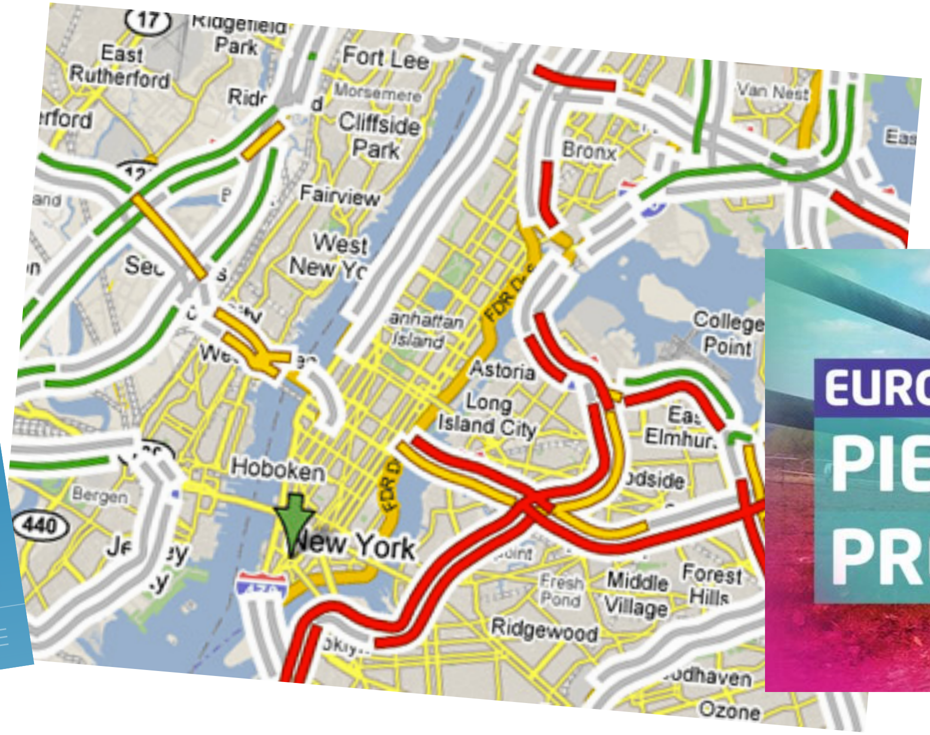
Using Predictions in Online Optimization: Looking Forward with an Eye on the Past

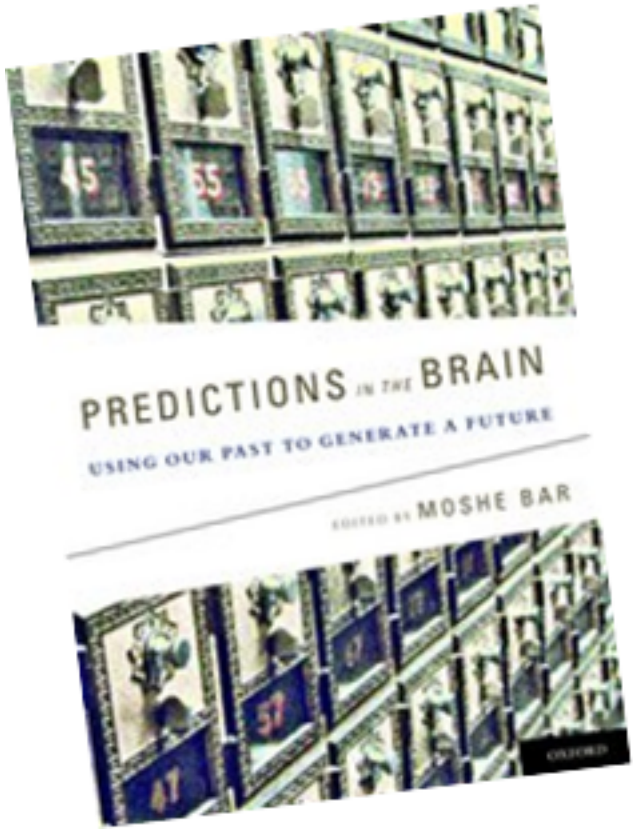
Niangjun Chen

Joint work with Joshua Comden, Zhenhua Liu, Anshul Gandhi, and Adam Wierman

Predictions are crucial for decision making

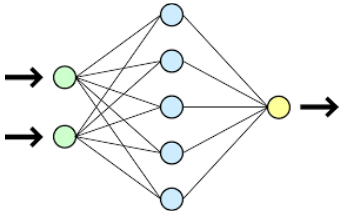


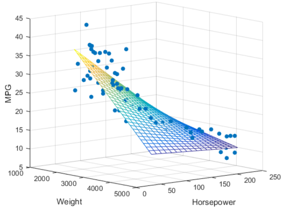
Predictions are crucial for decision making

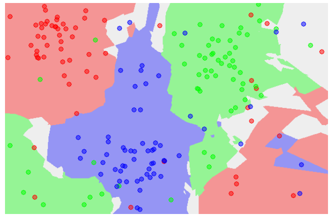


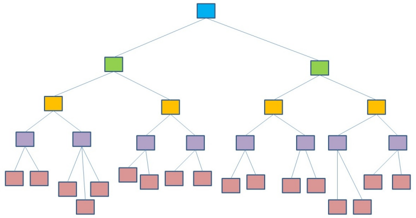
"The human brain, it is being increasingly argued in the scientific literature, is best viewed as an advanced prediction machine."

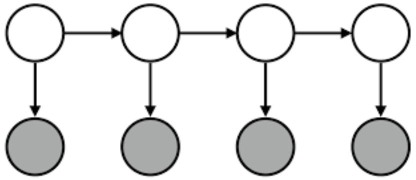
We know how to make predictions











We know how to make predictions

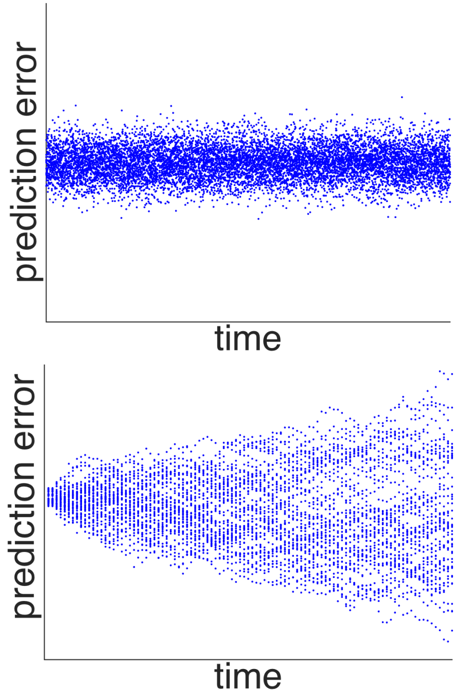
But not how to design algorithms to use prediction

