Beating Bandits in Gradually Evolving Worlds

Chao-Kai Chiang, **Chia-Jung Lee**, Chi-Jen Lu

COLT 2013

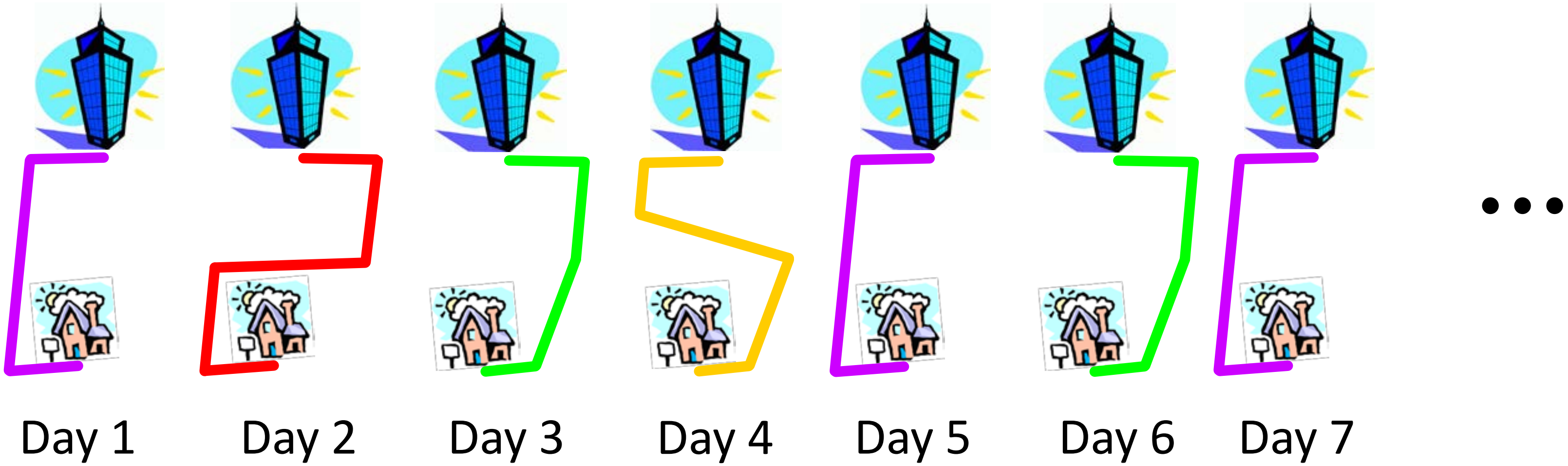
Outline

- INTRODUCTION
- DIFFICULTIES IN BANDIT SETTING
- TWO-POINT BANDIT SETTING
- RESULTS

Online Learning: Routing (Example)

- Go to your lab every day

Online Learning: Routing (Example)

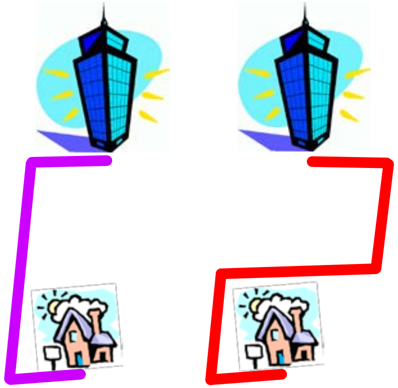


Minimize: total commuting time

Online Learning: Model

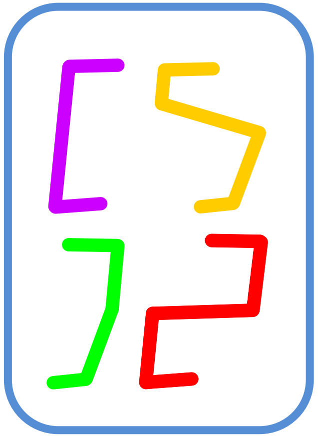
loss f_{t-1}	loss f_t
8	*
3	*
2	*
4	*

Round t-1 Round t



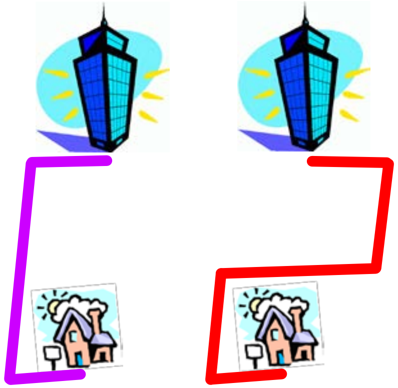
- In each round **t**
- **Environment**: $f_t(x)$, **convex loss** function
- **Player**: action $x_t \in \mathcal{X}$, **convex feasible set**

loss $f_t(x_t)$



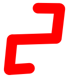
Online Learning: Model

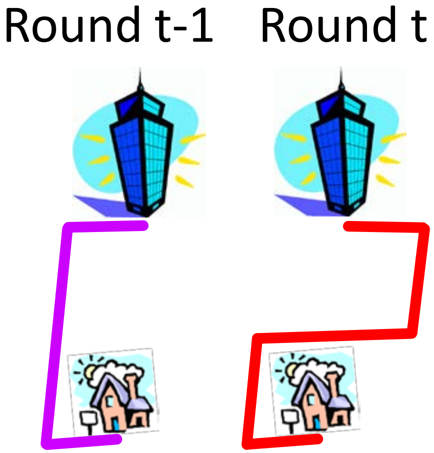
action	loss f_{t-1}	loss f_t
🟪	8	9
🟩	3	8
🟨	2	1
🟥	4	6



- **Full** information
  - See f_t after choosing x_t

Online Learning: Model

action	loss f_{t-1}	loss f_t
	8	*
	*	*
	*	*
	*	6



- **Full** information
  - See f_t after choosing x_t
- **Bandit**
  - **Only** see $f_t(x_t)$ after choosing x_t

Online Learning: Model

action	loss f_{t-1}	loss f_t	loss f_{t+1}	
🟪	8	9	5	
🟩	3	8	3	
🟨	2	1	2	…
🟥	4	6	8	

- Goal:
  Minimize regret

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

best fixed action

FULL INFORMATION SETTING

Online Problems

linear loss $f_t(x)$

Problems			
Online **Linear** Optimization			
Online **Convex** Optimization			
Online **Strongly** Convex Optimization			

Previous Results

Minimize regret

$$\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \min_{\boldsymbol{x}\in\mathcal{X}} \sum_{t=1}^{T} f_t(\boldsymbol{x})$$

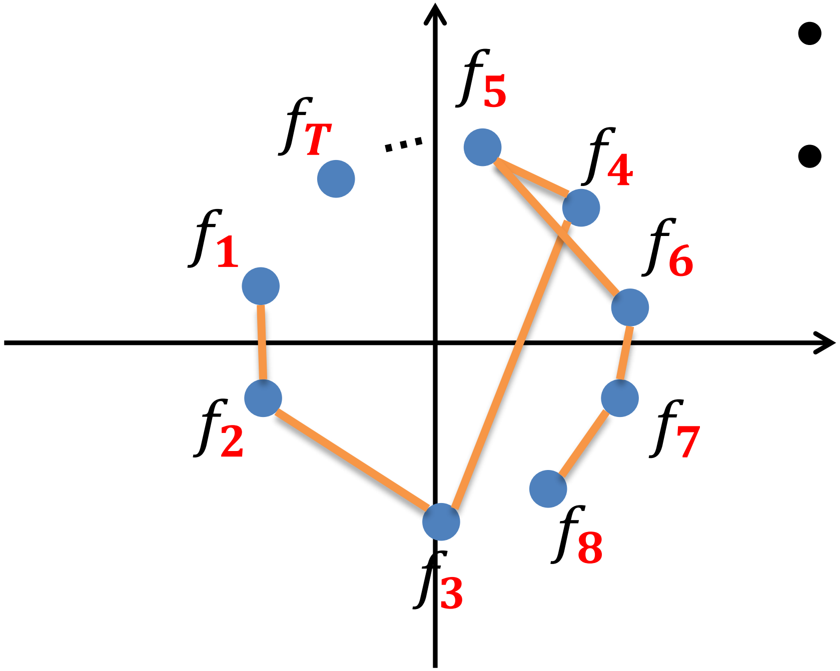
Problems	T rounds		Deviation D	
	Full	Bandit	Full	Bandit
Online **Linear** Optimization	$\sqrt{T}$ [Zin03]	$\sqrt{T}$ [BCK12]	$\sqrt{D}$ [CYL$^+$12]	
Online **Convex** Optimization	$\sqrt{T}$ [Zin03]	$T^{2/3}$ [ST11]	$\sqrt{D}$ [CYL$^+$12]	
Online **Strongly** Convex Optimization	$\log T$ [HAK07]	?	$\log D$ [CYL$^+$12]	

Previous Result: Deviation

- **Deviation**

$$D \triangleq \sum_{t=1}^{T} \|f_t - f_{t-1}\|_2^2$$

Ex. linear loss $f_t \in \mathbb{R}^2$



- Weather system
- Stock market

Previous Results

Problems	T rounds		**Deviation D**	
	Full	Bandit	Full	Bandit
Online **Linear** Optimization	$\sqrt{T}$ [Zin03]	$\sqrt{T}$ [BCK12]	$\sqrt{D}$ [CYL+12]	
Online **Convex** Optimization	$\sqrt{T}$ [Zin03]	$T^{2/3}$ [ST11]	$\sqrt{D}$ [CYL+12]	
Online **Strongly** Convex Optimization	$\log T$ [HAK07]	?	$\log D$ [CYL+12]	

BANDIT SETTING

Previous Results

Problems	T rounds		**Deviation D**	
	Full	Bandit	Full	Bandit
Online **Linear** Optimization	$\sqrt{T}$ [Zin03]	$\sqrt{T}$ [BCK12]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Convex** Optimization	$\sqrt{T}$ [Zin03]	$T^{2/3}$ [ST11]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Strongly** Convex Optimization	$\log T$ [HAK07]	?	$\log D$ [CYL$^+$12]	?

