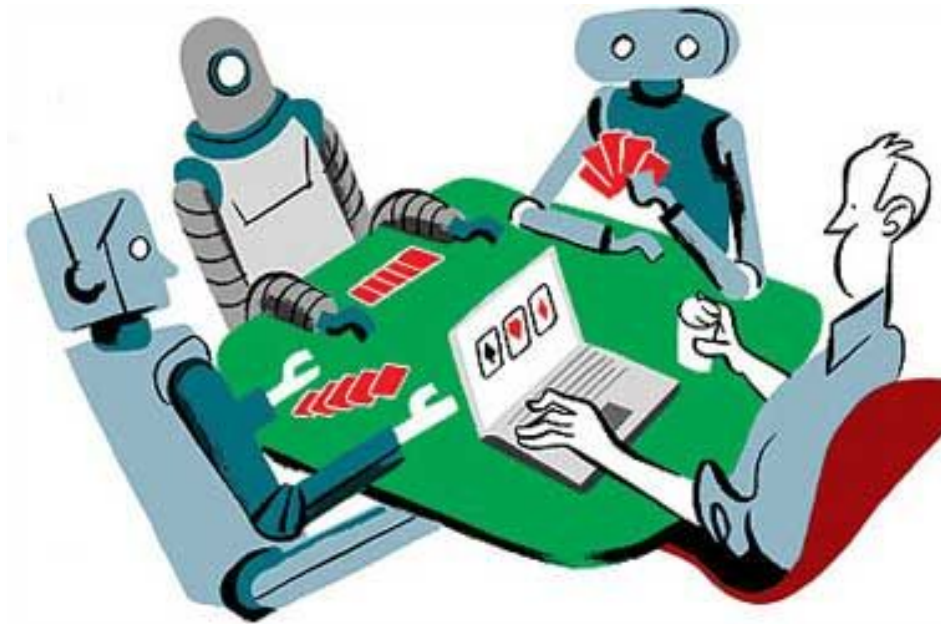


On the $O(1/T)$ Convergence of Alternating Gradient Descent-Ascent in Bilinear Games

Tianlong Nan , **Shuvomoy Das Gupta** , **Garud Iyengar** , **Christian Kroer**
Columbia University IEOR , **Rice University CMOR**

ICLR 2026



Applications:
Superhuman Poker AI

Bilinear Two-player Zero-sum Games

Payoff: Bilinear

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m}$$



Decision Sets: Probability Simplex

$$y^T A x$$

Bilinear Two-player Zero-sum Games

Payoff gradient for x-player

$$\min_{x \in \Delta_n} \max_{y \in \Delta_m}$$



x-player: minimizer

$$y^T A x$$

Bilinear Two-player Zero-sum Games

y-player: maximizer



min
 $x \in \Delta_n$

max
 $y \in \Delta_m$

$$y^T \underbrace{Ax}$$

Payoff gradient for y-player

Nash Equilibrium (NE)

$$(y^*)^T A x^* = \min_{x \in \Delta_n} (y^*)^T A x$$

$$(y^*)^T A x^* = \max_{y \in \Delta_m} y^T A x^*$$

Gradient Descent-Ascent (GDA)

minimizer

$$\mathbf{x}^{t+1} = \text{Proj}_{\Delta_n} (\mathbf{x}^t - \eta \mathbf{A}^T \mathbf{y}^t)$$

maximizer

$$\mathbf{y}^{t+1} = \text{Proj}_{\Delta_m} (\mathbf{y}^t + \eta \mathbf{A} \mathbf{x}^t)$$

Alternating Gradient Descent-Ascent (AltGDA)

$$\mathbf{x}^{t+1} = \text{Proj}_{\Delta_n} (\mathbf{x}^t - \eta \mathbf{A}^T \mathbf{y}^t)$$

