

6 Astra Review: English Translation

Provenance. The author supplied the Chinese report on 11 October 2026, identifying it as a “6 Astra review.” This document translates that report; it is not a new review. Residual main citation markers are retained from the received text and do not resolve to supplied page numbers or file links.

Reviewed version. The received message names `main(6).pdf`. That file was not supplied separately, so its identity with this repository’s `paper.pdf` cannot be confirmed by file hash.

Conclusion: I found no major mathematical gap in this version of `main(6).pdf`. Lemma 3.4, Theorems 4.1–4.5, and the key appendix derivations are consistent with one another. I found no issue requiring substantive changes to the main theorem assumptions, weakening of the principal conclusions, or reconstruction of the core proofs.

The following are the main checks.

Unbiasedness and moment control: the arguments hold

The Gaussian complex-correction unbiasedness proof is complete. Appendix A.1 uses

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

and the circular mean-value formula for an entire function of one variable. The exponential growth bound supplies the integrability needed to interchange expectation and angular integration. Invertibility of Ω is not required, so singular covariance is allowed. `main`

The passage from unbiasedness to the relative variance bound does not omit a cross term. Since

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

the correction error and the centered clean-sample gradient are orthogonal in the conditional second moment, giving

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

This uses the explicitly stated amplitude condition

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

without silently treating $M(\theta)$ as globally bounded. `main`

The higher moments needed for the CLT are derived rather than used without proof. In A.2, centering eliminates $g(\theta; 0)$; together with the input tail condition, this yields

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

The independently established almost-sure convergence then gives a finite random bound along the trajectory. This supports the conditional Lindeberg verification without an additional global deterministic higher-moment bound. `main`
`main` Conditional variance convergence together with conditional Lindeberg is indeed an applicable set of martingale CLT conditions. Galton reference.

Convergence and the two CLTs: the key steps close

The basic strongly convex recursion

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

is correct. Robbins–Siegmund gives almost-sure convergence; scalar recursions separately give the mean-square rates. The argument does not confuse almost-sure convergence with mean-square convergence. At $\alpha = 1$, the summation calculations for $2\mu c > 1$, $2\mu c = 1$, and $2\mu c < 1$ are also valid. `main`

The gradient-term cancellation in the convex objective bound and the descent inequality in the nonconvex result match their stated step-size restrictions. I found no coefficient or condition error. main main

I checked the $\alpha = 1$ CLT particularly carefully because it is a potential source of substantive gaps. The current treatment is valid. The paper requires only

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

not the stronger global condition $2\mu c > 1$. Through stopping-time localization, Appendix A.5 chooses

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

near the optimum, obtains a localized $O(t^{-1})$ second-moment bound, and controls the nonlinear remainder. The subsequent proof therefore does not improperly invoke an $O(t^{-1})$ rate unsupported by the global assumptions. main

In the last-iterate CLT, the matrix weights, remainder estimates, and covariance limits agree. The c and c^2 coefficients in the two Lyapunov equations are also correct. main

The averaged-iterate analysis keeps the two step-size regimes distinct. For $\alpha < 1$, it obtains $H^{-1}S_*H^{-1}$; for $\alpha = 1$, it separately derives

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

and its integral covariance. The appendix also controls the weight tail as $u \downarrow 0$, so this is more than a formal replacement by a Riemann sum. main main

Extensions and examples: no omission affecting the main conclusions was found

The non-Gaussian extension does not assert all optimization conclusions merely from the existence of a conjugate variable. It distinguishes the conditions for unbiasedness, second moments, and higher moments:

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

For the uniform and symmetric two-point noise examples, these conditions match the tails of the stated conjugate distributions and support the corresponding arguments. main

The random response Y in nonlinear-feature regression is not overlooked. A.9 directly uses

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

to verify the moment conditions and relative variance bound for the joint data, rather than applying a fixed-response result without checking it. main I recomputed the population matrix under the experimental distribution and obtained

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

consistent with the reported $\mu \approx 0.382$ and the theoretical $O(t^{-1})$ rate.

The neural-network experiment explicitly acknowledges that a parameter-independent exponential growth exponent and the amplitude condition have not been verified; it therefore does not invoke the nonconvex main theorem. **This is an explicit scope limitation, rather than an unsupported theoretical guarantee imposed on the experiment.** main

Final opinion: under the assumptions as currently written, this mathematical version passes a major-gap review. I found no substantive flaw requiring changes to the principal theorem conclusions. This judgment concerns the mathematical arguments in the current manuscript; it does not cover the unavailable full companion manuscript or reproduction of the experimental code.