Exploiting easy data in online optimization

Gergely Neu

INRIA Lille, SequeL team

Joint work with

Amir Sani & Alessandro Lazaric

Outline

- Online optimization
- Worst-case guarantees…
- … and beyond
- The best of both worlds:
  (AB)-Prod
- Applications
- (Proof, if there's time)

Online optimization

Parameters:

decision set S, set of loss functions $F \subseteq [0,1]^S$

For $t = 1,2, \ldots T$ **repeat**

- Learner picks decision $x_t \in S$
- Environment picks loss function $f_t \in F$
- Learner suffers loss $f_t(x_t)$
- Learner observes f_t

Online optimization – examples

Prediction with expert advice:

- $S = [N] \overset{\text{def}}{=} \{1,2,\ldots,N\}$
- $F = [0,1]^N$

Online convex optimization:

- S: a convex subset of R^d
- F: the set of bounded convex functions on R^d

Online combinatorial optimization:

- $S \subseteq \{0,1\}^d$
- $F = [0,1]^S$
- e.g., set of paths, spanning trees, matchings on a graph

+ pretty much anything with full-information feedback*

A central concept: Regret

$$R_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x \in S} \sum_{t=1}^{T} f_t(x)$$

A central concept: Regret

Learner aims to choose $x_1, x_2, \dots, x_T$ to **minimize** R_T

$$R_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x \in S} \sum_{t=1}^{T} f_t(x)$$

A central concept: Regret

Learner aims to choose $x_1, x_2, \dots, x_T$ to **minimize** R_T

$$R_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x \in S} \sum_{t=1}^{T} f_t(x)$$

Environment aims to choose $f_1, f_2, \dots, f_T$ to **maximize** R_T

A central concept: Regret

Learner aims to choose $x_1, x_2, \ldots, x_T$ to **minimize** R_T

R_T

A typical guarantee:

$$R_T = \Theta\left(C\sqrt{T}\right)$$

$x)$

Environment aims to choose $f_1, f_2, \ldots, f_T$ to **maximize** R_T

Beyond worst-case guarantees

- What if the environment is not that bad?

Beyond worst-case guarantees

- What if the environment is not that bad?

- Some known easy cases:
  - i.i.d. losses in the experts setting:

$$R_T = O(\log N)$$

  - Strongly convex losses in OCO:

$$R_T = O(\log T)$$

  - i.i.d. losses in bandits:

$$R_T = O(\log T)$$

  - …

Beyond worst-case guarantees

- What if the environment is not that bad?
- Some known easy cases:
  - i.i.d. losses in the ex rts ng:
  - Stron
  - i.i.d. lo

    $T)$
  - …

But you need to be aggressive to get these!

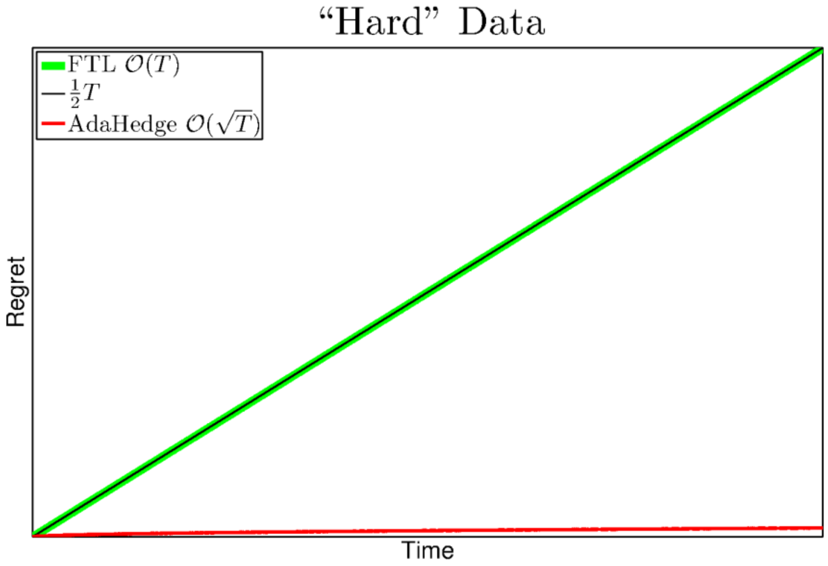
How bad can it be?

Classical counterexample:

- Two experts with losses

$$\begin{bmatrix} 1/2 & 0 & 1 & 0 & 1 & \dots \\ 0 & 1 & 0 & 1 & 0 & \dots \end{bmatrix}$$

- FTL suffers



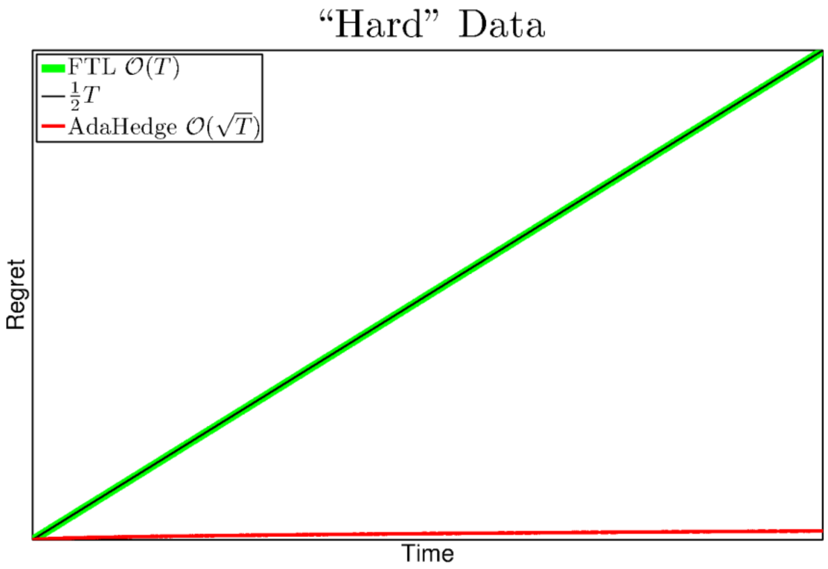
How bad can it be?