Previous Results

Problems	T rounds		**Deviation D**	
	Full	Bandit	Full	Bandit
Online **Linear** Optimization	$\sqrt{T}$ [Zin03]	$\sqrt{T}$ [BCK12]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Convex** Optimization	$\sqrt{T}$ [Zin03]	$T^{2/3}$ [ST11]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Strongly** Convex Optimization	$\log T$ [HAK07]	?	$\log D$ [CYL$^+$12]	?

Previous Results

Problems	T rounds		**Deviation D**	
	Full	Bandit	Full	Bandit
Online **Linear** Optimization	$\sqrt{T}$ [Zin03]	$\sqrt{T}$ [BCK12]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Convex** Optimization	$\sqrt{T}$ [Zin03]	$T^{2/3}$ [ST11]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Strongly** Convex Optimization	$\log T$ [HAK07]	?	$\log D$ [CYL$^+$12]	?

Lower bound: $\Omega(\sqrt{T})$ [JNR12]

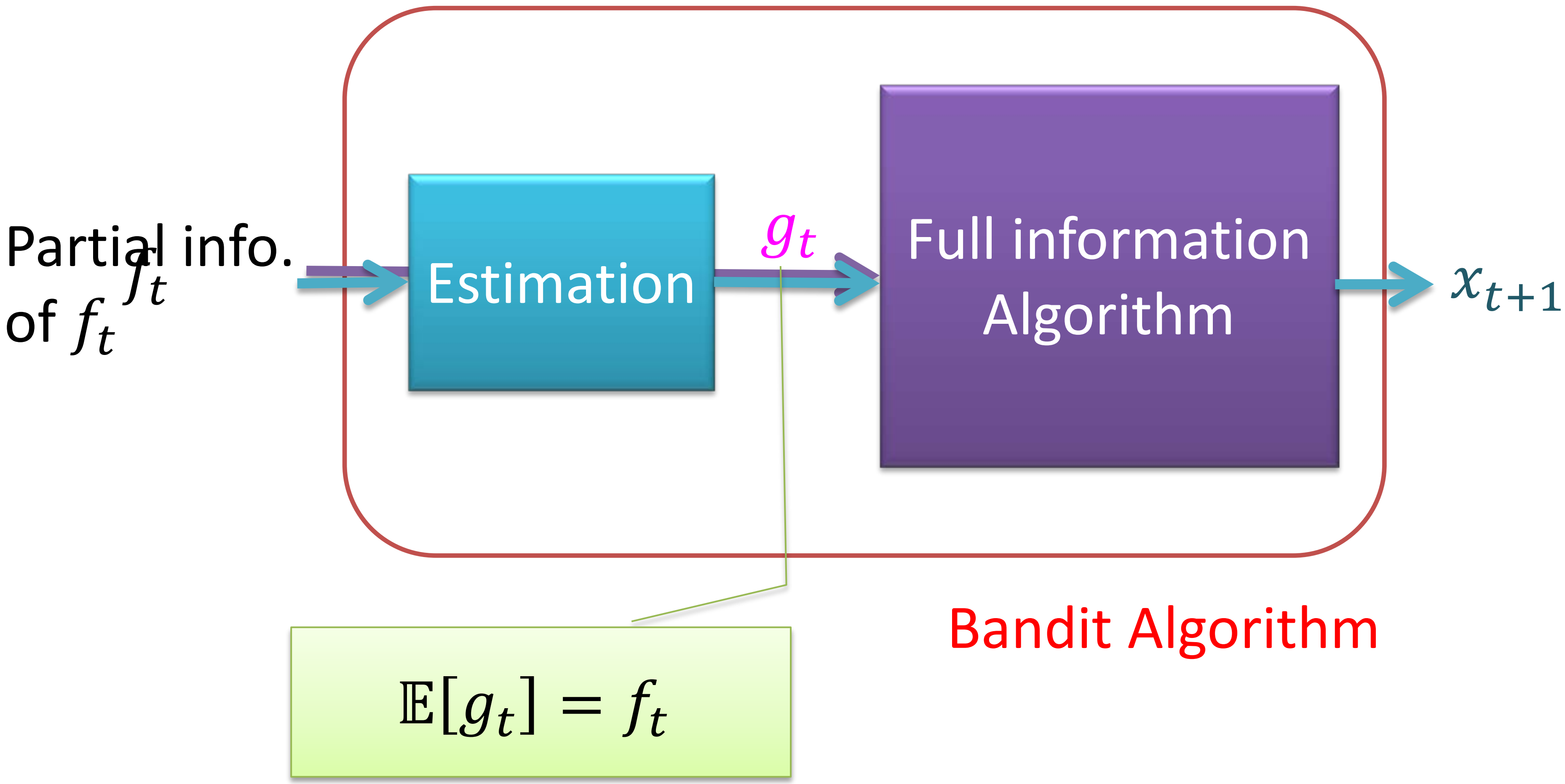
Previous Results

Problems	T rounds		**Deviation D**	
	Full	Bandit	Full	Bandit
Online **Linear** Optimization	$\sqrt{T}$ [Zin03]	$\sqrt{T}$ [BCK12]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Convex** Optimization	$\sqrt{T}$ [Zin03]	$T^{2/3}$ [ST11]	$\sqrt{D}$ [CYL$^+$12]	?
Online **Strongly** Convex Optimization	$\log T$ [HAK07]	?	$\log D$ [CYL$^+$12]	?

Lower bound: $\Omega(\sqrt{T})$ [JNR12]