How should an algorithm use predictions if errors are

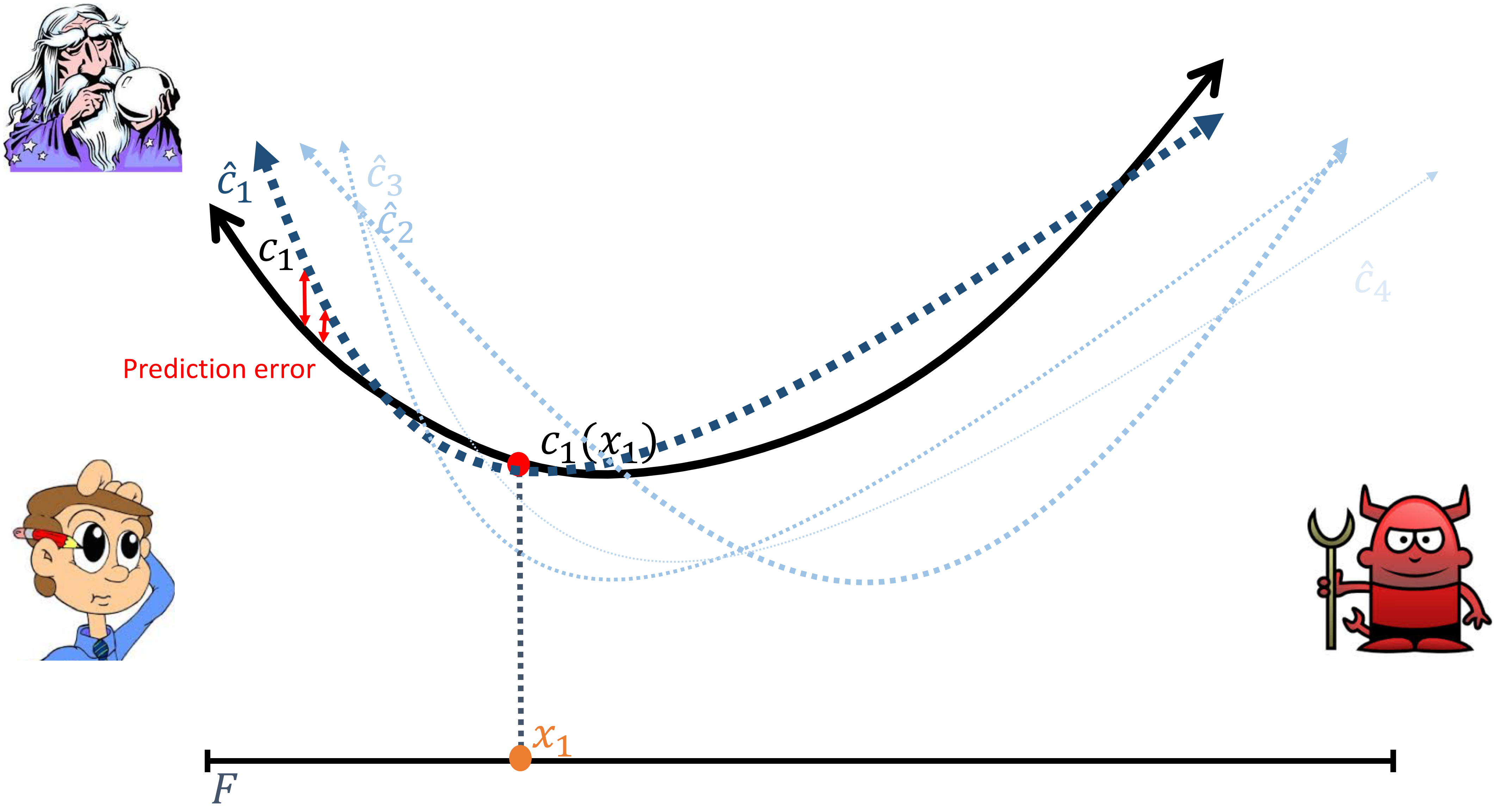
- independent

vs

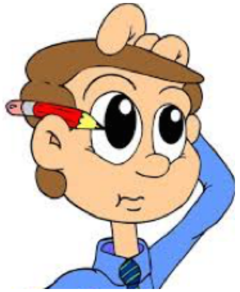
- correlated

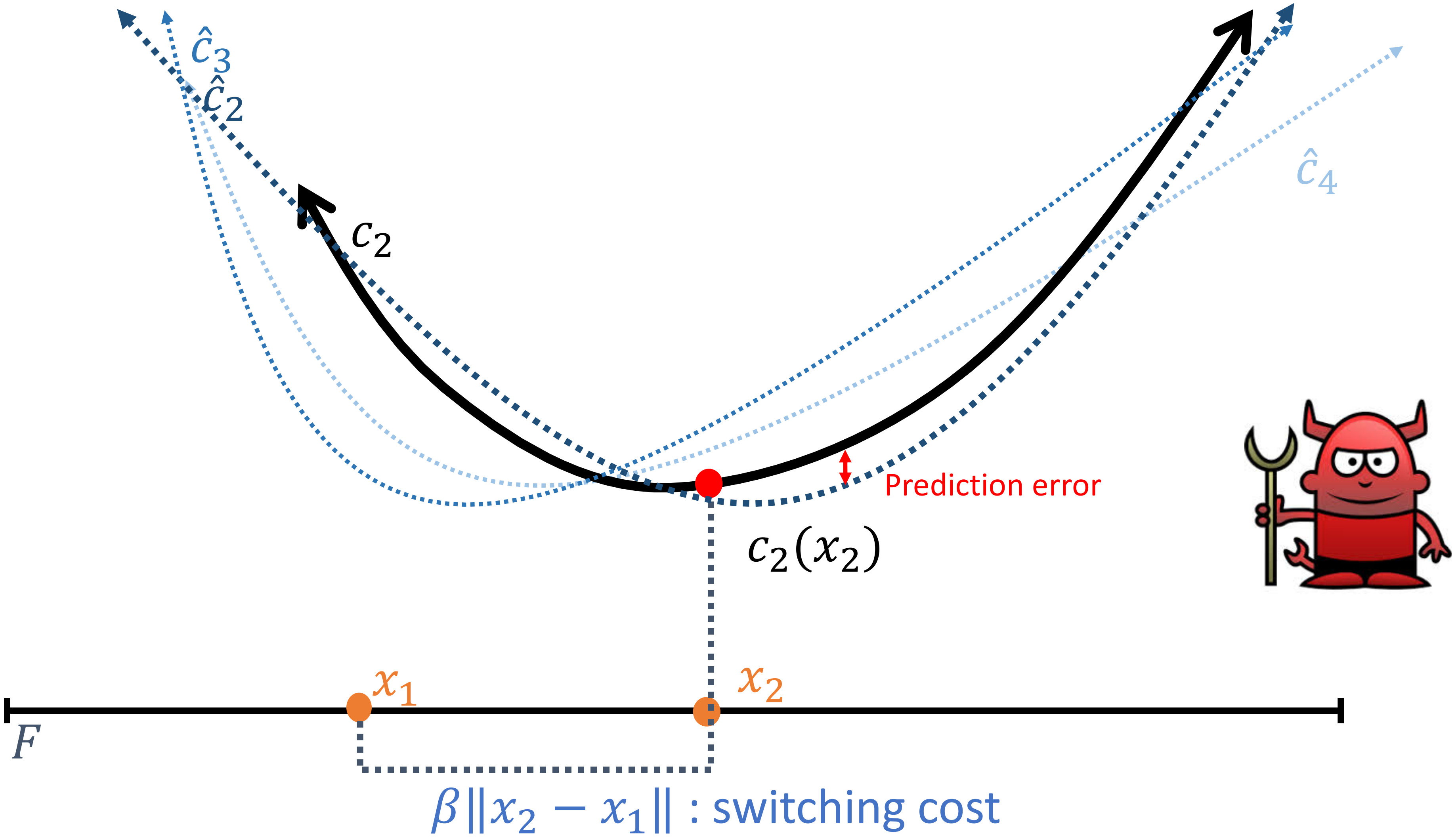


This paper: Online algorithm design with predictions in mind

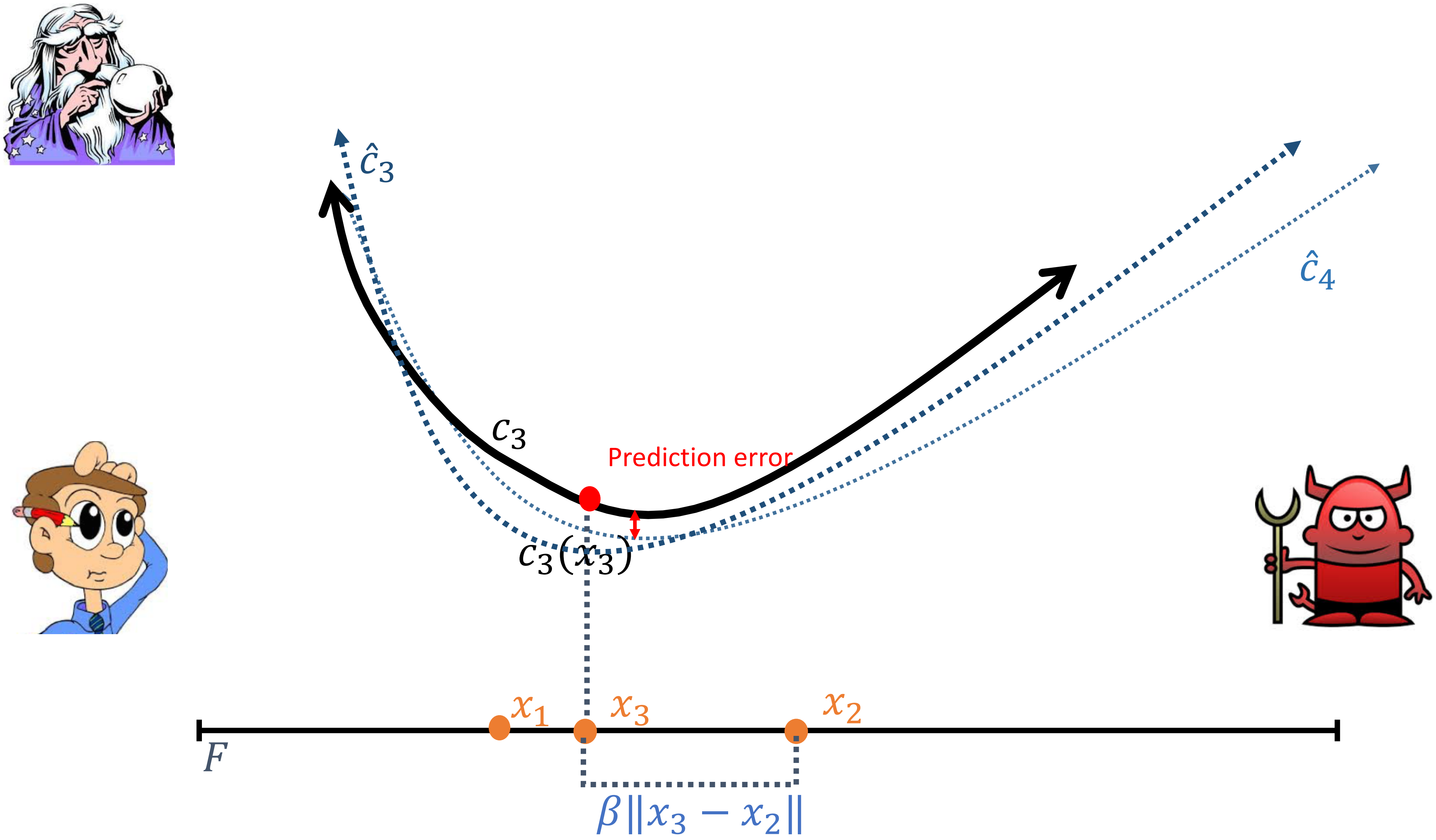


Cost = $c_1(x_1)$





Cost = $c_1(x_1) + \beta\left|\left|x_2 - x_1\right|\right| + c_2(x_2)$



Cost = $c_1(x_1) + \beta\|x_2 - x_1\| + c_2(x_2) + \beta\|x_3 - x_2\| + c_3(x_3)$ ...

Online convex optimization using predictions

$x_1, y_1, x_2, y_2, x_3, y_3 \cdots \leftarrow$ online

$$\min_{x_t \in F} \sum_t \boxed{c(x_t, y_t)} + \underbrace{\beta \| x_t - x_{t-1} \|}_{\text{switching cost}}$$

convex

e.g. online tracking cost $c(x_t, y_t)$

Given prediction of y_t at time τ, $y_{t|\tau}$

Goal: minimize competitive difference

$\text{cost}(ALG) - \text{cost}(OPT)$

Time	Information Available				Decision
1	$y_{1\|0}$	$y_{2\|0}$	$y_{3\|0}$	…	x_1

Prediction noise model

[Gan et al 2013] [Chen et al 2014] [Chen et al 2015]

$$y_t = y_{t|\tau} + \underbrace{\sum_{s=\tau+1}^{t} f(t-s)e(s)}_{\text{prediction error}}$$

Realization that algorithm is trying to track

prediction error

Prediction for time t given to algorithm at time τ

Prediction noise model

[Gan et al 2013] [Chen et al 2014] [Chen et al 2015]

Per-step noise

$$y_t = y_{t|\tau} + \sum_{s=\tau+1}^{t} f(t-s)\boxed{e(s)}$$

How much uncertainty is there one step ahead?