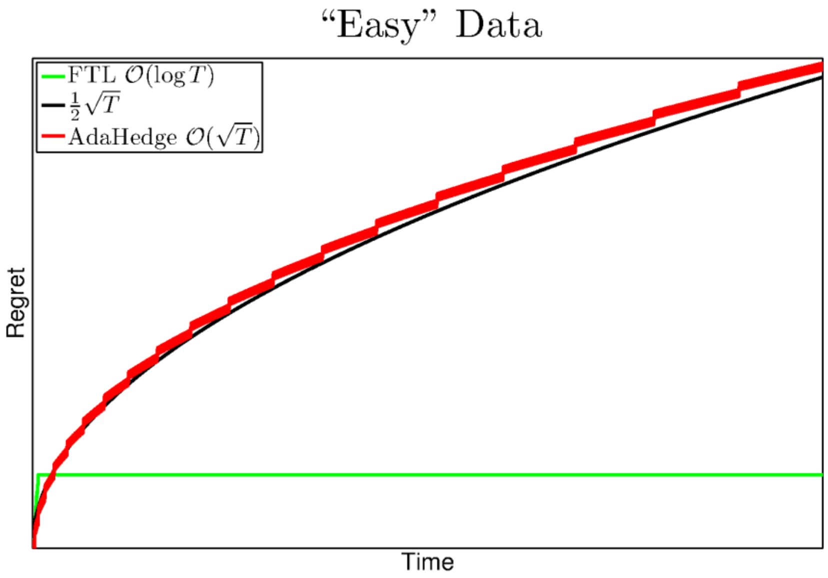
Classical counterexample:

- Two experts with losses

$$\begin{bmatrix} 1/2 & 0 & 1 & 0 & 1 & \ldots \\ 0 & 1 & 0 & 1 & 0 & \ldots \end{bmatrix}$$

- FTL suffers





Add a little twist:

- Two experts with losses

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 1 & \ldots \\ 1 & 1 & 0 & 1 & 0 & \ldots \end{bmatrix}$$

and a few more $[0\ 1]^T$ later

- FTL beats everything else

The best of both worlds

Some previous results:

	Easy data	Hard data
Experts (De Rooij, Van Erven, Grünwald and Koolen, 2014)	$\log N$ (for i.i.d.)	$\sqrt{T \log N}$ (standard worst-case)
OCO (Bartlett, Hazan and Rakhlin, 2007)	$\log T$ (for strongly convex)	$\sqrt{T}$ (standard worst-case)
Bandits (Seldin and Slivkins, 2014)	$\log^3 T$ (for i.i.d.)	$\sqrt{NT \log N}$ (standard worst-case)

Is there a generic way to combine two guarantees?

Is there a generic way to combine two guarantees?

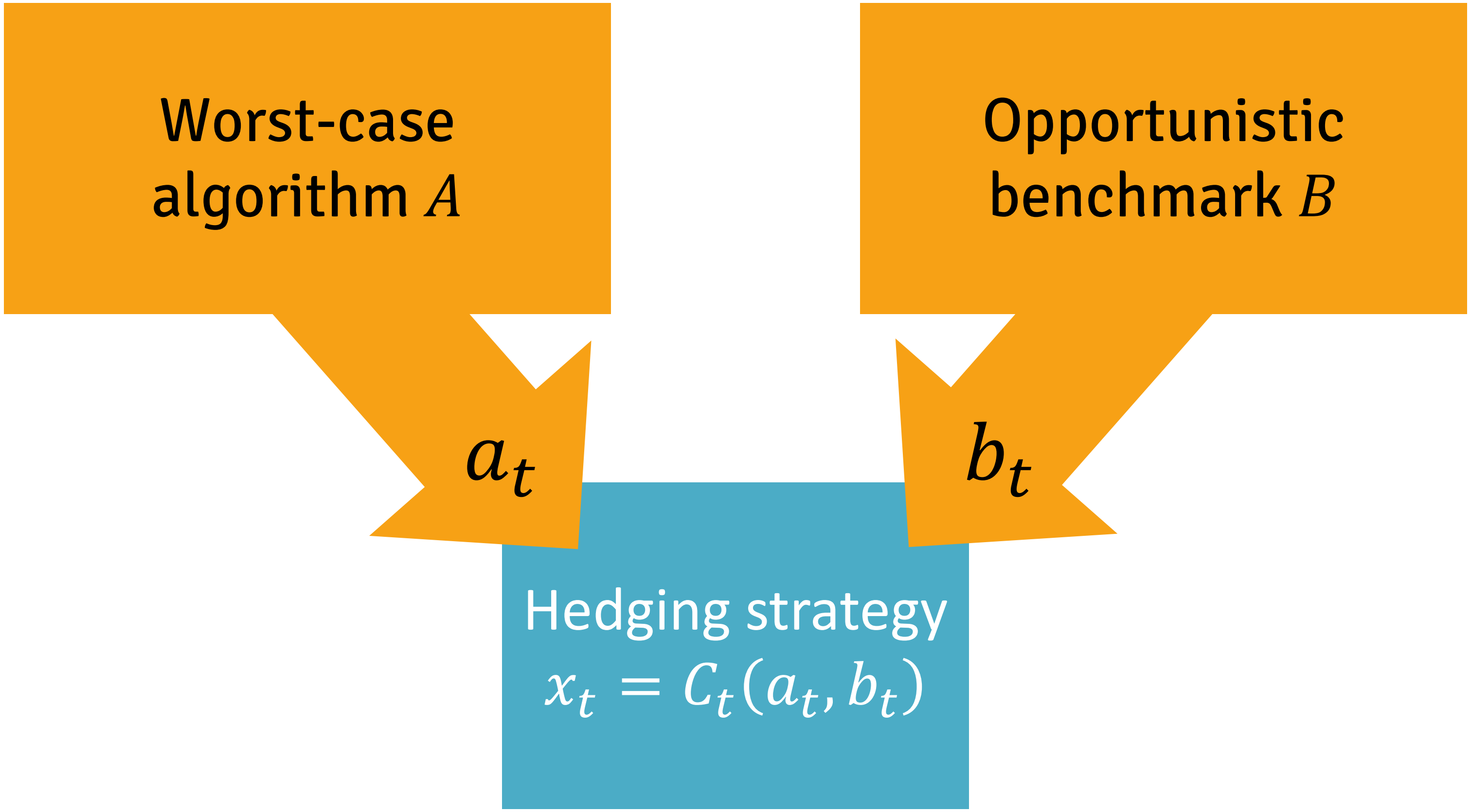
YES!!!!

