Adaptation in Online Learning and Bandits

Hédi HADIJI,
L2S, CNRS - CentraleSupélec - Université Paris-Saclay

May 2023, ILLS - DATAIA Workshop, U. Paris-Saclay

Online Convex Optimization

Protocol: Online Convex Optimization

1: **given**: (bounded) decision set $\mathcal{W} \subset \mathbb{R}^d$
2: **for** $t = 1, \ldots, T$
3: Player chooses $w_t \in \mathcal{W}$
4: Nature outputs convex loss $\ell_t : \mathcal{W} \to \mathbb{R}$

[Zinkevich ’03]

Goal: minimize **regret**

$$R_T(u) = \sum_{t=1}^{T} \ell_t(w_t) - \sum_{t=1}^{T} \ell_t(u)$$

No assumptions on how the losses are generated

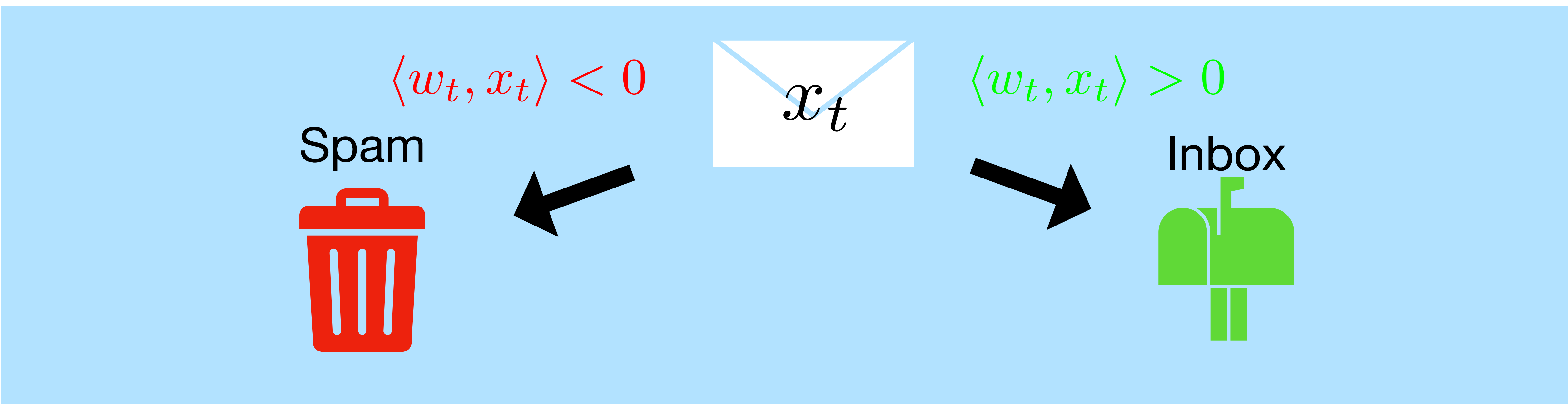
Nature can be an **adversary** who knows the algorithm!

Example: Online Spam Filtering

Training a linear model to filter spams

Email $x_t \in \mathbb{R}^d$, outcome $y_t = 1 - 2 \cdot \mathbf{1}\{x_t \text{ is spam}\} \in \{-1, +1\}$

Get w_t from an OCO algorithm



Proxy: Minimize a **convex** loss: e.g., $\ell_t : w \mapsto (y_t - \langle w, x_t \rangle)^2$

Regret: filter almost **as well as the best linear model in hindsight**

Plenty of Other Examples

see surveys by [Hazan ’16, Orabona ’19]

- Prediction with Expert Advice [Freund and Schapire ’97]
- Portfolio Selection [Cover ’91]
- … and many more

- With connections to batch optimisation, and practical impact: AdaGrad was introduced in the OCO framework.

he prequel. The ADAGRAD algorithm with full matrix divergences entertains bounds of the form

$$R_\phi(T) = O\left(\|x^*\|_2 \operatorname{tr}(G_T^{1/2})\right) \quad \text{and} \quad R_\phi(T) = O\left(\max_{t \le T} \|x_t - x^*\|_2 \operatorname{tr}(G_T^{1/2})\right).$$

[Duchi et al ‘11]

Online Gradient Descent

$$w_{t+1} = \textbf{Proj}_{\mathscr{W}}\Big(w_t - \eta \nabla \ell_t(w_t) \Big)$$

Theorem: OGD [Zinkevich '03]

If OGD is tuned with a constant step size $\eta = \sqrt{D/GT}$ where G is an upper bound on the gradient norms, and D is the diameter of the action set,

$$R_T(w^\star) \leqslant c\, GD\sqrt{T}$$

… and this is not improvable

Optimality of OGD

Theorem: OCO Minmax Lower Bound [Zinkevich '03]

For any algorithm, there exists a sequence of losses for which G is an upper bound on the gradient norms, D is the diameter of the action set, and

$$\max_{w^\star \in \mathscr{W}} R_T(w^\star) \geqslant c\, GD\sqrt{T}$$

Optimality of OGD

Theore

For any

bound

Optimality of ODG

Theorem: OCO Minmax Lower Bound [Zinkevich '03]