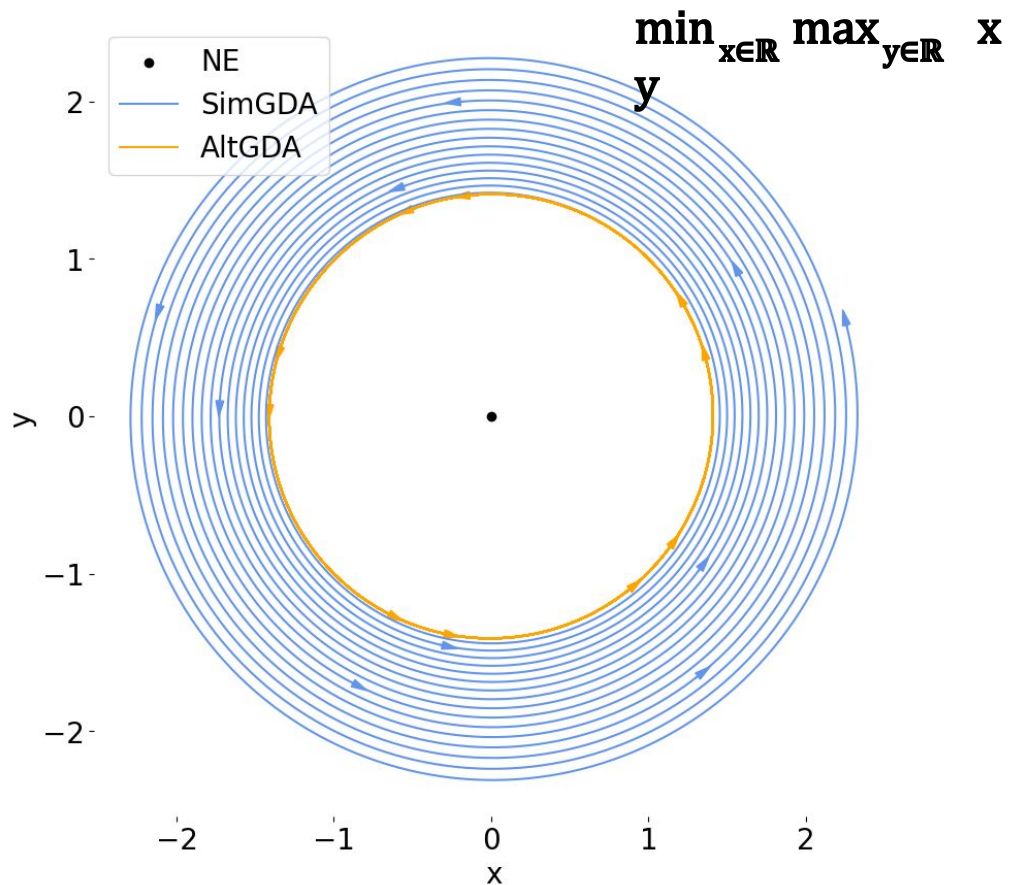
$$\mathbf{y}^{t+1} = \text{Proj}_{\Delta_m} (\mathbf{y}^t + \eta \mathbf{A} \mathbf{x}^{t+1})$$

Simultaneous Gradient Descent-Ascent (SimGDA)

$$\mathbf{x}^{t+1} = \text{Proj}_{\Delta_n} (\mathbf{x}^t - \eta \mathbf{A}^T \mathbf{y}^t)$$

$$\mathbf{y}^{t+1} = \text{Proj}_{\Delta_m} (\mathbf{y}^t + \eta \mathbf{A} \mathbf{x}^t)$$

AltGDA vs. SimGDA: Trajectory Behaviors



```
SimGDA = []  
x, y = 1, 1  
SimGDA.append((x, y))  
for _ in range(10000):  
    x_, y_ = x, y  
    x = x - 0.01 * y_  
    y = y + 0.01 * x_  
    SimGDA.append((x, y))
```

```
AltGDA = []  
x, y = 1, 1  
AltGDA.append((x, y))  
for _ in range(10000):  
    x = x - 0.01 * y  
    y = y + 0.01 * x  
    AltGDA.append((x, y))
```

