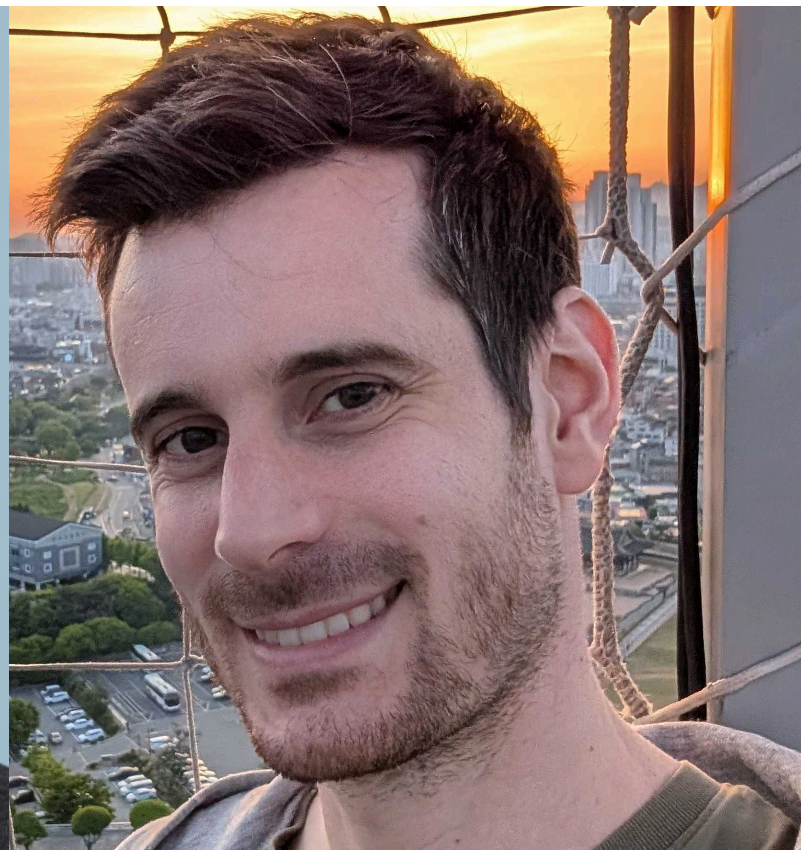
From Lyapunov Analysis to Algorithm Design in two-sided PL Minimax Optimization

