Online Learning with Optimism and Delay

Genevieve Flaspohler[1,2], Francesco Orabona[3], Judah Cohen[4],

Soukayna Mouatadid[5], Miruna Oprescu[7], Paulo Orenstein[6], and Lester Mackey[7]

MIT[1], Woods Hole Oceanographic Institution[2], Boston University[3],

Atmospheric and Environmental Research[4], University of Toronto[5], IMPA[6], Microsoft Research[7]

Sequential Decision-Making via Online Learning

At time t:

1. Make play $\mathbf{w}_t \in \mathbf{W}$
2. Receive loss function ℓ_t from an adversarial environment
3. Pay $\ell_t(\mathbf{w}_t)$

Our objective is to do *as well as the best constant play in retrospect.*

$$\mathrm{Regret}_T = \underset{\text{The best constant play}}{\underline{\sup_{\mathbf{u}\in\mathbf{U}}}} \sum_{t=1}^{T} \underset{\text{The loss of online learner plays}}{\underline{\ell_t(\mathbf{w}_t)}} - \overset{\text{The loss of a constant play u}}{\underline{\ell_t(\mathbf{u})}}$$

Challenges of Real-world Online Learning

Real-world online learning has:

Our algorithms support:

Challenges of Real-world Online Learning

Real-world online learning has:

❌ Delayed feedback

e.g., several plays before feedback for first is received.

Our algorithms support:

Challenges of Real-world Online Learning

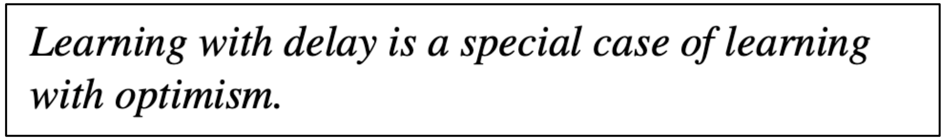
Real-world online learning has:

❌ Delayed feedback

e.g., several plays before feedback for first is received.

Our algorithms support:

✓ Unbounded or unknown delays



Learning with delay is a special case of learning with optimism.

Challenges of Real-world Online Learning

Real-world online learning has:

❌ Delayed feedback

e.g., several plays before feedback for first is received.

❌ Predictable/non-adversarial structure

e.g., contextual or side information.

Our algorithms support:

✓ Unbounded or unknown delays

via a novel delay-as-optimism reduction.

Challenges of Real-world Online Learning

Real-world online learning has:

- ❌ Delayed feedback

  e.g., several plays before feedback for first is received.

- ❌ Predictable/non-adversarial structure

  e.g., contextual or side information.

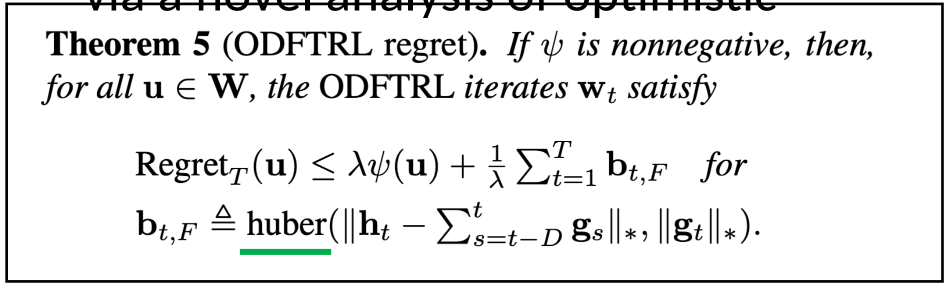
Our algorithms support:

- ✓ Unbounded or unknown delays

  via a novel delay-as-optimism reduction.