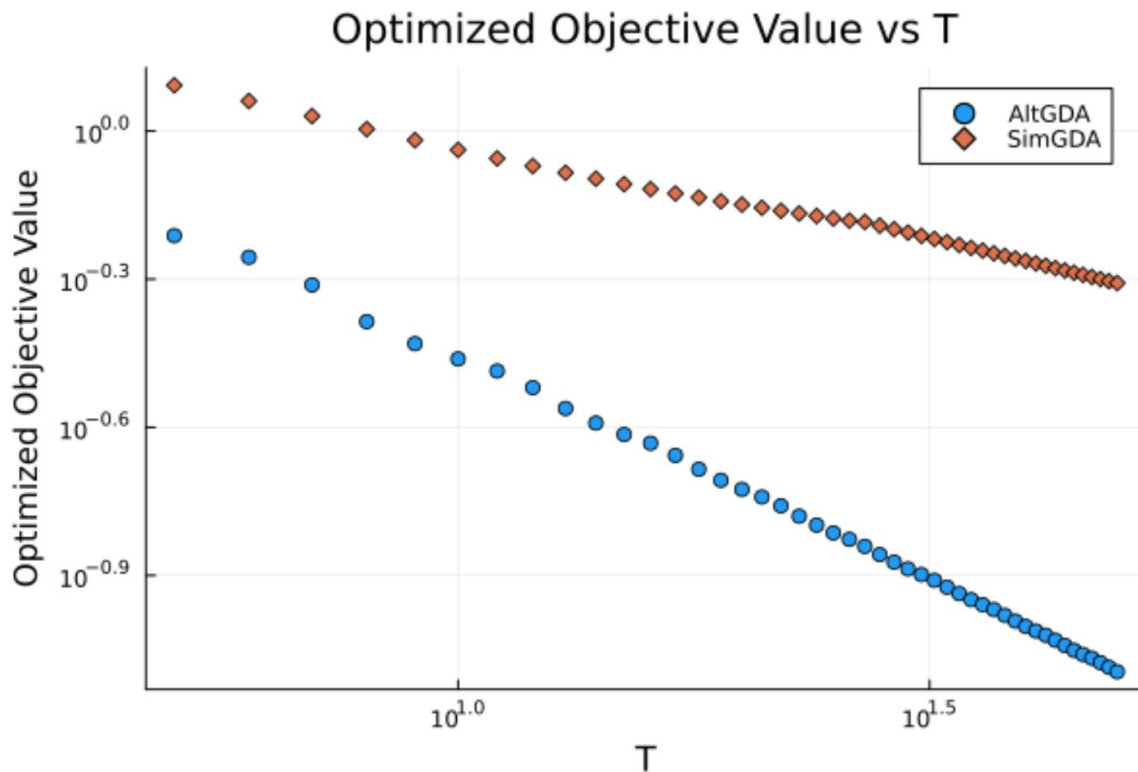
Theoretical Studies

- Bailey, J.P., Gidel, G. and Piliouras, G. (2020): In the unconstrained minimax problems, AltGDA with a fixed stepsize converges at a rate of $O(1/T)$
- Open Problem: Can we show a rate of $O(1/T)$ for AltGDA in constrained setting?
- Surprising because $O(1/T)$ convergence rate usually requires more complex techniques (e.g., optimism)

Performance Estimation Programming (PEP)

- Model worst-cases as solutions to optimization problems
- Those optimization problems are tractable for first-order methods in convex optimization
- In our case, for a fixed iteration budget T , by solving a semi-definite program (SDP), we can find the worst-case convergence rates over all possible instances

AltGDA vs. SimGDA: PEP Worst-case Analysis



Main Theoretical Results

Theorem 1. Assume that the bilinear game admits an interior NE. Let $\{(\mathbf{x}^t, \mathbf{y}^t)\}_{t=0,1,\dots}$ be a sequence of iterates generated by Algorithm 1 with $\eta \leq \frac{1}{\|A\|_2} \min\{\min_{i \in [n]} x_i^*, \min_{j \in [m]} y_j^*\}$. Then, we have

