Optimistic Online Non-stochastic Control (via FTRL)

Naram Mhaisen & George Iosifidis
Faculty of Electrical Engineering, Mathematics, and Computer Science

TU Delft Delft University of Technology

Objectives

- The Online Learning Framework
  - “Optimism” in Online Learning
- Relation to Control: Online Non-stochastic Control Framework (NSC)
  - Adding the optimism flavor: Optimistic Online NSC (this work)

Introduction

Online Learning

At each time step $t = 1, \cdots, T$

- Commit to an action $\boldsymbol{u}_t$ Based on Info available until $t-1$
- Incur a convex loss $c_t(\boldsymbol{u}_t)$

Objective: $\{\boldsymbol{u}_t\}_{t=1}^{T}$ that perform well under $\{c_t(\cdot)\}_{t=1}^{T}$

Hazan, Elad. "Introduction to online convex optimization." *Foundations and Trends in Optimization* 2.3-4 (2016): 157-325.

Introduction



At each time step $t = 1, \cdots, T$

- Receive an untrusted prediction of the loss $\tilde{c}_t(\cdot)$
- Commit to an action $\boldsymbol{u}_t$ Based on Info available until $t-1$
- Incur a convex loss $c_t(\boldsymbol{u}_t)$

Objective:- $\{\boldsymbol{u}_t\}_{t=1}^{T}$ that perform well under $\{c_t(\cdot)\}_{t=1}^{T}$

- The performance improves with accuracy *and never falls below a strong baseline.*

Rakhlin, Alexander, and Karthik Sridharan. "Online learning with predictable sequences." Conference on Learning Theory, 2013.

Introduction

Connection to Control



At each time step $t = 1, \cdots, T$

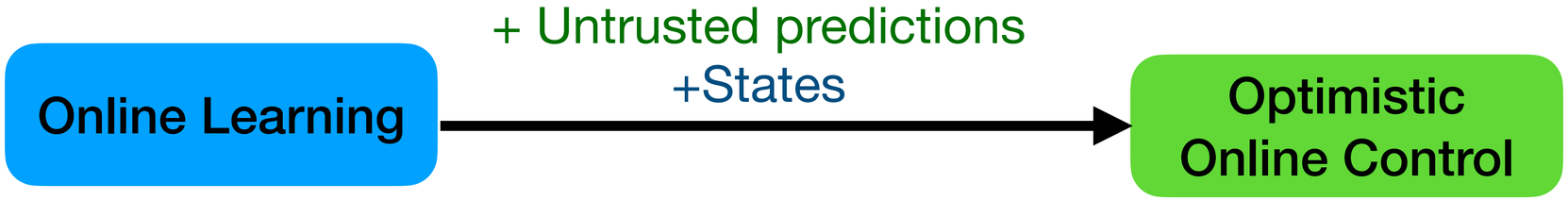
- Commit to an action $\boldsymbol{u}_t$ Based on Info available until $t-1$
- Transition to a new state $\boldsymbol{x}_{t+1} = A_t \boldsymbol{x}_t + B_t \boldsymbol{u}_t + \boldsymbol{w}_t$
- Incur a convex loss $c_t(\boldsymbol{x}_t, \boldsymbol{u}_t)$

Objective: $\{\boldsymbol{u}_t\}_{t=1}^T$ that perform well under $\{c_t(\cdot, \cdot)\}_{t=1}^T$.

Hazan, Elad, Sham Kakade, and Karan Singh. "The nonstochastic control problem." *Algorithmic Learning Theory*. PMLR, 2020.

Introduction

Parsing the title “Optimistic Online NSC”



At each time step $t = 1, \cdots, T$

- Receive an untrusted prediction of the loss $\tilde{c}_t(\cdot, \cdot)$
- Commit to an action $\boldsymbol{u}_t$ Based on Info available until $t-1$
- Transition to a new state $\boldsymbol{x}_{t+1} = A_t\boldsymbol{x}_t + B_t\boldsymbol{u}_t + \boldsymbol{w}_t$
- Incur a convex loss $c_t(\boldsymbol{x}_t, \boldsymbol{u}_t)$.