DIFFICULTIES IN BANDIT SETTING

Approach for Bandit



Partial info. of f_t

f_t

Estimation

g_t

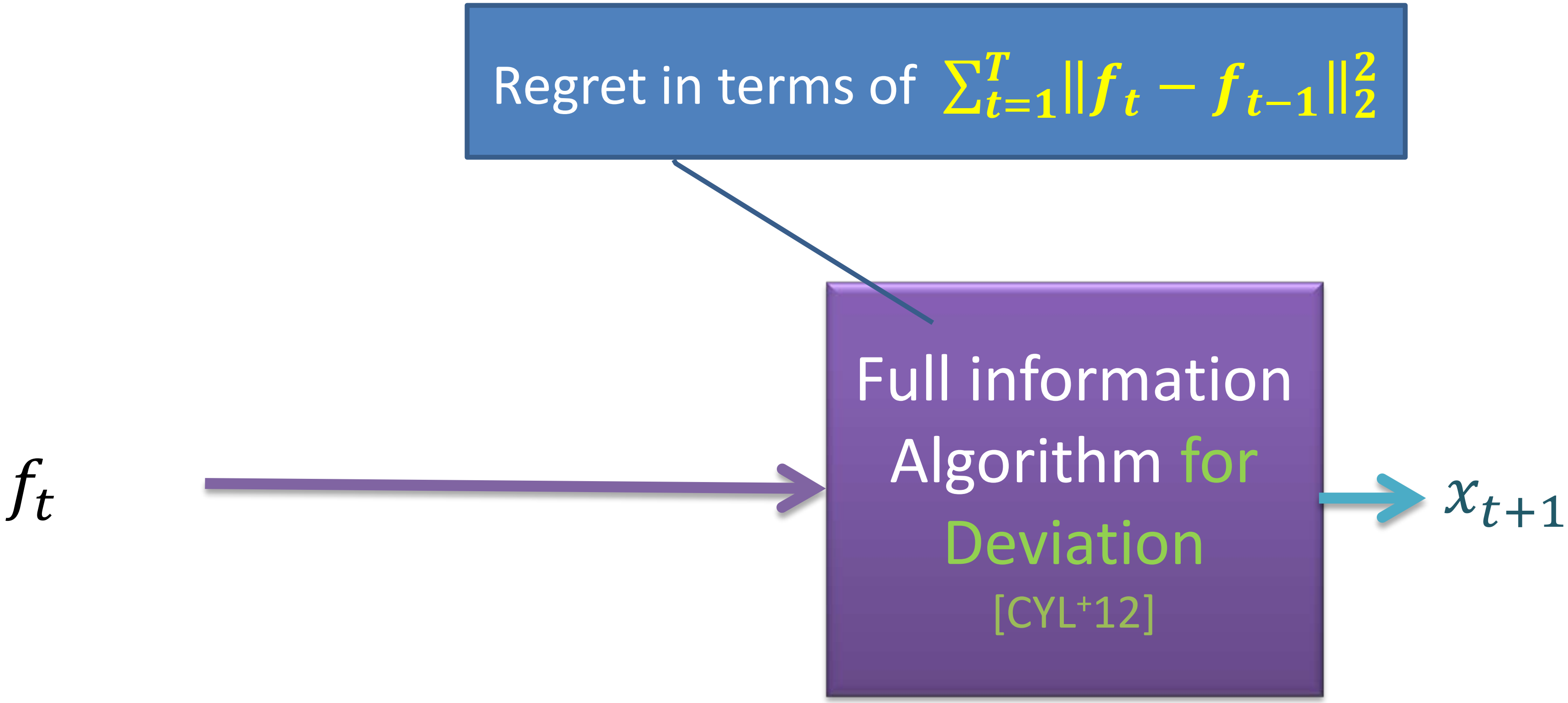
Full information Algorithm

x_{t+1}

Bandit Algorithm

$\mathbb{E}[g_t] = f_t$

Approach for Bandit



Regret in terms of $\sum_{t=1}^{T} \|f_t - f_{t-1}\|_2^2$

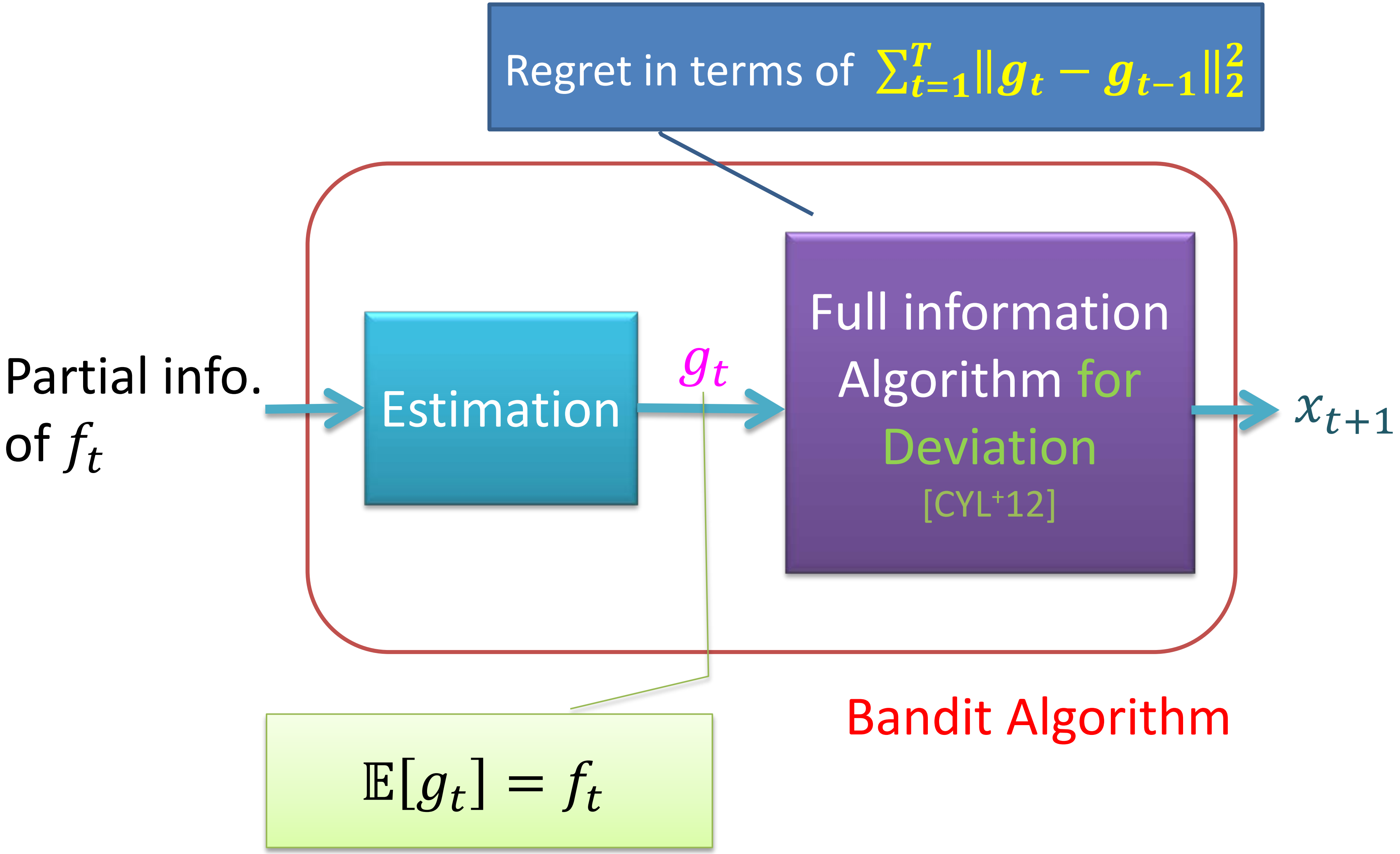
f_t

Full information Algorithm for Deviation

[CYL+12]

x_{t+1}

Approach for Bandit



Regret in terms of $\sum_{t=1}^{T}\|g_t - g_{t-1}\|_2^2$

Partial info. of f_t

Estimation

g_t

Full information Algorithm for Deviation

[CYL+12]

x_{t+1}

Bandit Algorithm

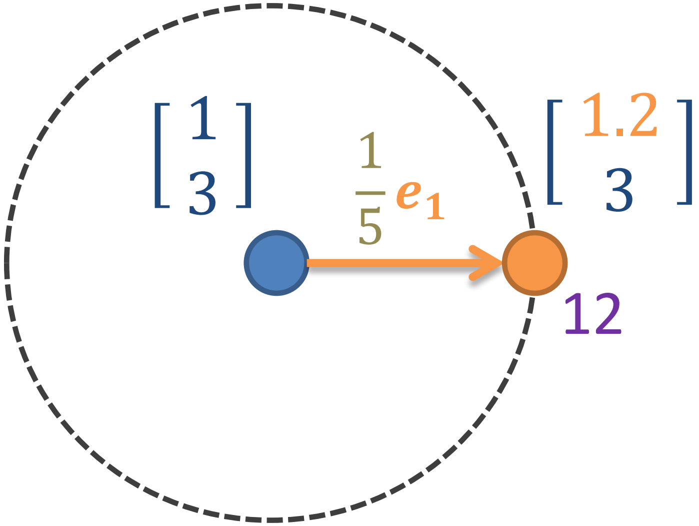
$\mathbb{E}[g_t] = f_t$

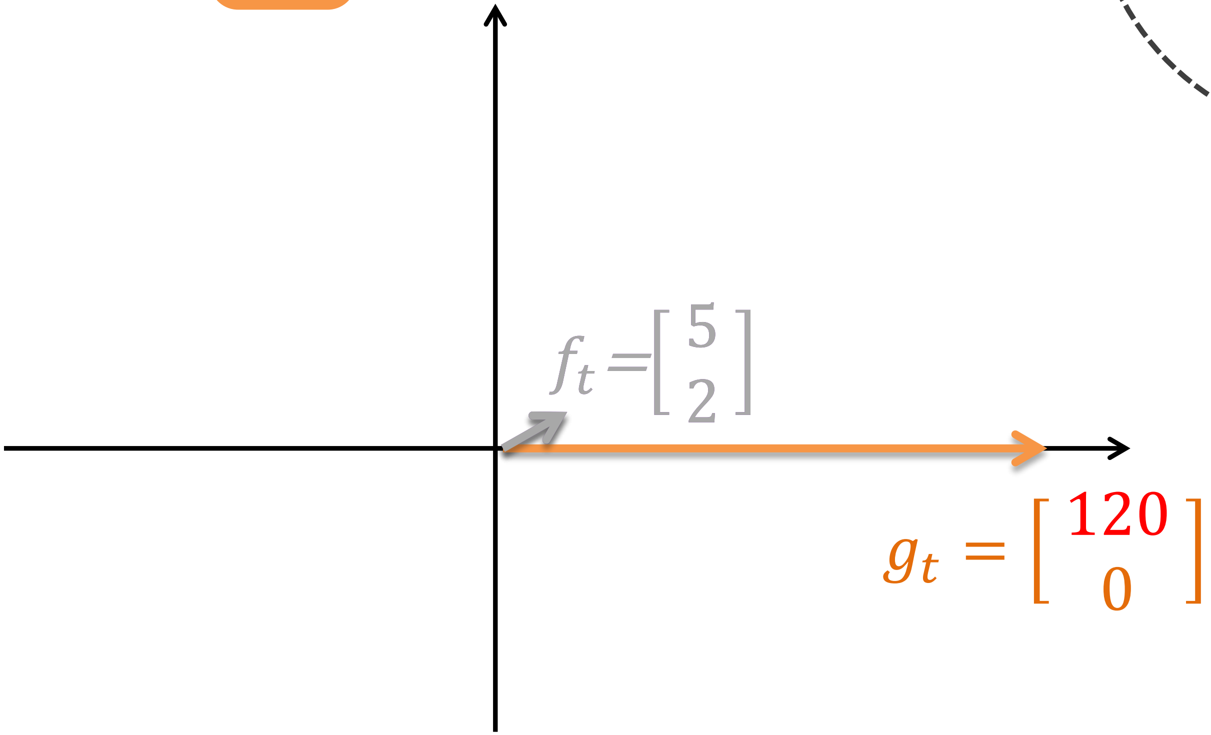
Estimation

Ex. linear loss $f_t \in \mathbb{R}^2$

- In round t

$$u_t \in_R \left\{ \begin{array}{cccc} e_1 & e_2 & -e_1 & -e_2 \\ \boxed{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}, & \begin{bmatrix} 0 \\ 1 \end{bmatrix}, & \begin{bmatrix} -1 \\ 0 \end{bmatrix}, & \begin{bmatrix} 0 \\ -1 \end{bmatrix} \end{array} \right\}$$





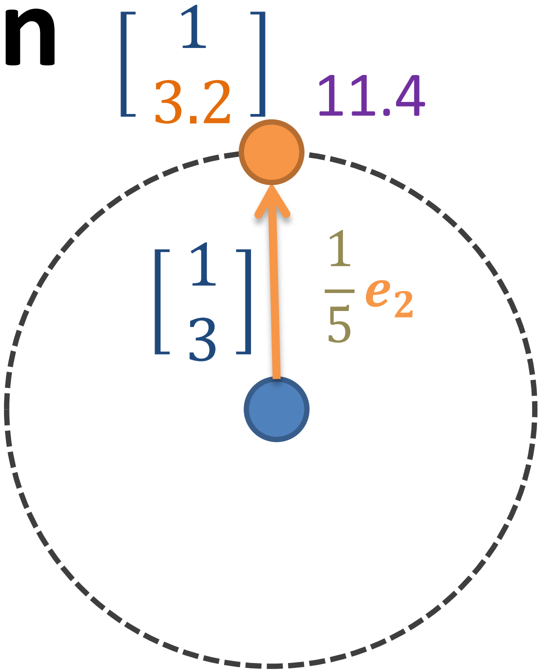
$$f_t = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