- ✓ Optimistic hints

  via a novel analysis of optimistic



Theorem 5 (ODFTRL regret)**.** *If ψ is nonnegative, then, for all $\mathbf{u} \in \mathbf{W}$, the* ODFTRL *iterates $\mathbf{w}_t$ satisfy*

$$\mathrm{Regret}_T(\mathbf{u}) \le \lambda\psi(\mathbf{u}) + \tfrac{1}{\lambda}\textstyle\sum_{t=1}^{T} \mathbf{b}_{t,F} \quad \textit{for}$$

$$\mathbf{b}_{t,F} \triangleq \underline{\mathrm{huber}}(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^{t} \mathbf{g}_s\|_*, \|\mathbf{g}_t\|_*).$$

Challenges of Real-world Online Learning

Real-world online learning has:

✗ Delayed feedback

e.g., several plays before feedback for first is received.

✗ Predictable/non-adversarial structure

e.g., contextual or side information.

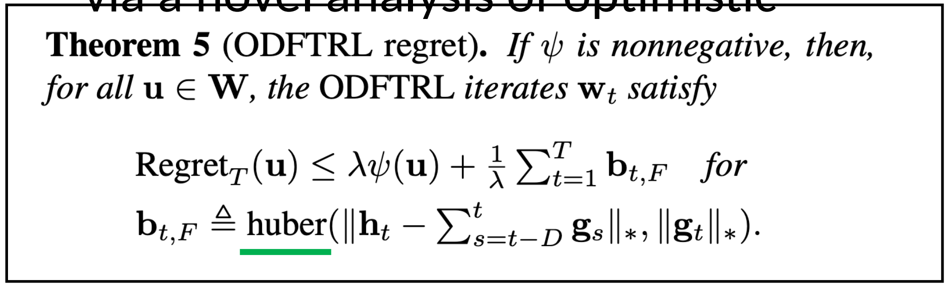
Our algorithms support:

✓ Unbounded or unknown delays

via a novel delay-as-optimism reduction.

✓ Optimistic hints

~~via a novel analysis of optimistic~~



Theorem 5 (ODFTRL regret)**.** *If ψ is nonnegative, then, for all $\mathbf{u} \in \mathbf{W}$, the* ODFTRL *iterates $\mathbf{w}_t$ satisfy*

$$\text{Regret}_T(\mathbf{u}) \le \lambda\psi(\mathbf{u}) + \tfrac{1}{\lambda}\sum_{t=1}^{T} \mathbf{b}_{t,F} \quad \textit{for}$$

$$\mathbf{b}_{t,F} \triangleq \text{huber}(\|\mathbf{h}_t - \textstyle\sum_{s=t-D}^{t} \mathbf{g}_s\|_*, \|\mathbf{g}_t\|_*).$$

The first optimal regret bound for general optimistic and delayed FTRL (and OMD).

Challenges of Real-world Online Learning

Real-world online learning has:

- ❌ Delayed feedback

  e.g., several plays before feedback for first is received.

- ❌ Predictable/non-adversarial structure

  e.g., contextual or side information.

- ❌ Real-time operational use

  e.g., continuous forecasting.

Our algorithms support:

- ✓ Unbounded or unknown delays

  via a novel delay-as-optimism reduction.

- ✓ Optimistic hints

  ~~via a novel analysis of~~ optimistic learning the reveals increased robustness to hint errors.

Challenges of Real-world Online Learning

Real-world online learning has:

✗ Delayed feedback

e.g., several plays before feedback for first is received.

✗ Predictable/non-adversarial structure

e.g., contextual or side information.

✗ Real-time operational use

e.g., continuous forecasting.

Our algorithms support:

✓ Unbounded or unknown delays

via a novel delay-as-optimism reduction.

✓ Optimistic hints

via a novel analysis of optimistic learning the reveals increased robustness to hint errors.

✓ Hyper-parameter free

with hint learning and no tuning

Challenges of Real-world Online Learning

Real-world online learning has:

- ✗ Delayed feedback

  e.g., several plays before feedback for first is received.

- ✗ Predictable/non-adversarial structure

  e.g., contextual or side information.

- ✗ Real-time operational use

  e.g., continuous forecasting.

- ✗ Short regret horizons

  e.g., weekly plays for a year period.

Our algorithms support:

- ✓ Unbounded or unknown delays

  via a novel delay-as-optimism reduction.

- ✓ Optimistic hints

  via a novel analysis of optimistic learning the reveals increased robustness to hint errors.

