Online Multi-Agent Control with Adversarial Disturbances

Anas Barakat

Joint work with John Lazarsfeld, Georgios Piliouras, Antonios Varvitsiotis

February 10th 2026 - Machine Learning and Dynamical Systems Symposium - Kyoto

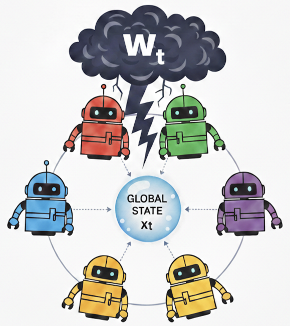
SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Multi-Agent Linear Dynamical Systems

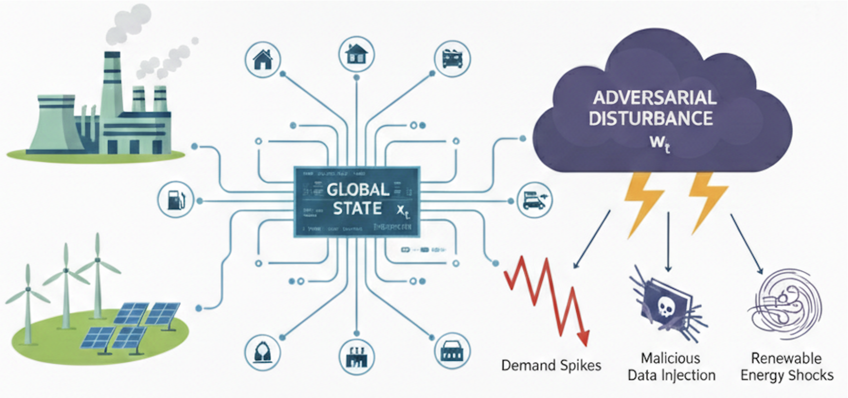
State Evolution

$$x_{t+1} = Ax_t + B_1 u_t^1 + \cdots + B_N u_t^N + w_t$$

- x_t: global state
- $(u_t^i)_{i \in \{1,\cdots,N\}}$: control inputs
- A and $(B_i)_{i \in \{1,\cdots,N\}}$: transition matrices
- w_t: *adversarial* disturbance



Example 1: Energy Grid Markets

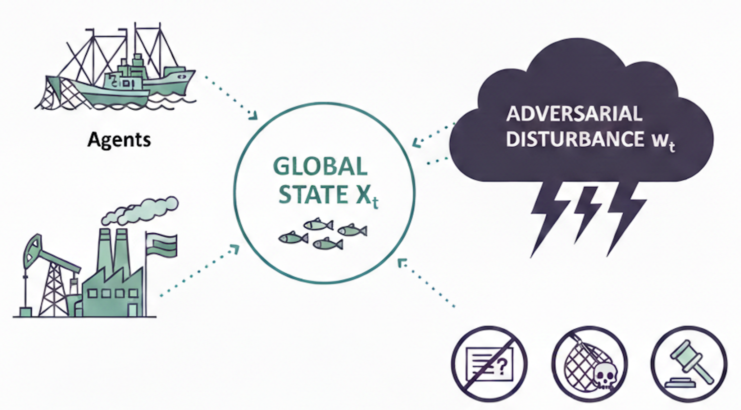


- **Adversarial disturbances:** strategic demand manipulation by large consumers
- **Costs:** local cost (e.g. fuel) + penalty for deviating from target state

Example 2: Formation Control

- **Adversarial disturbances:** wind gusts, magnetic interference, sensor spoofing
- **Costs:** formation error (maintained distance) and energy consumption

Example 3: Bioresource management



- **Adversarial disturbances:** misinformation about resource levels, illegal over-harvesting, policy shocks (e.g. sudden trade bans)
- **Costs:** max revenue while min exploitation cost