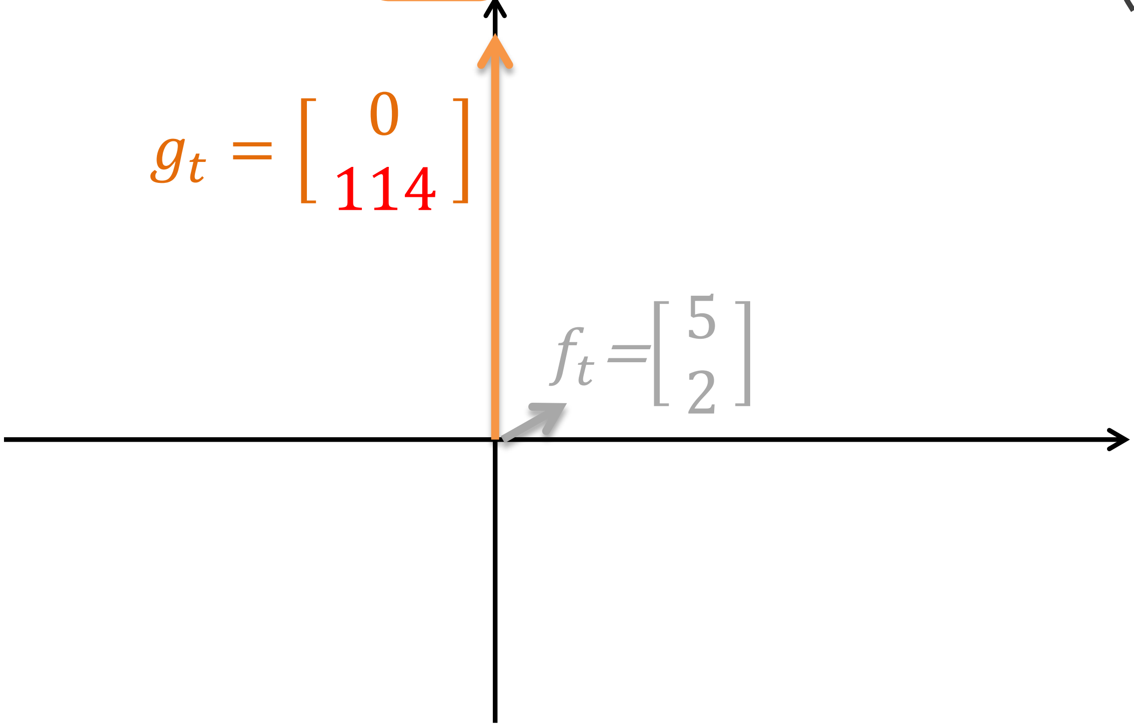
$$g_t = \begin{bmatrix} 120 \\ 0 \end{bmatrix}$$

Estimation

- In round t

$$u_t \in_R \left\{\begin{array}{cccc} e_1 & e_2 & -e_1 & -e_2 \\ \begin{bmatrix}1\\0\end{bmatrix}, & \boxed{\begin{bmatrix}0\\1\end{bmatrix}}, & \begin{bmatrix}-1\\0\end{bmatrix}, & \begin{bmatrix}0\\-1\end{bmatrix}\end{array}\right\}$$





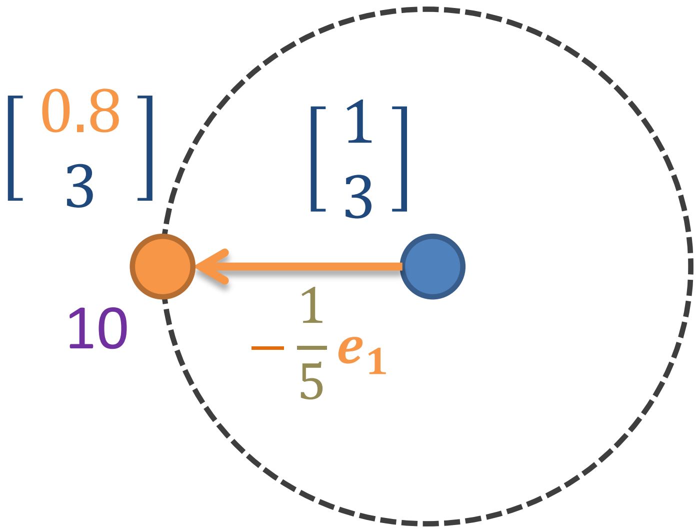
$$f_t = \begin{bmatrix}5\\2\end{bmatrix}$$

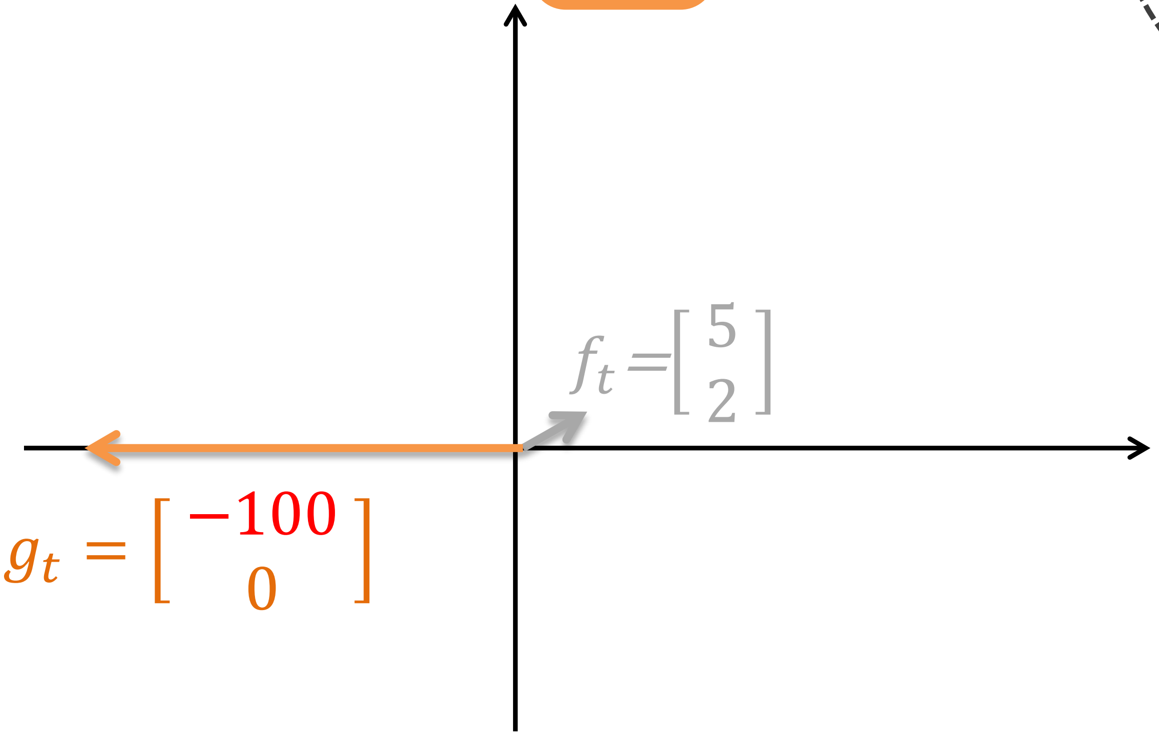
$$g_t = \begin{bmatrix}0\\114\end{bmatrix}$$

Estimation

- In round t

$$u_t \in_R \left\{ \overset{e_1}{\begin{bmatrix}1\\0\end{bmatrix}}, \overset{e_2}{\begin{bmatrix}0\\1\end{bmatrix}}, \boxed{\overset{-e_1}{\begin{bmatrix}-1\\0\end{bmatrix}}}, \overset{-e_2}{\begin{bmatrix}0\\-1\end{bmatrix}} \right\}$$





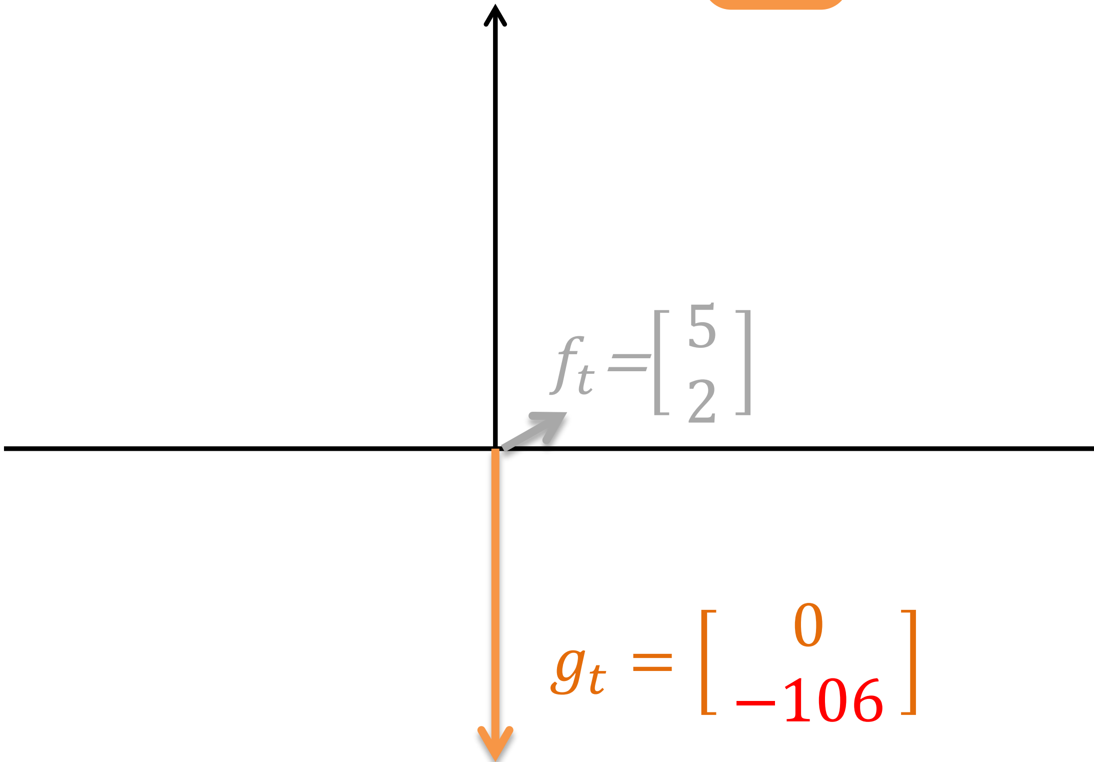
$$f_t = \begin{bmatrix}5\\2\end{bmatrix}$$