- ✓ Hyper-parameter free

  with hint learning and no tuning

Challenges of Real-world Online Learning

Real-world online learning has:

- ✗ Delayed feedback

  e.g., several plays before feedback for first is received.

- ✗ Predictable/non-adversarial structure

  e.g., contextual or side information.

- ✗ Real-time operational use

  e.g., continuous forecasting.

- ✗ Short regret horizons

  e.g., weekly plays for a year period.

Our algorithms support:

- ✓ Unbounded or unknown delays

  via a novel delay-as-optimism reduction.

- ✓ Optimistic hints

  via a novel analysis of optimistic learning the reveals increased robustness to hint errors.

- ✓ Hyper-parameter free

  with hint learning and no tuning

- ✓ Non-replicated design

  where a single learner observes every loss.

Delayed and Optimistic Online Learning

DORM

Delayed Optimistic Regret Matching

DORM+

Delayed Optimistic Regret Matching+

AdaHedgeD

Delayed Adaptive Hedge

Real-world online learning has:

- ✗ Delayed fe

  e.g., several
  first is receiv

- ✗ Predictabl
  structure

  e.g., context

- ✗ Real-time

  e.g., continu

- ✗ Short regr

  e.g., weekly plays for a year
  period.

Our algorithms support:

mism reduction.

timistic
eased

er free

ree algorithms

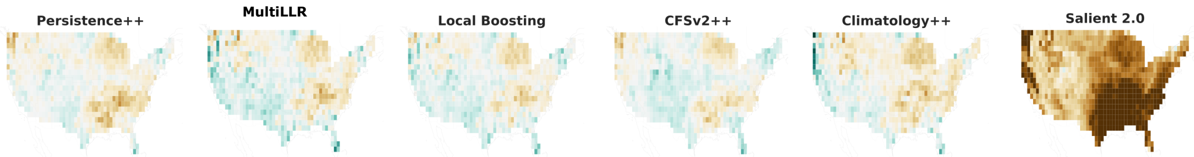
gn

where a single learner observes
every loss.

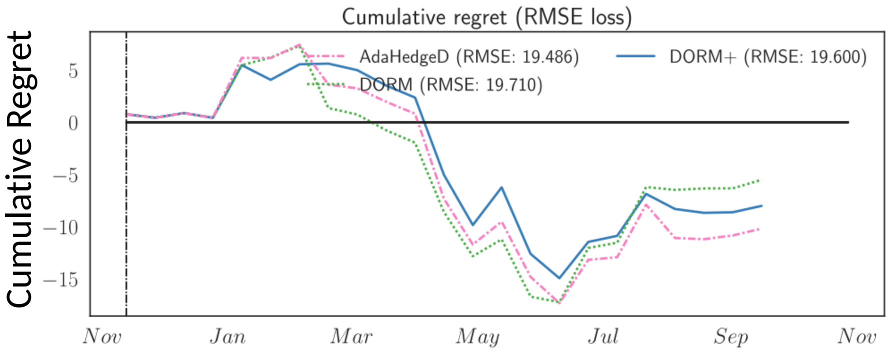
State-of-the-art Subseasonal Forecasting

What: Predicting the spatial distribution of temperature and precipitation 2 – 6 weeks out, w/applications in agriculture and energy [1].

Objective: Ensemble input models by playing weights: $\mathbf{w}_t \in \Delta$



Results: Using delayed and optimistic learners, we achieve negative regret in 3 of 4 subseasonal forecasting tasks.



T = 26

Positive regret

“Zero-Regret”
Performance of the
Best Input Model

Negative regret

[1] White et al., 2017, Meteorological Applications.

BOSTON UNIVERSITY

aer
A Verisk Business

Online Learning with Optimism and Delay

Genevieve Flaspohler, Francesco Orabona, Judah Cohen,
Soukayna Mouatadid, Miruna Oprescu, Paulo Orenstein, and Lester Mackey

Python implementation: https://github.com/geflaspohler/poold

Code:

Paper:

UNIVERSITY OF
TORONTO

impa
Instituto de Matemática
Pura e Aplicada

Microsoft