Unifying Distributed Optimization, Algorithmic Stability, and Privacy-Preserving Learning through Converse Lyapunov Theory

Guener Dilsad Er Michael Muehlebach
Max Planck Institute for Intelligent Systems, Tübingen, Germany

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Our approach quantifies the impact of disturbances on the performance of algorithms, offering fundamental insights for their theoretical analysis and understanding.

Summary

- **Objective:** Quantify disturbance-to-performance trade-offs for algorithms that converge linearly in a task-relevant metric.
- **Challenge:** Disturbances impact the rate of convergence.
  Dynamics are nonlinear.
- **Solution:** A systems-theoretic framework based on a rate-preserving converse Lyapunov function under mild regularity (Lipschitzness) and nominal convergence assumptions.
  Key mechanism: Construct a rate-matching Lyapunov function V where V is uniformly Lipschitz on compact sets
- **Impact:** Provides a unified, plug-and-play expression that
  (i) sharpens $\kappa-$dependence in event-triggered ADMM (accuracy vs. communication trade-off),
  (ii) recovers classical algorithmic stability/generalization bounds,
  (iii) matches optimal utility for noisy gradient descent.

Core Contributions

- Rate preserving Lyaponuv function:

$$V(\xi) := \sup_{k'\geq 0} \, \mathrm{d}(\phi(k',\xi), x_*)\, \tau^{-k'}.$$

V exists if the nominal algorithm dynamics are converging linearly at rate τ.
V is rate matching, which is non-standard in classical converse Lyapunov theory.

- Clear disturbance-to-performance bounds :

geometric decay + *a steady-state bias* $\propto$ *disturbance magnitude*

- Unified treatment across domains; avoids crafting Lyapunov functions for each algorithm

Main Result

Theorem: Let the unperturbed algorithm be represented by the nominal dynamics $x_{k+1} = f(x_k)$ and, satisfy the stability condition and the limit

$$\mathrm{d}\left(\phi(k', x_k), x_*\right) \leq c_0 \tau^{k'} \mathrm{d}\left(x_k, x_*\right), \qquad \lim_{k'\to\infty} \mathrm{d}\left(\phi(k', x_k), x_*\right) \tau^{-k'} = 0$$

with a linear rate τ. Then, there exists a constant $L_V > 0$ such that the following bound holds for the perturbed algorithm dynamics,

$$\mathrm{d}(z_k, x_*) \leq c_0 \tau^k \, \mathrm{d}(z_0, x_*) + L_V L_e \sum_{j=1}^{k} \tau^{k-j} |e_{j-1}|,$$

for all $z_0 \in \mathcal{S}$, where $\mathcal{S} \subset \mathbb{R}^n$ is compact, z_k denotes the state of the perturbed algorithm $z_{k+1} = g(z_k, e_k)$, e_k represents the disturbance, and $L_e > 0$ is the Lipschitz constant of disturbed dynamics with respect to e.

Scope, limitations, and assumptions

- **Linear convergence** required.
  – Extensions to sublinear/non-exponential rates are treated via time-varying rate schedules $\tau(i)$ in ongoing work [1].
- **Local Lipschitz continuity** in the disturbance (e) is assumed (excludes hard quantization and top-k without smoothing).
  – Regularized or randomized/smoothed operators (Lipschitz-in-expectation) can restore applicability.

Applications

Distributed Optimization with Event-triggered Communication

- For $\gamma-$strongly convex and $\beta-$ smooth objectives, ADMM admits a nominal linear rate $\tau = 1 - \frac{\alpha}{2\kappa^{\varepsilon+0.5}}$ (scaling in $\kappa = \beta/\gamma$ as in [2])
- Event-triggered updates yield $|e_k| \leq \Delta$, and $|\theta_k - \theta_*| \approx L_V L_e \Delta \kappa^{\epsilon+1/2}/\alpha$
- **Improved $\kappa-$dependence**: $\mathcal{O}(\kappa^{\frac{1}{2}+\epsilon}\Delta)$ vs $\mathcal{O}(\kappa^{1+\epsilon}\Delta)$ in prior analysis [2]