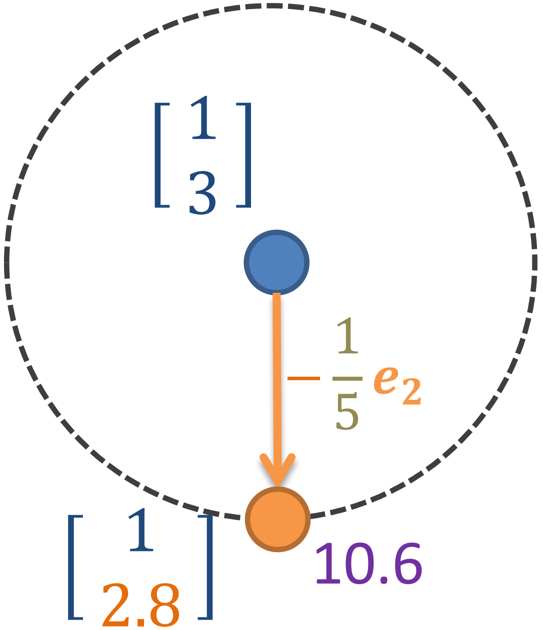
$$g_t = \begin{bmatrix}-100\\0\end{bmatrix}$$

Estimation

- In round t

$$u_t \in_R \left\{ \begin{array}{cccc} e_1 & e_2 & -e_1 & \boxed{\begin{array}{c} -e_2 \\ \begin{bmatrix} 0 \\ -1 \end{bmatrix} \end{array}} \\ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, & \begin{bmatrix} 0 \\ 1 \end{bmatrix}, & \begin{bmatrix} -1 \\ 0 \end{bmatrix}, & \end{array} \right\}$$





$$f_t = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

$$g_t = \begin{bmatrix} 0 \\ -106 \end{bmatrix}$$

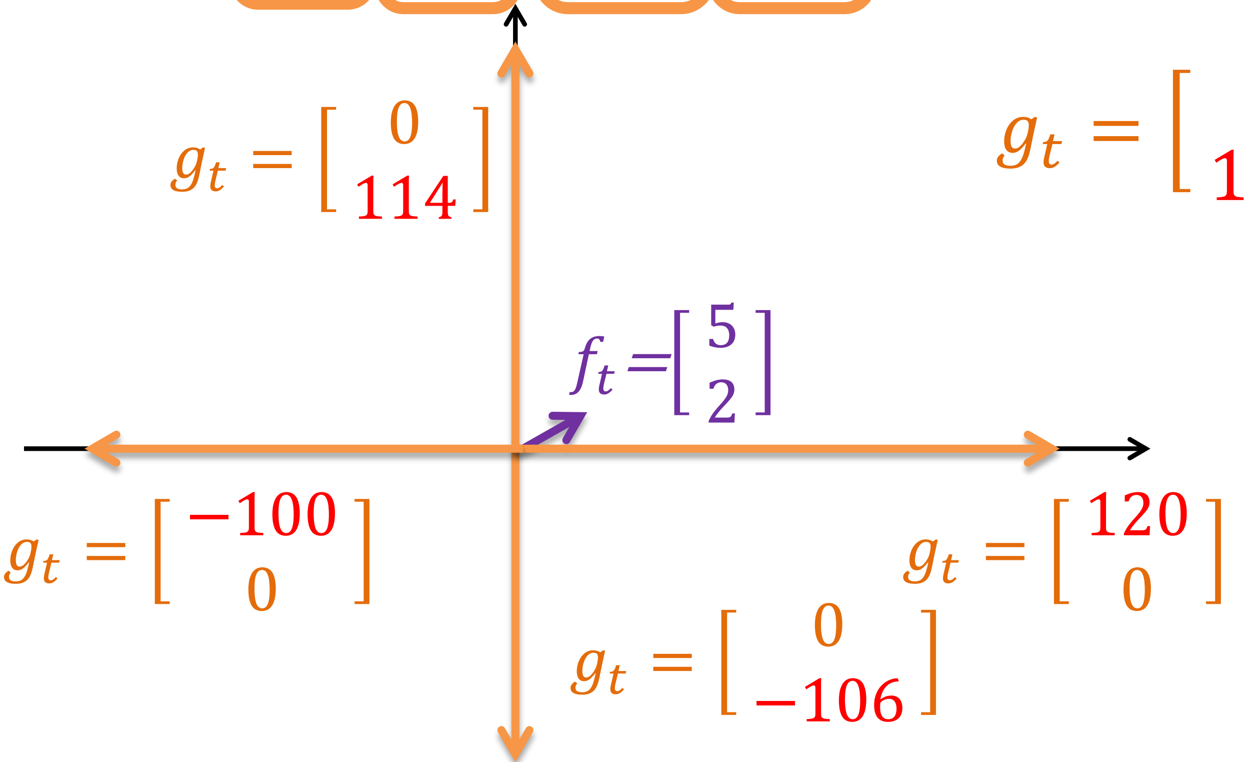
Estimation

- In round t

$$u_t \in_R \left\{ \boxed{\overset{e_1}{\begin{bmatrix}1\\0\end{bmatrix}}}, \boxed{\overset{e_2}{\begin{bmatrix}0\\1\end{bmatrix}}}, \boxed{\overset{-e_1}{\begin{bmatrix}-1\\0\end{bmatrix}}}, \boxed{\overset{-e_2}{\begin{bmatrix}0\\-1\end{bmatrix}}} \right\}$$

$$f_t = \begin{bmatrix}5\\2\end{bmatrix}$$

$$g_t = \begin{bmatrix}0\\114\end{bmatrix}$$



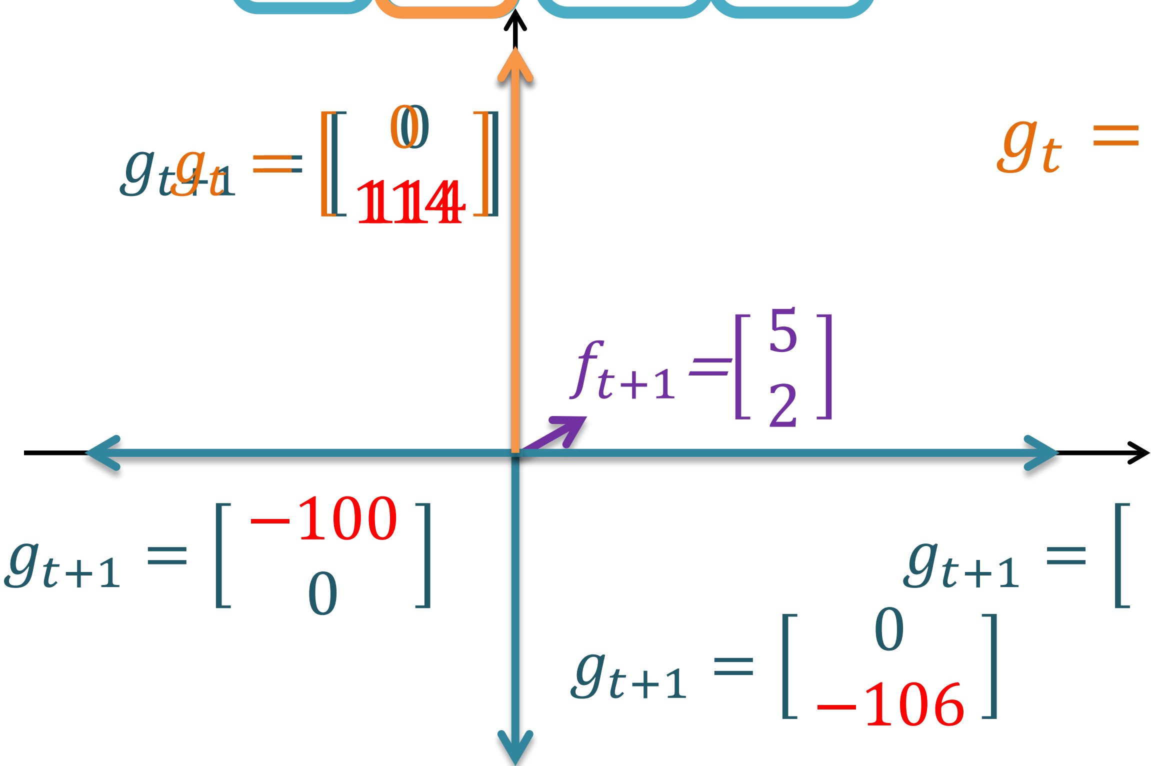
Estimation

- In round t+1

$$u_{t+1} \in_R \left\{ \begin{array}{cccc} e_1 & e_2 & -e_1 & -e_2 \\ \begin{bmatrix}1\\0\end{bmatrix}, & \begin{bmatrix}0\\1\end{bmatrix}, & \begin{bmatrix}-1\\0\end{bmatrix}, & \begin{bmatrix}0\\-1\end{bmatrix} \end{array} \right\}$$

$f_t = \begin{bmatrix}5\\2\end{bmatrix}$ $f_{t+1} = \begin{bmatrix}5\\2\end{bmatrix}$

$g_t = \begin{bmatrix}0\\114\end{bmatrix}$ $g_{t+1} = \begin{bmatrix}120\\0\end{bmatrix}$



$g_t = \begin{bmatrix}0\\114\end{bmatrix}$

$g_{t+1} = \begin{bmatrix}0\\114\end{bmatrix}$

$f_{t+1} = \begin{bmatrix}5\\2\end{bmatrix}$

$g_{t+1} = \begin{bmatrix}-100\\0\end{bmatrix}$

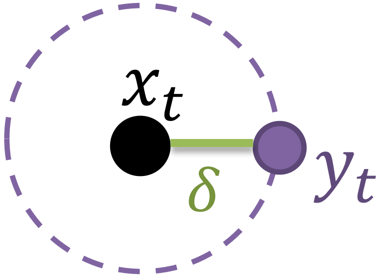
$g_{t+1} = \begin{bmatrix}120\\0\end{bmatrix}$

$g_{t+1} = \begin{bmatrix}0\\-106\end{bmatrix}$

$g_{t+1} \not\approx g_t$
even $f_{t+1} = f_t$

Another Issue: Exploration

- [FKM05] (Example)
- Exploration with parameter δ
- Regret $= \frac{c}{\delta^2} T + \delta T + \cdots$



$\|g_t\|_2 = O(1/\delta)$

- **In our case:** Regret $= \frac{c}{\delta^2} D + \delta T + \cdots$

In terms of D?

