Delta_n-u)\boldsymbol{A}_0}z\|^2\nu(\mathrm{d}z)\,\mathrm{d}s\Big)^{p/2}\Big)\\
&\le c_1\Delta_n\mathrm{e}^{p\Delta_n\|\boldsymbol{A}_0\|_{\mathrm{op}}}(\nu_p+\Delta_n^{p/2-1}\nu_2^{p/2}).
\end{aligned}
$$

将这些界代入（4.15）可得

$$
\begin{aligned}
&\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\|\ge\eta/2)\\
&\le\frac{2^{2p-1}}{\eta^p}\mathrm{e}^{p\Delta_n\|\boldsymbol{A}_0\|_{\mathrm{op}}}\Big(\Delta_n^{p/2-1}\mathrm{e}\lambda_{\max}(\mathbf{C})^{p/2}(p+d)^{p/2}+c_1(\nu_p+\Delta_n^{p/2-1}\nu_2^{p/2})\Big),
\end{aligned}
$$

因此，对任意

$$
\eta\ge\eta_0^{\mathrm{poly},p}:=T^{1/p}d^{1/2-1/p},
$$

均有

$$
\begin{aligned}
&\Delta_n^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\|\ge\eta/2)\\
&\le\frac{2^{2p-1}d}{T}\mathrm{e}^{p\Delta_n\|\boldsymbol{A}_0\|_{\mathrm{op}}}\Big(\Delta_n^{p/2-1}\big(\mathrm{e}\lambda_{\max}(\mathbf{C})^{p/2}(p/d+1)^{p/2}+c_1(\nu_2/d)^{p/2}\big)+c_1\nu_p d^{-p/2}\Big).
\end{aligned}
$$

■

4.4. 定理 3.1 的证明

现在我们可以给出主要结果的证明。

定理 3.1 的证明。 在整个证明过程中，我们假设事件

$$
\Lambda:=\Big\{\sup_{\mathbf{M}\in\mathcal{D}^*(\varepsilon)}|\widetilde{\mathcal{S}}(\mathbf{M})|\le c_\star\sqrt{\gamma(\delta_n)T}\Big\}\cap Q_T \tag{4.16}
$$

发生，这根据推论 4.9 以大于 $1-\varepsilon_0-\varepsilon$ 的概率成立。首先，由引理 4.1可得

$$
\begin{aligned}
&\|(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)X\|^2_{l_n^2(B,X,\eta)}\\
&\le\|(\boldsymbol{A}-\boldsymbol{A}_0)X\|^2_{l_n^2(B,X,\eta)}-\|(\hat{\boldsymbol{A}}-\boldsymbol{A})X\|^2_{l_n^2(B,X,\eta)}+2\Big(h(\boldsymbol{A})-h(\hat{\boldsymbol{A}})\Big)
\end{aligned}
$$

$$+\frac{2}{T}\Big(\mathcal{D}(\boldsymbol{A}-\hat{\boldsymbol{A}})+\mathcal{S}(\boldsymbol{A}-\hat{\boldsymbol{A}})\Big),$$

其中 $\mathcal{D}$ 和 $\mathcal{S}$ 定义于（4.1）中。现在，将命题 4.2 和引理 4.6 应用于 $c=3$可得

$$\begin{aligned}
&\|(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)}\\
&\le \|(\boldsymbol{A}-\boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)}-\frac{1}{3}\|(\hat{\boldsymbol{A}}-\boldsymbol{A})X\|^2_{l^2_n(B,X,\eta)}+2\Big(h(\boldsymbol{A})-h(\hat{\boldsymbol{A}})\Big)\\
&\quad+3\Delta_n^2\|\hat{\mathbf{C}}^\eta_n(B)\|\|\boldsymbol{A}_0\|_2^4\Big(\frac{1}{2}+\frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2\\
&\quad+\frac{3}{\Delta_n}\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\|\ge\eta_-)\\
&\quad+\frac{2}{T}|\widetilde{\mathcal{S}}(\boldsymbol{A}-\hat{\boldsymbol{A}})|.
\end{aligned}$$

由于事件 Λ 发生，我们可以按如下方式对 $\widetilde{\mathcal{S}}(\boldsymbol{A}-\hat{\boldsymbol{A}})$ 进行界定

$$\begin{aligned}
&\|(\hat{\boldsymbol{A}}-\boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)}+h(\boldsymbol{A}-\hat{\boldsymbol{A}})\\
&\le \|(\boldsymbol{A}-\boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)}-\frac{1}{3}\|(\hat{\boldsymbol{A}}-\boldsymbol{A})X\|^2_{l^2_n(B,X,\eta)}\\
&\quad+3\Delta_n^2\lambda_{\max}(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\Big(\frac{1}{2}+\frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2\\
&\quad+\frac{3}{\Delta_n}\lambda_{\max}(\mathbf{C}+\boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\|\ge\eta_-)\\
&\quad+2\left(c_\star\sqrt{\frac{\gamma(\Delta_n)}{T}}(\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_\star\vee\log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2)+\frac{1}{2}h(\boldsymbol{A}-\hat{\boldsymbol{A}})+h(\boldsymbol{A})-h(\hat{\boldsymbol{A}})\right),
\end{aligned}\tag{4.17}$$

其中，我们还在不等式两边加上了 $h(\boldsymbol{A}-\hat{\boldsymbol{A}})$。为了证明这些结论，现在只需对最后一项求界

$$\Upsilon:=2\left(c_\star\sqrt{\frac{\gamma(\Delta_n)}{T}}(\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_\star\vee\log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2)+\frac{1}{2}h(\boldsymbol{A}-\hat{\boldsymbol{A}})+h(\boldsymbol{A})-h(\hat{\boldsymbol{A}})\right).$$

为此，我们先考虑 Lasso 情形，此时 $h\equiv\|\cdot\|_1$。根据柯西–施瓦茨不等式和 [3] 中的公式 (2.7)，对于任意$\mathbf{M}\in\mathrm{R}^{d\times d}$，我们得到

$$\begin{aligned}
\|\mathbf{M}\|_\star&\le\sqrt{\sum_{i=1}^{s}\log\Big(\frac{2d^2}{i}\Big)}\|\mathbf{M}\|_2+\sum_{i=s+1}^{d^2}\sqrt{\log\Big(\frac{2d^2}{i}\Big)}\,\mathrm{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})^\#_i\\
&\le\sqrt{s\log\Big(\frac{2ed^2}{s}\Big)}\|\mathbf{M}\|_2+\sum_{i=s+1}^{d^2}\sqrt{\log\Big(\frac{2d^2}{i}\Big)}\,\mathrm{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})^\#_i,
\end{aligned}$$

并且由于事件 Λ 发生，因此该结论也成立

$$\begin{aligned}
\|\mathbf{M}\|_2^2&\le\lambda_{\min}(\hat{\mathbf{C}}^\eta_n)^{-1}\,\mathrm{tr}(\mathbf{M}\hat{\mathbf{C}}^\eta_n\mathbf{M}^\top)\\
&=\lambda_{\min}(\hat{\mathbf{C}}^\eta_n)^{-1}\|\mathbf{M}\|^2_{l^2_n(B,X,\eta)}\\
&\le 4\lambda_{\min}(\mathbf{C}_\infty(B))^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-)^{-1}\|\mathbf{M}\|^2_{l^2_n(B,X,\eta)}.
\end{aligned}\tag{4.18}$$

结合关于调节参数的假设 $\lambda_L(T) \geq 2c_\star\sqrt{\log(2ed^2/s)\gamma(\Delta_n)/T}$以及 [3] 中的引理 A.1 可得

$$
\begin{aligned}
&\Upsilon \\
&\leq \lambda_L\left(\log(2ed^2/s)^{-1/2}(\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_\star \vee \log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2) + 3\sqrt{s}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2 - \sum_{i=s+1}^{d^2}\mathrm{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#}\right) \\
&\leq \lambda_L\left(\left(\sqrt{s}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2 + \sum_{i=s+1}^{d^2}\mathrm{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#} \vee \frac{2\log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\log(2ed^2/s)\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-))^{1/2}}\right)\right. \\
&\quad \left. + 3\sqrt{s}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2 - \sum_{i=s+1}^{d^2}\mathrm{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#}\right),
\end{aligned}
\tag{4.19}
$$