Algorithmic Stability and Generalization

- Reproduces known bound [3] $\epsilon_{stab} \leq \frac{2L_\ell^2}{\gamma n}$ under strong convexity and step size $h = \frac{2}{\beta+\gamma}$.
  Under convexity with step size $h_l \leq \frac{2}{\beta}$, the bound becomes $\epsilon_{\text{stab}} \leq \frac{2L_\ell^2}{n} \sum_{l=1}^{k} h_l$.
- Stability emerges from the same Lyapunov template
- Change in one data point (D, D') → bounded disturbance → stability gap

Privacy-Preserving Learning (Noisy Gradient Descent)

- With additive noise of variance σ_n^2 and strong convexity, choosing step size $h = \frac{\log(N)}{2\beta N}$ yields $\mathbb{E}\left\{|\theta_N^{\text{priv}} - \theta_*|^2\right\} = \mathcal{O}\left(\frac{\sigma^2 \log(N)}{N}\right)$
  Under convexity, with $h = \frac{2}{\beta\sqrt{N+1}}$, the bound becomes $\mathbb{E}\{\ell(\theta_N^{\text{priv}}) - \ell(\theta_*)\} \approx \mathcal{O}\left(\sigma_n^2\right)$.

Broader Connections

Input to State Stability

ISS quantifies how disturbances bound the state of a system. Our disturbance–to–performance estimate is ISS-type with explicit decay and gain.

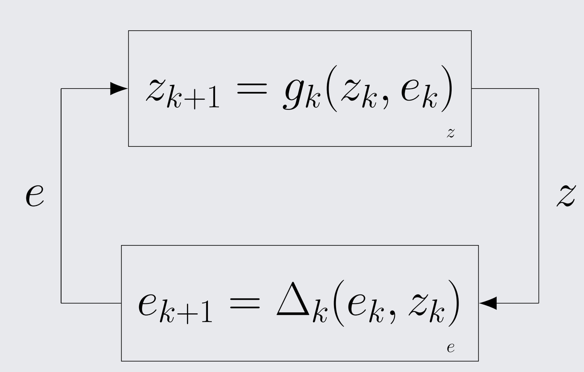
$$\mathrm{d}(z_k, x^*) \leq c_0 \tau^k \mathrm{d}(z_0, x^*) + \frac{L_V L_e}{1-\tau} \left(\sup_{j\leq k} |e_j|_\infty \right)$$

- Compared to classical ISS, our converse Lyapunov construction preserves the nominal rate and yields a quantitative steady-state "gain" $\sim (L_V L_e)/(1-\tau)$.

Performative Prediction

Algorithms often shape their environments; in performative prediction, the environment responds via a map Δ. This is modeled as a state-dependent disturbance,

$$e_k = \Delta(z_k) \quad \text{(static response)} \quad \text{or} \quad e_{k+1} = \Delta_k(e_k, z_k) \quad \text{(dynamic response)}.$$



- Performative prediction is a special case of our interconnected model; the same Lyapunov template certifies stability without constructing problem-specific Lyapunov functions.
- Continuity of Δ ensures existence of a performative fixed point (Brouwer). If Δ is Lipschitz with sufficiently small gain (relative to $L_V L_e$ and $(1-\tau)$), trajectories converge to this fixed point.

Paper

Prior Work [2]

References

[1] G. Dilsad Er, Sebastian Trimpe, and Michael Muehlebach. "A Systems-Theoretic View on the Convergence of Algorithms under Disturbances". In preparation.
[2] G. Dilsad Er, Sebastian Trimpe, and Michael Muehlebach. "Distributed Event-Based Learning via ADMM". In: *International Conference on Machine Learning* 267 (2025), pp. 15384–15418.
[3] Moritz Hardt, Benjamin Recht, and Yoram Singer. "Train faster, generalize better: Stability of stochastic gradient descent". In: *International Conference on Machine Learning* 48 (2016), pp. 1225–1234.