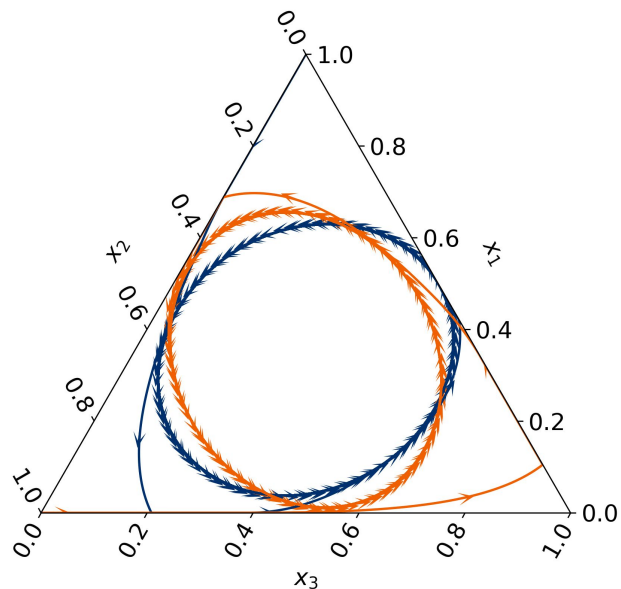
$$\text{DualityGap} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}^t, \frac{1}{T} \sum_{t=1}^T \mathbf{y}^t \right) \leq \frac{9 + 4\eta\|A\|_2}{\eta T}.$$

Theorem 2. Let $\{(\mathbf{x}^t, \mathbf{y}^t)\}_{t \geq 0}$ be a sequence of iterates generated by Algorithm 1 with stepsize $\eta \leq \frac{1}{2\|A\|_2}$ and an initial point $(\mathbf{x}^0, \mathbf{y}^0) \in S_0$, where S_0 is defined in Eq. (7). Then, we have that

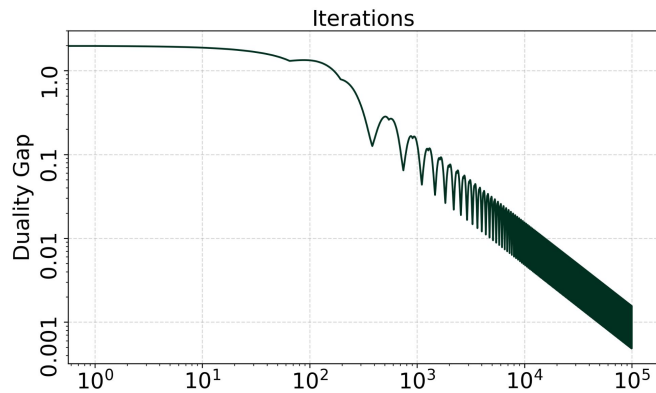
$$\text{DualityGap} \left(\frac{1}{T} \sum_{t=1}^T \mathbf{x}^t, \frac{1}{T} \sum_{t=1}^T \mathbf{y}^t \right) \leq \frac{9 + 7\eta\|A\|_2 + (\delta^2/128)}{\eta T},$$

where δ is defined in Eq. (6).