接下来，我们根据最大项继续讨论两种不同的情形。首先，若

$$
\sqrt{s}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2 + \sum_{i=s+1}^{d^2}\mathrm{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#} \geq 2\log(2/\varepsilon)^{1/2}\frac{\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\log(2ed^2/s)\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-))^{1/2}}
$$

则有

$$
\begin{aligned}
\Upsilon &\leq 4\sqrt{s}\lambda_L\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2 \\
&\leq 8\lambda_{\min}(\mathbf{C}_\infty(B))^{-1/2}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)^{-1/2}\sqrt{s}\lambda_L\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)} \\
&\leq 48\lambda_{\min}(\mathbf{C}_\infty(B))^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)^{-1}s\lambda_L^2 + \frac{1}{3}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}^2,
\end{aligned}
$$

最后一步用到了杨氏不等式。在相反的情形下，可得

$$
\sqrt{s}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2 \leq 2\log(2/\varepsilon)^{1/2}\frac{\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\log(2ed^2/s)\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-))^{1/2}},
$$

这意味着

$$
\begin{aligned}
\Upsilon &\leq \lambda_L\left(2\log(2/\varepsilon)^{1/2}\frac{\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\log(2ed^2/s)\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-))^{1/2}} + 3\sqrt{s}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2\right) \\
&\leq 8\lambda_L\log(2/\varepsilon)^{1/2}\frac{\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\log(2ed^2/s)\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-))^{1/2}} \\
&\leq 48\frac{\lambda_L^2\log(2/\varepsilon)}{\log(2ed^2/s)\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \leq \eta_-)} + \frac{1}{3}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}^2,
\end{aligned}
$$

其中我们再次应用了杨氏不等式。将 Υ的界代入（4.17），并注意到关于 η 的假设意味着 $\eta_- \geq \eta/2$，即完成了 Lasso 估计量的证明。接下来，我们为 Slope 估计量给出 Υ 的界。根据假设 $\lambda_S(T) \geq 2c_\star\sqrt{\gamma(\Delta_n)T}$，我们得到

$$
\Upsilon \leq \lambda_S\left((\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_\star \vee \log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2) + \|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_\star + 2\|\boldsymbol{A}\|_\star - 2\|\hat{\boldsymbol{A}}\|_\star\right),
$$

并且根据 [3] 中的引理 A.1，有

$$\begin{aligned}
&\frac{1}{2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_\star+\|\boldsymbol{A}\|_\star-\|\hat{\boldsymbol{A}}\|_\star \\
&\le \frac{3}{2}\sqrt{\sum_{i=1}^{s}\log\left(\frac{2d^2}{i}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2-\frac{1}{2}\sum_{i=s+1}^{d^2}\sqrt{\log\left(\frac{2d^2}{i}\right)}\operatorname{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#} \\
&\le \frac{3}{2}\sqrt{s\log\left(\frac{2ed^2}{s}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2-\frac{1}{2}\sum_{i=s+1}^{d^2}\sqrt{\log\left(\frac{2d^2}{i}\right)}\operatorname{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#}
\end{aligned}$$

其中我们在 [3] 最后一步使用了方程 (2.7)。仿照 (4.19) 的推导，我们由此得到

$$\begin{aligned}
\Upsilon \le \lambda_S\Bigg(&\left(\sqrt{s\log\left(\frac{2ed^2}{s}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2+\sum_{i=s+1}^{d^2}\operatorname{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#}\vee\frac{2\log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-))^{1/2}}\right) \\
&+3\sqrt{s\log\left(\frac{2ed^2}{s}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2-\sum_{i=s+1}^{d^2}\sqrt{\log\left(\frac{2d^2}{i}\right)}\operatorname{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#}\Bigg),
\end{aligned}$$

并且我们 再次考察由最大值算子产生的两种情形。如果

$$\sqrt{s\log\left(\frac{2ed^2}{s}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2+\sum_{i=s+1}^{d^2}\operatorname{vec}(\boldsymbol{A}-\hat{\boldsymbol{A}})_i^{\#}\ge\frac{2\log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-))^{1/2}},$$

仿照 Lasso 情形的推导，我们得到

$$\begin{aligned}
\Upsilon &\le 4\lambda_S\sqrt{s\log\left(\frac{2ed^2}{s}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2 \\
&\le 8\lambda_S\lambda_{\min}(\mathbf{C}_\infty(B))^{-1/2}\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-)^{-1/2}\sqrt{s\log\left(\frac{2ed^2}{s}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)} \\
&\le 48\lambda_S^2\lambda_{\min}(\mathbf{C}_\infty(B))^{-1}\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-)^{-1}s\log\left(\frac{2ed^2}{s}\right)+\frac{1}{3}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}^2.
\end{aligned}$$

与之前类似，另一种情形意味着

$$\sqrt{s\log\left(\frac{2ed^2}{s}\right)}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_2\le\frac{2\lambda_S\log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-))^{1/2}},$$

因此

$$\begin{aligned}
\Upsilon &\le \frac{8\lambda_S\log(2/\varepsilon)^{1/2}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}}{(\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-))^{1/2}} \\
&\le 48\frac{\lambda_S^2\log(2/\varepsilon)}{\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\|\le\eta_-)}+\frac{1}{3}\|\boldsymbol{A}-\hat{\boldsymbol{A}}\|_{l_n^2(B,X,\eta)}^2.
\end{aligned}$$

代入 将上述界代入（4.17），并结合关于 η 的假设，即完成证明。 ■

推论 3.3 几乎可直接由定理 3.1得到。

推论 3.3的证明。 我们仅给出 Slope 估计量的证明，Lasso 估计量的证明完全类似。由于定理 3.1的假设得到满足，因此对于任意 $\varepsilon > 0$ 在事件 Λ （4.16）中定义的界，我们有

$$
\begin{aligned}
&\|(\hat{\boldsymbol{A}}_{\mathsf{S}} - \boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)} + \lambda_{\mathsf{S}}\|\hat{\boldsymbol{A}}_{\mathsf{S}} - \boldsymbol{A}\|_\star \\
&\le \|(\boldsymbol{A} - \boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)} + 3\Delta_n^2\lambda_{\max}(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\Big(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2 \\
&\quad + \frac{3}{\Delta_n}\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\| \ge \eta/2) \\
&\quad + \frac{48\lambda_{\mathsf{S}}^2}{\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta/2)}\Bigg(s\log\Big(\frac{2\mathrm{e}d^2}{s}\Big) \vee \log(2/\varepsilon)\Bigg),
\end{aligned}
$$

以大于 $1 - \varepsilon_0 - \varepsilon$ 的概率对任意 s-稀疏矩阵 $\boldsymbol{A}$成立。现在，由于 $\boldsymbol{A}_0$ 假设为 s-稀疏的，我们可以选择 $\boldsymbol{A} = \boldsymbol{A}_0$，结合以下事实可得 η 满足条件 $(\mathcal{N})$

$$
\begin{aligned}
&\|(\hat{\boldsymbol{A}}_{\mathsf{S}} - \boldsymbol{A}_0)X\|^2_{l^2_n(B,X,\eta)} + \lambda_{\mathsf{S}}\|\hat{\boldsymbol{A}}_{\mathsf{S}} - \boldsymbol{A}_0\|_\star \\
&\le 3\Delta_n^2\lambda_{\max}(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\Big(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2 \\
&\quad + 3c_\eta\lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2)\exp(\Delta_n\|\boldsymbol{A}_0\|)\mathbb{P}(\|\Delta\widetilde{Z}_1\| \ge \eta/2) \\
&\quad + \frac{48\lambda_{\mathsf{S}}^2}{\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta/2)}\Bigg(s\log\Big(\frac{2\mathrm{e}d^2}{s}\Big) \vee \log(2/\varepsilon)\Bigg) \\
&\le 3\Delta_n^2\lambda_{\max}(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\Big(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2 \\
&\quad + \lambda_{\mathsf{S}}^2\Big(\frac{3c_\eta}{4c_\star^2} + \frac{48}{\lambda_{\min}(\mathbf{C}_\infty(B))\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta/2)}\Big)\Bigg(s\log\Big(\frac{2\mathrm{e}d^2}{s}\Big) \vee \log(2/\varepsilon)\Bigg),
\end{aligned}
$$