Objective: - $\{\boldsymbol{u}_t\}_{t=1}^T$ that perform well under $\{c_t(\cdot, \cdot)\}_{t=1}^T$

- The performance improves with accuracy *and never falls below a strong baseline.*

Online NSC

Metric

- **Policy Regret Metric:**

$$\mathcal{R}_T \doteq \sum_{t=1}^{T} c_t\left(\boldsymbol{x}_t, \boldsymbol{u}_t\right) - \min_{\pi \in \Pi} \sum_{t=1}^{T} c_t\big(\boldsymbol{x}_t(\pi), \boldsymbol{u}_t(\pi)\big),$$

  Difference between:

  our accumulated costs and that of **the optimal in hindsight**

  (For any $\{c_t(\,\cdot\,,\cdot\,)\}_{t=1}^{T}$, $\{\boldsymbol{w}_t\}_{t=1}^{T}$)

- If $R_T = \mathcal{O}(T^{\alpha}), \alpha < 1,$ Then, $\dfrac{R_T}{T} \to 0$
- i.e., we converge to the performance of the benchmark at the same rate.

Optimistic Online NSC

Design choices & assumption:

- **Transition:** LTI, known & stable (A, B)

$$\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t + B\boldsymbol{u}_t + \boldsymbol{w}_t\,.$$

  - Extensible to unknown (A_t, B_t)

- **Policy class:** DAC. A policy is parametrized by $M \doteq \left[M^{[j]}\right], j \in [p]$

$$\boldsymbol{u}_t = \sum_{j=1}^{p} M_t^{[j]} \boldsymbol{w}_{t-j}\,.$$

  - Approximates the class of linear controllers & has competitive ratio guarantees.

- **Cost function:** Linear (universal)

$$c_t(\boldsymbol{x}_t, \boldsymbol{u}_t) = \boldsymbol{\alpha}_t^\top \boldsymbol{x}_t + \boldsymbol{\beta}_t^\top \boldsymbol{u}_t$$

  - Extensible to strongly convex, exp-concave.
  - The same bounds apply to strongly convex costs (but loose)

Optimistic Online NSC

DAC motivation.

$$c_t(\boldsymbol{x}_t, \boldsymbol{u}_t) = g_t(M_{1,\cdots,t})$$

From the DAC class, depends on $M_{t-p,\cdots,t-1}$

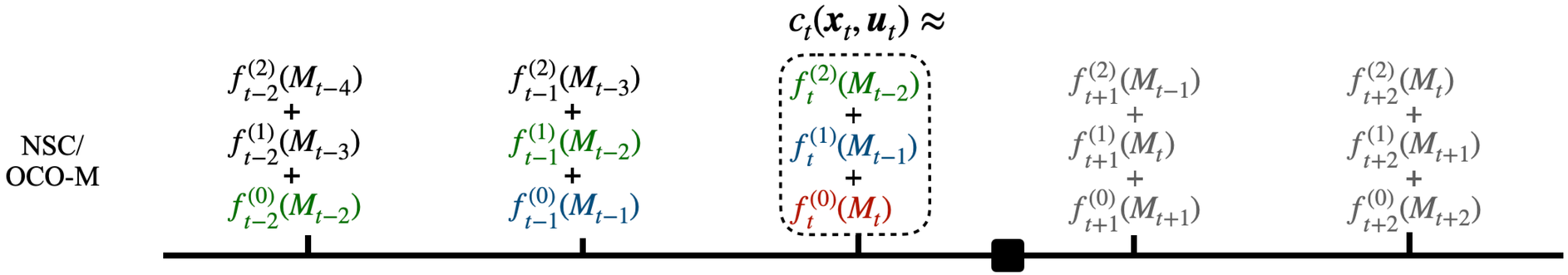
Depends on all previous actions $M_{1,\cdots,t-1}$

- $\boldsymbol{x}_{t+1} = \sum_{i=0}^{t} A^i \left(B \sum_{j=1}^{p} \left(M_{t-i}^{[j]} \, \boldsymbol{w}_{t-i-j} \right) + \boldsymbol{w}_{t-i} \right)$
- $g_t(\cdot)$ is convex in the policy parameters $M_{1,\cdots,t-1}$
- $g_t(\cdot)$ is mostly determined by $M_{t-p,\cdots,t-1}$

Optimistic Online Control

Online Control is OCO-*M*

- NSC's two main ingredients:
  - Take a memory of p.
  - Solution: perform OGD on those p parameters.