Mansi Rankawat, Michael Muehlebach, Simon Lacoste-Julien, Damien Scieur



ICML 2026

Mila, Université de Montréal, MPI-IS, Samsung AI Lab

Minmax optimization

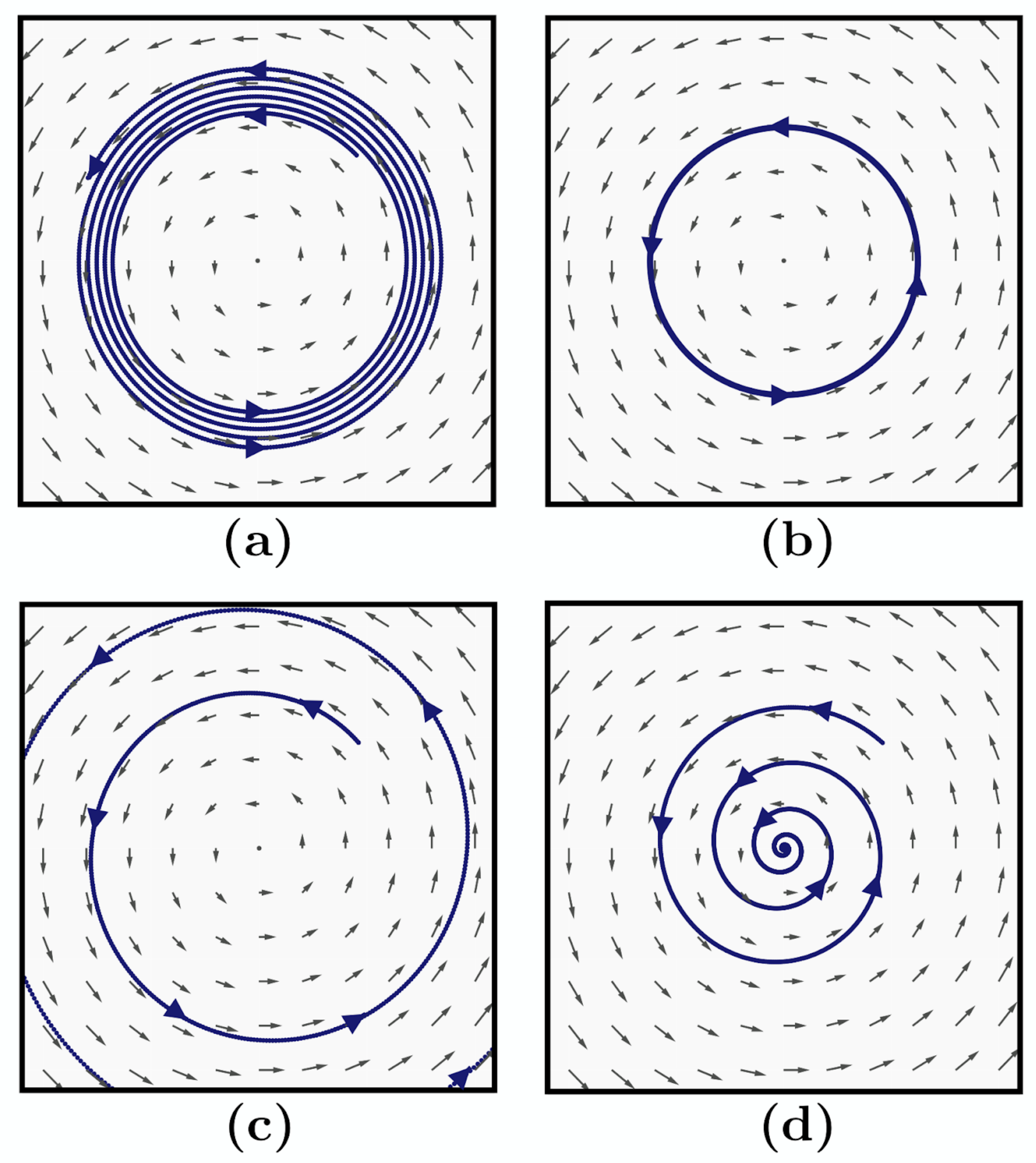
$$\min_{\mathbf{x}} \max_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$$

Goal: Find $\mathbf{x}_k$ and $\mathbf{y}_k$ such that

$$\max_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}) - \min_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_k) \leq \epsilon$$

- **Examples: GANs, Robust optimization, constrained optimization**

Minmax optimization



(a)	Simultaneous GDA,	(b)	Alternating GDA
(c)	Alternating GDA + momentum,	(d)	Negative momentum