这里我们用到了关于 λ_{S}的假设，并且利用了假设$(\mathcal{A}_0)$ 在最后一步推出 $s \ge d$这一事实。现在，由于 (4.18) 在 Λ 和 $\eta_- \ge \eta/2$ 上成立，我们得到

$$
\begin{aligned}
&\|(\hat{\boldsymbol{A}}_{\mathsf{S}} - \boldsymbol{A}_0)X\|_2^2 + 4\lambda_{\mathsf{S}}\lambda_{\min}(\mathbf{C}_\infty(B))^{-1}\|\hat{\boldsymbol{A}}_{\mathsf{S}} - \boldsymbol{A}_0\|_\star \\
&\le 24\Delta_n^2\kappa(\mathbf{C}_\infty(B))\|\boldsymbol{A}_0\|_2^4\Big(\frac{1}{2} + \frac{1}{6}\exp(\Delta_n\|\boldsymbol{A}_0\|_2)\Delta_n\|\boldsymbol{A}_0\|_2\Big)^2 \\
&\quad + 4\lambda_{\mathsf{S}}^2\Big(\frac{3c_\eta}{2c_\star^2\lambda_{\min}(\mathbf{C}_\infty(B))} + \frac{192}{\lambda_{\min}(\mathbf{C}_\infty(B))^2}\Big)\Bigg(s\log\Big(\frac{2\mathrm{e}d^2}{s}\Big) \vee \log(2/\varepsilon)\Bigg),
\end{aligned}
$$

这里我们还用到了 $\eta_- \ge \eta/2$ 以及关于 T 的假设可推出

$$
\mathbb{P}(\|\Delta\widetilde{Z}_1\| \le \eta_-) \ge \frac{1}{2}.
$$

注意到 $\|\mathbf{M}\|_\star \ge \sqrt{对数(2)}\|\mathbf{M}\|_1$, 对任意$\mathbf{M} \in \mathrm{R}^{d\times d}$, 即可完成证明。 ■

5. 模拟研究

本节旨在研究所提出估计量在合成数据上的实际表现，以及调节参数 b 和 η的影响。在以下所有实验中，分析均基于通过 Euler–Maruyama 格式生成的 OU 过程的离散观测，其离散步长等于 10^{-2}。我们根据数值分析选择这一数值，该分析见

中提供的 [17]。除非另有说明，我们假设已观测到所有生成的数据。

We c将 Lasso 和 Slope 估计量与两个基于似然的参考估计量进行比较 s.

1. 最小化截断负对数似然的估计量 $\mathcal{L}_n^D$(见 (2.6)) 用于理论分析。为简便起见，我们将此估计量称为“截断最大似然估计”。
2. 最小化以下函数的估计量

$$\frac{1}{T}\sum_{i=1}^{n}(\boldsymbol{A}X_{t_{i-1}})^\top \Delta X_i^c + \frac{\Delta_n}{2T}\sum_{i=1}^{n}\|\boldsymbol{A}X_{t_{i-1}}\|^2, \quad \text{where} \quad \Delta X_i^c = X_{t_i}^c - X_{t_{i-1}}^c, \tag{5.1}$$

其中 $(X_t^c)_{t\geq 0}$ 表示 $\mathbf{X}$ 的连续鞅部分，见第 2 节。由于 (5.1) 是对连续观测下负对数似然的一种自然离散化（见第 2.2 节），我们将此估计量称为“真实最大似然估计”。

后者仅用作参考，因为给定离散观测时，连续鞅部分是未知的。按照 [11]，的方法，Lasso 和 Slope 的调节参数 λ_{L} 和 λ_{S} 通过交叉验证选取。每条观测轨迹分为前后两个连续部分：前 80% 的样本用于计算候选估计量，剩余 20% 留作验证。针对对数网格上介于 10^{-3} 和 10 之间的各个相应调节参数值，在训练部分计算每个估计量，并通过计算验证部分上相应的截断负对数似然进行评估。随后选取使该验证准则最小的参数，并将其作为最终估计量。

每个估计量的表现将通过相对于真实漂移矩阵 $\boldsymbol{A}_0$ 的 L_1 和 L_2 误差来评估。

为了生成既是 s-稀疏矩阵又属于 $M_+(\mathrm{R}^d)$ 的 $\boldsymbol{A}_0$，我们按以下方式进行：首先选取 $s-d$ 个非对角位置，除非另有说明，这些位置的值均从区间 $[-0.5,0.5]$. 中均匀抽样。随后，将对角线元素设为对应行中非对角元素绝对值之和，再加上一个较小的偏移量 0.1，即

$$(\boldsymbol{A}_0)_{ii} = \sum_{j\neq i}|(\boldsymbol{A}_0)_{ij}| + 0.1, \quad i = 1,\dots,d.$$

这会生成一个稀疏且对角占优的矩阵，从而确保模拟过程的稳定性。

我们的估计量还要求选择集合 B 和截断参数 η，以便在估计过程中分别排除$\mathbf{X}$ 及其增量中的极端值。对于 B，我们始终选择 $B = B(0,b)$，其中 $b>0$ 是一个待确定的参数。根据 b 或 η 的取值，我们可以计算被过滤掉的观测值所占的比例，见图5.5 和5.6。我们观察到，当参数达到一定水平后，估计量的表现便趋于稳定。因此，除非另有说明，我们始终选择能够过滤掉不超过 10% 观测值的 b 和 η 的最小值。

不同估计量的初步比较见图 5.1，其中漂移矩阵$\boldsymbol{A}_0$ 及不同估计量均以热图呈现。本例中的维数为$d=25$，矩阵 $\boldsymbol{A}_0$ 是 s-稀疏的，其中 $s=60$ 且可获得截至时间 $T=100$ 的观测值。 $\boldsymbol{A}_0$ 的所有非零非对角元素均从区间 $[-0$中均匀抽取。8，0。8]。 本例中的 BDLP 是由标准布朗运动与一个跳跃强度为 1、跳跃大小服从拉普拉斯分布的复合泊松过程之和构成的。