$y_t - y_{t|t-1} = f(0)e(t)$

where $e(t)$ are white, mean zero (unbiased)

and $f(0)=1$, $\mathbb{E}e(t)^2 = \sigma^2$

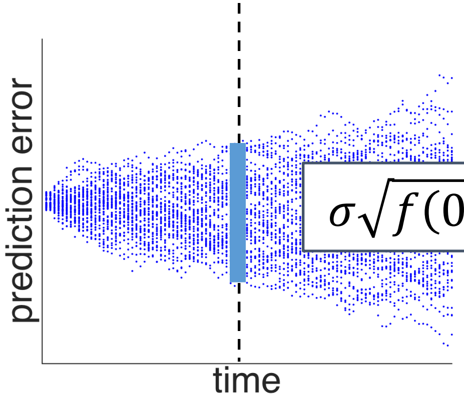
Prediction noise model

[Gan et al 2013] [Chen et al 2014] [Chen et al 2015]

Weighting factor

$$y_t = y_{t|\tau} + \sum_{s=\tau+1}^{t} \boxed{f(t-s)}\, e(s)$$

How important is the noise at time $t - s$ for the prediction of t?



$$\sigma\sqrt{f(0)^2 + \cdots + f(s)^2}$$

$t = s$

Prediction noise model

[Gan et al 2013] [Chen et al 2014] [Chen et al 2015]

$$y_t = y_{t|\tau} + \underbrace{\sum_{s=\tau+1}^{t} f(t-s)e(s)}_{\text{prediction error}}$$

- Predictions are “refined” as time moves forward
- Predictions are more noisy as you look further ahead
- Prediction errors can be correlated
- **Form of errors matches many classical models**

Prediction of wide-sense stationary process using Wiener filter

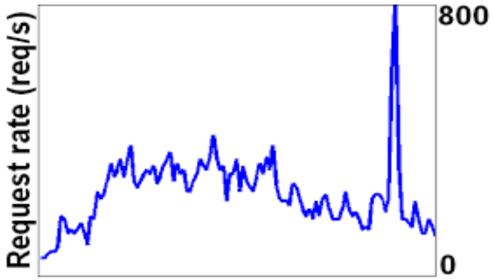
Prediction of linear dynamical system using Kalman filter

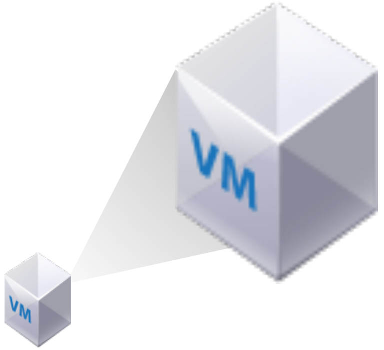
Lots of applications …

Dynamic capacity management in data centers [Gandhi et al. 2012][Lin et al 2013]
Power system generation/load scheduling[Lu et al. 2013]
Portfolio management [Cover 1991][Boyd et al. 2012]
Video streaming [Sen et al. 2000][Liu et al. 2008]
Network routing [Bansal et al. 2003][Kodialam et al. 2003]
Geographical load balancing [Hindman et al. 2011] [Lin et al. 2012]
Visual speech generation [Kim et al. 2015]
…

Lots of applications …

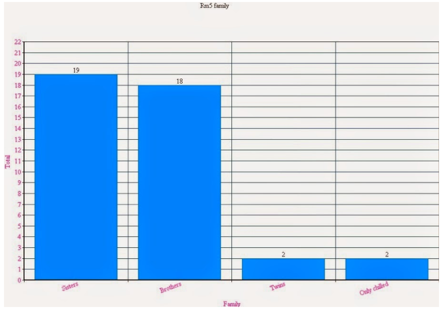
Dynamic capacity management in data centers [Gandhi et al. 2012][Lin et al 2013]
Power system generation/load scheduling[Lu et al. 2013]
Portfolio management [Cover 1991][Boyd et al. 2012]
Video streaming [Sen et al. 2000][Liu et al. 2008]
Network routing [Bansal et al. 2003][Kodialam et al. 2003]
Geographical load balancing [Hindman et al. 2011] [Lin et al. 2012]
Visual speech generation [Kim et al. 2015]
…





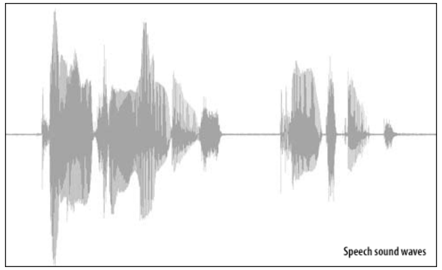
Lots of applications …

Dynamic capacity management in data centers [Gandhi et al. 2012][Lin et al 2013]
Power system generation/load scheduling[Lu et al. 2013]
Portfolio management [Cover 1991][Boyd et al. 2012]
Video streaming [Sen et al. 2000][Liu et al. 2008]
Network routing [Bansal et al. 2003][Kodialam et al. 2003]
Geographical load balancing [Hindman et al. 2011] [Lin et al. 2012]
Visual speech generation [Kim et al. 2015]
…



Lots of applications …

Dynamic capacity management in data centers [Gandhi et al. 2012][Lin et al 2013]
Power system generation/load scheduling[Lu et al. 2013]
Portfolio management [Cover 1991][Boyd et al. 2012]
Video streaming [Sen et al. 2000][Liu et al. 2008]
Network routing [Bansal et al. 2003][Kodialam et al. 2003]
Geographical load balancing [Hindman et al. 2011] [Lin et al. 2012]
Visual speech generation [Kim et al. 2015]
…



Lots of applications… lots of algorithms

Most popular choice by far: Receding Horizon Control (RHC)

[Morari et al 1989][Mayne 1990][Rawling et al 2000][Camacho 2013]…

$$\underbrace{y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t}}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$x_{t+1}, x_{t+2}, \dots, x_{t+w} = \operatorname{argmin}\left\{ \sum_{s=t+1}^{t+w} c(x_t, y_{t|s}) + \beta \|x_t - x_{t-1}\|_1 \right\}$$

Lots of applications… lots of algorithms

Most popular choice by far: Receding Horizon Control (RHC)

[Morari et al 1989][Mayne 1990][Rawling et al 2000][Camacho 2013]…

$$y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$y_{t+2|t+1}, y_{t+3|t+1}, \dots, y_{t+w+1|t+1}, y_{t+w+2|t+1}, y_{t+w+3|t+1}, \cdots$$

$$x_{t+2}, x_{t+3}, \dots x_{t+w+1}$$

Lots of applications… lots of algorithms

Most popular choice by far: Receding Horizon Control (RHC)

[Morari et al 1989][Mayne 1990][Rawling et al 2000][Camacho 2013]…

$$y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$y_{t+2|t+1}, y_{t+3|t+1}, \dots, y_{t+w+1|t+1}, y_{t+w+2|t+1}, y_{t+w+3|t+1}, \cdots$$

$$\underbrace{y_{t+3|t+2}, y_{t+4|t+2}, \dots, y_{t+w+2|t+2}}_{\downarrow}, y_{t+w+3|t+2}, y_{t+w+4|t+2}, \cdots$$

$$x_{t+3}, x_{t+4}, \dots x_{t+w+2}$$

Lots of applications… lots of algorithms

Most popular choice by far: Receding Horizon Control (RHC)

[Morari et al 1989][Mayne 1990][Rawling et al 2000][Camacho 2013]…

Recent suggestion: Averaging Fixed Horizon Control (AFHC)
[Lin et al 2012] [Chen et al 2015] [Kim et al 2015]

Averaging Fixed Horizon Control

Fixed Horizon Control (FHC)

$$\boxed{y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t}}$$

$$x_{t+1}, x_{t+2}, \dots, x_{t+w} = \operatorname{argmin}\left\{\sum_{s=t+1}^{t+w} c(x_t, y_{t|s}) + \beta \|x_t - x_{t-1}\|_1\right\}$$

Averaging Fixed Horizon Control

Fixed Horizon Control (FHC)

$y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t}, y_{t+w+1|t+w}, y_{t+w+2|t+w}, \cdots$

$x_{t+1}, x_{t+2}, \dots, x_{t+w} \quad x_{t+w+1}, x_{t+w+2}, \dots, x_{t+2w}$

Averaging Fixed Horizon Control

Average choices of FHC algorithms

$$x_{AFHC} = \frac{1}{\mathrm{w}} \sum_{k=1}^{w} x_{FHC}^{(k)}$$

w FHC algorithms

$$\left\{ \begin{array}{l} x_{t-2}^1, x_{t-1}^1, x_t^1 \dots, x_{t+w-2}^1 \quad x_{t+w-1}^1, \dots, \\ x_{t-1}^2, x_t^2, x_{t+4}^2 \dots, x_{t+w-1}^2 \quad x_{t+w}^2, \dots, \\ x_t^3, x_{t+4}^3, x_{t+5}^3 \dots, x_{t+w}^3 \quad x_{t+w+1}^3, \dots, \\ \vdots \\ x_t^w \end{array} \right.$$

Algorithms Using Noisy Prediction

Most popular choice by far: Receding Horizon Control (RHC)

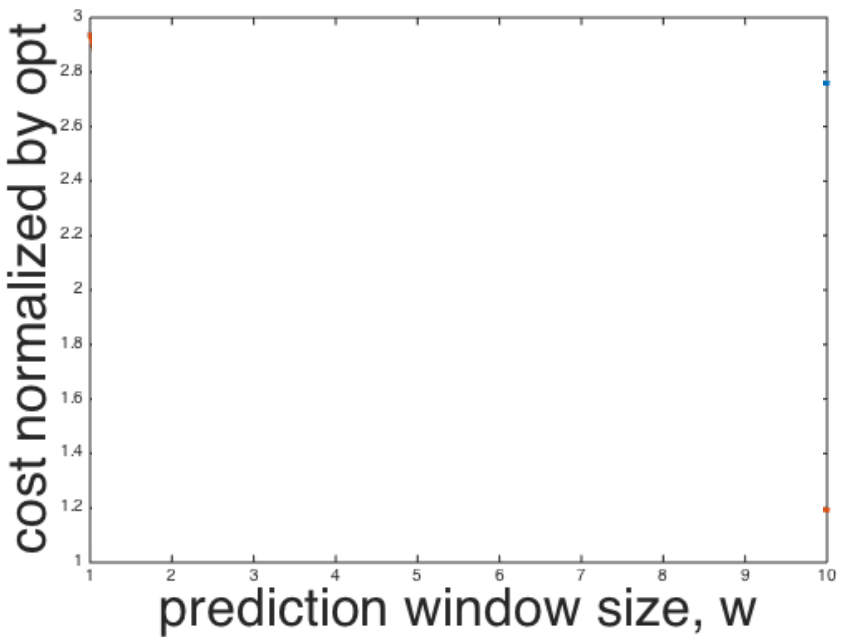
[Morari et al 1989][Mayne 1990][Rawling et al 2000][Camacho 2013]…