Our setup

Worst-case algorithm A

Opportunistic benchmark B

Our setup



Worst-case algorithm A

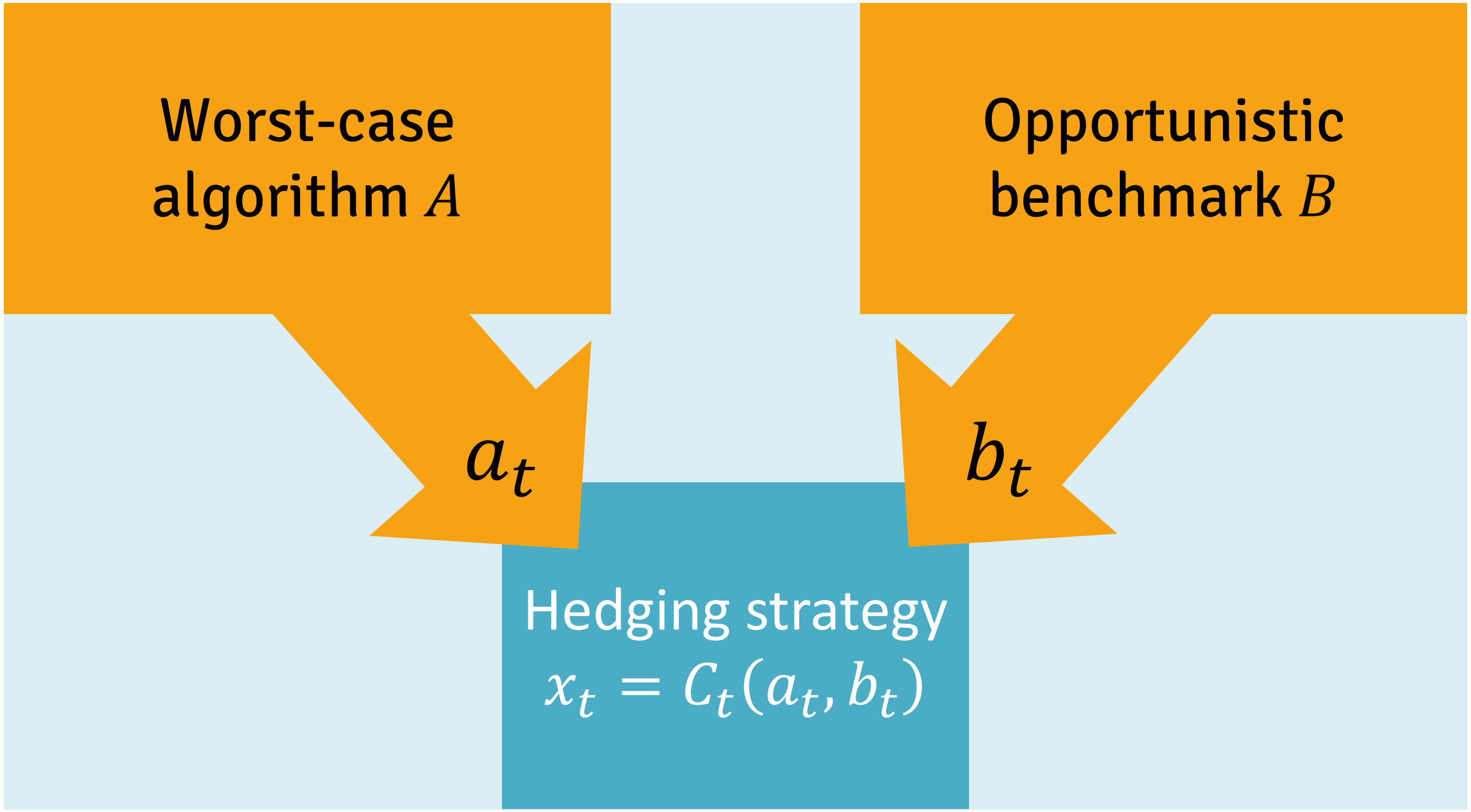
Opportunistic benchmark B

a_t

b_t

Hedging strategy $x_t = C_t(a_t, b_t)$

Our setup



Worst-case algorithm A

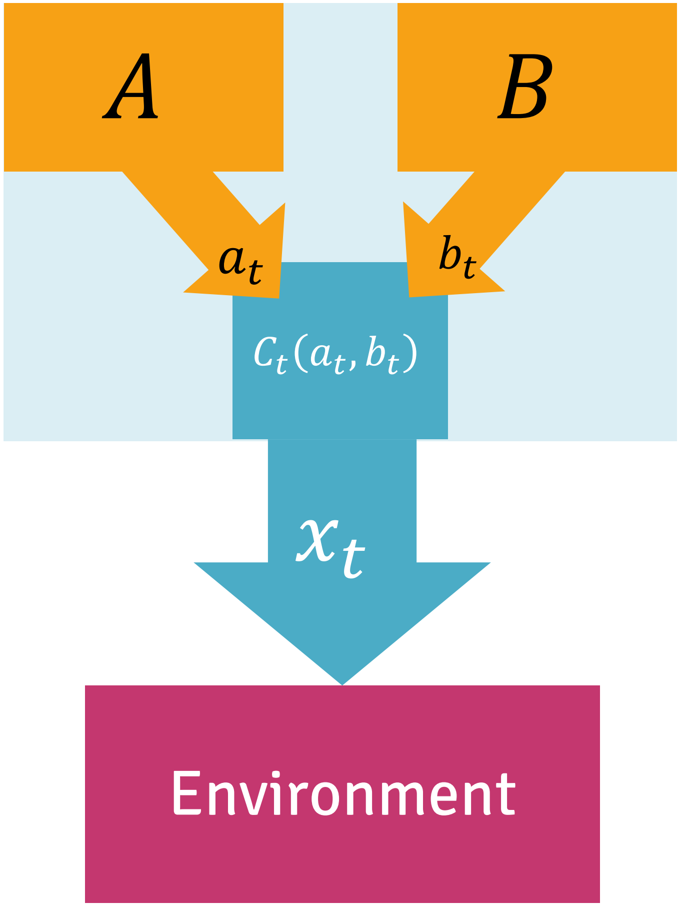
Opportunistic benchmark B

a_t

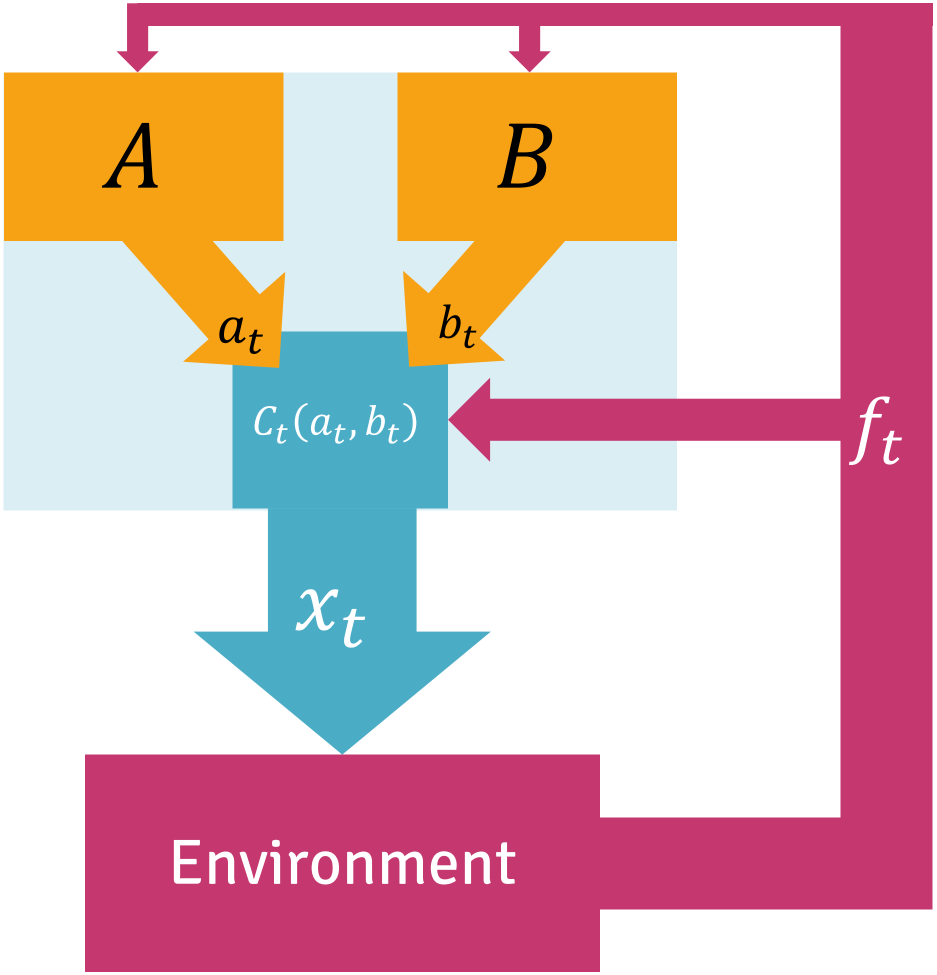
b_t

Hedging strategy $x_t = C_t(a_t, b_t)$

Our setup



Our setup



A naïve approach

Let’s treat A and B as experts and use Hedge* on top of them!

* Vovk(1990), Littlestone and Warmuth (1994), Freund and Schapire (1997)

A naïve approach

Let’s treat A and B as experts and use Hedge* on top of them!

```
Initialize $\eta > 0,\ w_{1,A} = w_{1,B} = 1/2$
For $t = 1,2,\ldots,T$ repeat
    • $s_t = \frac{w_{t,A}}{w_{t,A}+w_{t,B}}$
    • Predict $x_t = a_t$ w.p. $s_t$ and $x_t = b_t$otherwise
    • Observe $f_t$ and suffer loss $f_t(x_t)$
    • Compute $\delta_t = f_t(a_t) - f_t(b_t)$
    • Update $w_{t+1,A} = w_{t,A}e^{-\eta\delta_t}$
```