我们发现，与最大似然估计型估计量相比，Lasso 和 Slope 估计量能更好地恢复真实漂移矩阵的稀疏性。图5.2详细展示了所讨论估计量的支撑恢复结果。

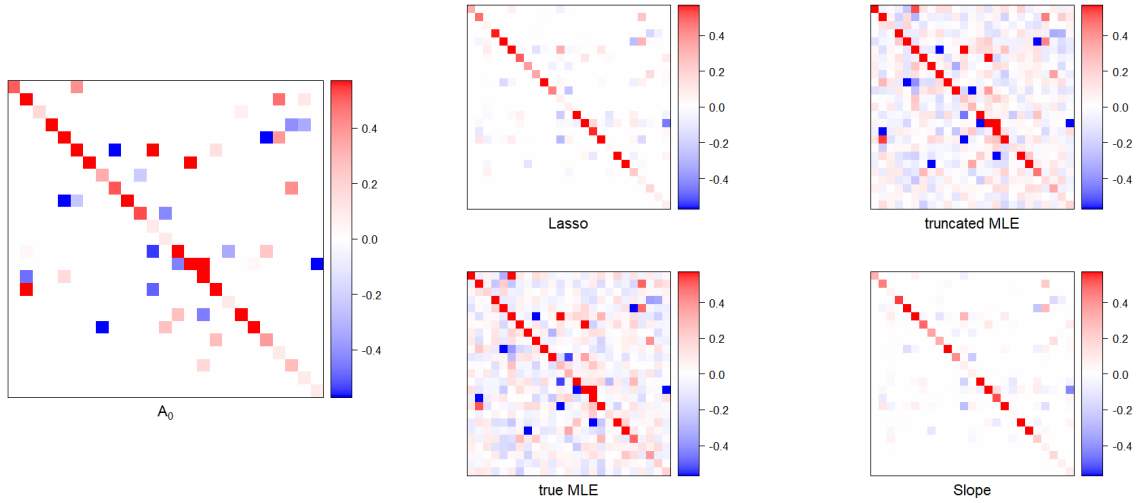


图 5.1：真实参数 A_0 与 Lasso 和 Slope 估计量，以及真实最大似然估计和截断最大似然估计的比较

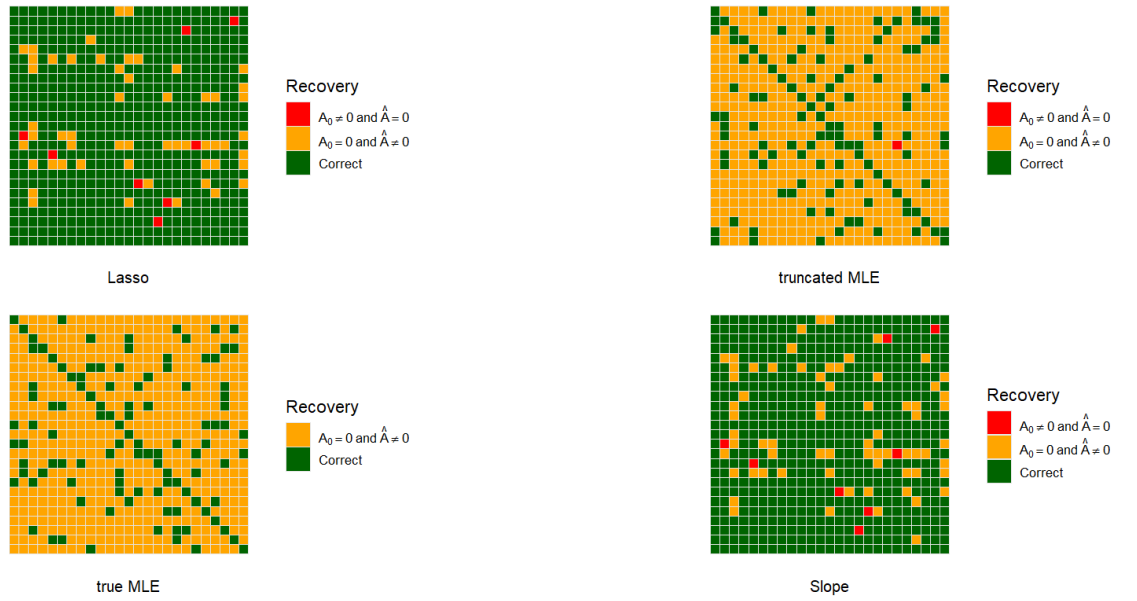


图 5.2：所研究估计量的支撑恢复。绿色条目表示对 A_0 系数的正确分类，即真实值和估计值要么均为零，要么均非零。红色表示估计量将某个系数设为零，但其真实值非零。橙色表示估计出了非零值，但真实系数为零。

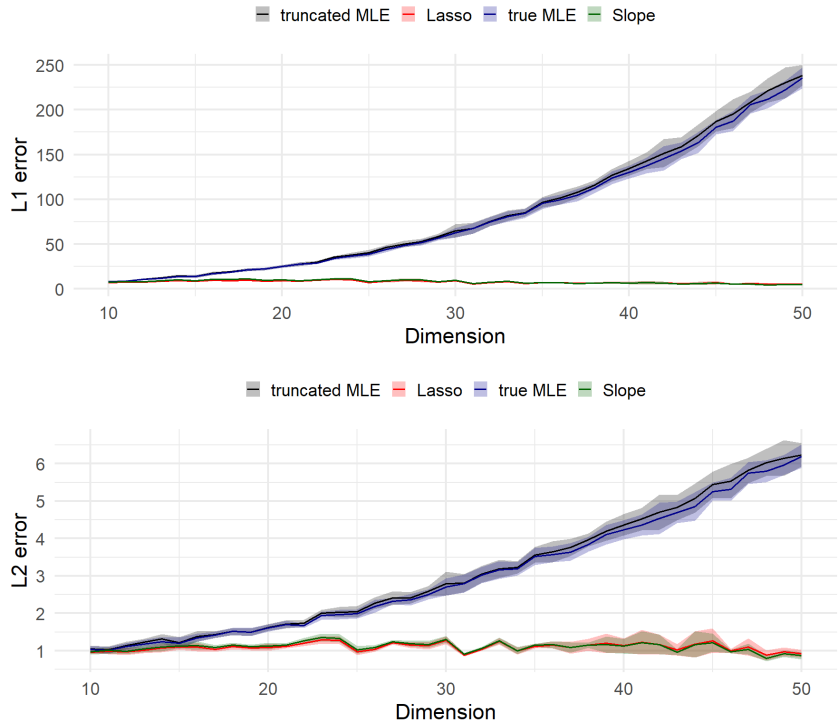


图 5.3： L_1 （上）和 L_2 （下）为真实值与截断最大似然估计、Lasso 和 Slope ± 的误差， ± 表示一个标准差。

为评估所研究的估计量对维数的依赖性， d 在稀疏设置下，我们计算其相应的 L_1 和 L_2 误差，考察 $d = 10$ 至 50 ，同时将稀疏性 s 固定为 $s = 55$。对于每个 d 的取值，我们生成一个时间跨度为 $T = 100$ 的 Lévy 驱动的 OU 过程，并重复此过程 10 次。BDLP 的选择与图 5.1 和 5.2中的相同。本研究的结果见图 5.3。我们观察到，Lasso 和 Slope 的误差显著小于 MLE 型估计量。此外，对于 Lasso 和 Slope，误差在所有维数下几乎保持恒定，这与我们的理论结果一致。相比之下，MLE 型估计量的误差随着维数增加而增大，因此无法利用漂移矩阵的稀疏性。

我们还在低频设置下进行了同样的实验，即采用更大的步长Δ_n。更具体地说，我们仍然模拟$\mathbf{X}$于一个网格上，该网格具有$T \cdot 10^2 = 10^4$个点，以确保 Euler–Maruyama 格式的精度，但仅使用 $n = 80$ 个近似等距的观测值进行估计。因此，步长 $\Delta_n = T/n$ 由上一实验中的0.01 增加到1.25。该模拟的结果见图5.4。尽管处于低频设置，我们并未观察到 Lasso 和 Slope 估计量的行为有显著变化。关于观测间距 Δ_n 之影响的更详尽分析见图5.7。

我们通过一项额外研究来补充该模拟研究，考察关键模型参数如何影响估计量的表现。具体而言，我们分析截断参数 b 和 η以及离散化水平 Δ_n的影响。由于 Lasso 与 Slope 估计量在先前实验中的表现几乎相同，我们仅对 Lasso 和 MLE 型估计量开展实验。

在这些实验中，我们预期会观察到稳定化效应。一旦 b 和 η 超过一定阈值，就不会再剔除观测值，因此估计量的性能应趋于平稳。同样，当 Δ_n足够小时，离散化误差会被随机误差和截断偏差所主导，从而再次趋于稳定。这些理论预期

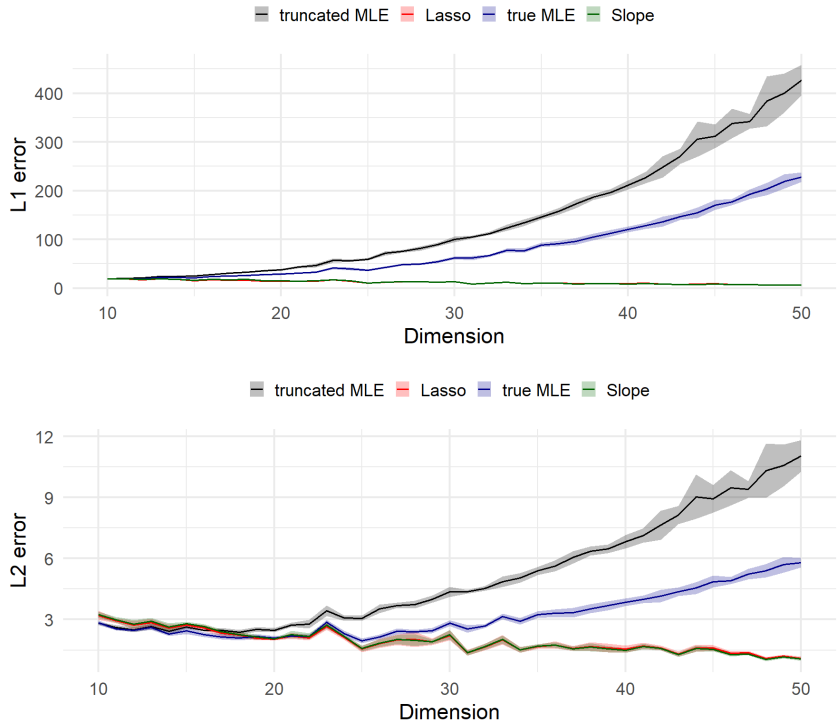


图 5.4： L_1 （上）和 L_2 （下）真实值与截断最大似然估计、Lasso 和 Slope 的误差 ± 一个标准差。模拟在低频设置下进行。

得到了实证结果的证实。此外，模拟使我们能够考察稳定化之前的过渡阶段，展示估计量在达到其极限性能之前的表现。

所有后续模拟均在维数为 $d = 25$的情况下进行。真实漂移矩阵 A_0被构造为具有 $s = 125$ 个非零元素，对应的稀疏性为 20%。所有报告结果均为 10 次独立模拟运行的平均值。

