Online Convex Optimization with Unbounded Memory

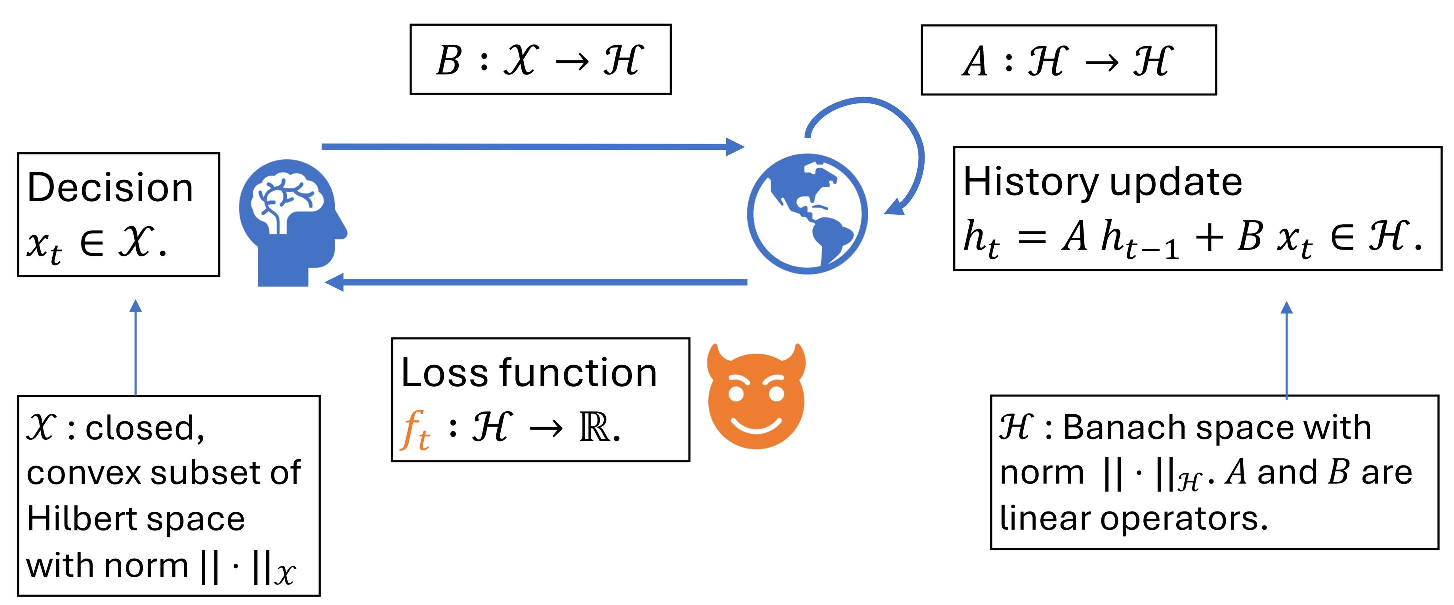
Raunak Kumar (Cornell), Sarah Dean (Cornell), Robert Kleinberg (Cornell)

NEURAL INFORMATION PROCESSING SYSTEMS

TL;DR

- **Generalization of OCO** capturing complete dependence of present losses on the **entire history of decisions**.
- Matching upper and lower regret bounds, including first **non-trivial lower bound** for OCO with finite memory [1].
- Applications to **online performative prediction** and **online linear control** [2] improving existing regret bounds by a factor of dimension.