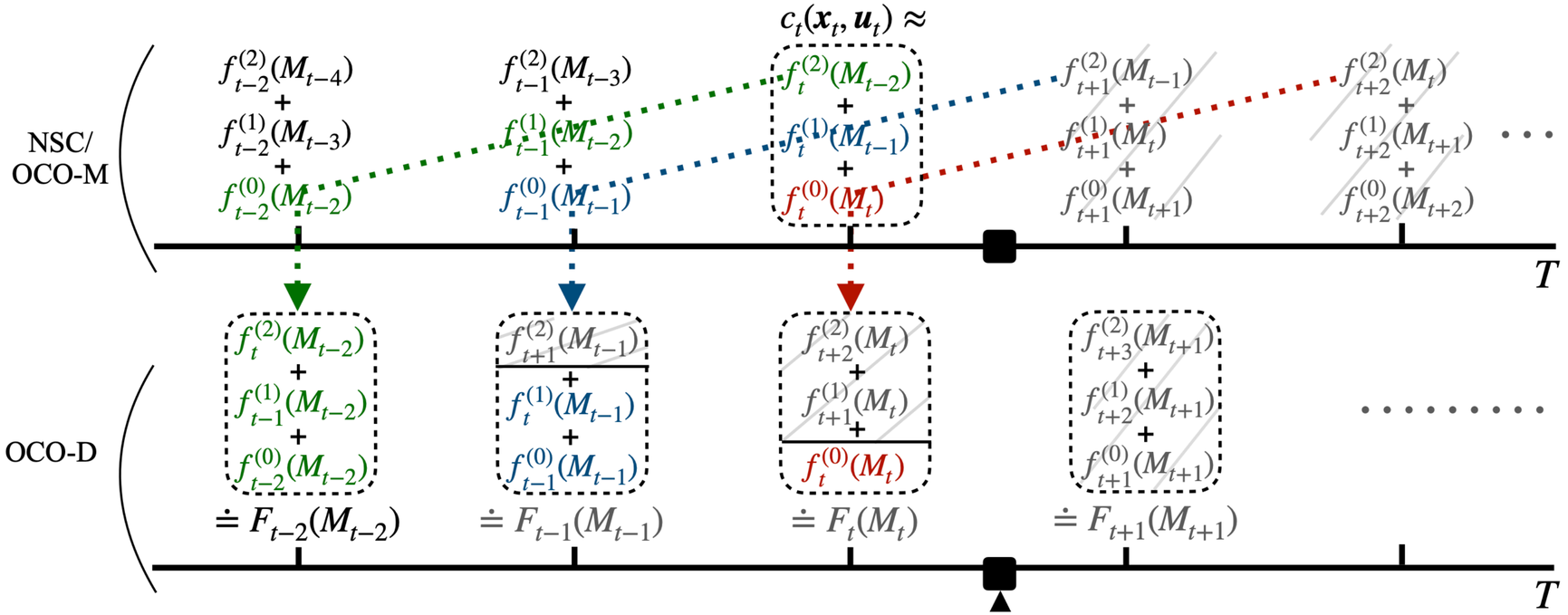


GPC: $M_{t+1} = M_t - \eta_t \nabla f_t(M_t)$ $\longrightarrow$ $R_T = \mathcal{O}(G\sqrt{T})$

Anava, Oren, Elad Hazan, and Shie Mannor. "Online learning for adversaries with memory: price of past mistakes." *NeurIPS 2015*

Optimistic Online Control

OCO-M is OCO-*D* (Paper idea 1)



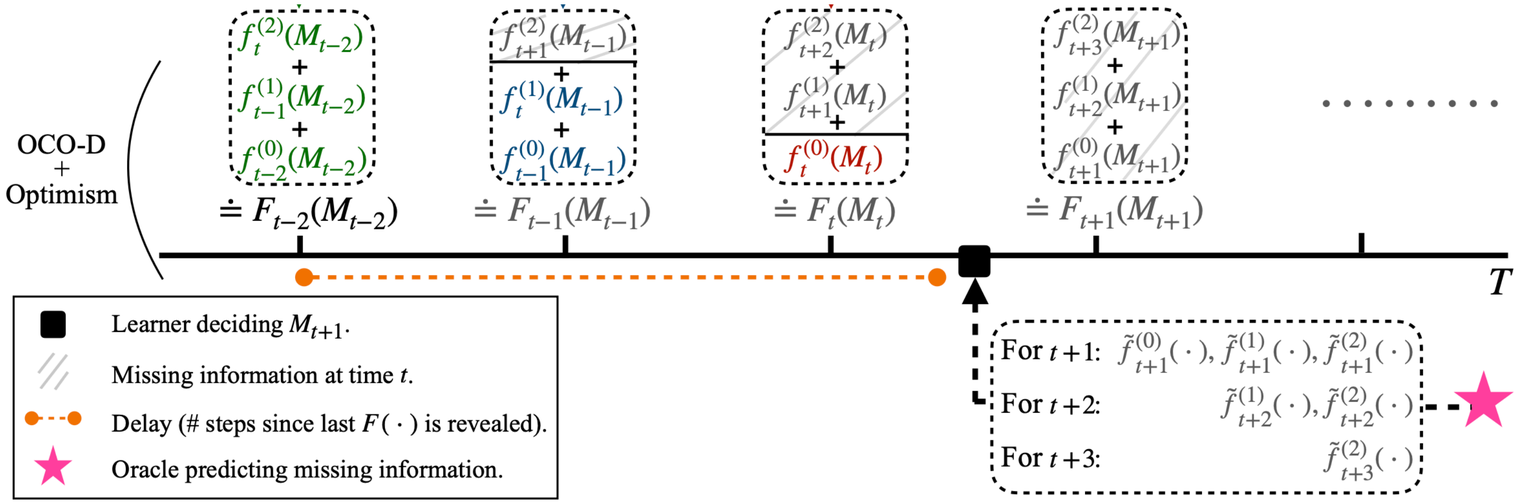
While deciding M_{t+1}, feedback only up to F_{t-p} is available