对于截断参数 b，它用于移除范数较大的观测值，图 5.5 显示，Lasso 和最大似然估计的估计误差均随 b 增大而减小。这一现象符合预期，因为较大的 b 使估计量能够利用更多可用数据。最大似然估计的这一效应明显更强，表明 Lasso 估计量对极端观测值具有更强的稳健性。将误差与保留观测值的比例进行比较可知，当使用约 75% 的数据后， b 截断的影响变得可以忽略。所讨论模拟中的 BDLP 由标准布朗运动与强度为1 的复合泊松过程之和构成，跳跃服从参数为 4.5的对称帕累托分布。为了聚焦于 b的影响，我们取 $\eta = 1000$，即取一个非常大的值。

我们现在重点考察 η这一截断偏差的主要成因所带来的影响。我们设置 $b = 1000$ 以减小 b 截断的影响，并将 BDLP 设为布朗运动与复合泊松过程之和，其中复合泊松过程具有独立的拉普拉斯分布跳跃，强度为 10。结果见图5.6 ，可概括如下。经过初期的嘈杂阶段后，出现了两个明显的阶段。第一阶段约使用 90% 的观测值（截至约 $\eta = 5$），第二阶段则几乎纳入全部观测值（$\eta > 8$）。在前一阶段，所有误差都显著更小，且截断最大似然估计优于 Lasso。尤其是，其表现与真实最大似然估计相当。然而，过渡阶段之后，截断最大似然估计的误差显著增大，而 Lasso 估计量的误差仅略有增长。这

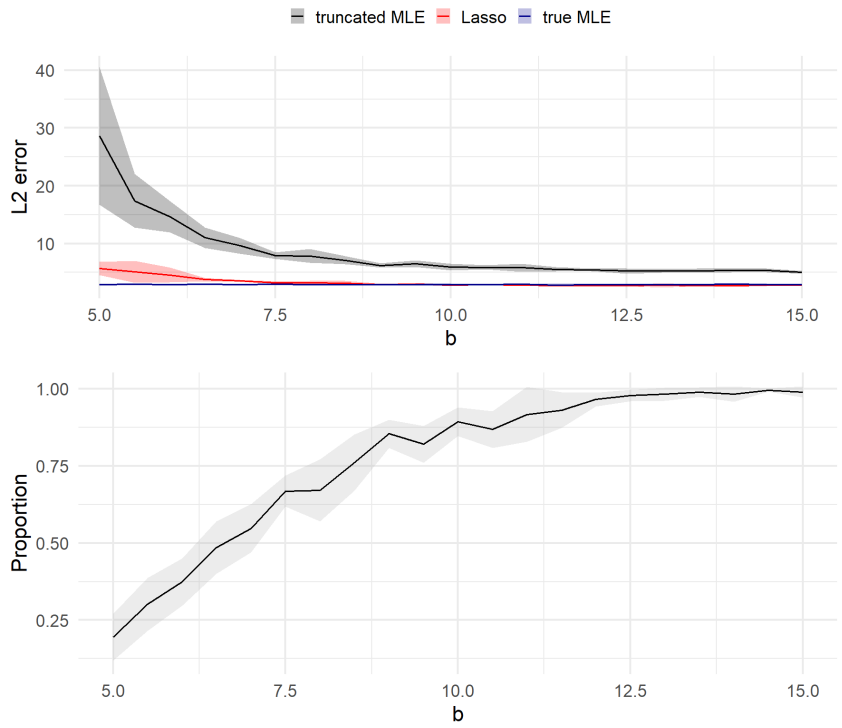


图 5.5： L_2 误差（上图）以及截断最大似然估计和 Lasso 的使用观测值比例（下图）（± 一个标准差），对应于不同的 b取值。上图还绘出了真实最大似然估计的 L_2 误差作为参照。

这再次表明，与截断最大似然估计相比，Lasso 对底层参数变化的敏感度较低。

最后，我们考察观测间距 Δ_n的影响。为此，我们再次选择一个非常宽松的截断设置 $b = \eta = 1000$。与之前一样，我们沿着该过程的轨迹模拟 $T \cdot 10^2$ 个点，但现在假设只观测其中 n 个近似等间距的点。因此，将 n 从 50 增加到 1000 意味着观测频率更高，且 Δ_n更小。

BDLP 再次选为标准布朗运动与强度为 1 且具有拉普拉斯分布跳跃的复合泊松过程之和；我们考察 Δ_n 在 $T = 10$ 和 $T = 100$下的影响。该研究的结果见图 5.7。对于 $T = 10$，我们看到最大似然估计和 Lasso 估计量的误差都几乎保持不变，即使 n 非常小、对应于 Δ_n较大的取值时也是如此。这表明，在此范围内，随机误差占主导地位，正如我们的理论所预测的那样。类似地，对于 $T = 100$，我们观察到，只要 $n \geq 300$ 或 $\Delta_n \leq 1/3$，误差就保持稳定。然而，当 Δ_n 取值较大时，误差会减小，这也与我们的理论结果一致。

6. 结论与展望

本文研究了基于离散观测的 Lévy 驱动的 OU 过程的高维漂移估计。在漂移矩阵满足稀疏性假设的条件下，我们通过推导尖锐的非渐近预言不等式，证明了 Lasso 和 Slope 估计量在高频情形下能够达到极小极大最优结果。这些界量化了偏差、截断误差、离散化误差和随机波动。我们还在非常一般的假设下，确定了样本复杂度与 BDLP 的 Lévy 测度尾部之间的关系。

还有　若干种可能的推广方向。一种自然的方法是扩展　g

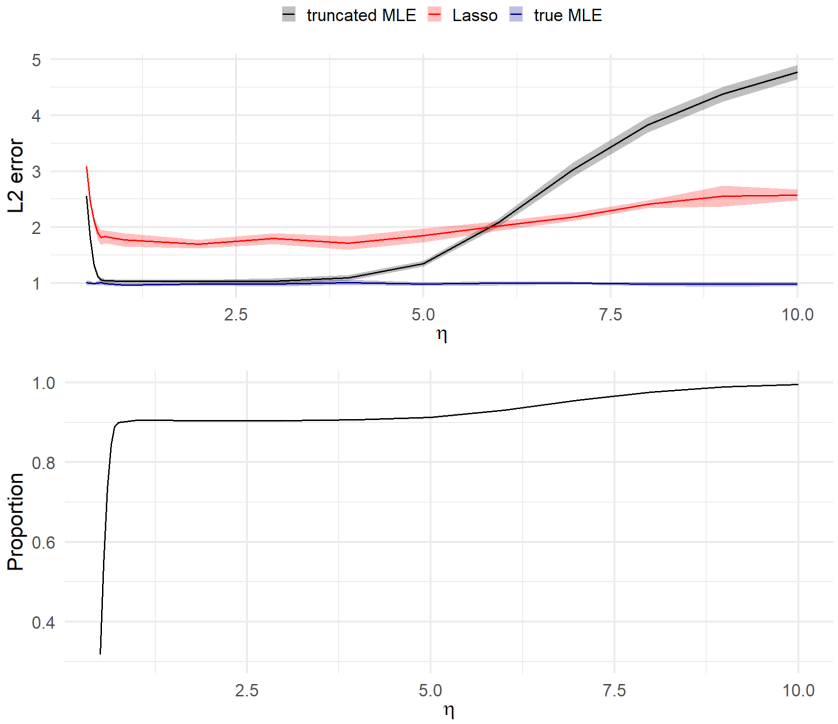


图 5.6: L_2 截断最大似然估计和 Lasso 的误差 ± 一个标准差（上图），以及不同 η 值下使用的观测值比例（下图）。作为参考，上图中还绘出了真实最大似然估计的 L_2 误差。