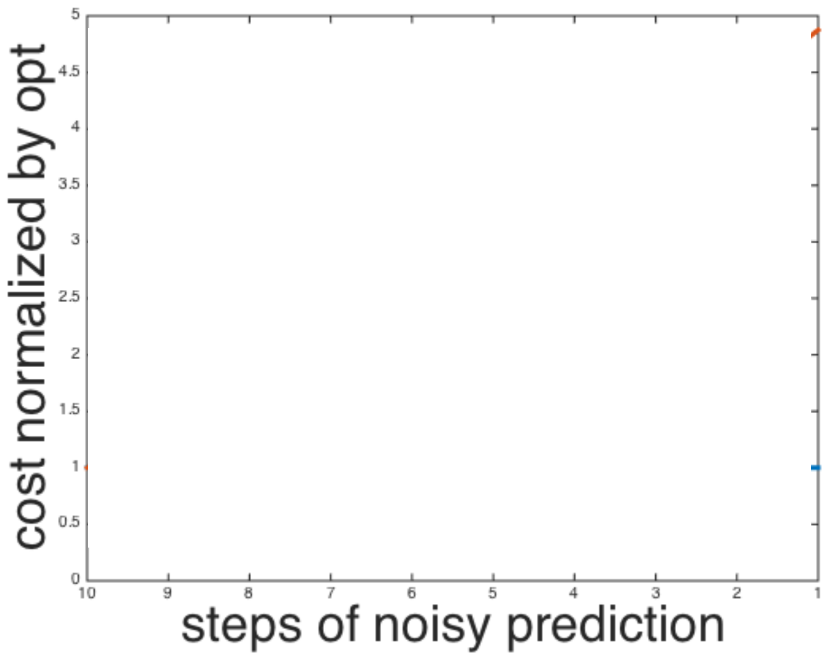
Recent suggestion: Averaging Fixed Horizon Control (AFHC)

[Lin et al 2012] [Chen et al 2015] [Kim et al 2015]

Which algorithm is better? Unclear…

AFHC and RHC have vastly different behavior

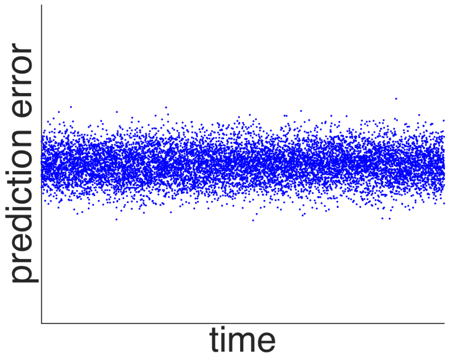


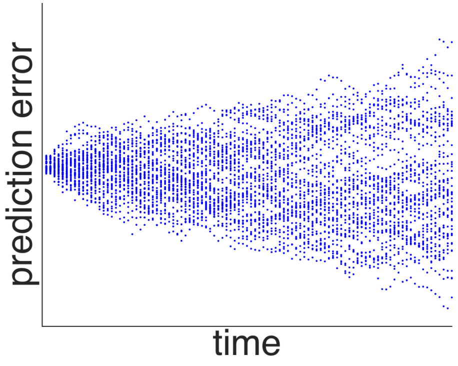


Empirically
- AFHC is better in worst case under perfect predictions
- RHC is better in stochastic case when prediction errors are correlated

This paper: Online algorithm design with predictions in mind

➡ How to design algorithm optimal for prediction noise?





Three key design choices

1. How far to look-ahead in making decisions?

$$\boxed{y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t},} y_{t+w+1|t}, y_{t+w+2|t}, \dots$$

Lookahead w steps

Three key design choices

1. How far to look-ahead in making decisions?
2. How many actions to commit?

$$\boxed{y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t},} y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$x_{t+1}, x_{t+2}, \dots, x_{t+w} = \text{argmin}\left\{\sum_{s=t+1}^{t+w} c(x_t, y_{t|s}) + \beta\left|\left|x_t - x_{t-1}\right|\right|_1\right\}$$

Three key design choices

1. How far to look-ahead in making decisions?
2. How many actions to commit?

$$\boxed{y_{t+1|t}, y_{t+2|t}, \dots, y_{t+w|t}}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$\underbrace{x_{t+1}, x_{t+2}, \dots}_{\text{commits } v \text{ steps}}, x_{t+w} = \operatorname{argmin}\left\{\sum_{s=t+1}^{t+w} c(x_t, y_{t|s}) + \beta \left\|x_t - x_{t-1}\right\|_1\right\}$$

Three key design choices

1. How far to look-ahead in making decisions?
2. How many actions to commit?
3. How to aggregate action plans?

Our focus: what is the optimal commitment level given the structure of prediction noise?

$$
\begin{array}{l}
x_{t-2}^1, x_{t-1}^1, x_t^1 \dots, x_{t+w-2}^1 \\
\quad x_{t-1}^2, x_t^2, x_{t+4}^2 \dots, x_{t+w-1}^2 \\
\qquad x_t^3, x_{t+4}^3, x_{t+5}^3 \dots, x_{t+w}^3
\end{array}
$$

$$x_t = g(x_t^1, x_t^2, x_t^3)$$

Key: commitment balances switching cost and prediction errors

Committed Horizon Control

FHC with limited commitment v, for $t \equiv k \bmod w$

$$\boxed{y_{t+1|t}, y_{t+2|t}, \dots y_{t+v|t}, y_{t+v+1|t}, \dots, y_{t+w|t}}, y_{t+w+1|t}, y_{t+w+2|t}, \dots$$

$$x_{t+1}, x_{t+2}, \dots, x_{t+v}, x_{t+v+1}, \dots, x_{t+w} = \operatorname{argmin}\left\{\sum_{s=t+1}^{t+w} c(x_t, y_{t|s}) + \beta\left|\left|x_t - x_{t-1}\right|\right|_1\right\}$$

$$x^{(k)} = (\dots, x_{t+1}^{k}, x_{t+2}^{k}, \dots, x_{t+v}^{k},)$$

Committed Horizon Control

FHC with limited commitment v, for $t \equiv k \bmod w$

$y_{t+1|t}, y_{t+2|t}, \cdots y_{t+v|t}, y_{t+v+1|t}, \cdots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$

$\underbrace{y_{t+v+1|t+v}, \cdots y_{t+2v|t+v}, y_{t+2v+1|t+v} \cdots, y_{t+v+w|t+v}}, y_{t+v+w+1|t+v}, \cdots$

$x_{t+v+1}, x_{t+v+2}, \dots, x_{t+2v}, x_{t+2v+1}, \dots, x_{t+w+v}$

$$x^{(k)} = (\dots, x^k_{t+1}, x^k_{t+2}, \dots, x^k_{t+v}, x^k_{t+v+1}, x^k_{t+v+2}, \dots, x^k_{t+2v})$$

Committed Horizon Control

FHC with limited commitment v, for $t \equiv k \bmod v$

$y_{t+1|t}, y_{t+2|t}, \cdots y_{t+v|t}, y_{t+v+1|t}, \cdots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$

$y_{t+v+1|t+v}, \cdots y_{t+2v|t+v}, y_{t+2v+1|t+v} \cdots, y_{t+v+w|t+v}, y_{t+v+w+1|t+v}, \cdots$

$\underbrace{y_{t+2v+1|t+2v}, \cdots y_{t+3v|t+2v}, \cdots, y_{t+2v+w|t+v}}, y_{t+2v+w+1|t+2v}, \cdots$

$x_{t+2v+1}, x_{t+2v+2}, \ldots, x_{t+3v}, x_{t+3v+1}, \ldots, x_{t+w+2v}$

$$x^{(k)} = (\ldots, x^k_{t+1}, x^k_{t+2}, \ldots, x^k_{t+v}, x^k_{t+v+1}, x^k_{t+v+2}, \ldots, x^k_{t+2v}, x^k_{t+2v+1}, x^k_{t+2v+2}, \ldots, x^k_{t+3v})$$

Committed Horizon Control

FHC with limited commitment v, for $t \equiv k \bmod v$

$y_{t+1|t}, y_{t+2|t}, \cdots y_{t+v|t}, y_{t+v+1|t}, \cdots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$

$y_{t+v+1|t+v}, \cdots y_{t+2v|t+v}, y_{t+2v+1|t+v} \cdots, y_{t+v+w|t+v}, y_{t+v+w+1|t+v}, \cdots$

$\underbrace{y_{t+2v+1|t+2v}, \cdots y_{t+3v|t+2v}, \cdots, y_{t+2v+w|t+v}}, y_{t+2v+w+1|t+2v}, \cdots$

$\downarrow$

$x_{t+2v+1}, x_{t+2v+2}, \dots, x_{t+3v}, x_{t+3v+1}, \dots, x_{t+w+2v}$

$$x^{(k)} = (\dots, x_{t+1}^k, x_{t+2}^k, \dots, x_{t+v}^k, x_{t+v+1}^k, x_{t+v+2}^k, \dots, x_{t+2v}^k, x_{t+2v+1}^k, x_{t+2v+2}^k, \dots, x_{t+3v}^k, \dots)$$

Committed Horizon Control

FHC with limited commitment v, for $t \equiv k \bmod v$

$$y_{t+1|t}, y_{t+2|t}, \cdots y_{t+v|t}, y_{t+v+1|t}, \cdots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$y_{t+v+1|t+v}, \cdots y_{t+2v|t+v}, y_{t+2v+1|t+v} \cdots, y_{t+v+w|t+v}, y_{t+v+w+1|t+v}, \cdots$$

$$y_{t+2v+1|t+2v}, \cdots y_{t+3v|t+2v}, \cdots, y_{t+2v+w|t+v}, y_{t+2v+w+1|t+2v}, \cdots$$

$$x_{t+2v+1}, x_{t+2v+2}, \dots, x_{t+3v}, x_{t+3v+1}, \dots, x_{t+w+2v}$$

$$\left.\begin{aligned}
x^{(1)} &= (\dots, x^1_{t+1}, x^1_{t+2}, \dots, x^1_{t+v}, x^1_{t+v+1}, x^1_{t+v+2}, \dots, x^1_{t+2v}, x^1_{t+2v+1}, x^1_{t+2v+2}, \dots, x^1_{t+3v}, \dots) \\
&\vdots \\
x^{(k)} &= (\dots, x^k_{t+1}, x^k_{t+2}, \dots, x^k_{t+v}, x^k_{t+v+1}, x^k_{t+v+2}, \dots, x^k_{t+2v}, x^k_{t+2v+1}, x^k_{t+2v+2}, \dots, x^k_{t+3v}, \dots) \\
&\vdots \\
x^{(v)} &= (\dots, x^v_{t+1}, x^v_{t+2}, \dots, x^v_{t+v}, x^v_{t+v+1}, x^v_{t+v+2}, \dots, x^v_{t+2v}, x^v_{t+2v+1}, x^v_{t+2v+2}, \dots, x^v_{t+3v}, \dots)
\end{aligned}\right\} v \text{ FHC(v) algorithms}$$