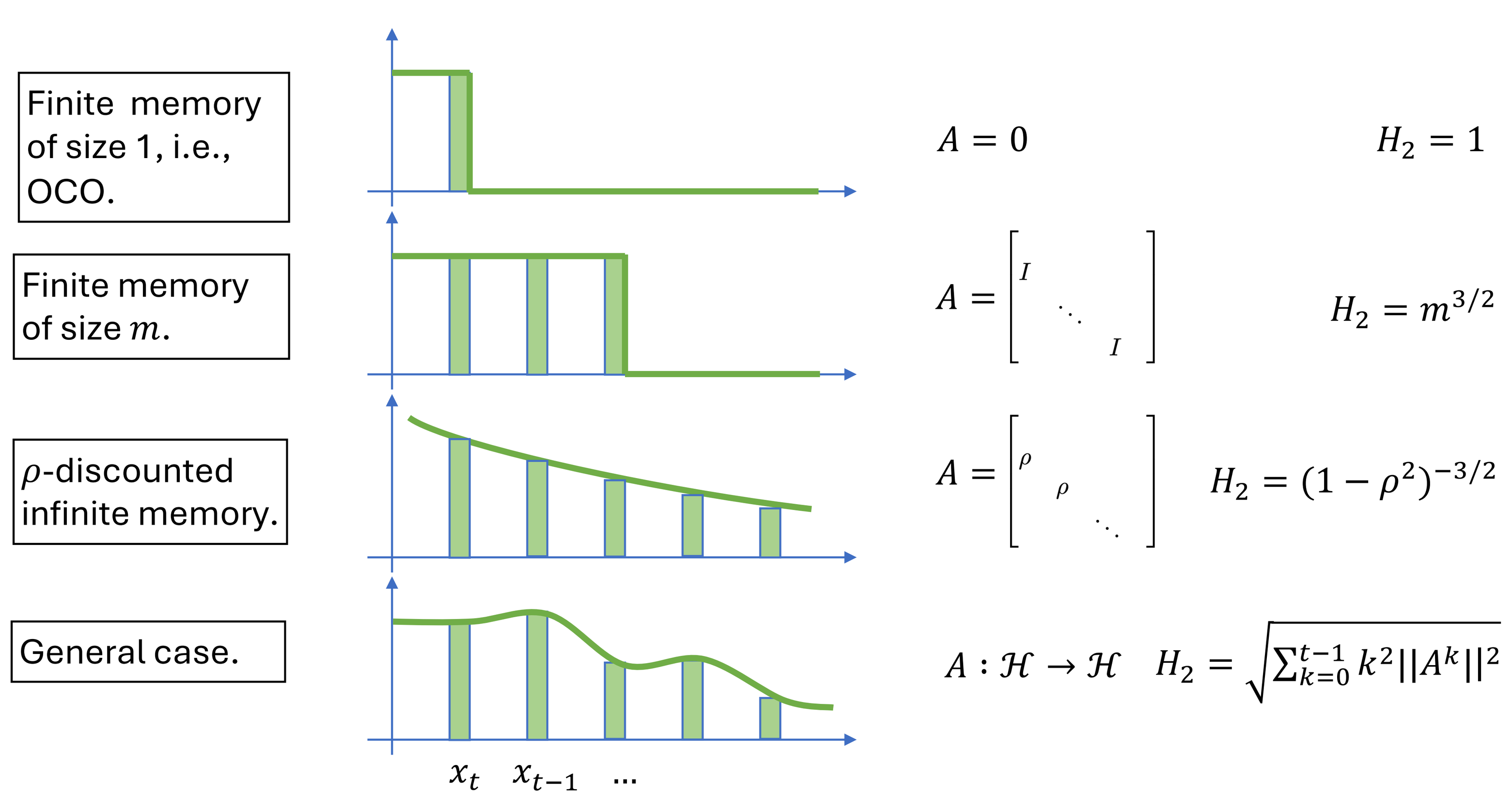
Framework



Goal: Minimize regret, $R_T = \sum_{t=1}^T f_t(h_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^T f_t\left(\sum_{k=0}^{t-1} A^k B\, x\right)$.

Definition (*p*-effective memory capacity): $H_p = \left(\sum_{k=0}^\infty k^p ||A^k||^p\right)^{1/p}$ bounds distance b/w histories generated by (y_k) and $(\tilde{y}_k)$, where $||y_k - \tilde{y}_k|| \le k$.

Examples



Assumptions

1. Learner knows A and B, observes f_t at the end of round t.
2. Norm of B is at most 1, i.e., $||B|| \le 1$.
3. Functions f_t are convex and L-Lipschitz continuous.
4. The 1-effective memory capacity is finite. ($\Rightarrow H_p < \infty$ for all $p \ge 1$.)

Regret Bounds

Theorem (upper bound): We design an algorithm with regret $O(\sqrt{T}\sqrt{H_p}\sqrt{L\tilde{L}})$.

Theorem (lower bound): We construct instances with regret $\Omega(\sqrt{T}\sqrt{H_p}\sqrt{L\tilde{L}})$.

Specialization finite memory: We improve the upper bound by a factor of $m^{1/4}$ to $O(m\sqrt{T})$ and provide **first non-trivial, matching lower bound** of $\Omega(m\sqrt{T})$.

By assumption 3, convex and $\tilde{L}$-Lipschitz for $\tilde{L} \le L \sum_{k=0}^\infty ||A^k||$.