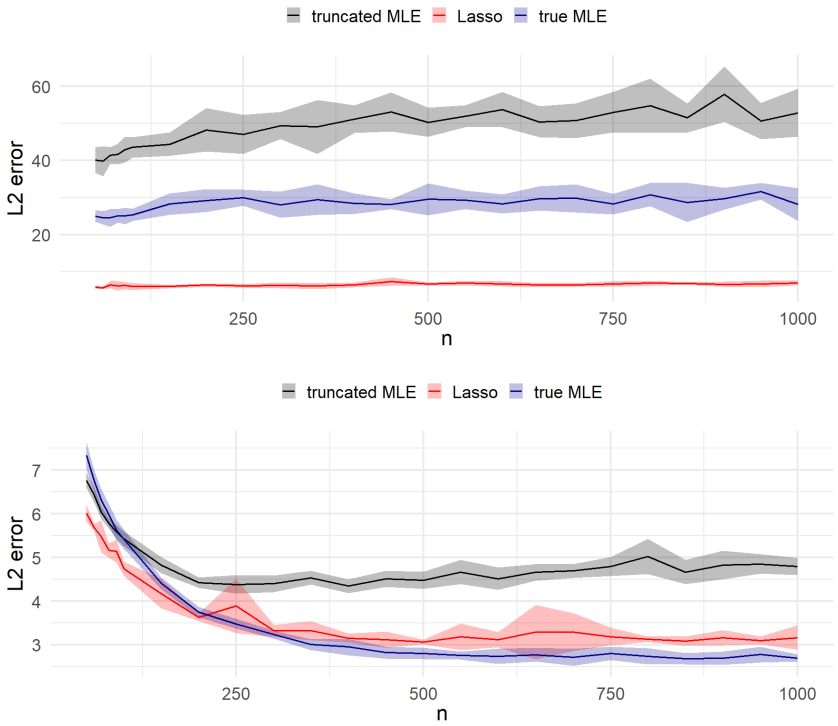


图 5.7: L_2 截断最大似然估计、真实最大似然估计和 Lasso 的误差 ± 一个标准差（上图），以及不同 n 和 $T = 10$ 值（上图）与 $T = 100$ （下图）。

将分析从 Lévy 驱动的 OU 过程推广到更一般的莱维驱动的扩散过程，类似于 [1] 对连续扩散过程的做法。这需要对离散化误差进行更深入的分析；由于我们聚焦于 OU 过程，因而可以利用其动力学被显式已知这一事实。另一个可能的推广是非遍历情形。事实上，遍历性假设 $(\mathcal{A}_0)$ 要求稀疏性 s 大于维数 d，从而不允许出现极端稀疏性的设定。对于标量 OU 过程，非遍历情形例如已在第 3.5 章中得到研究，见 [18]。然而，高维分析需要更为精细的方法；协方差矩阵的集中性可能是最具挑战性的部分。针对非遍历情形的一个具体结构假设是漂移矩阵的低秩结构。在经典的独立同分布情形中，这一设定已由 [21] 加以研究，并且对 OU 过程或许可以获得类似的结果。

由于将经典 OU 过程的结果推广到 Lévy 驱动的 OU 过程，与将具有次高斯噪声的统计方法推广到重尾误差密切相关，因此，推广我们结果的另一种有希望的方法是借鉴这一领域的研究。该领域的一个常见做法是替换 Lasso 估计量定义中的常规 L_2 准则，因为它对重尾噪声导致的离群值惩罚过于严厉。一个例子是 [15],，它采用 L_1惩罚形式的分位数回归，不需要对噪声的尾部作任何假设。然而，与这项研究不同的是，在我们的设定中，不仅观测噪声具有重尾特性，设计分布也是如此。据我们所知，这一点仅在近期论文 [30], 中得到探讨；我们在撰写本文时尚未留意到该论文。

7. 致谢

本研究得到 ERC A2B 项目资助（资助协议编号：101124751）。作者在伯克利访问西蒙斯计算理论研究所期间，完成了本研究的部分工作。

A. 辅助结果

引理 A.1. 对于任意时间点 $0 \le t_1 < t_2 \le T$，均有

$$X_{t_2} = \mathrm{e}^{-(t_2-t_1)A_0} X_{t_1} + \int_{t_1}^{t_2} \mathrm{e}^{-(t_2-u)A_0}\,\mathrm{d}Z_u.$$

证明。 回顾奥恩斯坦–乌伦贝克方程（2.3）解的显式形式，由此可得

$$X_0 = \mathrm{e}^{t_1 A_0}\left(X_{t_1} - \int_0^{t_1} \mathrm{e}^{-(t_1-u)A_0}\,\mathrm{d}Z_u\right).$$

再次将上述结果代入解的公式（2.3），但这次取点 t_2，得到

$$\begin{aligned} X_{t_2} &= \mathrm{e}^{-t_2 A_0}\mathrm{e}^{t_1 A_0}\left(X_{t_1} - \int_0^{t_1} \mathrm{e}^{-(t_1-u)A_0}\,\mathrm{d}Z_u\right) + \int_0^{t_2} \mathrm{e}^{-(t_2-u)A_0}\,\mathrm{d}Z_u \\ &= \mathrm{e}^{-(t_2-t_1)A_0} X_{t_1} + \int_{t_1}^{t_2} \mathrm{e}^{-(t_2-u)A_0}\,\mathrm{d}Z_u. \end{aligned}$$

■

引理 A.2。 对于任意 $\eta > 0$，均有

$$\left\{\|\Delta\widetilde{Z}_i\| < \eta_-\right\} \cap \{X_{t_{i-1}} \in B\} \subseteq \{\|\Delta X_i\| < \eta\} \cap \{X_{t_{i-1}} \in B\} \subseteq \left\{\|\Delta\widetilde{Z}_i\| < \eta_+\right\} \cap \{X_{t_{i-1}} \in B\}$$

and

$$\{\|\Delta X_i\| \geq \eta\} \cap \{X_{t_{i-1}} \in B\} \subseteq \left\{\|\Delta \widetilde{Z}_i\| \geq \eta_-\right\},$$

其中

$$\begin{aligned} \eta_+ &:= \eta + b(\exp(\Delta_n \|\boldsymbol{A}_0\|_{\mathrm{op}}) - 1), \\ \eta_- &:= \eta - b(\exp(\Delta_n \|\boldsymbol{A}_0\|_{\mathrm{op}}) - 1). \end{aligned} \tag{A.1}$$

证明。 注意，根据引理A.1可知

$$\begin{aligned} \|\Delta X_i\| &= \|X_{t_i} - X_{t_{i-1}}\| \\ &= \|(\exp(-\Delta_n \boldsymbol{A}_0) - \mathbb{I}_d) X_{t_{i-1}} + \int_{t_{i-1}}^{t_i} \exp(-(t_i - u))\boldsymbol{A}_0 \, \mathrm{d}Z_u\| \\ &\geq \|\int_{t_{i-1}}^{t_i} \exp(-(t_i - u)\boldsymbol{A}_0) \, \mathrm{d}Z_u\| - \|(\exp(-\Delta_n \boldsymbol{A}_0) - \mathbb{I}_d) X_{t_{i-1}}\| \\ &= \|\Delta \widetilde{Z}_i\| - \|(\exp(-\Delta_n \boldsymbol{A}_0) - \mathbb{I}_d) X_{t_{i-1}}\| \end{aligned}$$

这意味着

$$\begin{aligned} &\{\|\Delta X_i\| < \eta\} \cap \{X_{t_{i-1}} \in B\} \\ &\subseteq \left\{\|\Delta \widetilde{Z}_i\| < \eta + \|(\exp(-\Delta_n \boldsymbol{A}_0) - \mathbb{I}_d) X_{t_{i-1}}\|\right\} \cap \{X_{t_{i-1}} \in B\} \\ &\subseteq \left\{\|\Delta \widetilde{Z}_i\| < \eta + b\|(\exp(-\Delta_n \boldsymbol{A}_0) - \mathbb{I}_d)\|_{\mathrm{op}}\right\} \cap \{X_{t_{i-1}} \in B\} \\ &\subseteq \left\{\|\Delta \widetilde{Z}_i\| < \eta + b(\exp(\Delta_n \|\boldsymbol{A}_0\|_{\mathrm{op}}) - 1)\right\} \cap \{X_{t_{i-1}} \in B\}. \end{aligned}$$

其余的集合包含关系可用类似的论证得出。 ∎

引理 A.3。 对于任意 $i \in [n]$，均有

$$\lambda_{\max}\left(\mathbb{E}[(\Delta \widetilde{Z}_i)(\Delta \widetilde{Z}_i)^\top | \mathcal{F}_{t_{i-1}}]\right) \leq \Delta_n \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2) \exp(\Delta_n \|\boldsymbol{A}_0\|).$$