TWO-POINT BANDIT SETTING

Two-Point Bandit [ADX10]

- See $\boldsymbol{f_t}(\boldsymbol{x}_{t1})\&\boldsymbol{f_t}(\boldsymbol{x}_{t2})$ after choosing $\boldsymbol{x}_{t1}\&\boldsymbol{x}_{t2}$

- Goal: **Minimize regret**

$$\sum_{t=1}^{T} \frac{f_t(\boldsymbol{x}_{t1}) + f_t(\boldsymbol{x}_{t2})}{\mathbf{2}} - \min_{\boldsymbol{x}\in\mathcal{X}} \sum_{t=1}^{T} f_t(\boldsymbol{x})$$

Motivation

- Online **Strongly** Convex Optimization

	Full Info.	**Bandit**
Online **Strongly** Convex Optimization	$\mathbf{O}(\log \boldsymbol{T})$ [HAK07]	$\mathbf{\Omega}(\sqrt{\boldsymbol{T}})$ [JNR12]

Previous Results

$\sum_{t=1}^{T} \|f_t - f_{t-1}\|^2$

Problems	T rounds		Deviation D	
	Full	Two-point Bandit	Full	Two-point Bandit
Online Linear Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL$^+$12]	?
Online Convex Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL$^+$12]	?
Online Strongly Convex Optimization	$O(\log T)$ [HAK07]	$O(\log T)$ [ADX10]	$O(\log D)$ [CYL$^+$12]	?

Previous Results

$$\sum_{t=1}^{T}\|f_t - f_{t-1}\|^2$$

Problems	T rounds		Deviation D	
	Full	Two-point Bandit	Full	Two-point Bandit
Online Linear Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL+12]	?
Online Convex Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL+12]	?
Online Strongly Convex Optimization	$O(\log T)$ [HAK07]	$O(\log T)$ [ADX10]	$O(\log D)$ [CYL+12]	?

Our Results

$\sum_{t=1}^{T} \|f_t - f_{t-1}\|^2$

Problems	T rounds		Deviation D	
	Full	Two-point Bandit	Full	Two-point Bandit
Online Linear Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL+12]	$O(\sqrt{D})$
Online Convex Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL+12]	$O(\sqrt{D})$
Online Strongly Convex Optimization	$O(\log T)$ [HAK07]	$O(\log T)$ [ADX10]	$O(\log D)$ [CYL+12]	$O(\log D)$

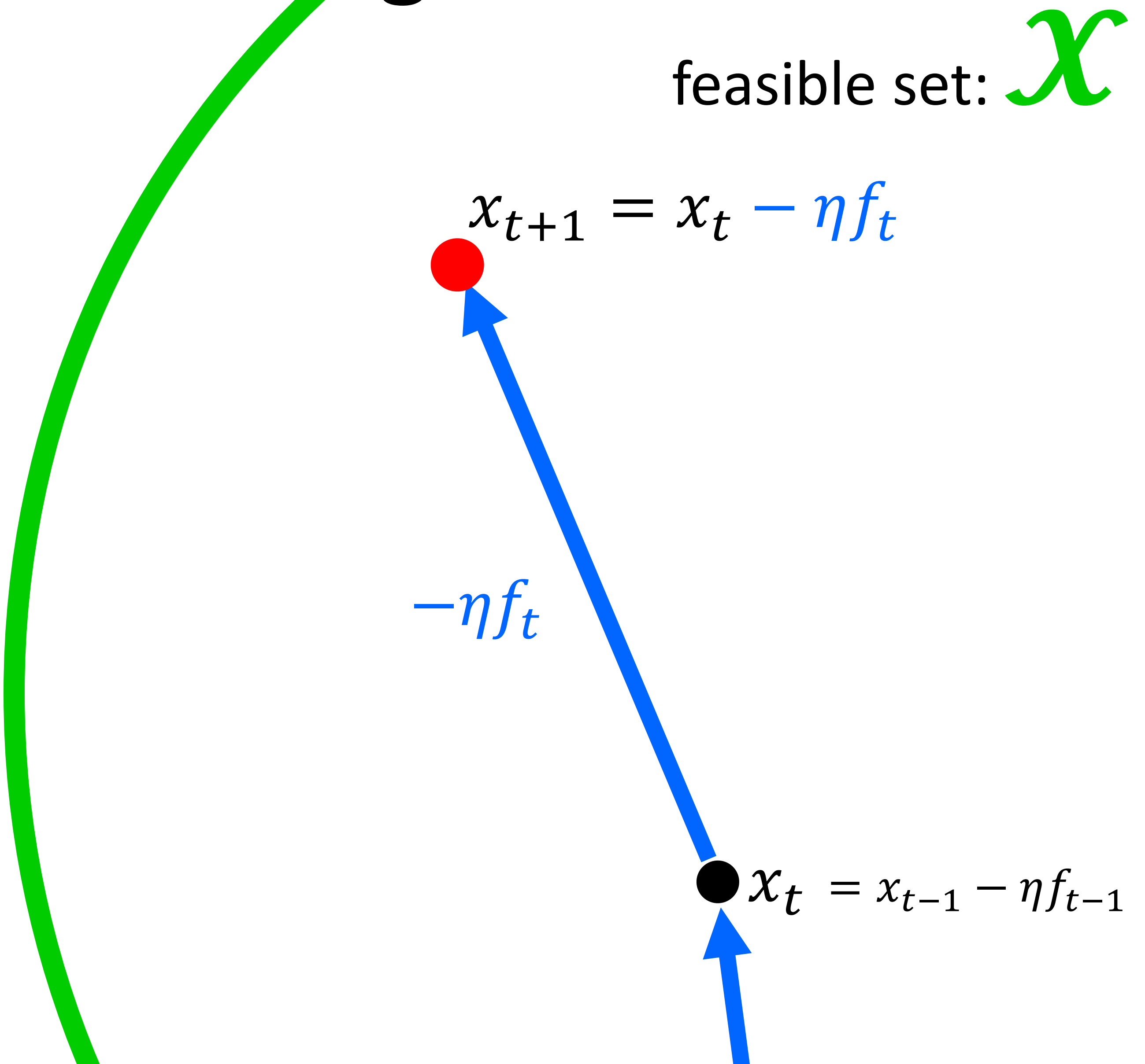
MAIN ALGORITHM

Full-information Algorithm for Deviation

Full information Algorithm's idea

In round t

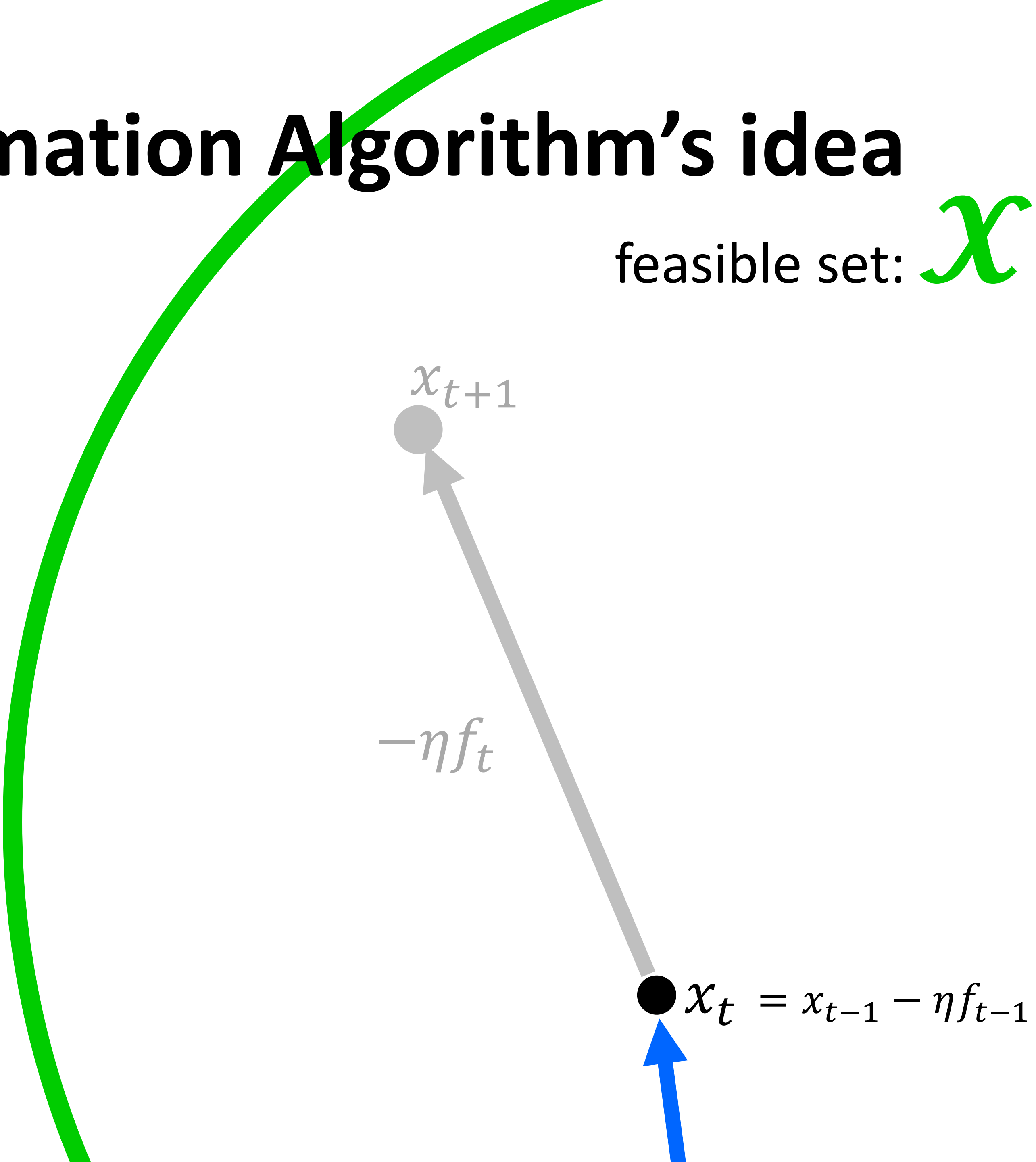
- Fact:
  if *play* $\boldsymbol{x_{t+1}}$
  in round ***t***, …