Online Control Setting

Multi-Agent LDS

$$x_{t+1} = Ax_t + B_1 u_t^1 + \cdots + B_N u_t^N + w_t$$

Online Setting

At each time step t, each agent $i \in \{1, \cdots, N\}$:

- observes the state x_t,
- selects a control input u_t^i mapping states to controls,
- incurs an individual time-varying cost $c_t^i(x_t, u_t^i)$

Goal and Challenges

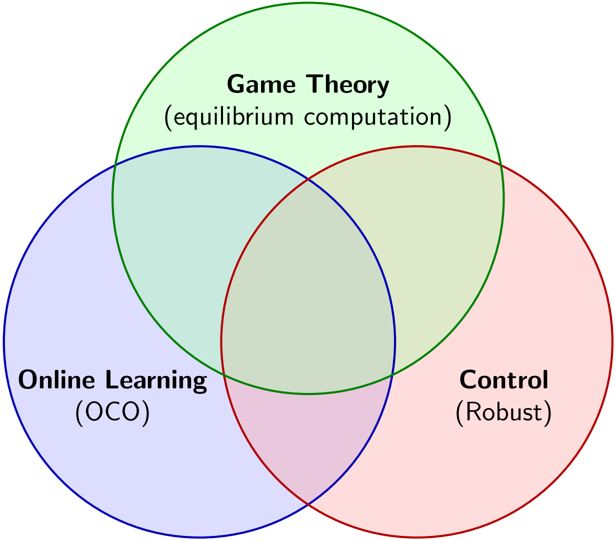
Goal

decentralized online control algorithms under adversarial disturbances

- Individual (per-agent) guarantees?
- Collective (equilibrium) behavior?

- **Decentralization:** Agents act locally without access to others' policies.
- **Scaling:** Dependence on the number of agents in the system?
- **Equilibrium behavior:** Global equilibrium tracking under identical interest?

Prior Work and Positioning



Related Work

- Online non-stochastic control
[Agarwal et al., 2019, Hazan et al., 2020, Foster and Simchowitz, 2020, Simchowitz et al., 2020, Simchowitz, 2020, Gradu et al., 2020, Ghai et al., 2023, Cai et al., 2024, Tsiamis et al., 2024, Golowich et al., 2024].
- Multi-agent control
  - Control $\cap$ game theory [Marden and Shamma, 2018, Chen and Ren, 2019]
  - linear-quadratic games [Başar and Olsder, 1998, Zhang et al., 2019, Bu et al., 2019, Zhang et al., 2021, Mazumdar et al., 2020, Hambly et al., 2023] ...
  - **mainly QUADRATIC COSTS + PROBABILISTIC or MINMAX (worst-case)**
- Online setting
  - *distributed* control [Chang and Shahrampour, 2023] $\neq$ *strategic* agents $+$ 1 LDS.
  - online control for population dynamics [Golowich et al., 2024].

(Single-Agent) Online Control: A Brief Overview

$$\min_{u(x)} \sum_{t=1}^{T} c_t(x_t, u_t)$$

$$\text{s.t.} \quad x_{t+1} = A_t x_t + B_t u_t + w_t$$

x_t : state,

u_t : control input,

w_t : perturbation.

From Optimal to Online Control

- LQR – Gaussian noise & quadratic costs only: $u_t = Kx_t$ (where $K(A, B)$).
- H_∞-control:
  $$\min_{K_{1:T}} \max_{\|w_{1:T}\|_2 \leq C} \sum_t c_t(u_t, x_t)$$
  Pessimistic, computationally ill-behaved for non-quadratics (even convex costs!), non-adaptive

What is missing?

1. Adaptive performance metric to handle adversarial costs and disturbances.
2. Efficient methods for general losses for which optimal policy is complicated.

Motivation for Online Control

- Flying drone from S to D/ unknown weather
  - Optimal control: optimistic (stochastic)
  - Robust control: pessimistic (worst-case)
- Online Control: instance-optimal, i.e., fast when possible, careful otherwise.