Committed Horizon Control

FHC with limited commitment v, for $t \equiv k \bmod v$

$$y_{t+1|t}, y_{t+2|t}, \cdots y_{t+v|t}, y_{t+v+1|t}, \cdots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$y_{t+v+1|t+v}, \cdots y_{t+2v|t+v}, y_{t+2v+1|t+v} \cdots, y_{t+v+w|t+v}, y_{t+v+w+1|t+v}, \cdots$$

$$\underbrace{y_{t+2v+1|t+2v}, \cdots y_{t+3v|t+2v}, \cdots, y_{t+2v+w|t+v}}, y_{t+2v+w+1|t+2v}, \cdots$$

$$\downarrow$$

$$x_{t+2v+1}, x_{t+2v+2}, \ldots, x_{t+3v}, x_{t+3v+1}, \ldots, x_{t+w+2v}$$

$$\left.\begin{aligned}
x^{(1)} &= (\ldots, x^1_{t+1}, x^1_{t+2}, \ldots, x^1_{t+v}, x^1_{t+v+1}, x^1_{t+v+2}, \ldots, x^1_{t+2v}, x^1_{t+2v+1}, x^1_{t+2v+2}, \ldots, x^1_{t+3v}, \ldots) \\
&\vdots \\
x^{(k)} &= (\ldots, x^k_{t+1}, x^k_{t+2}, \ldots, x^k_{t+v}, x^k_{t+v+1}, x^k_{t+v+2}, \ldots, x^k_{t+2v}, x^k_{t+2v+1}, x^k_{t+2v+2}, \ldots, x^k_{t+3v}, \ldots) \\
&\vdots \\
x^{(v)} &= (\ldots, x^v_{t+1}, x^v_{t+2}, \ldots, x^v_{t+v}, x^v_{t+v+1}, x^v_{t+v+2}, \ldots, x^v_{t+2v}, x^v_{t+2v+1}, x^v_{t+2v+2}, \ldots, x^v_{t+3v}, \ldots)
\end{aligned}\right\} v \text{ FHC(v) algorithms}$$

$$x_{CHC}(t) = \frac{1}{v} \sum_{k=1}^{v} x_t^k$$

Committed Horizon Control

FHC with limited commitment v, for $t \equiv k \bmod v$

$$y_{t+1|t}, y_{t+2|t}, \cdots y_{t+v|t}, y_{t+v+1|t}, \cdots, y_{t+w|t}, y_{t+w+1|t}, y_{t+w+2|t}, \cdots$$

$$y_{t+v+1|t+v}, \cdots y_{t+2v|t+v}, y_{t+2v+1|t+v} \cdots, y_{t+v+w|t+v}, y_{t+v+w+1|t+v}, \cdots$$

$$y_{t+2v+1|t+2v}, \cdots y_{t+3v|t+2v}, \cdots, y_{t+2v+w|t+v}, y_{t+2v+w+1|t+2v}, \cdots$$

$$x_{t+2v+1}, x_{t+2v+2}, \dots, x_{t+3v}, x_{t+3v+1}, \dots, x_{t+w+2v}$$

$$x^{(1)} = (\dots, \boxed{x^1_{t+1}}, x^1_{t+2}, \dots, x^1_{t+v}, x^1_{t+v+1}, x^1_{t+v+2}, \dots, x^1_{t+2v}, x^1_{t+2v+1}, x^1_{t+2v+2}, \dots, x^1_{t+3v}, \dots)$$

$$\vdots$$

$$x^{(k)} = (\dots, \boxed{x^k_{t+1}}, x^k_{t+2}, \dots, x^k_{t+v}, x^k_{t+v+1}, x^k_{t+v+2}, \dots, x^k_{t+2v}, x^k_{t+2v+1}, x^k_{t+2v+2}, \dots, x^k_{t+3v}, \dots)$$

$$\vdots$$

$$x^{(v)} = (\dots, \boxed{x^v_{t+1}}, x^v_{t+2}, \dots, x^v_{t+v}, x^v_{t+v+1}, x^v_{t+v+2}, \dots, x^v_{t+2v}, x^v_{t+2v+1}, x^v_{t+2v+2}, \dots, x^v_{t+3v}, \dots)$$

v FHC(v) algorithms

$$x_{CHC}(t) = \frac{1}{v} \sum_{k=1}^{v} x^k_t$$

$v = 1 \Rightarrow$ RHC,

$v = w \Rightarrow$ AFHC

Main Result

Theorem

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\text{cost}(CHC) - \mathbf{E}\text{cost}(OPT) \leq \frac{2T\beta D}{v} + \frac{2GT}{v}\sum_{k=1}^{v}||f_k||^{\alpha}$$

Main Result

<u>Theorem</u>

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\underbrace{\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT)}_{\text{Competitive difference}} \leq \frac{2T\beta D}{v} + \frac{2GT}{v}\sum_{k=1}^{v} ||f_k||^{\alpha}$$

Main Result

<u>Theorem</u>

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \leq \frac{2T\beta D}{\boxed{v}} + \frac{2GT}{\boxed{v}} \sum_{k=1}^{\boxed{v}} ||f_k||^{\alpha}$$

Commitment level v $\uparrow$ $\downarrow$?

Main Result

Theorem

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \leq \underbrace{\frac{2T\beta D}{v}}_{\text{Due to switching cost}} + \frac{2GT}{v}\sum_{k=1}^{v}||f_k||^{\alpha}$$

Main Result

<u>Theorem</u>

$||x_1 - x_2|| \leq D, \forall x_1, x_2 \in F$

For c that is α-Hölder continuous in the second argument and feasible set $\boxed{F \text{ is bounded,}}$

$$\mathbf{E}\text{cost}(CHC) - \mathbf{E}\text{cost}(OPT) \leq \underbrace{\frac{2T\beta \boxed{D}}{v}}_{\text{Due to switching cost}} + \frac{2GT}{v}\sum_{k=1}^{v}||f_k||^{\alpha}$$

Main Result

Theorem

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\text{cost}(CHC) - \mathbf{E}\text{cost}(OPT) \leq \underbrace{\frac{2T\beta D}{v}}_{\substack{\text{Due to}\\\text{switching cost}}} + \underbrace{\frac{2GT}{v}\sum_{k=1}^{v}||f_k||^{\alpha}}_{\substack{\text{Due to}\\\text{prediction error}}}$$

Main Result

Theorem

$|c(x, y_1) - c(x, y_2)| \leq G||y_1 - y_2||^{\alpha}, \forall x, y_1, y_2$

For c that is $\boxed{\alpha\text{-Hölder continuous}}$ in the second argument and feasible set F is bounded,

$$\mathbf{E}\text{cost}(CHC) - \mathbf{E}\text{cost}(OPT) \leq \underbrace{\frac{2T\beta D}{v}}_{\text{Due to switching cost}} + \underbrace{\frac{2\boxed{G}T}{v} \sum_{k=1}^{v} ||f_k||^{\boxed{\alpha}}}_{\text{Due to prediction error}}$$

Main Result

$$||f_k||^2 \triangleq \mathbf{E}\left|\left|y_{t+k} - y_{t+k|t}\right|\right|^2 = \sigma^2 \sum_{s=1}^{k} f(s)^2$$

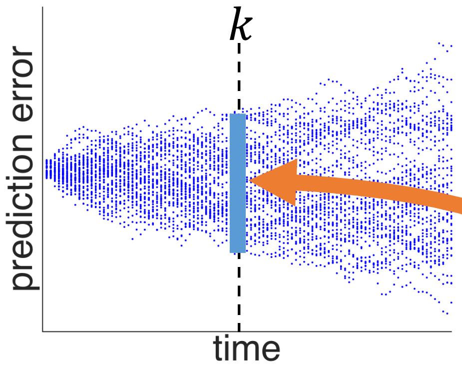
Prediction error k-steps away

<u>Theorem</u>

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \leq \underbrace{\frac{2T\beta D}{v}}_{\text{Due to switching cost}} + \underbrace{\frac{2GT}{v} \sum_{k=1}^{v} \boxed{||f_k||}^{\alpha}}_{\text{Due to prediction error}}$$

Main Result



Theorem

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \le \underbrace{\frac{2T\beta D}{v}}_{\text{Due to switching cost}} + \underbrace{\frac{2GT}{v}\sum_{k=1}^{v}\boxed{||f_k||}^{\alpha}}_{\text{Due to prediction error}}$$

Main Result

<u>Theorem</u>

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \le \underbrace{\frac{2T\beta D}{\boxed{v}}}_{\text{Due to switching cost}} + \underbrace{\frac{2GT}{\boxed{v}} \sum_{k=1}^{\boxed{v}} ||f_k||^{\alpha}}_{\text{Due to prediction error}}$$

Commitment level v ↑ ↓ ?

Main Result

Theorem

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\text{cost}(CHC) - \mathbf{E}\text{cost}(OPT) \le \underbrace{\frac{2T\beta D}{\boxed{v}}}_{\text{Due to switching cost}} + \underbrace{\frac{2GT}{\boxed{v}} \sum_{k=1}^{\boxed{v}} ||f_k||^{\alpha}}_{\text{Due to prediction error}}$$

Commitment level v ⬆ ⬇ ?