* Vovk(1990), Littlestone and Warmuth (1994), Freund and Schapire (1997)

A naïve approach

Let’s treat A and B as experts and use Hedge* on top of them!

Initiali… $\eta > 0,\ w_{1,}$ … $w_{1,B}$ … $/2$

For $t = 1,2,$ …

- s_t …
- Predict … ise
- Observe f_t a… $t)$
- Compute $\delta_t = f_t($ … $- f_t$
- Update $w_{t+1,A} = w_{t,A}e^{-\eta\delta_t}$

This only guarantees $O(\sqrt{T})$ regret against both A and B!!

* Vovk(1990), Littlestone and Warmuth (1994), Freund and Schapire (1997)

Our algorithm: (AB)-Prod

Let's replace Hedge by Prod*!

Initialize $\eta > 0,\ w_{1,A} = w_{1,B} = 1/2$

For $t = 1,2,\ldots,T$ **repeat**

- $s_t = \frac{w_{t,A}}{w_{t,A}+w_{t,B}}$
- Predict $x_t = a_t$ w.p. s_t and $x_t = b_t$otherwise
- Observe f_t and suffer loss $f_t(x_t)$
- Compute $\delta_t = f_t(a_t) - f_t(b_t)$
- Update $w_{t+1,A} = w_{t,A}(1 - \eta\delta_t)$

* Cesa-Bianchi, Mansour and Stoltz (2007), Even-Dar, Kearns, Mansour, Wortman (2008)

Our algorithm: (AB)-Prod

Let's replace Hedge by Prod*!