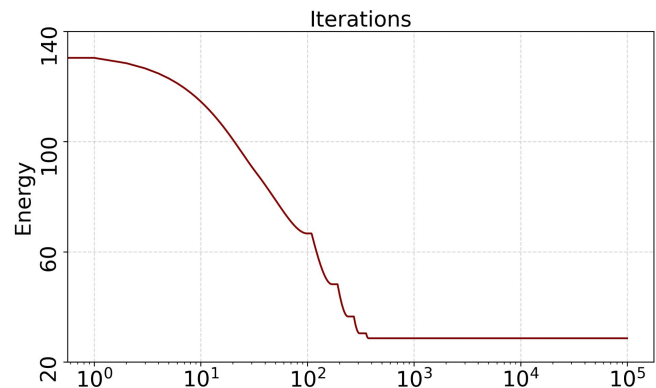
Key Observations



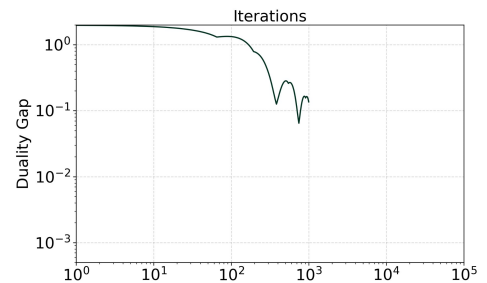
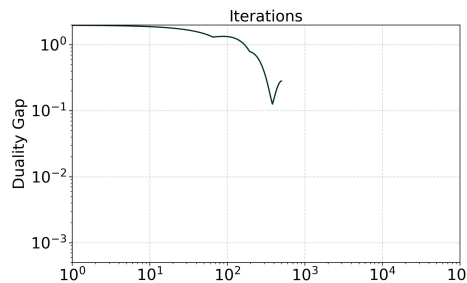
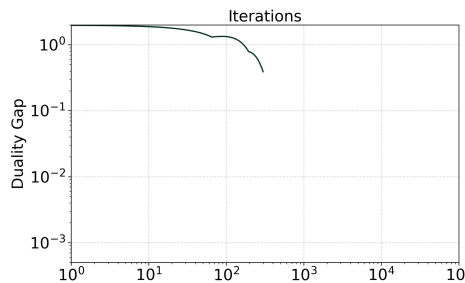
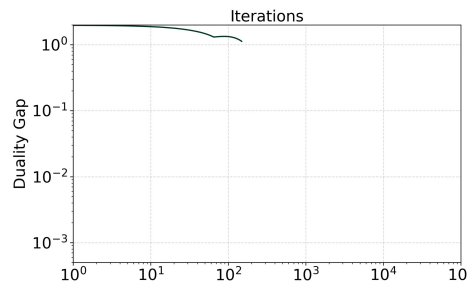
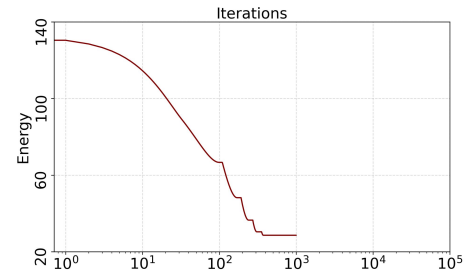
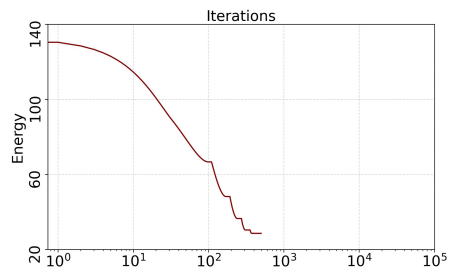
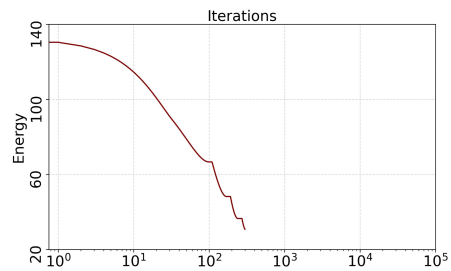
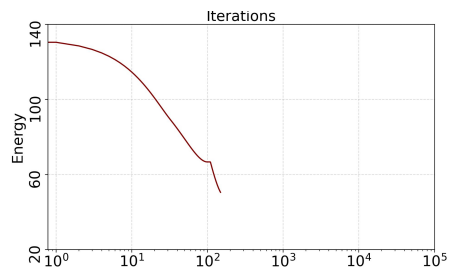
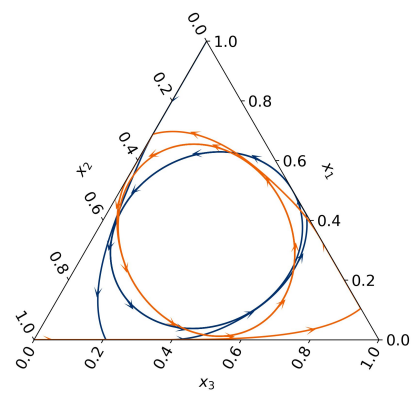
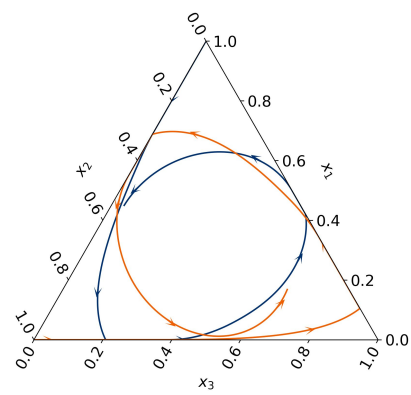
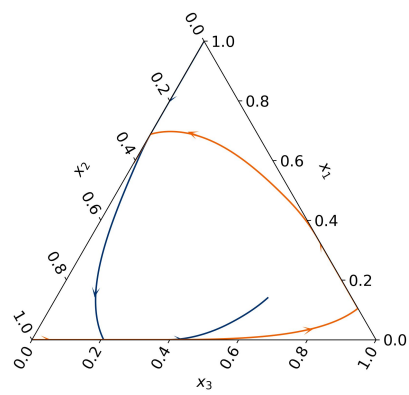
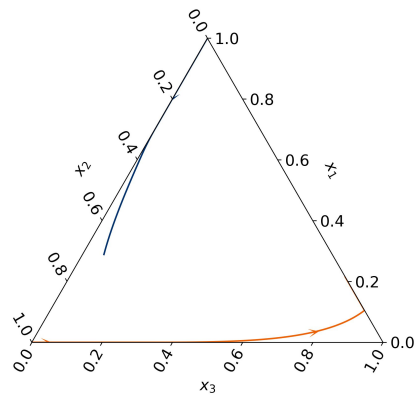
Trajectories