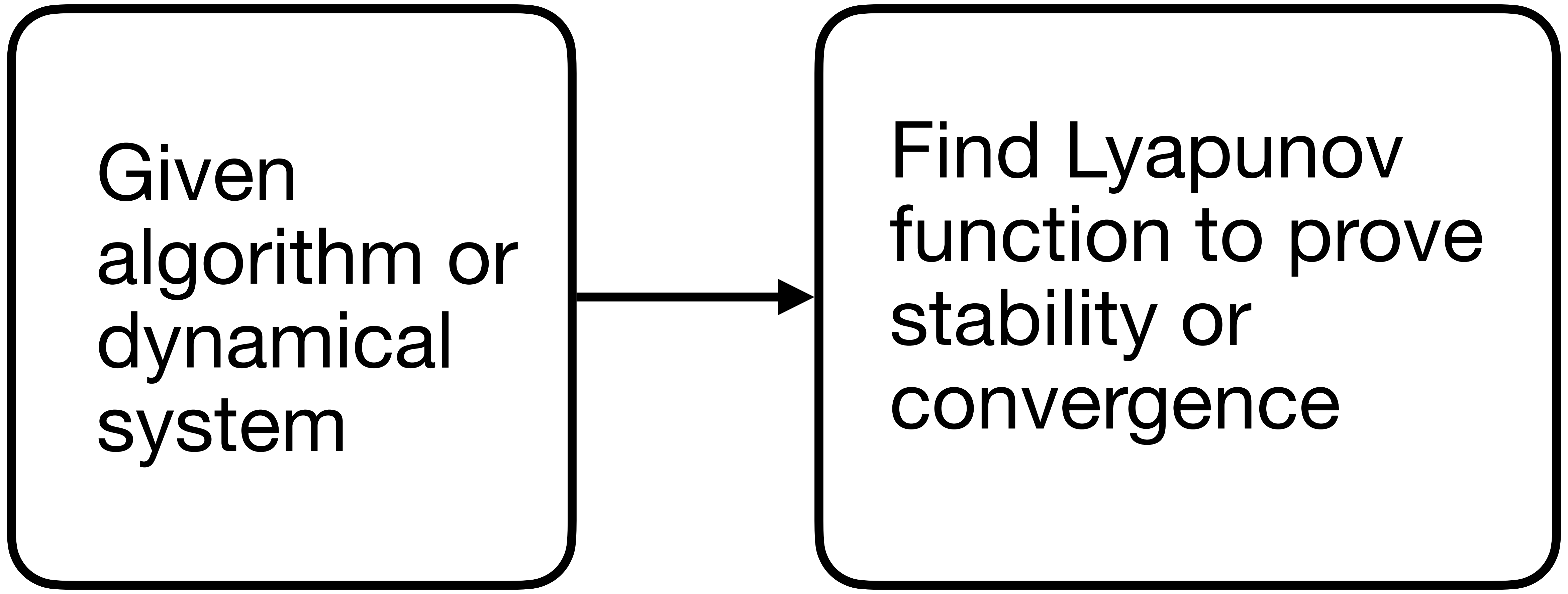
- **Complicated dynamics:** convergence in min-max demand non-intuitive updates, and intuitive updates often diverge.
- **The Challenge:** Principled framework for algorithm design in Min-Max