… and put a large weight on B!

```
**Initialize** $\eta > 0,\ w_{1,A} = \eta,\ w_{1,B} = 1 - \eta$
**For** $t = 1,2,\ldots,T$ **repeat**
    • $s_t = \frac{w_{t,A}}{w_{t,A}+w_{t,B}}$
    • Predict $x_t = a_t$ w.p. $s_t$ and $x_t = b_t$otherwise
    • Observe $f_t$ and suffer loss $f_t(x_t)$
    • Compute $\delta_t = f_t(a_t) - f_t(b_t)$
    • Update $w_{t+1,A} = w_{t,A}(1 - \eta\delta_t)$
```

* Cesa-Bianchi, Mansour and Stoltz (2007), Even-Dar, Kearns, Mansour, Wortman (2008)

Our main result

- Define
  - $R_T(C, x) = \mathbf{E}\left[\sum_{t=1}^{T}\left(f_t(x_t) - f_t(x)\right)\right]$
  - $R_T(A, x) = \mathbf{E}\left[\sum_{t=1}^{T}\left(f_t(x_t) - f_t(a_t)\right)\right]$
  - $R_T(C, A) = \mathbf{E}\left[\sum_{t=1}^{T}\left(f_t(x_t) - f_t(a_t)\right)\right]$

Our main result

- Define
  - $R_T(C, x) = \mathbf{E}\left[\sum_{t=1}^{T}\left(f_t(x_t) - f_t(x)\right)\right]$
  - $R_T(A, x) = \mathbf{E}\left[\sum_{t=1}^{T}\left(f_t(x_t) - f_t(a_t)\right)\right]$
  - $R_T(C, A) = \mathbf{E}\left[\sum_{t=1}^{T}\left(f_t(x_t) - f_t(a_t)\right)\right]$

Theorem:

$$R_T\big((AB)\text{–Prod}, x\big) \leq R_T(A, x) + 2\sqrt{T \log T}$$

$$R_T\big((AB)\text{–Prod}, x\big) \leq R_T(B, x) + 2 \log 2$$

Why does it work?

- The asymmetry of $(1 - \eta x)$ is key, e.g., $\eta = \frac{1}{2}$

Why does it work?

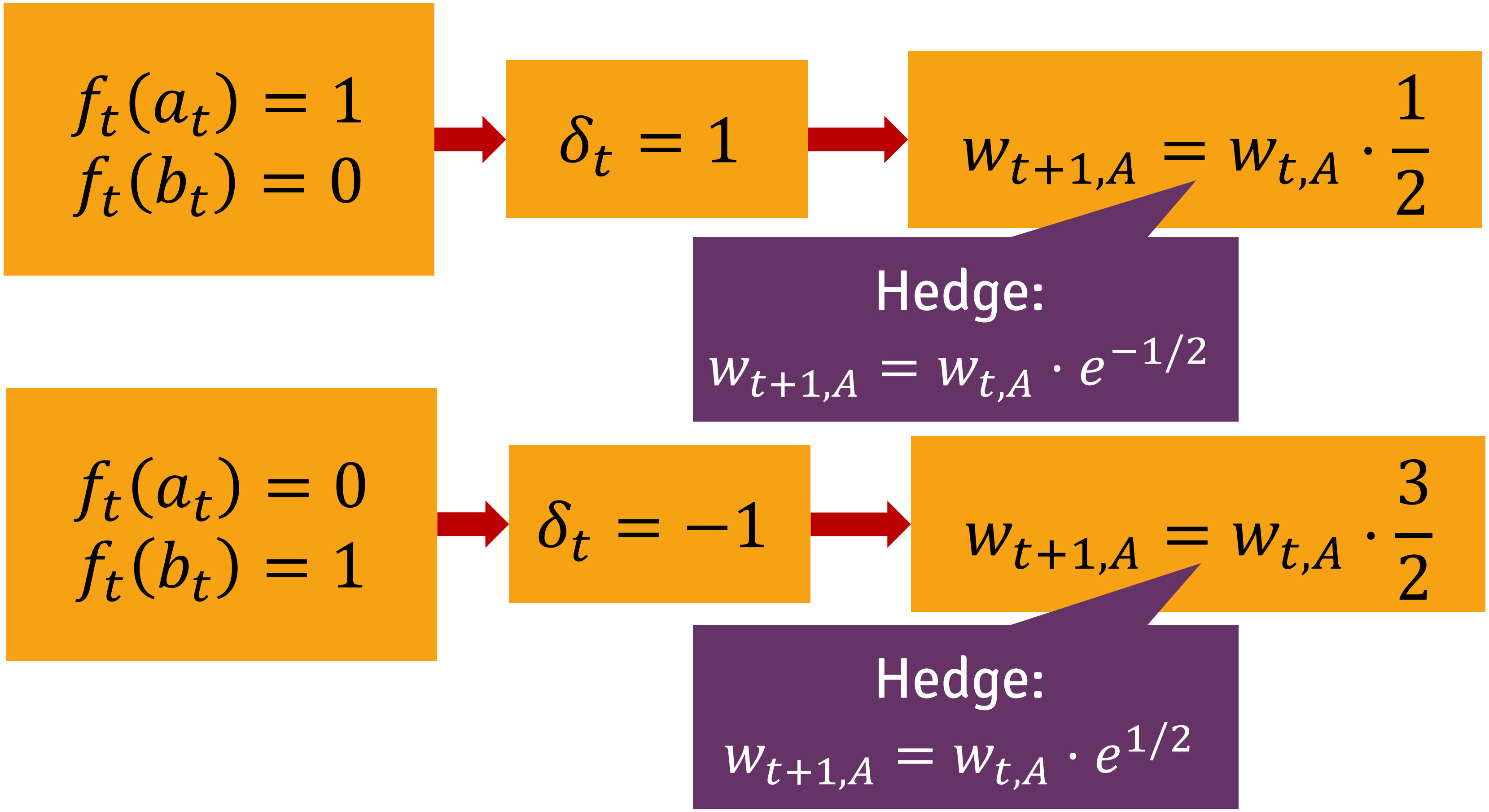
- The asymmetry of $(1 - \eta x)$ is key, e.g., $\eta = \frac{1}{2}$

$$\begin{tikzcd}
\begin{array}{c} f_t(a_t) = 1 \\ f_t(b_t) = 0 \end{array} \arrow[r] & \delta_t = 1 \arrow[r] & w_{t+1,A} = w_{t,A} \cdot \frac{1}{2}
\end{tikzcd}$$

$$\begin{tikzcd}
\begin{array}{c} f_t(a_t) = 0 \\ f_t(b_t) = 1 \end{array} \arrow[r] & \delta_t = -1 \arrow[r] & w_{t+1,A} = w_{t,A} \cdot \frac{3}{2}
\end{tikzcd}$$

Why does it work?