Key: choose commitment level v to balance these two terms

Main Result

<u>Theorem</u>

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

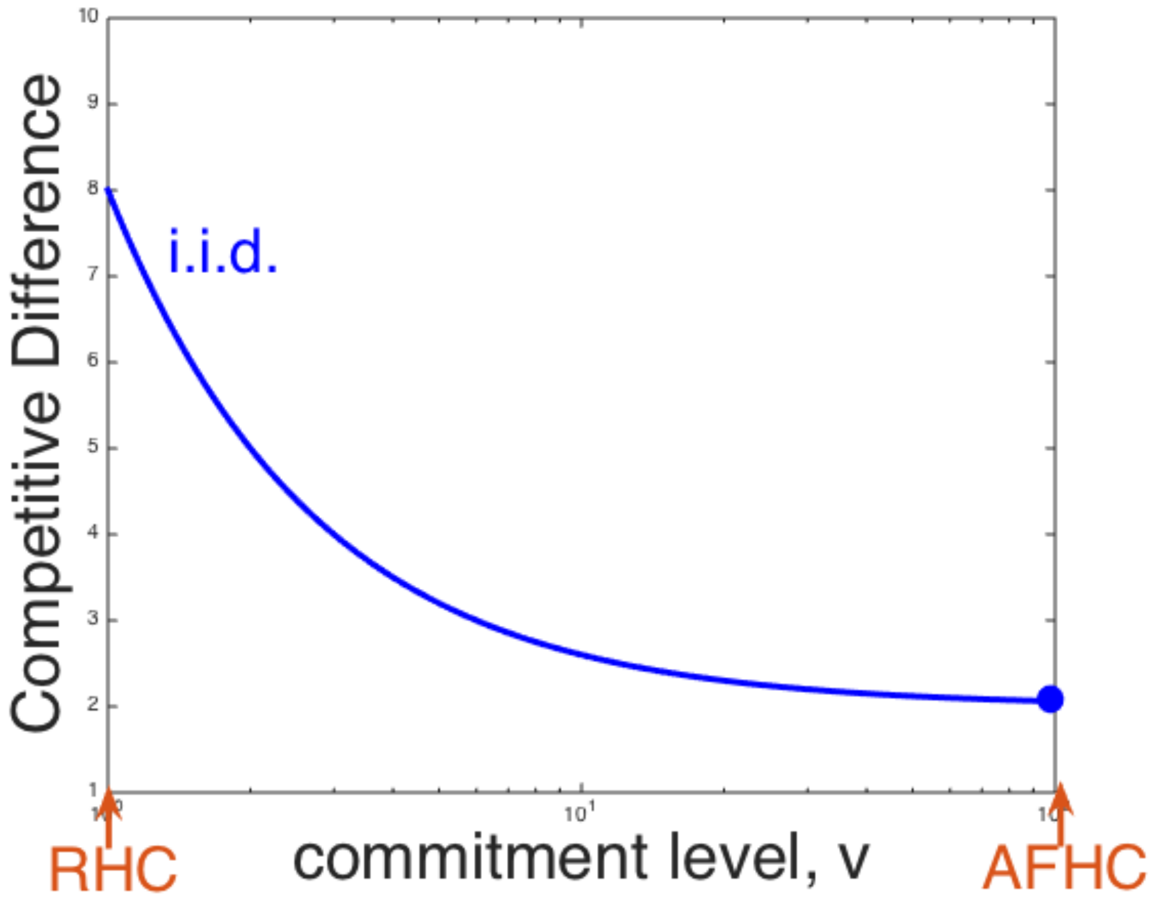
$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \leq \frac{2T\beta D}{v} + \frac{2GT}{v}\sum_{k=1}^{v} ||f_k||^{\alpha}$$

$$\boxed{\text{e.g. i.i.d. noise } f(s) = \begin{cases} 1, s = 0 \\ 0, s > 0 \end{cases}} \quad = \underbrace{\frac{2T\beta D}{v} + 2GT\sigma^{\alpha}}_{\text{Decreasing function of } v}$$

AFHC is best when noise is i.i.d

i.i.d. prediction noise

$$f(s) = \begin{cases} 1, s = 0 \\ 0, s > 0 \end{cases}$$

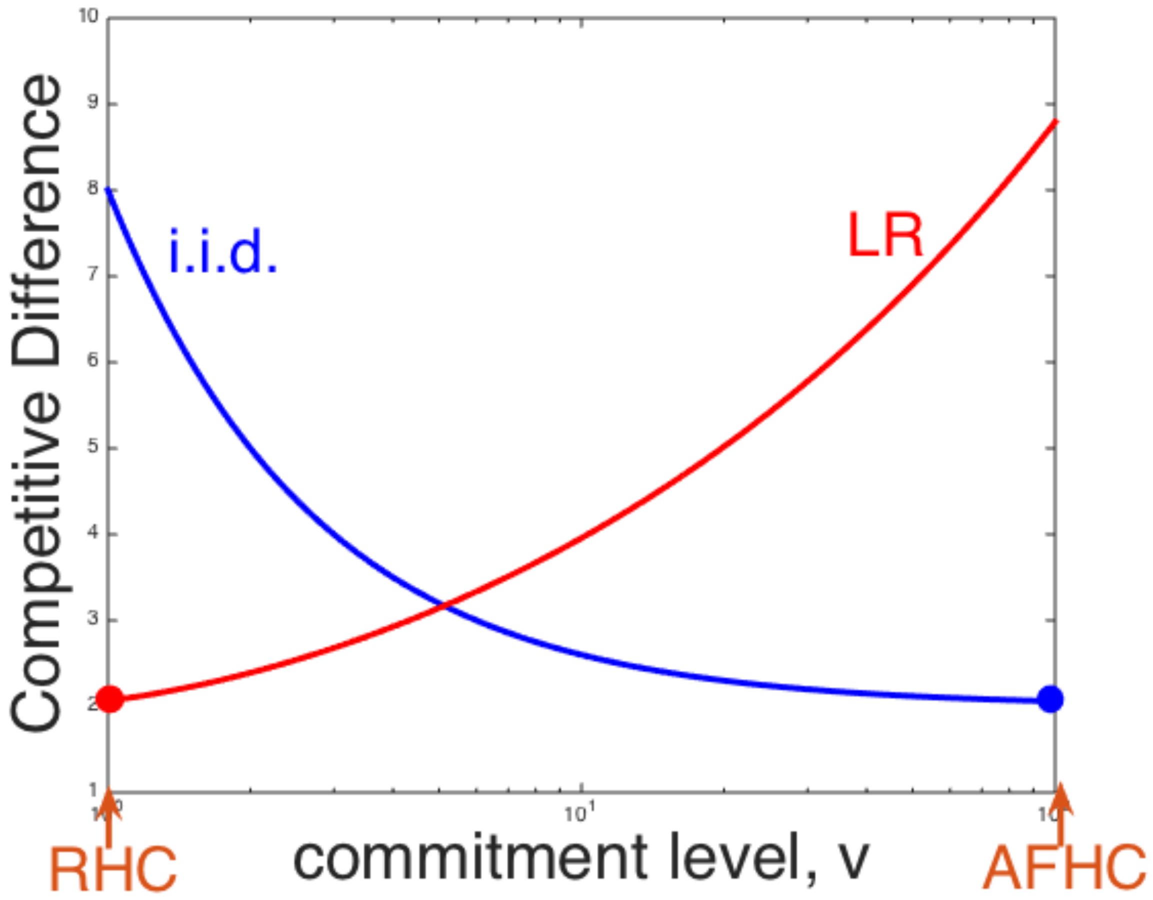


i.i.d. prediction noise

$$f(s) = \begin{cases} 1, s = 0 \\ 0, s > 0 \end{cases}$$

Long range correlated

$$f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L > w$$



i.i.d. prediction noise

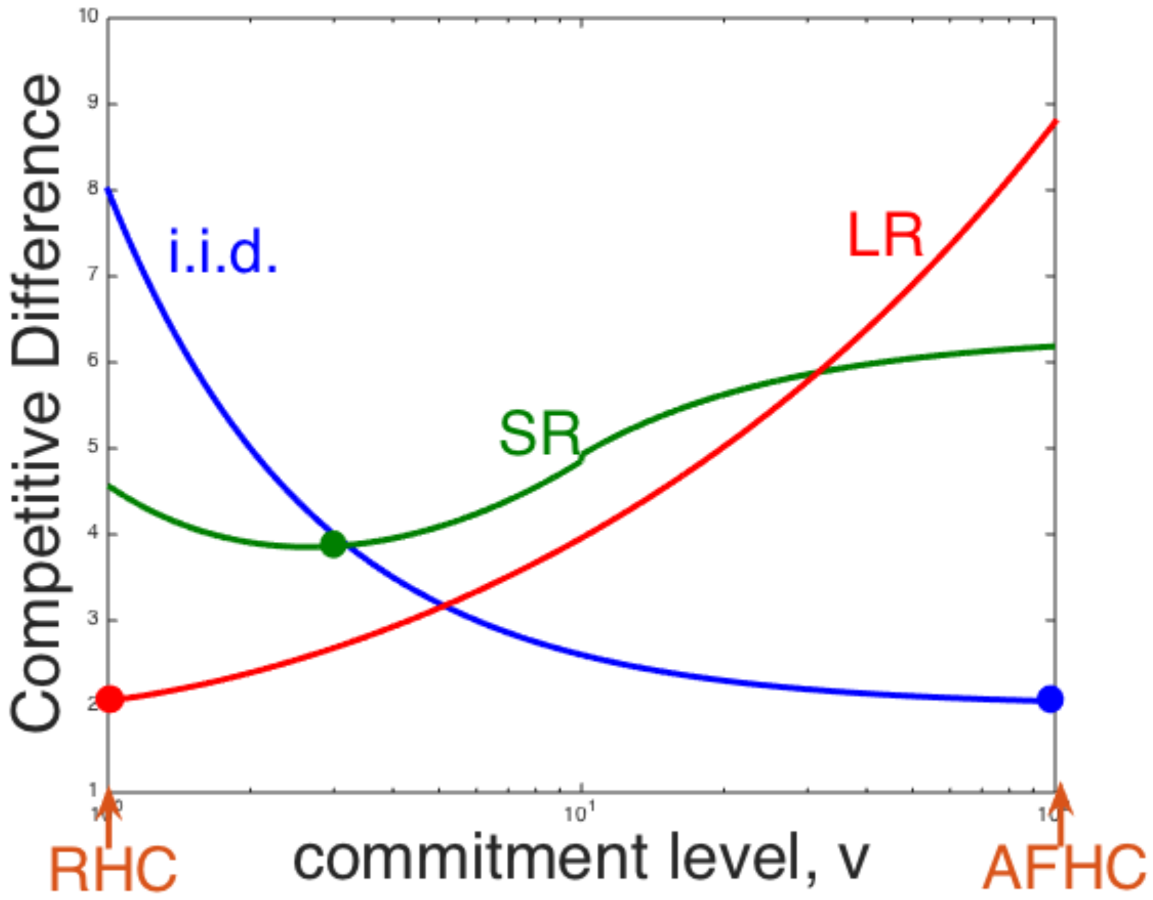
$$f(s) = \begin{cases} 1, s = 0 \\ 0, s > 0 \end{cases}$$