Algorithm: Follow-the-regularized-leader (FTRL) on $\tilde{f}_t(x) = f_t\left(\sum_{k=0}^{t-1} A^k B\, x\right)$.
With step-size η and strongly-convex regularizer $R : \mathcal{X} \to \mathbb{R}$, choose

$$x_t \in \operatorname{argmin}_{x \in \mathcal{X}} \sum_{s=1}^{t-1} \tilde{f}_t(x) + {R(x)}/{\eta}.$$

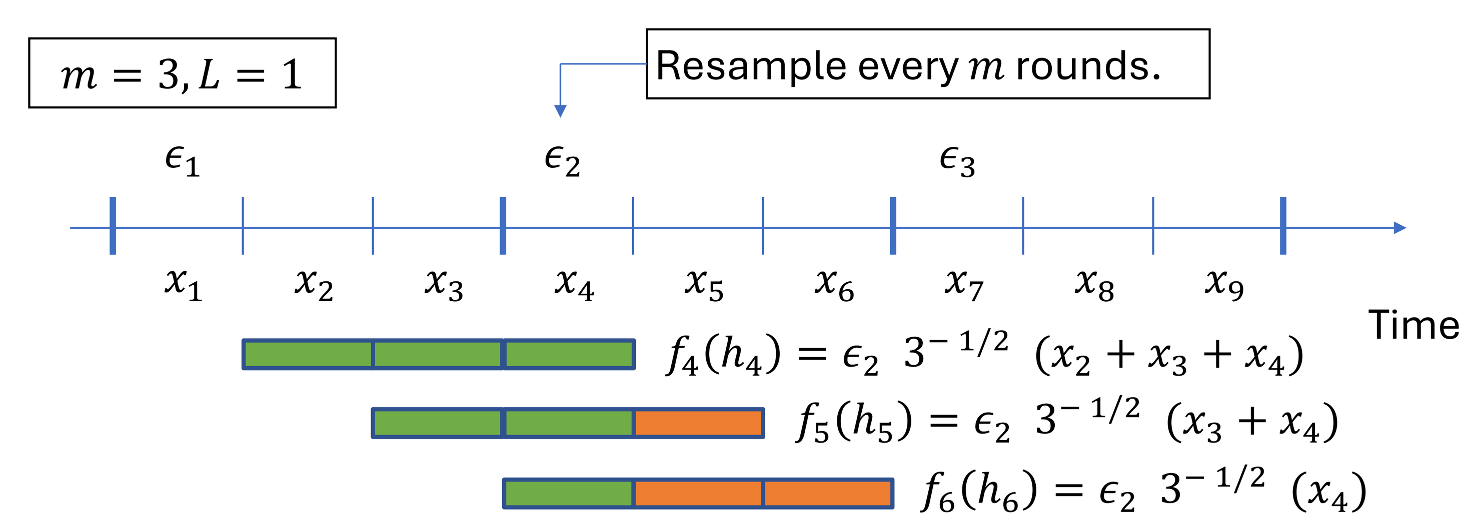
See paper for efficient implementation and version with low switches.

Analysis Sketch: $R_T = \underbrace{\sum_{t=1}^T f_t(h_t) - \tilde{f}_t(x_t)}_{\text{(i) FTRL movement cost}} - \underbrace{\min_{x \in \mathcal{X}} \sum_{t=1}^T \tilde{f}_t(x_t) - \tilde{f}_t(x)}_{\text{(ii) FTRL regret}}$.

Key Innovation: Bound FTRL movement cost with weighted norms, allowing us to derive non-trivial regret bounds for unbounded-length histories.

Lower Bound: Consider OCO with m-finite memory.

- Let $\mathcal{X} = [-1,1]$ and $\mathcal{H} = \mathcal{X}^m$.
- Let $\epsilon_1, \epsilon_2, \ldots, \epsilon_{\frac{T}{m}}$ be equal to $+1$ or -1 with probability $\frac{1}{2}$ each.
- Set $f_t(h_t) = \epsilon_{\lceil \frac{t}{m} \rceil}\, m^{-1/2} \left(x_{t-m+1} + \cdots + x_{m \lfloor \frac{t}{m} \rfloor + 1}\right)$.