~~证明。 结合 Courant–Fischer 定理与伊藤等距，由于 ΔZ_i，是~~独立同分布的，且 ΔZ_i 与 $\mathcal{F}_{t_{i-1}}$相互独立，得到

$$\begin{aligned} &\lambda_{\max}\left(\mathbb{E}[(\Delta \widetilde{Z}_i)(\Delta \widetilde{Z}_i)^\top | \mathcal{F}_{t_{i-1}}]\right) \\ &= \max_{u \in \mathbb{R}^d : \|u\| = 1} \mathbb{E}[(u^\top \Delta \widetilde{Z}_1)^2] \\ &= \max_{u \in \mathbb{R}^d : \|u\| = 1} \left(\int_0^{\Delta_n} \|u^\top \exp(-(\Delta_n - s)\boldsymbol{A}_0)\boldsymbol{\Sigma}\|^2 \, \mathrm{d}s + \int_0^{\Delta_n} \int_{\mathbb{R}^d} \|u^\top \exp(-(\Delta_n - s)\boldsymbol{A}_0) z\|^2 \nu(\mathrm{d}z) \, \mathrm{d}s \right) \\ &= \max_{u \in \mathbb{R}^d : \|u\| = 1} u^\top \left(\int_0^{\Delta_n} \exp(-(\Delta_n - s)\boldsymbol{A}_0) \left(\mathbf{C} + \int_{\mathbb{R}^d} zz^\top \nu(\mathrm{d}z) \right) \exp(-(\Delta_n - s)\boldsymbol{A}_0^\top) \, \mathrm{d}s \right) u \\ &\leq \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2) \max_{u \in \mathbb{R}^d : \|u\| = 1} \int_0^{\Delta_n} \|\exp(-(\Delta_n - s)\boldsymbol{A}_0^\top) u\|^2 \, \mathrm{d}s \\ &\leq \Delta_n \lambda_{\max}(\mathbf{C} + \boldsymbol{\nu}_2) \exp(\Delta_n \|\boldsymbol{A}_0\|), \end{aligned}$$

至此证明完毕。 ∎

命题 A.4. 假设 *BDLP* $\mathbf{Z}$ 的莱维三元组为 $(0,\mathbf{C},0)$。那么，对于 $T \geq \lambda_{\max}(\mathbf{C}_\infty)\lambda_{\min}(\mathbf{C})^{-1}\log(2d\kappa(\mathbf{C}_\infty)^2\kappa(\mathbf{C}))$，有

$$\beta_{\mathbf{X}}(T) \leq \frac{3}{2}d\kappa(\mathbf{C}_\infty)^2\kappa(\mathbf{C})\exp\left(-\frac{T\lambda_{\min}(\mathbf{C})}{\lambda_{\max}(\mathbf{C}_\infty)}\right) + \frac{1}{\sqrt{2}}d\kappa(\mathbf{C}_\infty)\exp\left(-\frac{T\lambda_{\min}(\mathbf{C})}{2\lambda_{\max}(\mathbf{C}_\infty)}\right).$$

证明。 首先回顾一下，$\mathbf{C}_\infty = \int_0^\infty \mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}\,\mathrm{d}s$，满足李雅普诺夫方程

$$-A_0\mathbf{C}_\infty - \mathbf{C}_\infty A_0^\top = -\mathbf{C},$$

因此， [29] 中的推论 5 给出，对于任意 $T > 0$

$$\begin{aligned}\left\|\mathrm{e}^{-TA_0}\right\|_2^2 &\leq \kappa(\mathbf{C}_\infty)\,\mathrm{tr}\left(\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\mathbf{C}\right)\right)\\ &\leq d\kappa(\mathbf{C}_\infty)\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right).\end{aligned} \tag{A.2}$$

根据 [20], 中的命题 2.1 和 2.2，以及 [22]、μ 和 P_T 中的定理 3.1、4.1 和 4.2，x，$\cdot$) 对任意 $x \in \mathrm{R}^d$ 都是无限可分的，其相应的莱维三元组分别为 $(0,\ \mathbf{C}_\infty,\ 0)$ 和 $(\mathrm{e}^{-TA_0}x,\ \mathbf{C}_T,\ 0)$，其中

$$\mathbf{C}_T := \int_0^T \mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}\,\mathrm{d}s,$$

也就是说，它们服从正态分布。现在，对于任意 $x \in \mathrm{R}^d$，$T \geq 0$ 有

$$\|P_T(x,\cdot) - \mu\|_{\mathrm{TV}} \leq \|P_T(x,\cdot) - P_T(0,\cdot)\|_{\mathrm{TV}} + \|P_T(0,\cdot) - \mu\|_{\mathrm{TV}}, \tag{A.3}$$

并且根据 [9]中的定理 1.1 和公式 (2)A.2，可得

$$\begin{aligned}&\|P_T(0,\cdot) - \mu\|_{\mathrm{TV}}\\ &\leq \frac{3}{2}\left\|\mathbf{C}_\infty^{-1/2}\int_T^\infty \mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}\,\mathrm{d}s\mathbf{C}_\infty^{-1/2}\right\|_2\\ &\leq \frac{3}{2}\|\mathbf{C}_\infty^{-1/2}\|^2\left\|\int_T^\infty \mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}\,\mathrm{d}s\right\|_2\\ &\leq \frac{3}{2}\lambda_{\min}(\mathbf{C}_\infty)^{-1}\int_T^\infty \left\|\mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}\right\|_2\,\mathrm{d}s\\ &\leq \frac{3}{2}\lambda_{\min}(\mathbf{C}_\infty)^{-1}\lambda_{\max}(\mathbf{C})\int_T^\infty \left\|\mathrm{e}^{-sA_0}\right\|_2^2\,\mathrm{d}s\\ &\leq \frac{3}{2}d\lambda_{\min}(\mathbf{C}_\infty)^{-1}\lambda_{\max}(\mathbf{C})\kappa(\mathbf{C}_\infty)\int_T^\infty \exp\left(-s\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right)\mathrm{d}s\\ &= \frac{3}{2}d\kappa(\mathbf{C})\kappa(\mathbf{C}_\infty)^2\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right).\end{aligned} \tag{A.4}$$

此外，平克斯克不等式以及关于正态分布的库尔贝克–莱布勒散度的经典结果给出，连同(A.2)

$$\begin{aligned}&\|P_T(x,\cdot) - P_T(0,\cdot)\|_{\mathrm{TV}}\\ &\leq \sqrt{\frac{1}{2}\,\mathrm{KL}(P_T(x,\cdot), P_T(0,\cdot))}\\ &= \frac{1}{2}\sqrt{x^\top \mathrm{e}^{-TA_0^\top}\mathbf{C}_T^{-1}\mathrm{e}^{-TA_0}x}\\ &= \frac{1}{2}\left\|\mathbf{C}_T^{-1/2}\mathrm{e}^{-TA_0}x\right\|\end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{2}\|x\|\left\|\mathbf{C}_T^{-1/2}\mathrm{e}^{-TA_0}\right\| \\
&\le \frac{1}{2}\|x\|\lambda_{\min}(\mathbf{C}_T)^{-1/2}\left\|\mathrm{e}^{-TA_0}\right\|_2 \\
&\le \frac{1}{2}\|x\|\lambda_{\min}(\mathbf{C}_T)^{-1/2}\left(d\kappa(\mathbf{C}_\infty)\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right)\right)^{1/2}.
\end{aligned}
$$

进而，极小极大定理蕴含，连同(A.2)

$$
\begin{aligned}
\lambda_{\min}(\mathbf{C}_T) &= \min_{x\in\mathbb{R}^d:\|x\|=1}\int_0^T x^\top \mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}x\,\mathrm{d}s \\
&\ge \min_{x\in\mathbb{R}^d:\|x\|=1}\int_0^\infty x^\top \mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}x\,\mathrm{d}s - \max_{x\in\mathbb{R}^d:\|x\|=1}\int_T^\infty x^\top \mathrm{e}^{-sA_0}\mathbf{C}\mathrm{e}^{-sA_0^\top}x\,\mathrm{d}s \\
&\ge \lambda_{\min}(\mathbf{C}_\infty) - \lambda_{\max}(\mathbf{C})\int_T^\infty \|\mathrm{e}^{-sA_0^\top}\|_2^2\,\mathrm{d}s \\
&\ge \lambda_{\min}(\mathbf{C}_\infty) - d\kappa(\mathbf{C}_\infty)\lambda_{\max}(\mathbf{C})\int_T^\infty \exp\left(-s\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right)\mathrm{d}s \\
&= \lambda_{\min}(\mathbf{C}_\infty) - d\lambda_{\max}(\mathbf{C}_\infty)\kappa(\mathbf{C}_\infty)\kappa(\mathbf{C})\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right),
\end{aligned}
$$