---

Joulani, Pooria, *et. al.* "Delay-tolerant online convex optimization: Unified analysis and…", *AAAI*. 2016.

Optimistic Online Control

OCO-*D* learning via Optimistic FTRL (Paper idea 2)



$$M_{t+1} = \arg\min_M \underbrace{F_{1:t-p}(M)}_{\text{Past costs}} + \underbrace{\tilde{F}_{t-p+1:t+1}(M)}_{\text{Future pred.}} + \underbrace{\frac{\lambda_{t+1}}{2}\|M\|^2}_{\text{Regs.}}$$

$$\lambda_{t+1} \propto \sqrt{\underbrace{\Delta_{1:t-p}}_{\text{Witnessed pred. error}}} \quad , \Delta_t = \|\nabla F_{t-p}(M_{t-p}) - \nabla \tilde{F}_{t-p}(M_{t-p})\|^2$$

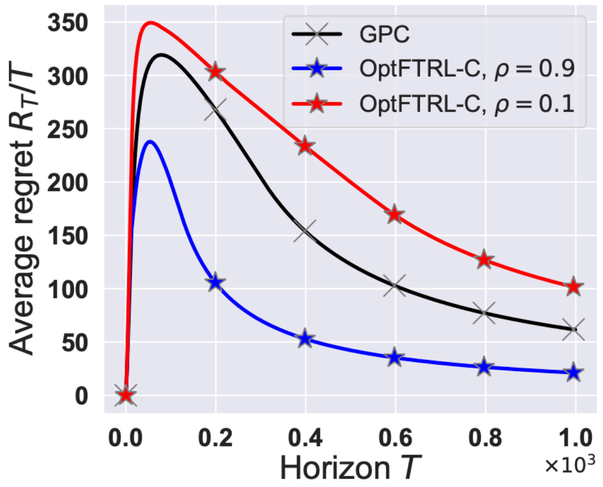
$$R_T = \mathcal{O}\left(G\sqrt{\sum_t \Delta_t}\right)$$

Flaspohler, Genevieve E., et al. "Online learning with optimism and delay." *ICML*. PMLR, 2021.

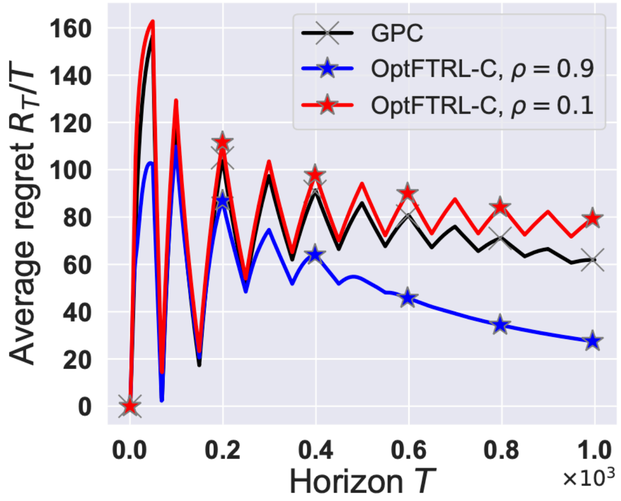
Optimistic Online Control

Numerical Example & Conclusion

$$R_T = \mathcal{O}\left(G\sqrt{\sum_t \Delta_t}\right) \quad \begin{cases} \mathcal{O}(1) \\ \mathcal{O}(\sqrt{T}) \end{cases}$$



(a) Stationary costs.



(b) Alternating costs.

Thanks

- More info in the paper:



- NSC Tutorial (video): https://sites.google.com/view/nsc-tutorial/home

- Optimistic Learning Tutorial (slides): https://www.futurenetworkslab.net/