Full information Algorithm’s idea

In round t

- Fact:
  ***if** play* $\boldsymbol{x_{t+1}}$
  in round ***t***, …
- ***DON’T know***
  $\boldsymbol{f_t}$ in round ***t***



Gradient Descent

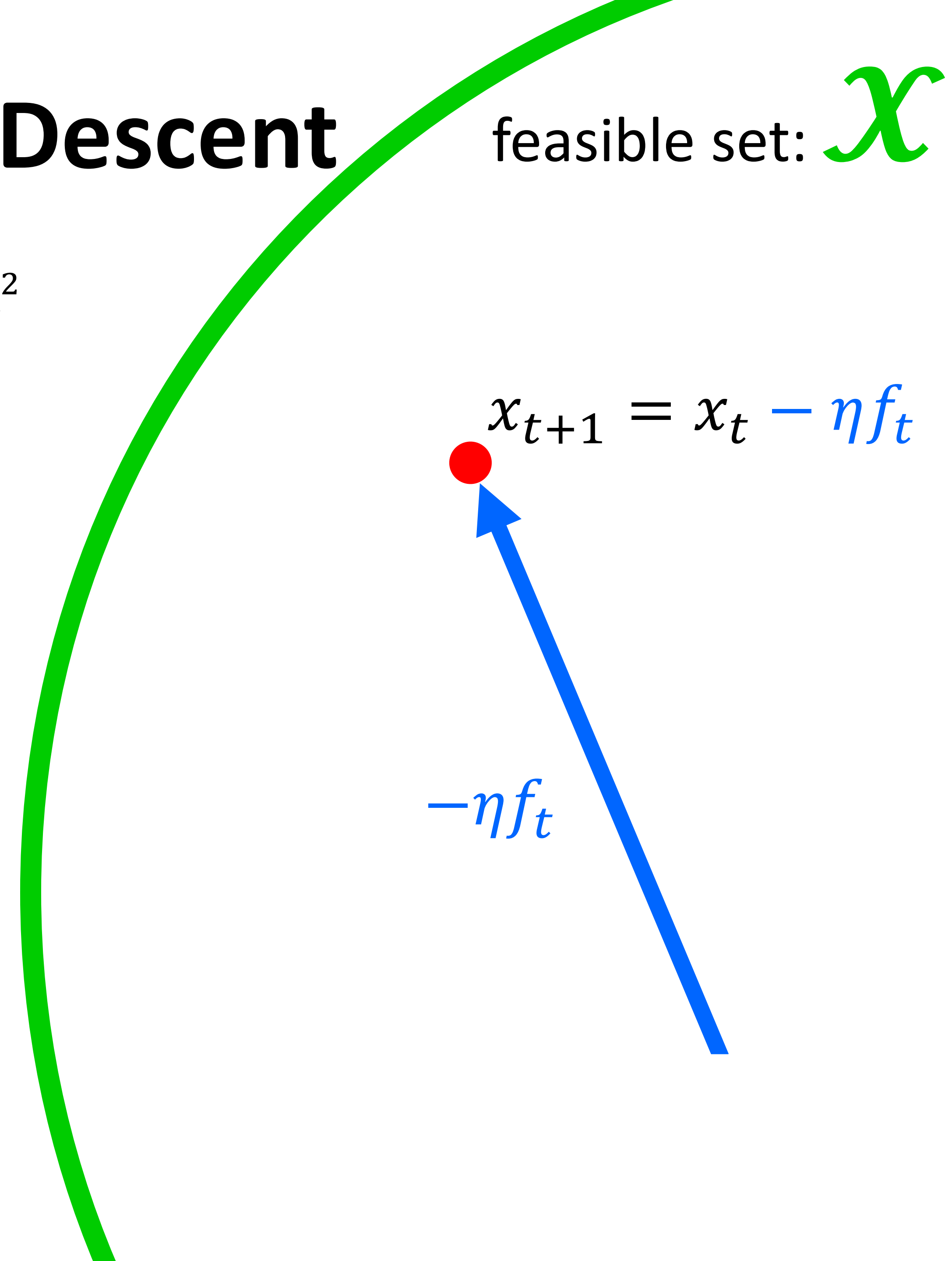
[Zink03]