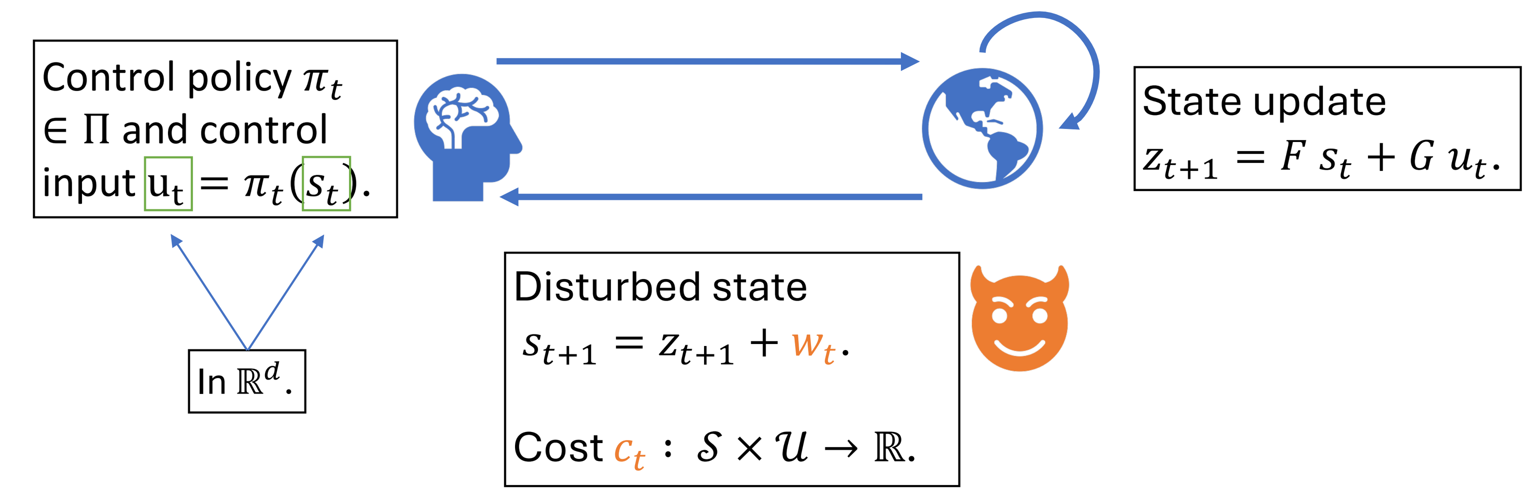


- Expected loss of any algorithm is 0.
- Expected loss of the optimal decision is $-\Omega(m\sqrt{T})$.

See paper for a similar construction for ρ-discounted infinite memory.

Application to Online Linear Control

Consider online linear control with adversarial disturbances [2].



Goal: Minimize regret $\sum_{t=1}^T c_t(s_t, u_t) - \min_{\pi \in \Pi} \sum_{t=1}^T c_t(s_t^\pi, u_t^\pi)$.

Benchmark policy class: linear controllers $u_t = K\, s_t$, where K is (κ, ρ)-stable, i.e., $||F - GK||^2 \le \kappa^2 \rho$ with $\rho < 1$.
Direct parameterization with a linear controller leads to non-convexity.

Convex reparameterization: Disturbance-action controller $(K, M_t = (M_t^{[s]}))$

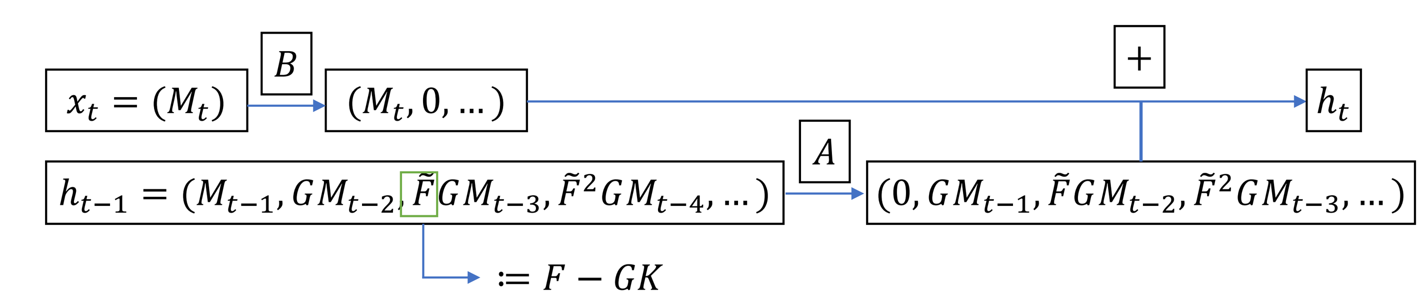
$$u_t = -K\, s_t + \sum_{s=1}^{t+1} M_t^{[s]}\, w_{t-s}.$$

Fixed matrix. (K) — (Parameters) A sequence of matrices. ($M_t^{[s]}$)

State is a *linear* function of past parameters $(M_1, M_2, \ldots)$.

Formulation as OCO with Unbounded Memory:

1. Decisions are disturbance-action controllers.
2. History is a *transformed* sequence of past decisions.
3. Linear operators A and B defined by dynamics.



4. Loss functions f_t parameterized by past disturbances and cost.

Key Innovation: Weighted norms on the history and decision spaces. This captures the dimension of infinite-dimensional spaces and improves the regret bound in existing works [2] by $O(d\, (\log T)^{3.5} \kappa^5 (1-\rho)^{-1})$.

References

[1] Anava et al. (2015). “Online Learning for Adversaries with Memory: Price of Past Mistakes.” In: NeurIPS.

[2] Agarwal et al. (2019). “Online Control with Adversarial Disturbances.” In: ICML.

[3] Anderson et al. (2019). “System Level Synthesis.” In: Annual Reviews in Control.

[4] Youla et al. (1976). “Modern Wiener-Hopf Design of Optimal Controllers—Par II: The Multivariable Case.” In: IEEE Transactions on Automatic Control.