从而对于 $T \ge \lambda_{\max}(\mathbf{C}_\infty)\lambda_{\min}(\mathbf{C})^{-1}\log(2d\kappa(\mathbf{C}_\infty)^2\kappa(\mathbf{C})) =: T_0$,

$$
\lambda_{\min}(\mathbf{C}_T) \ge \frac{1}{2}\lambda_{\min}(\mathbf{C}_\infty),
$$

因此，对于 $T \ge T_0$，有

$$
\begin{aligned}
&\|P_T(x,\cdot) - P_T(0,\cdot)\|_{\mathrm{TV}} \\
&\le \|x\|(2\lambda_{\min}(\mathbf{C}_\infty))^{-1/2}\left(d\kappa(\mathbf{C}_\infty)\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right)\right)^{1/2}.
\end{aligned}
\tag{A.5}
$$

结合式 (A.3)、(A.4) 和 (A.5)，由 $T \ge T_0$ Hölder不等式

$$
\begin{aligned}
\beta_{\mathbf{X}}(T) &= \int \|P_T(x,\cdot) - \mu\|_{\mathrm{TV}}\mu(\mathrm{d}x) \\
&\le \frac{3}{2}d\kappa(\mathbf{C})\kappa(\mathbf{C}_\infty)^2\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right) \\
&\quad + (2\lambda_{\min}(\mathbf{C}_\infty))^{-1/2}\left(d\kappa(\mathbf{C}_\infty)\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right)\right)^{1/2}\mathbb{E}[\|X_0\|^2]^{1/2} \\
&\le \frac{3}{2}d\kappa(\mathbf{C})\kappa(\mathbf{C}_\infty)^2\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})\right) \\
&\quad + \frac{1}{\sqrt{2}}d\kappa(\mathbf{C}_\infty)\exp\left(-T\lambda_{\max}(\mathbf{C}_\infty)^{-1}\lambda_{\min}(\mathbf{C})/2\right),
\end{aligned}
$$

至此，证明完成。 ■

参考文献

[1] C. Amorino、F. Pina 和 M. Podolskij。《高维扩散过程中漂移函数的 Lasso 估计所受的采样影响》。载于：*Electronic Journal of Statistics* 19.2 (2025)，第 5068–5116 页。[2] D. Applebaum。*Lévy* 过程与随机微积分。第二版。第 116 卷。Cambridge Studies in Advanced Mathematics。剑桥：剑桥大学出版社，2009 年。ISBN：978-0-521-73865-1。

[3] P. C. Bellec、G. Lecué 和 A. B. Tsybakov。《Slope 与 Lasso 相遇：改进的预言界与最优性》。载于：《统计年刊》46.6B (2018)，第 3603–3642 页。ISSN：0090-5364。[4] M. Bogdan、E. van den Berg、C. Sabatti、W. Su 和 E. J. Candés。《SLOPE——通过凸优化实现自适应变量选择》。载于：《应用统计年刊》9.3 (2015)，第 1103–1140 页。ISSN：1932-6157。[5] R. Carmona、J.-P. Fouque 和 L.-H. Sun。《平均场博弈与系统性风险》。载于：《数学科学通讯》13.4 (2015)，第 911–933 页。ISSN：1539-6746。[6] S. Christensen、C. Strauch 和 L. Trottner。《学会反射：数据驱动随机控制策略的统一方法》。载于：《伯努利》30.3 (2024)，第 2074–2101 页。[7] G. Ciołek、D. Marushkevych 和 M. Podolskij。《高维奥恩斯坦–乌伦贝克模型中漂移的 Dantzig 与 Lasso 估计量》。载于：《电子统计学杂志》14.2 (2020)，第 4395–4420 页。[8] A. De Gregorio、D. Frisardi、S. Iacus 和 F. Iafrate。《稀疏扩散过程的自适应弹性网络估计》。载于：《随机过程统计推断》28.3 (2025 年 12 月)，第 22 页。[9] L. Devroye、A. Mehrabian 和 T. Reddad。《均值相同的高维高斯分布之间的全变差距离》。2023 年。arXiv：1810.08693 [math.ST]。[10] N. Dexheimer、S. Gaudlitz 和 J. Schmidt-Hieber。《脉冲时序依赖赫布学习作为带噪梯度下降》。载于：第三十九届神经信息处理系统年会。2025 年。[11] N. Dexheimer 和 C. Strauch。《Lévy 驱动的 Ornstein–Uhlenbeck 过程的 Lasso 与 Slope 漂移估计量》。载于：《伯努利》30.1 (2024)，第 88–116 页。[12] N. Dexheimer、C. Strauch 和 L. Trottner。《连续时间混合马尔可夫过程在上确界范数风险下的自适应不变密度估计》。载于：《亨利·庞加莱研究所概率与统计年刊》58.4 (2022)，第 2029–2064 页。ISSN：0246-0203,1778-7017。[13] S. Dirksen。《通过泛链得到尾界》。载于：《电子概率学杂志》20 (2015)，第 53 号，29 页。[14] K. Dzhaparidze 和 J. van Zanten。《关于鞅的伯恩斯坦型不等式》。载于：《随机过程及其应用》93.1 (2001)，第 109–117 页。ISSN：0304-4149。[15] J. Fan、Y. Fan 和 E. Barut。《自适应稳健变量选择》。载于：《统计年刊》42.1 (2014 年 2 月)，第 324–351 页。[16] J.-P. Fouque 和 T. Ichiba。《银行间借贷模型中的稳定性》。载于：《SIAM金融数学杂志》4.1 (2013)，第 784–803 页。[17] S. Gaïffas 和 G. Matulewicz。《高维奥恩斯坦–乌伦贝克过程漂移的稀疏推断》。载于：《多元分析杂志》169 (2019)，第 1–20 页。ISSN：0047- 259X。[18] Y. A. Kutoyants。《遍历扩散过程的统计推断》。载于：施普林格统计学丛书 (2004)。[19] H. Mai。《Lévy 驱动的 Ornstein–Uhlenbeck 过程的高效最大似然估计》。载于：《伯努利》20.2 (2014)，第 919–957 页。ISSN：1350-7265。[20] H. Masuda。《由一般Lévy过程驱动的多维奥恩斯坦–乌伦贝克过程》。载于：《伯努利》10.1 (2004)，第 97–120 页。ISSN：1350-7265。[21] A. Rohde 和 A. B. Tsybakov。《高维低秩矩阵的估计》。载于：《统计年刊》39.2 (2011)，第 887–930 页。ISSN：00905364, 21688966。[22] K.-i. Sato 和 M. Yamazato。《算子自分解分布作为奥恩斯坦–乌伦贝克型过程的极限分布》。载于：《随机过程应用》17.1 (1984)，第 73–100 页。ISSN：0304-4149。

[23] M. Sørensen。《带跳扩散的似然方法》。载于：随机过程中的统计推断。第6卷。概率与纯应用。Dekker，纽约，1991年，第67–105页。[24] R. Tibshirani。《通过Lasso进行回归收缩与选择》。载于：皇家统计学会期刊。B辑（方法论）58.1（1996年），第267–288页。issn：00359246。[25] J. A. Tropp。《矩阵鞅的Freedman不等式》。载于：概率电子通讯16（2011年），第262–270页。[26] J. A. Tropp。《随机矩阵之和的用户友好型尾界》。载于：计算与数学基础12.4（2012年），第389–434页。issn：1615-3375,1615-3383。[27] R. Vershynin。《高维概率》。第47卷。剑桥统计与概率数学丛书。剑桥大学出版社，剑桥，2018年。isbn：978-1- 108-41519-4。[28] K. C.Wong、Z. Li和A.Tewari。《Lasso保证：针对β混合重尾时间序列》。载于：《统计年刊》48.2（2020年），第1124–1142页。[29] H. Xu。《关于矩阵指数的两个结果》。载于：线性代数及其应用262（1997年），第99–109页。issn：0024-3795,1873-1856。[30] K. Zhou、X. Zhang、W. Zhang和D. Wang。《重尾设计下极小极大最优稳健稀疏回归：一种基于梯度的方法》。2026年。arXiv：2601。05669 [统计学（方法论）]。