For any algorithm, there exists a sequence of losses for which G is an upper bound on the gradient norms, D is the diameter of the action set, and

$$\max_{w^\star \in \mathscr{W}} R_T(w^\star) \geqslant c\, GD\sqrt{T}$$

Proof:

$$\ell_t(w) = \langle w, g_t \rangle \quad \text{where } g_t = \pm\, G\, e_1 \text{ with probability } 1/2$$

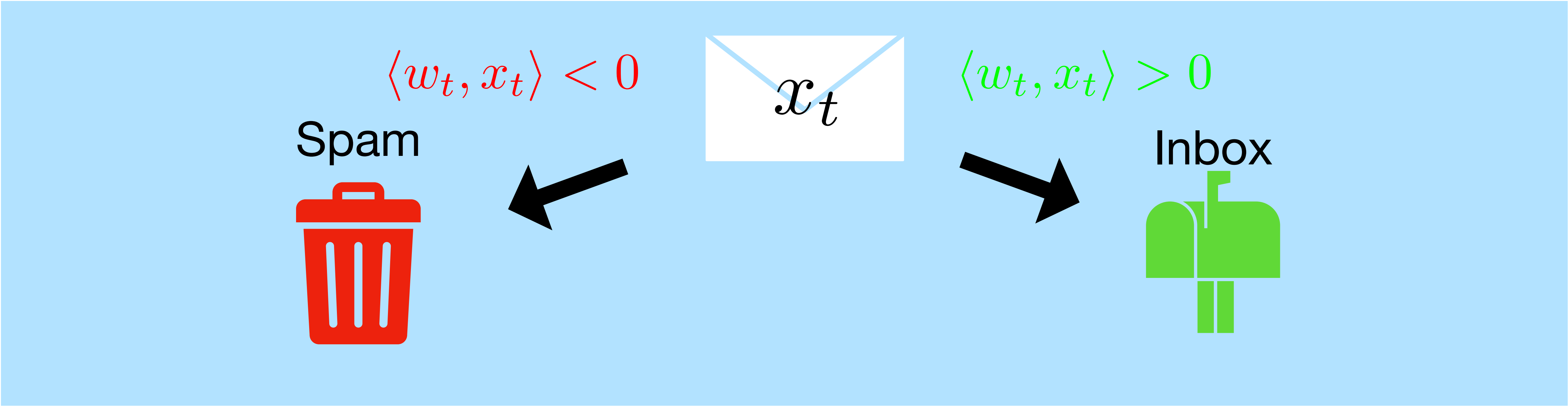
The expected regret of any player against these losses is $c\, GD\sqrt{T}$,
so there is at least one sequence for which the regret is lower bounded.

Construction involves:

- **Linear losses**
- **Pure Noise**

Online Spam Filtering II

Email $x_t \in \mathbb{R}^d$, outcome $y_t = 1 - 2 \cdot \mathbf{1}\{x_t \text{ is spam}\} \in \{-1, +1\}$



Do we expect the data to be like in the lower bound: pure noise?

To what extent is it adversarial?

Adapting to Easy Data

Losses are often far from worst-case

- Small gradients [Zinkevich'03, Duchi'10], Simple optimal comparator [Orabona, Cutkosky]
- Both [Mhammedi, Koolen '20, Mayo, Hadiji, van Erven '22]
- Predictable gradients [Rakhlin, Sridharan'13]
- Many more… (curved losses, extra information available, etc.)

- **…? Applications can inspire theory here!**

Adapting to Stochastic Data

Theorem: [Sachs, Hadiji, Van Erven, Guzmán '23]

There exists an algorithm such that, **if the losses come from i.i.d. data** with $\mathbb{E}\left[\ell_t(w)\right] = F(w)$ and F is L-smooth, then

$$\mathbb{E}\left[R_T(w^\star)\right] \leqslant c\,\sigma D\sqrt{T} + LD^2$$

where $\sigma^2 = \max_{1 \leq t \leq T} \mathbb{E}\left[\mathrm{Var}\left(\nabla \ell_t(w_t)\right)\right]$.

In the worst-case, this same algorithm enjoys the optimal rate

$$R_T(w^\star) \leqslant c'\,GD\sqrt{T}$$

- The algorithm is Optimistic Follow-the-Regularized-Leader
- Generalizes the linear ($L = 0$) case, known by [Rakhlin, Sridharan '13]
- Exploits iid-ness if available, but does not assume it
- Actually interpolates between fully stochastic and adversarial cases

Bandits

Action space $\mathscr{A}$

Losses $\ell_t : \mathscr{A} \to \mathbb{R}$

For $t = 1, \ldots, T, \ldots$:

- Player selects **action** $a_t \in \mathscr{A}$
- Player observes **loss** $\ell_t(a_t)$

Regret

$$R_T = \sum_{t=1}^{T} \ell_t(a_t) - \min_{a^\star \in \mathscr{A}} \sum_{t=1}^{T} \ell_t(a^\star)$$

Need to balance between:

- **Exploring:** acquiring information about actions
- **Exploiting** available information to optimise losses

Adaptation in (Stochastic) Bandits

Same question as in OL:
Can we have nice guarantees for nice data, while staying (close to) optimal in the worst-case?

The cost of exploration can make adaptation difficult, or even impossible

Adapting to regular loss functions when the action set is continuous [Hadiji '19]

Adapting to the range of the losses [Hadiji, Stoltz '22]

Again, a close look at the lower bounds hints that common assumptions are not relevant to practice

Thank you!