Online control of LDS

- **Online setting**, $t = 1, \ldots, T$:
  - Select input $u_t \in \mathbb{R}^n$
  - Observe x_t, incur loss: $c_t(u_t, x_t)$
- Performance metric: **POLICY REGRET**

$$\max_{w_{1:T}} \left(\sum_{t=1}^{T} c_t(x_t, u_t) - \min_{\pi \in \Pi} \sum_{t=1}^{T} c_t(x_t^\pi, u_t^\pi) \right)$$

- x_t^π = **counterfactual** state under $u_t^\pi = \pi(x_t^\pi)$, $x_{t+1}^\pi = Ax_t^\pi + Bu_t^\pi + w_t$
- Bounded noise: $\|w_t\| \leq W$

What is a reasonable policy comparator class?

- For Gaussian LDS with quadratic costs: linear policies are optimal (Riccati)

$$\Pi_K = \{\pi_K \mid u_t = Kx_t\}$$

- **More general:** Disturbance-Action Controllers:

$$\Pi_{\mathrm{DAC}} = \left\{ \pi_{M_{1:H}} \;\middle|\; u_t = \sum_{i=1}^{H} M_i w_{t-i} \right\}$$

Single-Agent Online Control Guarantee [Agarwal et al., 2019]

Agarwal, N., Bullins, B., Hazan, E., Kakade, S., Singh, K. *Online control with adversarial disturbances.* ICML 2019.

Online Control Regret Guarantee

Efficient algorithm s.t.

$$\sum_{t=1}^{T} c_t(x_t, u_t) \; - \; \min_{\pi \in \Pi} \left(\sum_{t=1}^{T} c_t(x_t^{\pi}, u_t^{\pi}) \right) \; \leq \; O(\sqrt{T}) \tag{1}$$

- Efficient $\rightarrow$ Polynomial in system parameters, logarithmic in T

Main Argument: Convex Relaxation of Π_K

$$\min_K \sum_{t=1}^{T} c_t(x_t, u_t) \quad \text{s.t.} \quad x_{t+1} = Ax_t + Bu_t + w_t, \quad u_t = Kx_t$$

Unrolling the control law gives:

$$u_{t+1}(K) = Kx_{t+1} = \sum_{i=0}^{t} K(A+BK)^i w_{t-i} \Rightarrow \text{ optimization is non-convex in } K.$$

Relaxation (Disturbance-Action parameterization):

$$u_{t+1}(M) = \sum_{i=0}^{H} M_i\, w_{t-i}, \quad H = \mathcal{O}(\log T)$$

Relaxed (convex in M!) problem:

$$\min_{M \in \mathcal{M}} \sum_{t=1}^{T} c_t\big(x_t(M), u_t(M)\big)$$

Gradient Perturbation Controller

Input: memory H, step size η, $M = M_1, \dots, M_H$

Compute a stabilizing linear controller K knowing (A, B)

For $t = 1, \dots, T$ **do**

1. Use control $u_t = Kx_t + \sum_{i \le H} M_i w_{t-i}$
2. Observe state x_{t+1}, compute noise $w_t = x_{t+1} - Ax_t - Bu_t$
3. Construct cost function: $\ell_t(M) = c_t\big(x_t(M_{1:H}), u_t(M_{1:H})\big)$
4. Update $\bar{M}$: $M \leftarrow M - \eta \nabla_M \ell_t(M)$

Information Settings: What do agents have access to?

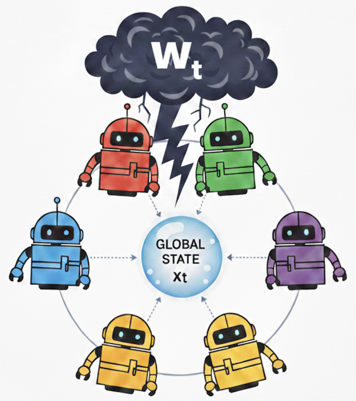
Information Setting 1: Independent Learning

At each time step t, agent $i \in [N]$ observes only the state x_t and their own cost.