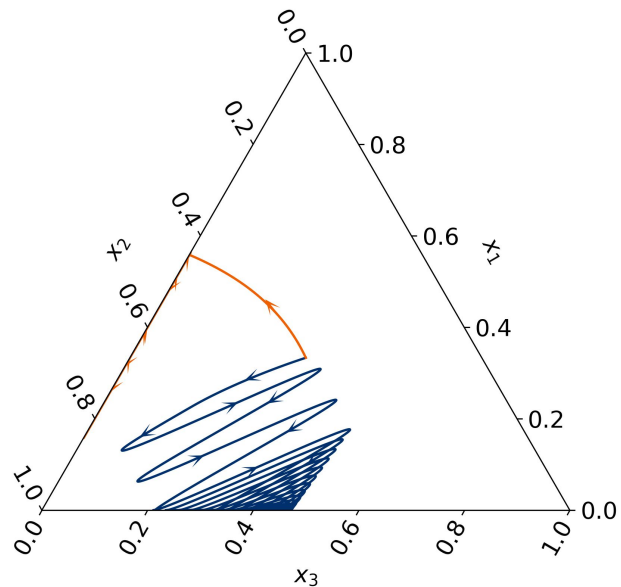
Convergence Rate



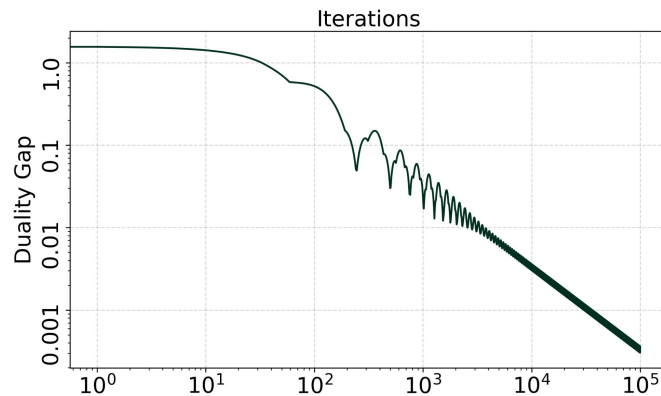
An Energy Function



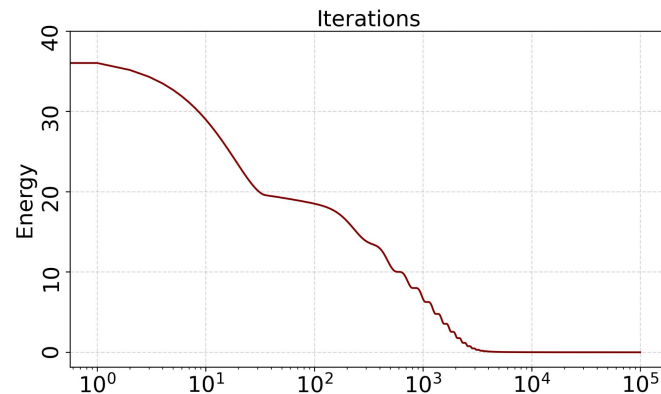
Key Observations



Trajectories



**Duality
Gap**



**An Energy
Function**

Thank you for listening!

Any discussion would be welcome!