Gidel et al., Negative Momentum for Improved Game Dynamics

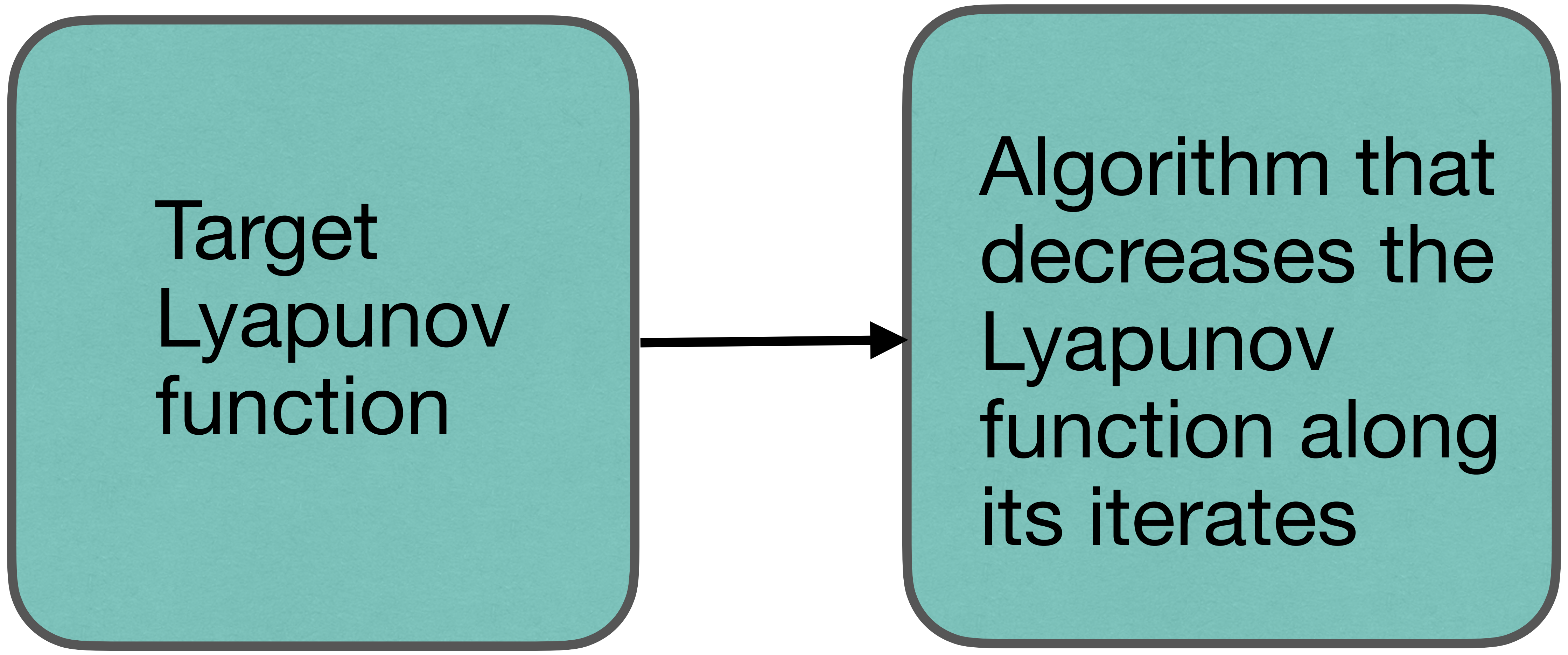
Lyapunov functions

- Lyapunov functions: Generalized energy for dynamical systems.

Lyapunov function to analyze stability and convergence



Our approach



Idealized Lyapunov function for minimax problems

$$V^{Exact}(\mathbf{x}_k, \mathbf{y}_k) = \overbrace{\max_{\mathbf{y}} f(\mathbf{x}_k, \mathbf{y}) - \min_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}_k)}^{\text{primal-dual gap}}$$

- Not exactly computable

Assumption: Two-sided Polyak-Lojaseiwicz (PL) inequality

$$f(\mathbf{x}, \mathbf{y}) - \min_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) \leq \frac{1}{2\mu} \|\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})\|^2$$

$$\max_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) - f(\mathbf{x}, \mathbf{y}) \leq \frac{1}{2\mu} \|\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})\|^2$$

Computable Lyapunov function

- Using two-sided PL inequality, we have

$$0 \le V^{Exact}(\mathbf{x}, \mathbf{y}) \le V(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}, \tilde{\mathbf{y}})$$

where

$$V(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = f(\mathbf{x}, \tilde{\mathbf{y}}) - f(\tilde{\mathbf{x}}, \mathbf{y}) + \frac{1}{2\mu}\|\nabla_{\mathbf{x}} f(\tilde{\mathbf{x}}, \mathbf{y})\|^2 + \frac{1}{2\mu}\|\nabla_{\mathbf{y}} f(\mathbf{x}, \tilde{\mathbf{y}})\|^2$$

- Introduce two new variables $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$

If $\tilde{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})$ and $\tilde{\mathbf{y}} = \operatorname{argmax}_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$, then $V(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = V^{Exact}(\mathbf{x}, \mathbf{y})$

Algorithm design

- Goal is to find updates for $\mathbf{x}_{k+1}$, $\mathbf{y}_{k+1}$, $\tilde{\mathbf{x}}_{k+1}$, and $\tilde{\mathbf{y}}_{k+1}$, such that

$$V(\mathbf{x}_{k+1}, \mathbf{y}_{k+1}, \tilde{\mathbf{x}}_{k+1}, \tilde{\mathbf{y}}_{k+1}) \leq (1-r)V(\mathbf{x}_k, \mathbf{y}_k, \tilde{\mathbf{x}}_k, \tilde{\mathbf{y}}_k) + \text{ some negative terms}$$

for some rate $r \in (0,1)$

Algorithm: TALDA-Tri Action Lyapunov Descent Ascent

- **Single loop algorithm**, thanks to Lyapunov approach

Comparison

Alternating gradient descent ascent:

$$T = \kappa_{\mathbf{x}}\kappa_{\mathbf{y}}^2 \log(\epsilon^{-1})$$

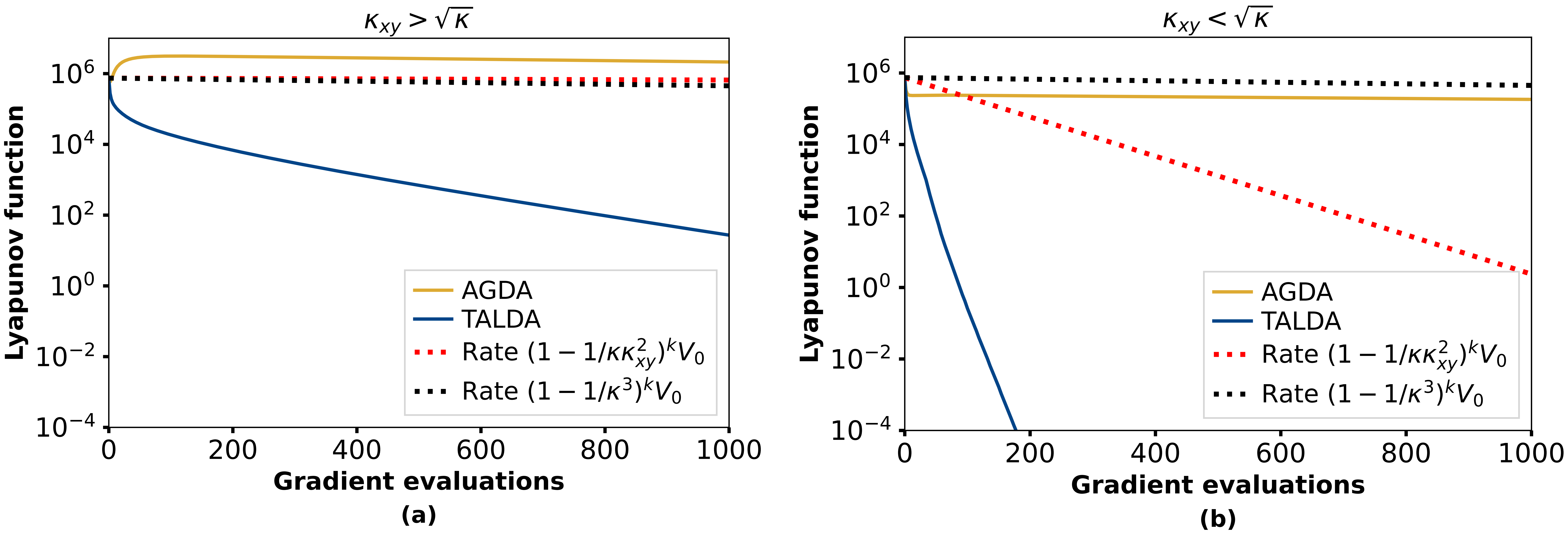
TALDA:

$$T = O(\kappa_f(\kappa_f + \kappa_{\mathbf{xy}}^2)\log^2(\epsilon^{-1}))$$

TALDA is better than AGDA

when $\kappa_{\mathbf{xy}} \leq \sqrt{\kappa_f}$

Numerical experiments (Synthetic)



$$\min_{\mathbf{x}} \max_{\mathbf{y}} \frac{1}{2}\mathbf{x}^T A \mathbf{x} + \mathbf{y}^T B \mathbf{x} - \frac{1}{2}\mathbf{y}^T C \mathbf{y} + d^T \mathbf{x} - e^T \mathbf{y}$$

Numerical experiments (Robust Least Squares)

Theoretical learning rates

Finetuned learning rates

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \|A\mathbf{x} - \mathbf{y}\|_M^2 - \lambda \|\mathbf{y} - \mathbf{y}_0\|_M^2$$

Key Takeaways

- **Lyapunov functions as a design tool.**
- **TALDA:** Single loop first order method with linear convergence under smooth two sided PL setting.
- **Swap procedure** as a result of using Lyapunov functions as design tool.
- **Coupling matters**: Rate scales with cross-smoothness not worst case condition number.

Please check out our paper and visit our poster on July 8th 2:30 pm - 4:15 pm