(no access to control inputs of other agents $j \neq i$).

Information Setting 2: Aggregated Control

Information Setting 1
$+$ **aggregated feedback** $\sum_{j \neq i} B_j u_t^j$.



Performance Metric: Regret for Online Multi-Agent Control

Regret for agent i

$$\mathrm{Reg}_i^T(\mathcal{A}_i, \{u_t^{-i}\}, \Pi_i) = \sum_{t=0}^{T} c_t^i(x_t, u_t^i) - \min_{\pi^i \in \Pi_i} \sum_{t=0}^{T} c_t^i(x_t^{\pi^i}, u_t^{\pi^i})$$

- $\mathcal{A}_i$: learning algorithm used by the i'th agent to select its control u_t^i.
- $(x_t^{\pi^i}, u_t^{\pi^i})$: *counterfactual* state-control pair ($\{u_t^{-i}\}$ fixed), i.e.,

$$u_t^{\pi^i} = \pi^i(x_t^{\pi^i}), \quad x_{t+1}^{\pi^i} = A x_t^{\pi^i} + B_i u_t^{\pi^i} + \sum\nolimits_{j \neq i} B_j u_t^j + w_t \,.$$

Online Gradient Perturbation Controller (Agent $i \in [N]$)

Input: memory H, step size η, initialization $M_{i,1}^{[0:H-1]}$.

Compute a stabilizing linear controller K_i knowing (A, B_i).

For $t = 1, \ldots, T$ **do**

1. Observe state x_t.
2. Update under Info. Setting 2.1:
   $\tilde{w}_{t-1} = x_t - Ax_{t-1} - B_i u_{t-1}^i$.
   $u_t^i = K_i x_t + \sum_{p=1}^{H} M_{i,t}^{[p]} \tilde{w}_{t-p}$.

2′. Update under Info. Setting 2.2:
   Observe $\sum_{j \neq i} B_j u_{t-1}^j$.
   $w_{t-1} = x_t - Ax_{t-1} - \sum_{k=1}^{N} B_k u_{t-1}^k$.
   $u_t^i = K_i x_t + \sum_{p=1}^{H} M_{i,t}^{[p]} w_{t-p}$.

3. Record instantaneous cost $c_t^i(x_t, u_t^i)$.
4. Construct loss $\ell_t^i(M_i) = c_t^i\big(y_t^{K_i}(M_i), v_t^{i,K_i}(M_i)\big)$.
5. Update $M_{i,t+1} = \Pi_{\mathcal{M}_i}\left[M_{i,t} - \eta \nabla \ell_t^i(M_{i,t})\right]$.

end for

Individual Regret Guarantee - Independent Learning

From the viewpoint of a given agent i, (5) can be rewritten:

$$x_{t+1} = Ax_t + B_i u_t^i + \widetilde{w}_t \,, \quad \widetilde{w}_t = \sum\nolimits_{j \neq i} B_j u_t^j + w_t \,.$$

- Agent i executes a DAC policy with disturbance sequence $\widetilde{w}_t$.

Theorem - Independent Learning

Assumption: each $i \in [N]$ knows a strongly stable linear controller K_i for LDS (A, B_i),

$$\mathrm{Reg}_i^T(\mathcal{A}_i, \{u_t^{-i}\}, \Pi_i^{\mathrm{lin}}) = \widetilde{\mathcal{O}}(N^2\sqrt{T})$$

- Lower bound: $\sqrt{T}$ is tight.
- What about dependence on number of agents N?

Improved Regret Guarantee - Aggregated Control Learning

- Under stronger feedback model ACL, $\sum_{j \neq i} B_j u_t^j$ accessible.
- Agent i executes a DAC policy with disturbance sequence w_t.