- The asymmetry of $(1 - \eta x)$ is key, e.g., $\eta = \frac{1}{2}$



$f_t(a_t) = 1$
$f_t(b_t) = 0$

$\delta_t = 1$

$w_{t+1,A} = w_{t,A} \cdot \frac{1}{2}$

Hedge:
$w_{t+1,A} = w_{t,A} \cdot e^{-1/2}$

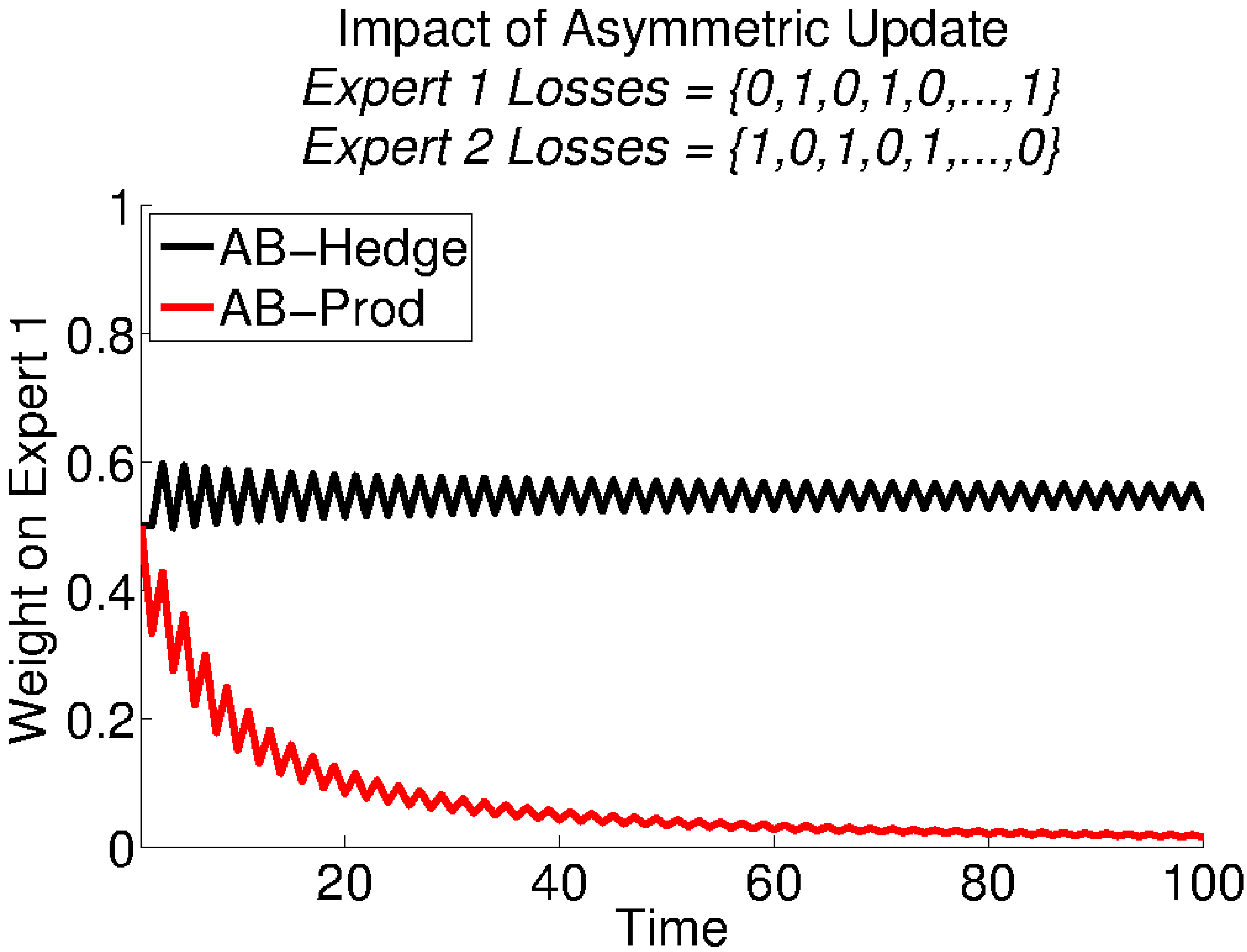
$f_t(a_t) = 0$
$f_t(b_t) = 1$

$\delta_t = -1$

$w_{t+1,A} = w_{t,A} \cdot \frac{3}{2}$

Hedge:
$w_{t+1,A} = w_{t,A} \cdot e^{1/2}$

Why does it work?



Applications

Prediction with expert advice:

- A = Hedge, B = Follow the Leader
- **Regret:** $O(\log N)$ against i.i.d. losses, $O\left(\sqrt{T \log N} + \sqrt{T \log T}\right)$ in worst case

Applications

Prediction with expert advice:

- A = Hedge, B = Follow the Leader
- **Regret:** $O(\log N)$ against i.i.d. losses, $O\left(\sqrt{T \log N} + \sqrt{T \log T}\right)$ in worst case

Matches De Rooij et al. (2014) up to $\sqrt{\log T}$

Applications

Prediction with expert advice:

- $A =$ Hedge, $B =$ Follow the Leader
- **Regret:** $O(\log N)$ against i.i.d. losses, $O\left(\sqrt{T \log N} + \sqrt{T \log T}\right)$ in worst case

Matches De Rooij et al. (2014) up to $\sqrt{\log T}$

Online convex optimization:

- $A =$ OGD* with $\eta = \frac{1}{\sqrt{t}}$, $B =$ OGD with $\eta = \frac{1}{t}$
- **Regret:** $O(\log T)$ against strongly convex losses, $O\left(\sqrt{T \log T}\right)$ in worst case

* Zinkevich (2003), Hazan, Agarwal and Kale (2007)

Applications

Prediction with expert advice:

- A = Hedge, B = Follow the Leader
- **Regret:** $O(\log N)$ against i.i.d. losses, $O\left(\sqrt{T \log N} + \sqrt{T \log T}\right)$ in worst case

Matches De Rooij et al. (2014) up to $\sqrt{\log T}$

Online convex optimization:

- A = OGD* with $\eta = \frac{1}{\sqrt{t}}$, B = OGD with $\eta = \frac{1}{t}$
- **Regret:** $O(\log T)$ against strongly convex losses, $O\left(\sqrt{T \log T}\right)$ in worst case

Matches Bartlett et al. (2007) up to $\sqrt{\log T}$

* Zinkevich (2003), Hazan, Agarwal and Kale (2007)

Applications

Tracking the best expert:

- A = Fixed Share*, B = windowed FTL
- **Regret** measured against best sequence $x_{1:T}$ with K switches: $O(K \log(T/K))$ for piecewise i.i.d., $O\left(\sqrt{KT \log N} + \sqrt{T \log T}\right)$ in worst case

* Herbster and Warmuth (1998)

Applications

Tracking the best expert:

- $A =$ Fixed Share*, $B =$ windowed FTL
- **Regret** measured against best sequence $x_{1:T}$ with K switches: $O(K \log(T/K))$ for piecewise i.i.d., $O\left(\sqrt{KT \log N} + \sqrt{T \log T}\right)$ in worst case

Solves COLT 2014 open problem by Koolen and Warmuth

* Herbster and Warmuth (1998)

Applications

Two-points bandit feedback:

- Observing $f_t(a_t)$ and $f_t(b_t)$ is enough!
- A = EXP3*, B = UCB*
- **Regret:** $O(\log T)$ against i.i.d. losses, $O\left(\sqrt{NT \log N} + \sqrt{T \log T}\right)$ in worst case

* Auer, Cesa-Bianchi, Freund and Schapire (2002), Auer, Cesa-Bianchi and Fischer (2002)

Applications

Two-points bandit feedback:

- Observing $f_t(a_t)$ and $f_t(b_t)$ is enough!
- $A = \text{EXP3*}, B = \text{UCB*}$
- **Regret:** $O(\log T)$ against i.i.d. losses, $O\left(\sqrt{NT \log N} + \sqrt{T \log T}\right)$ in worst case

Much better than the $\log^3 T$ of Seldin and Slivkins (2014), although much less general

* Auer, Cesa-Bianchi, Freund and Schapire (2002), Auer, Cesa-Bianchi and Fischer (2002)

Conclusions

- A generic scheme to combine aggressive and principled algorithms
- Also guarantees that your new solution is essentially always better than your old one

Conclusions

- A generic scheme to combine aggressive and principled algorithms
- Also guarantees that your new solution is essentially always better than your old one
- Open problems:
  - extending to real partial information?
  - reinforcement learning?
  - optimality in every single time window?

Thanks!

Proof

- Directly follows from the Prod analysis:
  - Let $S = \{1,2,\dots,N\}$, $w_{1,i} = \mu_i$ for all i
  - In every round t, choose i w.p. $p_{t,i} \propto w_{t,i}$
  - Loss of expert i in round t: $\ell_{t,i}$
  - $w_{t+1,i} = w_{t,i} \cdot \left(1 - \eta \ell_{t,i}\right)$

Proof

- Directly follows from the Prod analysis:
  - Let $S = \{1,2,\dots,N\}$, $w_{1,i} = \mu_i$ for all i
  - In every round t, choose i w.p. $p_{t,i} \propto w_{t,i}$
  - Loss of expert i in round t: $\ell_{t,i}$
  - $w_{t+1,i} = w_{t,i} \cdot \left(1 - \eta\ell_{t,i}\right)$

Theorem (Cesa-Bianchi, Mansour and Stoltz, 2007):
If $\sum_{i=1}^{N} \mu_i = 1$, then for all i