Long range correlated

$$f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L > w$$

Short range correlated

$$f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L \leq w$$



i.i.d. prediction noise

$$f(s) = \begin{cases} 1, s = 0 \\ 0, s > 0 \end{cases}$$

Long range correlated

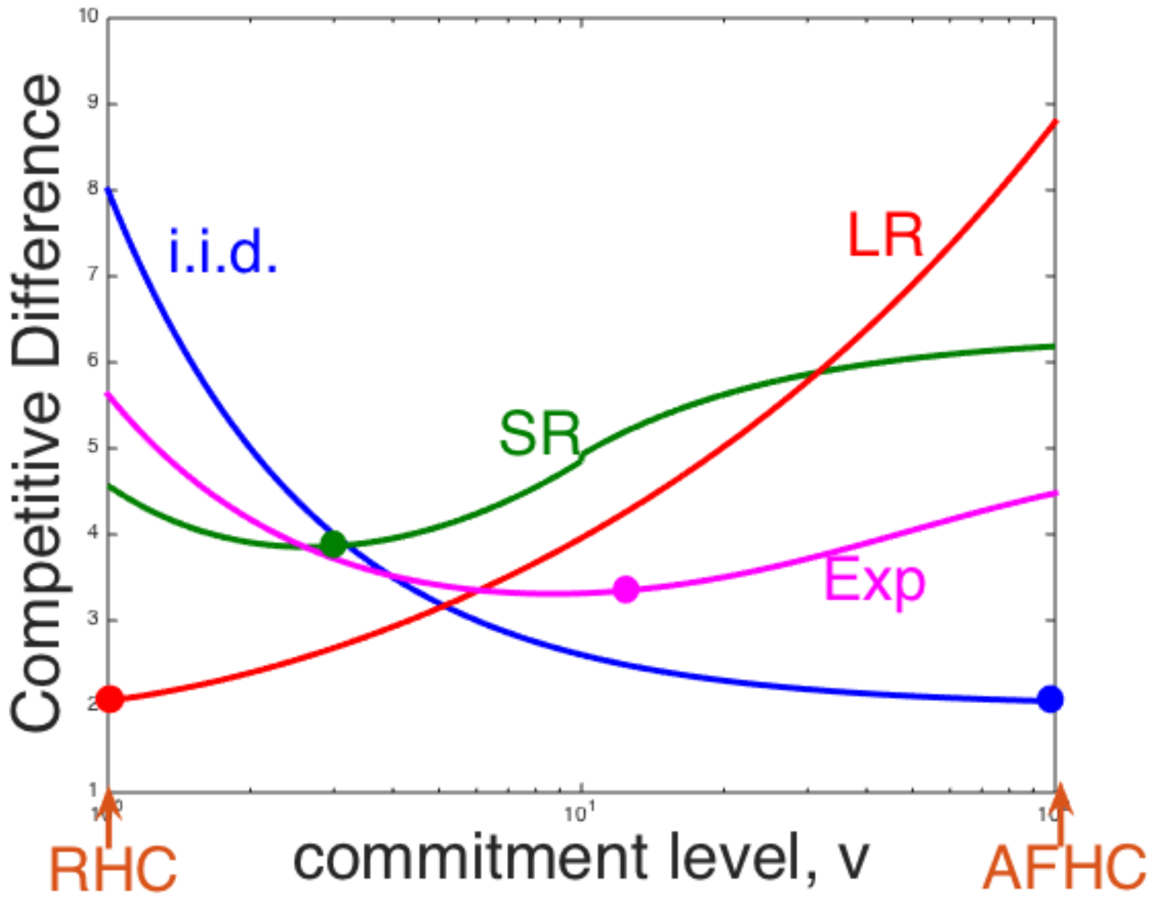
$$f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L > w$$

Short range correlated

$$f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L \leq w$$

Exponentially decaying

$$f(s) = a^s, \qquad a < 1$$



Optimal commitment level depends on prediction noise structure

i.i.d. prediction noise

$$f(s) = \begin{cases} 1, s = 0 \\ 0, s > 0 \end{cases}$$

Long range correlated

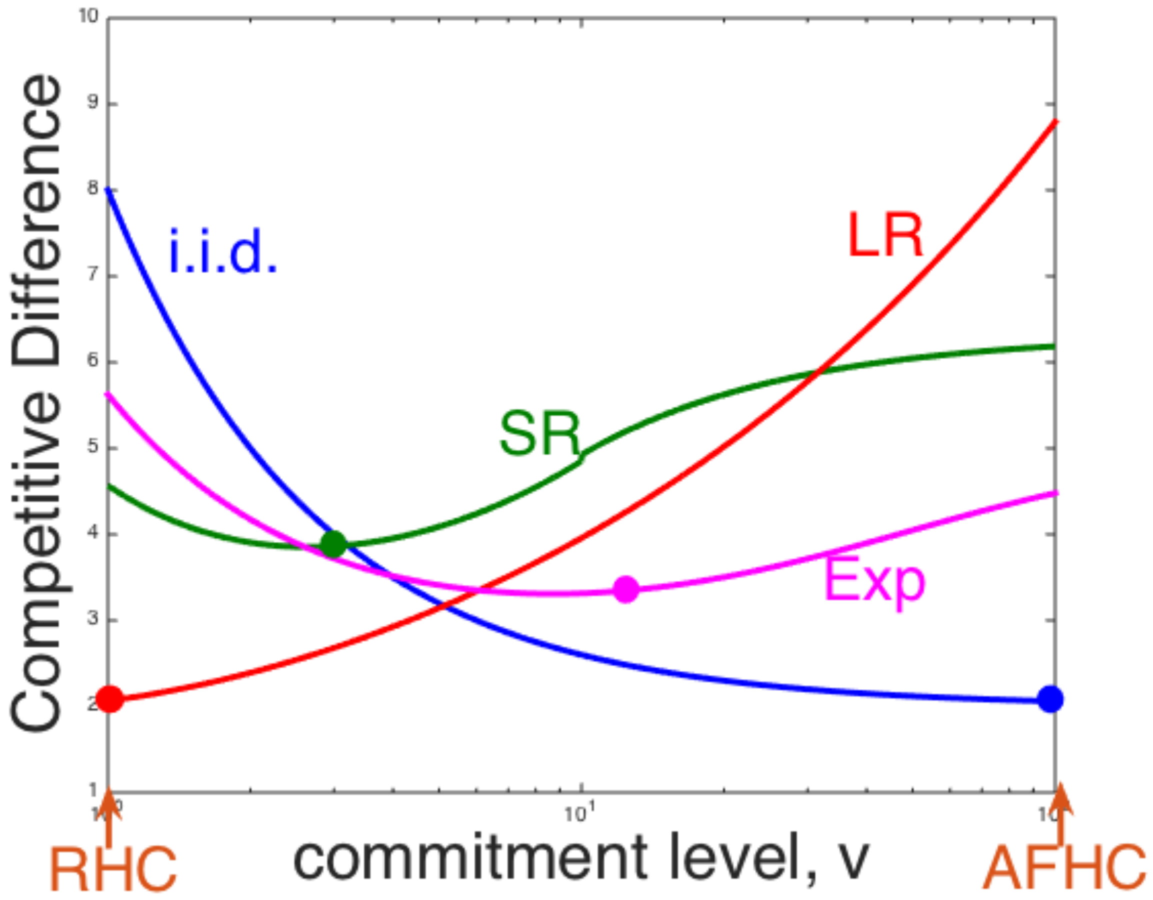
$$f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L > w$$

Short range correlated

$$f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L \leq w$$

Exponentially decaying

$$f(s) = a^s, \qquad a < 1$$



More detail: long-range correlated noise

Theorem

If prediction noise is long

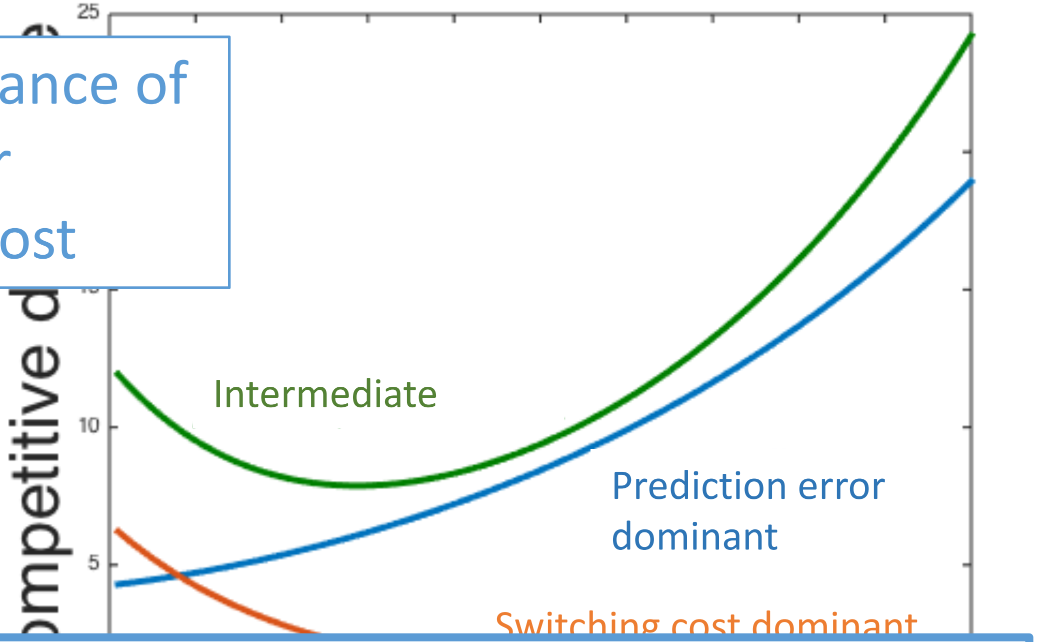
correlated, $f(s) = \begin{cases} 1, s \\ 0, s \end{cases}$

Relative importance of prediction error and switching cost

AFHC is optimal if $\boxed{\frac{\beta D}{G\sigma^{\alpha}}} > \alpha(2w)^{1+\frac{\alpha}{2}}$

