

6 Astra 评审原文

来源：作者于 2026-10-11 在会话中提供，标注为“6 Astra 的审稿”。下文保留收到的评审内容；其中 main 等引用标记来自原消息，未附带可解析的页码或文件链接。

评审对象：消息中称《main(6).pdf》。该文件未单独提供，尚不能通过文件哈希确认与本仓库 paper.pdf 完全一致。

结论：这版《main(6).pdf》，我没有发现大的数学 gap。正文的 Lemma 3.4、Theorems 4.1–4.5 与附录中的关键推导能够对应起来，没有发现需要实质性修改主定理假设、削弱主结论或重做核心证明的问题。

下面说明几个重点核查结果。

无偏性与矩控制：成立

高斯复修正的无偏性证明是完整的。附录 A.1 利用

$$U + iV \stackrel{d}{=} e^{i\varphi}(U + iV)$$

和一元整函数的圆周平均公式完成证明；指数增长界提供了交换期望与角积分所需的可积性。这个论证也不要求 Ω 可逆，因此允许奇异协方差没有问题。main

从无偏性到相对方差界，也没有遗漏关键交叉项。因为

$$\mathbb{E}[\Delta_t \mid \mathcal{F}_{t-1}, Z_t] = 0,$$

修正误差与干净样本的中心化梯度在条件二阶矩中正交，因而确实得到

$$\mathbb{E}_t \|\xi_t\|^2 \leq \sigma^2 + \rho \|\nabla F(\theta_t)\|^2.$$

这里依靠的是明确写出的幅度条件

$$M(\theta)^2 \leq A + B \|\nabla F(\theta)\|^2,$$

没有暗中把 $M(\theta)$ 当作全局有界常数。main

CLT 所需的高阶矩，也不是未经证明直接使用的。A.2 中，中心化消掉了 $g(\theta; 0)$ ，结合输入的尾部条件，可以推出

$$\mathbb{E}_t \|\xi_t\|^{2+\delta} \leq C_\delta (A + B \|\nabla F(\theta_t)\|^2)^{1+\delta/2}.$$

再利用已经独立证明的几乎处处收敛，就得到沿轨迹的有限随机上界。这足够支持文中的条件 Lindeberg 验证，不必另外增加一个全局、确定常数的高阶矩上界。main main 所用的“条件方差收敛 + 条件 Lindeberg”确实是适用的鞅中心极限定理条件。Galton

收敛性与两类 CLT：关键步骤闭合

强凸部分的基础递推

$$\mathbb{E}_t \|e_{t+1}\|^2 \leq (1 - 2\mu\eta_t + (1 + \rho)L^2\eta_t^2) \|e_t\|^2 + \sigma^2\eta_t^2$$

是正确的。由此使用 Robbins–Siegmund 得到几乎处处收敛，再通过标量递推分别得到均方速率，逻辑上没有把“几乎处处收敛”错误地直接当成“均方收敛”。 $\alpha = 1$ 时， $2\mu c > 1$ 、 $= 1$ 、 < 1 三种速率的求和计算也成立。main

凸目标界中的梯度项抵消，以及非凸结果中的下降不等式，与各自声明的步长限制一致；没有发现系数或条件错误。main main

我特别检查了 $\alpha = 1$ 的 CLT，因为这里容易出现实质性漏洞。当前版本的处理成立。文中只要求

$$c\lambda_{\min}(H) > \frac{1}{2},$$

而没有要求更强的全局条件 $2\mu c > 1$ 。附录 A.5 通过停时局部化，在最优点附近选取

$$\frac{1}{2c} < m < \lambda_{\min}(H),$$

得到局部化二阶矩 $O(t^{-1})$ ，再控制非线性余项。因此，后续证明没有错误调用一个未被全局条件保证的 $O(t^{-1})$ 速率。main

末次迭代 CLT 中，矩阵权重、余项估计和协方差极限能够对应上；两条 Lyapunov 方程中的 c 与 c^2 系数也正确。main

平均迭代部分同样没有把两个步长区间混淆。 $\alpha < 1$ 得到 $H^{-1}S_*H^{-1}$ ； $\alpha = 1$ 则单独推导了核

$$K_c(u) = cu^{cH-I} \int_u^1 v^{-cH} dv$$

及其积分协方差。附录还控制了 $u \downarrow 0$ 时的权重尾部，因此这里不只是形式上的黎曼和替换。main main

扩展与例子：没有发现影响主结论的遗漏

非高斯扩展没有仅凭“存在共轭变量”就直接宣称所有优化结论成立，而是区分了无偏性、二阶矩和高阶矩所需条件：

$$a < \lambda, \quad 2a < \lambda, \quad (2 + \delta)a < \lambda.$$

对于文中的均匀噪声与对称两点噪声，这些条件与所给共轭分布的尾部相符，能够支持相应论证。main

非线性特征回归中的随机响应 Y ，也没有被遗漏。A.9 直接使用

$$G(\theta) = A\theta - b$$

验证联合数据下的矩条件和相对方差界，而不是把固定响应的结论未经检查地套过去。main 我按实验给定分布重新计算了总体矩阵，得到

$$\lambda_{\min}(H) \approx 0.38190577, \quad 2c\lambda_{\min}(H) \approx 1.98591 > 1,$$

与文中给出的 $\mu \approx 0.382$ 及 $O(t^{-1})$ 理论速率一致。

神经网络实验则明确承认：尚未验证参数无关的指数增长率与幅度条件，因此不调用非凸主定理。这属于明确限定适用范围，不是把未经证明的理论保证强加给实验。main

最终意见：按当前写出的假设，这版数学部分可以通过“大 gap”审核。没有发现需要修改主定理结论的实质性漏洞。这个判断针对当前稿件的数学论证，不包含未提供的伴随论文全文或实验代码复现。