$$R_{T,i} \leq \frac{\log \mu_i}{\eta} + \eta \sum_{t=1}^{T} \ell_{t,i}^2$$

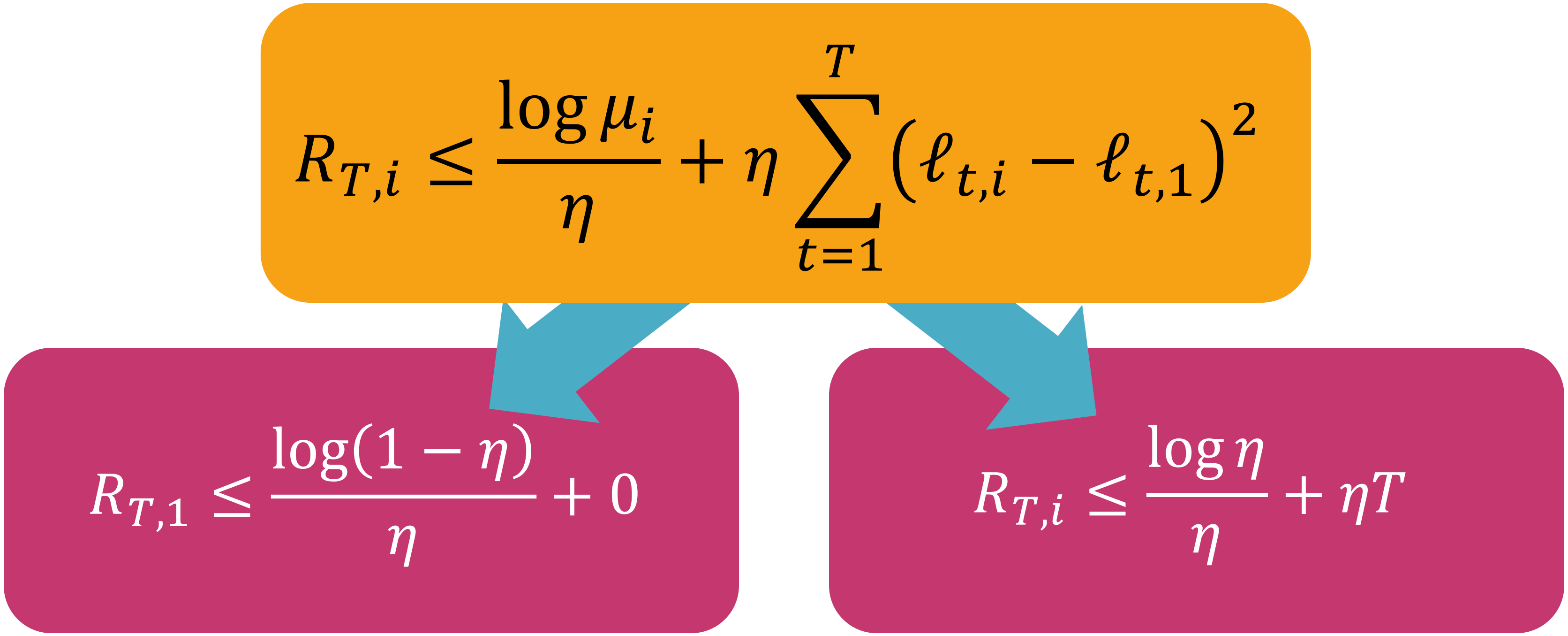
Proof

- **Idea** (Even-Dar, Kearns, Mansour and Wortman, 2008):
  - set $\widehat{\ell}_{t,i} = \ell_{t,i} - \ell_{t,1}$
  - $\mu_1 = 1 - \eta$ and $\mu_i = \frac{\eta}{N-1}$ for $i > 1$

$$\boxed{R_{T,i} \le \frac{\log \mu_i}{\eta} + \eta \sum_{t=1}^{T} \left(\ell_{t,i} - \ell_{t,1}\right)^2}$$

Proof

- **Idea** (Even-Dar, Kearns, Mansour and Wortman, 2008):
  - set $\hat{\ell}_{t,i} = \ell_{t,i} - \ell_{t,1}$
  - $\mu_1 = 1 - \eta$ and $\mu_i = \frac{\eta}{N-1}$ for $i > 1$



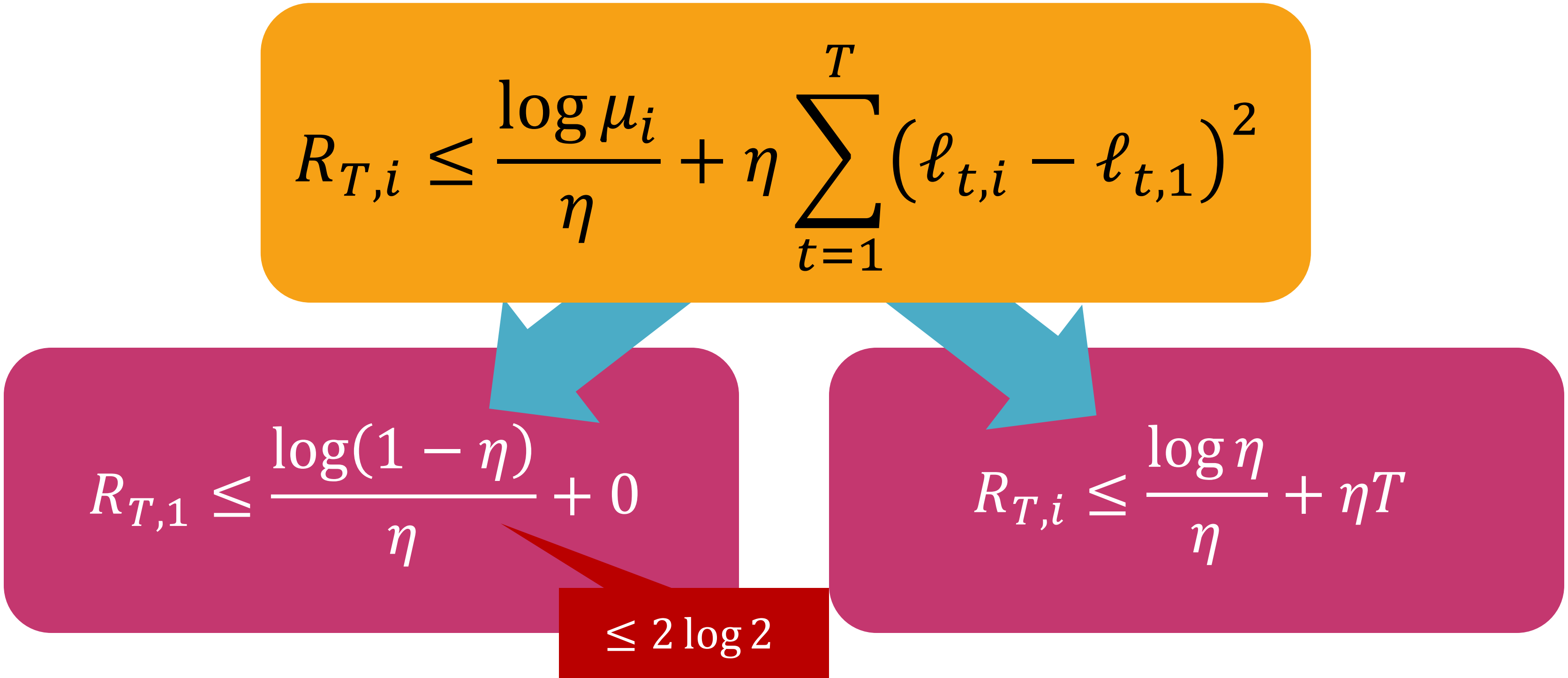
$$R_{T,i} \le \frac{\log \mu_i}{\eta} + \eta \sum_{t=1}^{T} \left(\ell_{t,i} - \ell_{t,1}\right)^2$$

$$R_{T,1} \le \frac{\log(1-\eta)}{\eta} + 0$$

$$R_{T,i} \le \frac{\log \eta}{\eta} + \eta T$$

Proof

- **Idea** (Even-Dar, Kearns, Mansour and Wortman, 2008):
  - set $\hat{\ell}_{t,i} = \ell_{t,i} - \ell_{t,1}$
  - $\mu_1 = 1 - \eta$ and $\mu_i = \frac{\eta}{N-1}$ for $i > 1$



$$R_{T,i} \leq \frac{\log \mu_i}{\eta} + \eta \sum_{t=1}^{T} \left(\ell_{t,i} - \ell_{t,1}\right)^2$$

$$R_{T,1} \leq \frac{\log(1-\eta)}{\eta} + 0$$

$\leq 2 \log 2$

$$R_{T,i} \leq \frac{\log \eta}{\eta} + \eta T$$