RHC is optimal if $\frac{\beta D}{G\sigma^{\alpha}} < \frac{2}{\alpha+2}$

CHC is optimal with $v \in (1, w)$ o/w



$$\boxed{v^{*} = \operatorname{argmin} \frac{2T\beta D(\alpha+2) - 4GT\sigma^{\alpha}}{v(\alpha+2)} + \frac{2^{3+\frac{\alpha}{2}}GT\sigma^{\alpha}}{\alpha+2} v^{\frac{\alpha}{2}}}$$

More detail: short-range correlated noise

<u>Theorem</u>

If prediction noise is short-range correlated, $f(s) = \begin{cases} 1, s \leq L \\ 0, s > L \end{cases}, L$

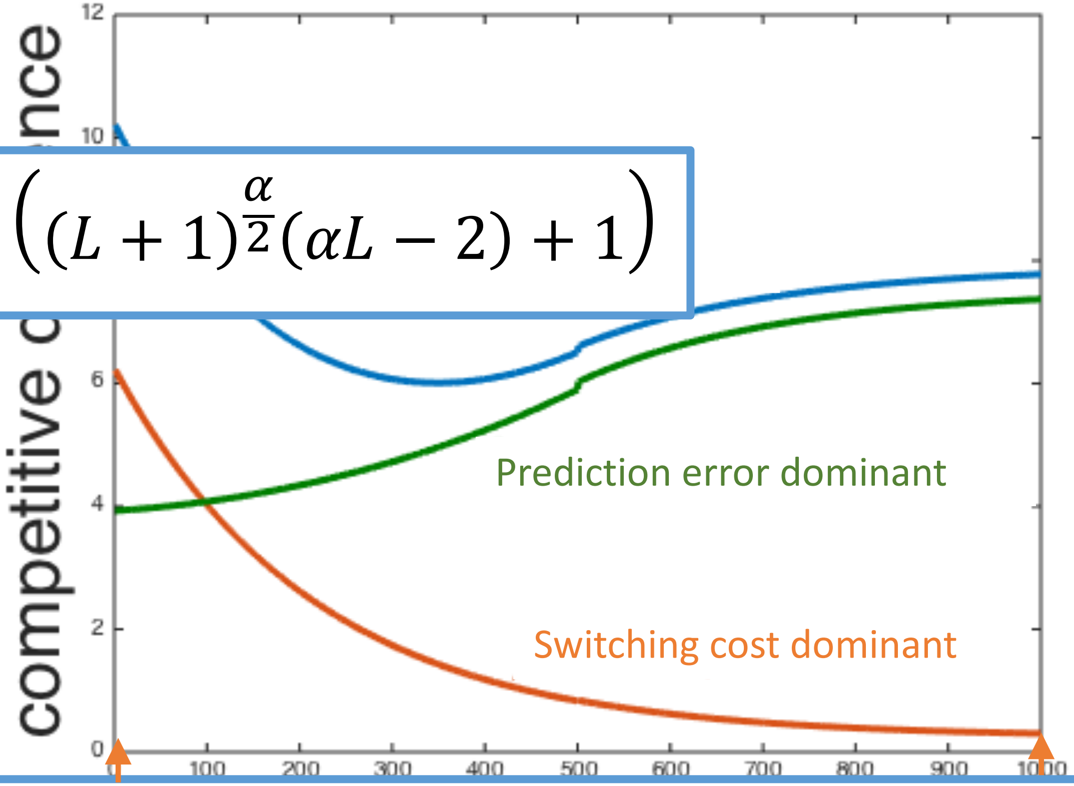
$$\frac{1}{\alpha + 2}\left((L+1)^{\frac{\alpha}{2}}(\alpha L - 2) + 1\right)$$

AFHC is optimal if $\frac{\beta D}{G\sigma^{\alpha}} > H(L)$

RHC is optimal if $\frac{\beta D}{G\sigma^{\alpha}} < \frac{2}{\alpha+2}$

CHC is optimal with $v \in (1, w)$ o/w

$$v^* = \operatorname{argmin} \frac{2T\beta D}{v} + 2GT\sigma^{\alpha}(L+1)^{\alpha/2} - \frac{2GT\sigma^{\alpha}}{v}H(L)$$



More detail: exponentially decaying noise

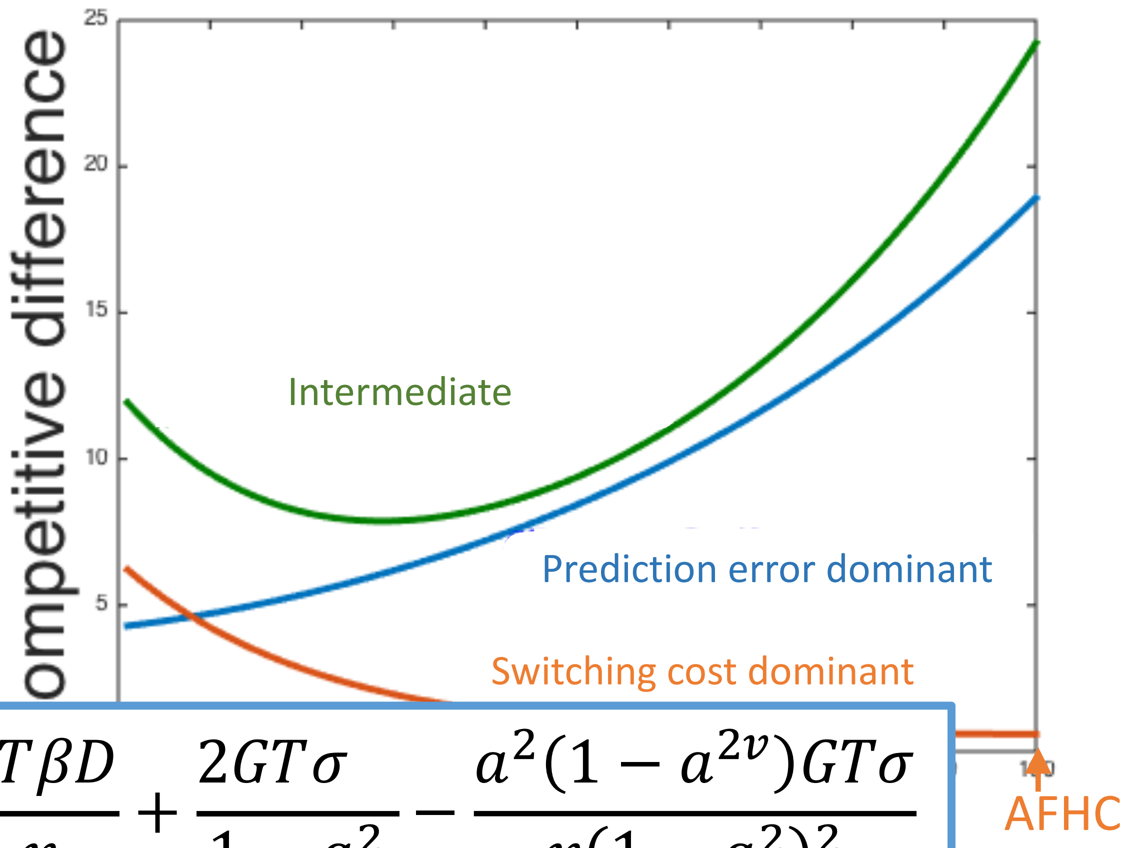
<u>Theorem</u>

If prediction noise is exponentially decaying, $f(s) = a^s,\ 0 < a < 1$

AFHC is optimal if $\frac{\beta \mathrm{D}}{G\sigma^{\alpha}} > \frac{a^2}{2(1-a^2)}$

RHC is optimal if $\frac{\beta D}{G\sigma^{\alpha}} < \frac{a^2}{2(1+a)}$

CHC is optimal with $v \in (1, w)$ o/w



$$\boxed{v^* = \operatorname{argmin} \frac{2T\beta D}{v} + \frac{2GT\sigma}{1-a^2} - \frac{a^2(1-a^{2v})GT\sigma}{v(1-a^2)^2}}$$

Main Result

<u>Theorem</u>

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \leq \frac{2T\beta D}{v} + \frac{2GT}{v}\sum_{k=1}^{v}||f_k||^{\alpha}$$

➡ We can use prediction error structure to guide design of online algorithm

Main Result

Theorem

For c that is α-Hölder continuous in the second argument and feasible set F is bounded,

$$\mathbf{E}\mathrm{cost}(CHC) - \mathbf{E}\mathrm{cost}(OPT) \leq \boxed{\frac{2T\beta D}{v} + \frac{2GT}{v}\sum_{k=1}^{v} ||f_k||^{\alpha} =: V}$$

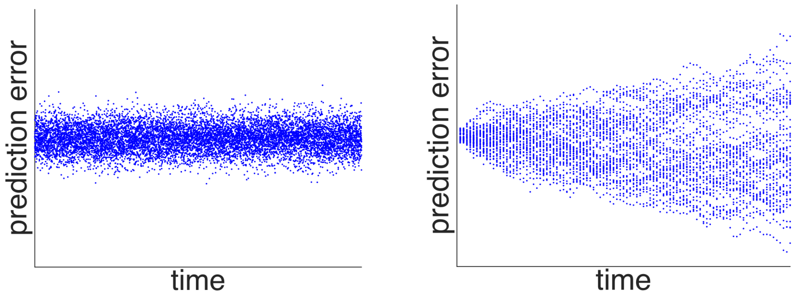
Competitive difference holds with high probability

$$\mathbf{P}(\mathrm{cost}(CHC) - \mathrm{cost}(OPT) > V + u) > \exp\left(-\frac{u^2}{F(v)}\right)$$

Conclusion

Design of optimal algorithm depends on

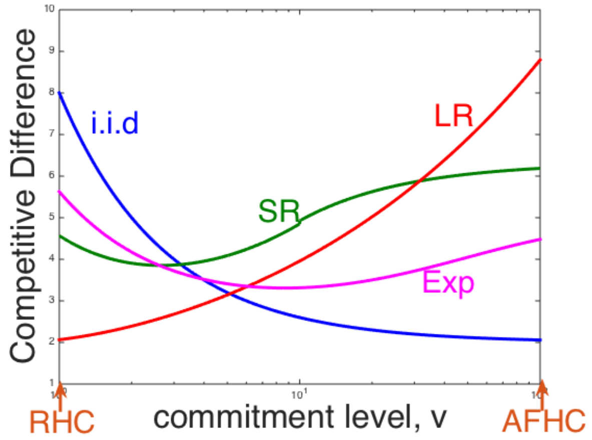
structure of prediction error



This talk: OCO with prediction

“Commitment” should be optimal to

prediction noise



Future: can we extend this framework to other online problems?