Theorem - Aggregated Control Learning

Assumption: each $i \in [N]$ knows a linear controller K_i s.t. $(K_1, \cdots, K_N)^T$ is strongly stable for LDS $(A, [B_1, \cdots, B_N])$ and all other agents use a DAC policy w.r.t. (w_t).

$$\mathrm{Reg}_i^T(\mathcal{A}_i, \{u_t^{-i}\}, \Pi_i^{\mathrm{DAC}}) = \widetilde{\mathcal{O}}(N\sqrt{T}).$$

- Under a stronger Lipschitzness assumption, $\mathrm{Reg}_i^T = \widetilde{\mathcal{O}}(\sqrt{T})$.

Equilibrium Tracking

Global guarantees when all agents run the same algorithm independently

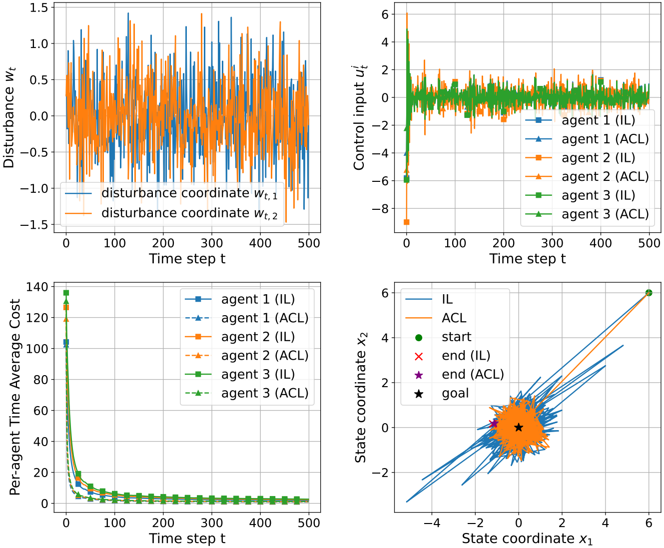
- All cost functions identical (i.e., $c_t^i = c_t^j := c_t$ for any $i, j \in [N]$ for every t).
- **Time-varying game**

Equilibrium Tracking Guarantee

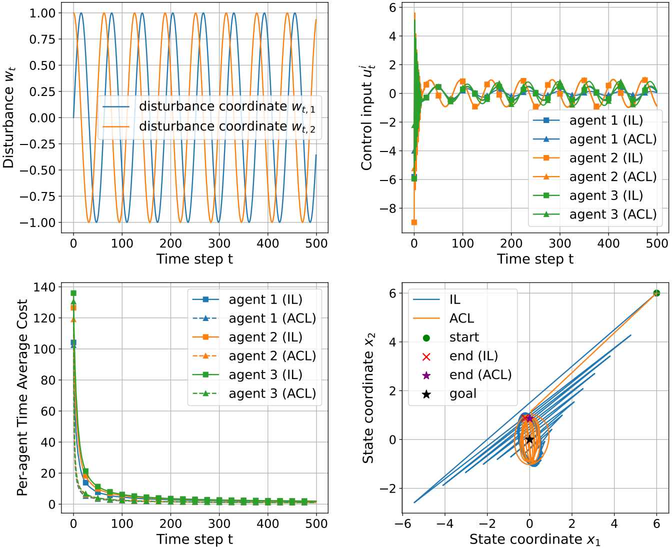
$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathrm{EQGAP}^{(t)}(M_t)\right)^2 = \mathcal{O}\left(\frac{\ell_1(M_1) - c_{\mathrm{inf}}}{T} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}\Delta_{c_t}}_{\text{cost variability}} + \underbrace{\frac{1}{T}\sum_{t=1}^{T}\|w_{t+1} - w_t\|}_{\text{disturbance variability}}\right),$$

where $\Delta_{c_t} := \max_{\|x\|,\|u\| \le D}\{c_{t+1}(x, u) - c_t(x, u)\}$ for every t.

Numerical Simulations (Gaussian Disturbance)



Numerical Simulations (Sinusoidal Disturbance)



Conclusion and Future Work

- Online Multi-Agent Control with Adversarial Disturbances:

  Individual regret + global equilibrium guarantees.

