

CCSGD：中文阅读导读

与英文版逐节对应 | 2026 年 10 月 11 日

1. 论文解决什么问题？

目标是优化干净输入下的目标函数 $F(\theta) = \mathbb{E}\ell(\theta; Z)$ ，实际只能观测 $W = Z + U$ 。即使输入噪声均值为零， $g(\theta; W)$ 通常仍有偏；降低学习率不能消除这个总体偏差。论文构造可直接执行、面向干净目标的随机梯度更新。

主体理论假设观测误差为已知半正定协方差的高斯噪声、潜在输入亚高斯，且每步样本新鲜并独立于历史。全纯延拓和输入增长条件针对梯度；凸性针对目标函数关于参数的几何结构。这两类要求需分别检查。

2. 如何分类问题？

先分清修正问题与优化问题。高斯精确修正需要梯度关于输入的 entire 延拓及可积性，不需要凸性。要获得优化保证，先推导相对方差界，再按目标函数类别选用结果。

凸目标：定理 4.1 控制按步长加权平均的目标函数差。强凸目标：定理 4.2 证明到 θ^* 的几乎处处及均方收敛；定理 4.3–4.4 在额外局部条件下给出最后迭代和均匀平均的 CLT。非凸目标：定理 4.5 控制加权平均梯度平方，不保证到全局最优解的距离。

再按噪声分类。高斯噪声对应主体定理；均匀和对称双点噪声有明确的共轭扰动，见第 7.1 节，但共轭分布的指数尾部限制梯度增长。增长过快的 entire 梯度，以及任意非 entire 梯度，均不能自动获得保证。

3. 核心机制是什么？

独立生成 $V_t \sim N(0, \Omega)$ ，令 $G_t = \operatorname{Re} g^{\mathbb{C}}(\theta_t; W_t + iV_t)$ ，再更新 $\theta_{t+1} = \theta_t - \eta_t G_t$ 。经典复数 corrected-score 恒等式使每步梯度精确无偏。本篇通过明确的矩估计将它接入优化理论，计算时无需先求闭式积分。

关键增长界为 $\|D_z g^{\mathbb{C}}(\theta; q)\| \leq M(\theta)e^{a\|q\|}$ ，并要求 $M(\theta)^2 \leq A + B\|\nabla F(\theta)\|^2$ 。条件正交性分离原始采样噪声与修正误差，得到 $\mathbb{E}_t \|\xi_t\|^2 \leq \sigma^2 + \rho\|\nabla F(\theta_t)\|^2$ 。高阶矩由增长界和分布尾部推出，不能仅从二阶矩假设推出。

4. 哪些结论便于复用?

取 $\eta_t = c/(t + \gamma)^\alpha$, $1/2 < \alpha \leq 1$ 。在规定的步长上界下, 凸目标差和非凸加权平稳性指标在 $\alpha < 1$ 时为 $O(T^{\alpha-1})$, 在 $\alpha = 1$ 时为 $O(1/\log T)$ 。强凸 MSE 在 $\alpha < 1$ 时为 $O(t^{-\alpha})$; 在 $\alpha = 1$ 且 $2\mu c > 1$ 时为 $O(t^{-1})$, 临界和低于临界时分别有对数和较慢速率。

CLT 先使用收敛结论, 再局部化, 并用推导得到的 $(2 + \delta)$ 阶矩验证 Lindeberg 条件; 仍需协方差连续性和局部 Hessian 正则性。端点 $\alpha = 1$ 要求 $c\lambda_{\min}(H) > 1/2$, 且 PR 协方差不同, 不能直接代入区间内部的公式 $H^{-1}S_*H^{-1}$ 。

非高斯共轭情形中, $a < \lambda$ 足以保证修正恒等式, $2a < \lambda$ 足以完成矩控制, 其中均匀噪声的 $\lambda = \pi/b$, 双点噪声的 $\lambda = \pi/(2b)$ 。后一严格不等式还能容许适当的 $\delta > 0$ 。多项式梯度容易满足输入矩条件, 但仍需验证幅度控制及目标函数假设。

5. 实验说明什么?

非线性特征回归关于系数是强凸的; 附录 A.9 单独验证联合响应模型的矩条件。参数 MSE 拟合斜率为 -0.978 , 与 $O(t^{-1})$ 一致。虚线是指定更新规则的总体渐近平方偏差, 不是每个有限时刻的下界。

正弦网络实验衡量干净输入的预测 MSE, 属于经验研究。固定潜在训练样本每次使用时重抽观测噪声, 无偏性针对干净的经验目标; 不宣称网络全局收敛或预测误差趋零。到 200,000 步时, SIMEX 报告的 MSE 略小。WH 比较对象是其确定性式 (5), 不代表该文所有方法。

6. 建议阅读路径

了解算法: 第 3 节及算法 1, 再读第 5 节例子。复用证明: 引理 3.4 $\rightarrow$ 附录 A.2 $\rightarrow$ 第 4 节对应定理。理解 CLT: 定理 4.2 $\rightarrow$ 附录 A.5 $\rightarrow$ A.6 或 A.7。实现非高斯修正: 第 7.1 节及算法 2。核对实验范围: 第 6 节、附录 A.9 和附录 B。

Agent 应先读五页 AI 版本, 再用定理图中的稳定 ID 定位结果。原论文是数学叙述的权威版本; 共轭变量判据属于外部依赖, 作者提供的 6 Astra 评审未发现 main(6).pdf 存在大的数学缺口, 但该文件与仓库快照的文件级对应尚未确认。