Future work:

- Unknown dynamics (system identification).
- Time-varying dynamics.
- Bandit setting, partial observability.
- Network of communicating agents with local state observations.

References I

Agarwal, N., Bullins, B., Hazan, E., Kakade, S., and Singh, K. (2019).
Online control with adversarial disturbances.
In *International Conference on Machine Learning*, pages 111–119. PMLR.

Başar, T. and Olsder, G. J. (1998).
Dynamic noncooperative game theory.
SIAM.

Bu, J., Ratliff, L. J., and Mesbahi, M. (2019).
Global convergence of policy gradient for sequential zero-sum linear quadratic dynamic games.
arXiv preprint arXiv:1911.04672.

Cai, S., Han, F., and Cao, X. (2024).
Performative control for linear dynamical systems.
Advances in Neural Information Processing Systems.

References II

Chang, T.-J. and Shahrampour, S. (2023).
Regret analysis of distributed online control for lti systems with adversarial disturbances.
arXiv preprint arXiv:2310.03206.

Chen, F. and Ren, W. (2019).
On the control of multi-agent systems: A survey.
Foundations and Trends® in Systems and Control, 6(4):339–499.

Foster, D. and Simchowitz, M. (2020).
Logarithmic regret for adversarial online control.
In *International Conference on Machine Learning*, pages 3211–3221. PMLR.

Ghai, U., Gupta, A., Xia, W., Singh, K., and Hazan, E. (2023).
Online nonstochastic model-free reinforcement learning.
Advances in Neural Information Processing Systems, 36:23362–23388.

References III

Golowich, N., Hazan, E., Lu, Z., Rohatgi, D., and Sun, Y. J. (2024).
Online control in population dynamics.
In *Advances in Neural Information Processing Systems*, volume 37, pages 111571–111613.

Gradu, P., Hallman, J., and Hazan, E. (2020).
Non-stochastic control with bandit feedback.
In *Advances in Neural Information Processing Systems*, volume 33, pages 10764–10774.

Hambly, B., Xu, R., and Yang, H. (2023).
Policy gradient methods find the nash equilibrium in n-player general-sum linear-quadratic games.
Journal of Machine Learning Research, 24(139):1–56.

References IV

Hazan, E., Kakade, S., and Singh, K. (2020).
The nonstochastic control problem.
In *Algorithmic Learning Theory*, pages 408–421. PMLR.

Marden, J. R. and Shamma, J. S. (2018).
Game theory and control.
Annual review of control, robotics, and autonomous systems, 1(1):105–134.

Mazumdar, E., Ratliff, L. J., Jordan, M. I., and Sastry, S. S. (2020).
Policy-gradient algorithms have no guarantees of convergence in linear quadratic games.
In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, page 860–868, Richland, SC. International Foundation for Autonomous Agents and Multiagent Systems.

References V

Simchowitz, M. (2020).
Making non-stochastic control (almost) as easy as stochastic.
Advances in Neural Information Processing Systems, 33:18318–18329.

Simchowitz, M., Singh, K., and Hazan, E. (2020).
Improper learning for non-stochastic control.
In *Conference on Learning Theory*, pages 3320–3436. PMLR.

Tsiamis, A., Karapetyan, A., Li, Y., Balta, E. C., and Lygeros, J. (2024).
Predictive linear online tracking for unknown targets.
In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 48657–48694. PMLR.

References VI

Zhang, K., Yang, Z., and Basar, T. (2019).
Policy optimization provably converges to nash equilibria in zero-sum linear quadratic games.
Advances in Neural Information Processing Systems, 32.

Zhang, K., Zhang, X., Hu, B., and Basar, T. (2021).
Derivative-free policy optimization for linear risk-sensitive and robust control design: Implicit regularization and sample complexity.
Advances in neural information processing systems, 34:2949–2964.