Ex. linear loss $f_t \in \mathbb{R}^2$

In round t

- f_t revealed

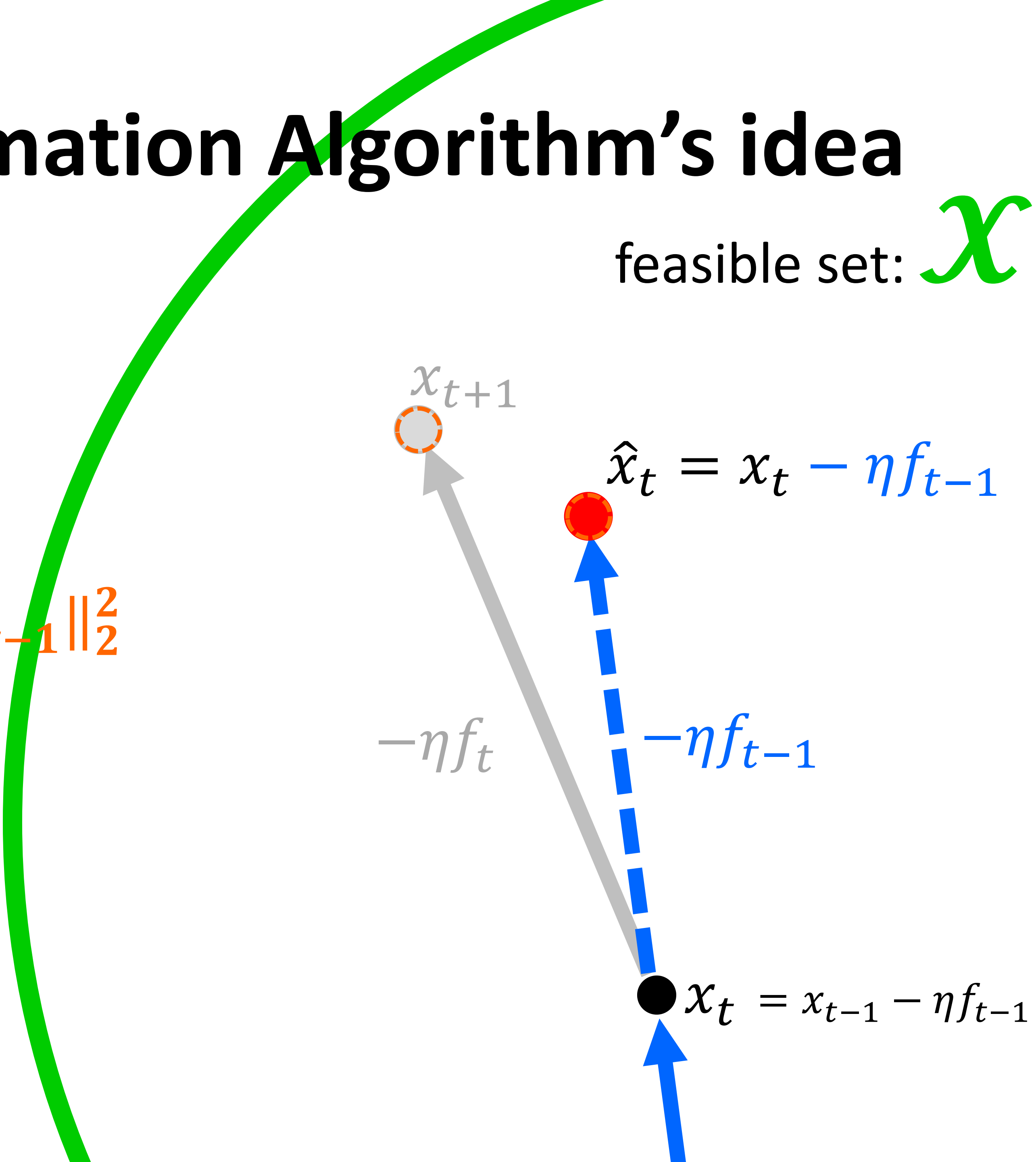
feasible set: $\mathcal{X}$



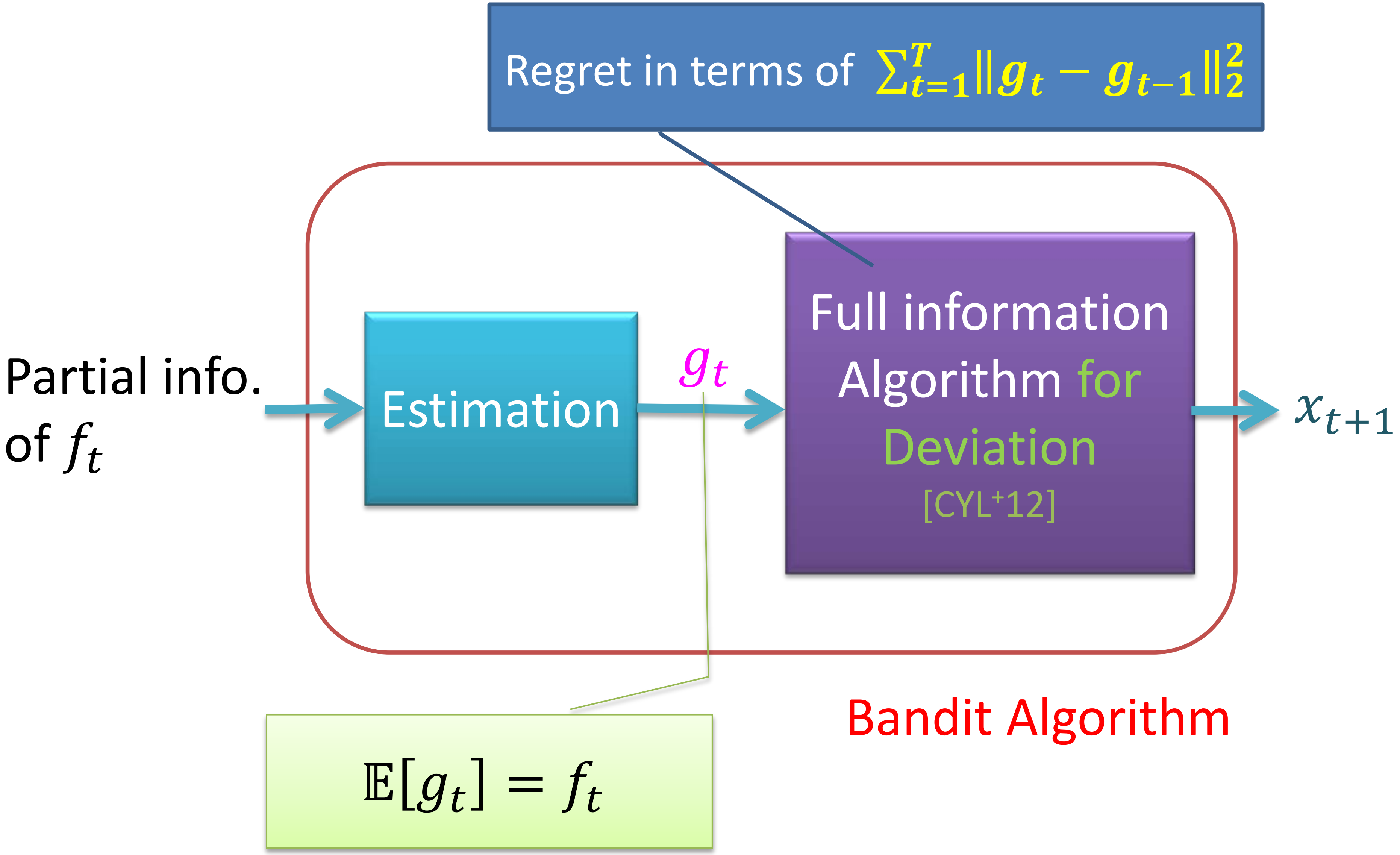
Full information Algorithm's idea

In round t

- Know f_{t-1}
- Deviation $=$ $\sum_{t=1}^{T}\|f_t - f_{t-1}\|_2^2$
- $f_{t-1} \approx f_t$
- **Estimate** x_{t+1} by $\hat{x}_t$ in round t



Approach for Bandit



Regret in terms of $\sum_{t=1}^{T} \|g_t - g_{t-1}\|_2^2$

Partial info. of f_t

Estimation

g_t

Full information Algorithm for Deviation

[CYL+12]

x_{t+1}

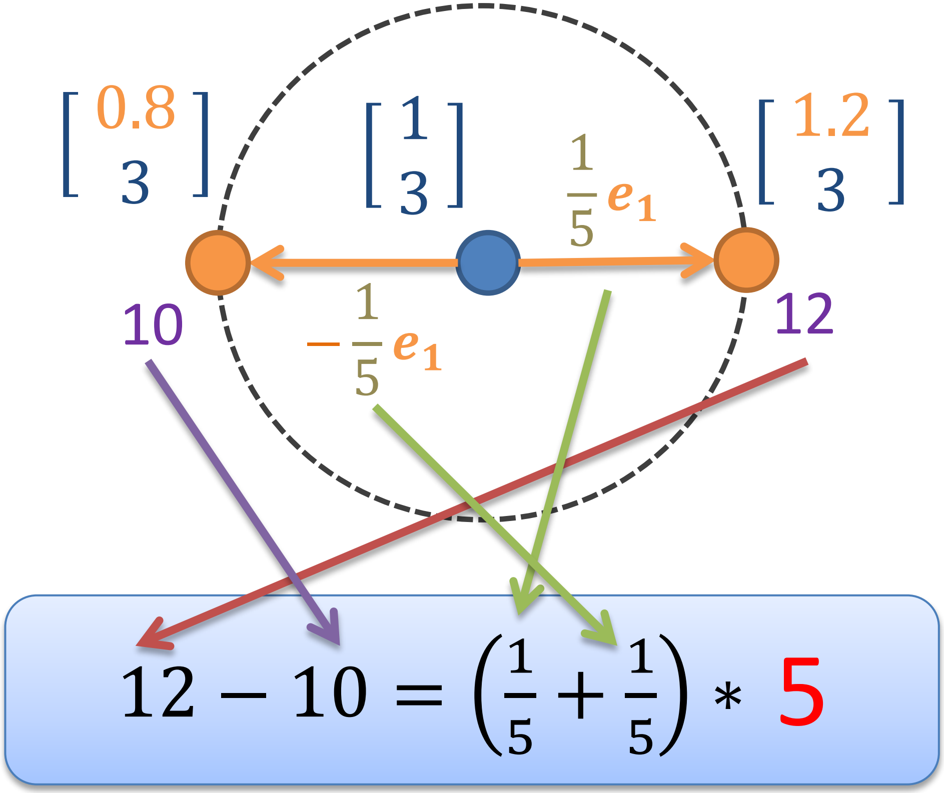
Bandit Algorithm

$\mathbb{E}[g_t] = f_t$

Observation

Ex. linear loss $f_t \in \mathbb{R}^2$

$$\overset{f_t}{\begin{bmatrix} \boxed{5} \\ 2 \end{bmatrix}}$$

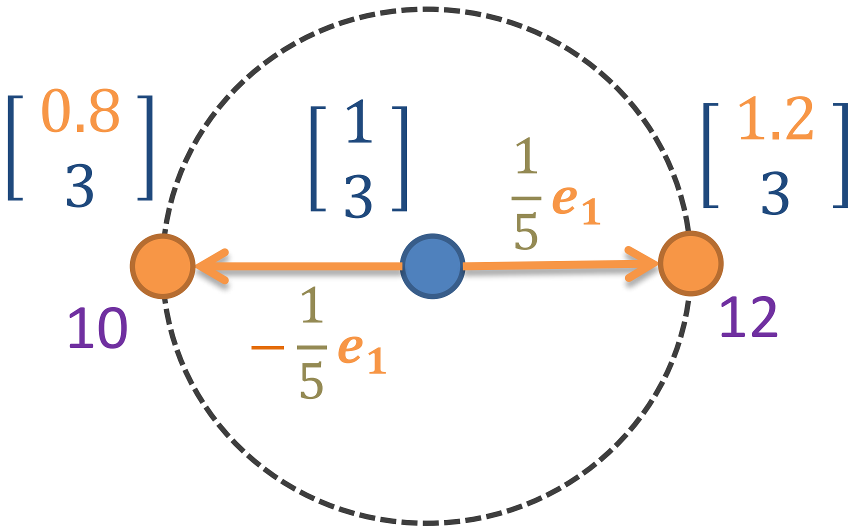


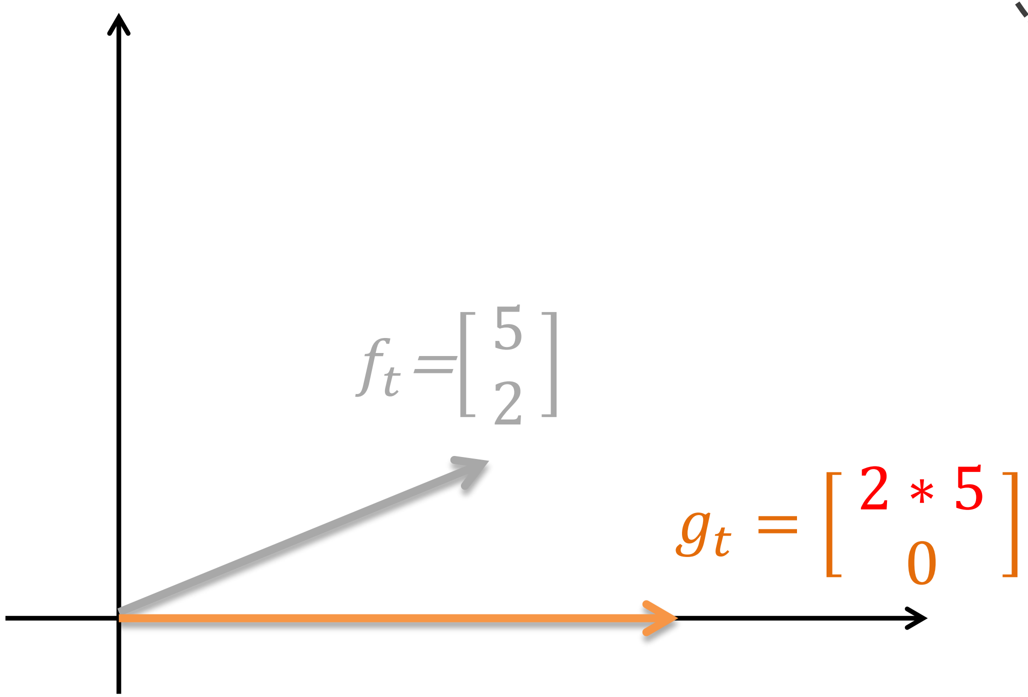
$$12 - 10 = \left(\frac{1}{5} + \frac{1}{5}\right) * 5$$

First Try

- In round t

$$\begin{array}{ccc} & e_1 & e_2 \\ u_t \in_R & \left\{\begin{bmatrix}1\\0\end{bmatrix}\right., & \left.\begin{bmatrix}0\\1\end{bmatrix}\right\} \end{array}$$





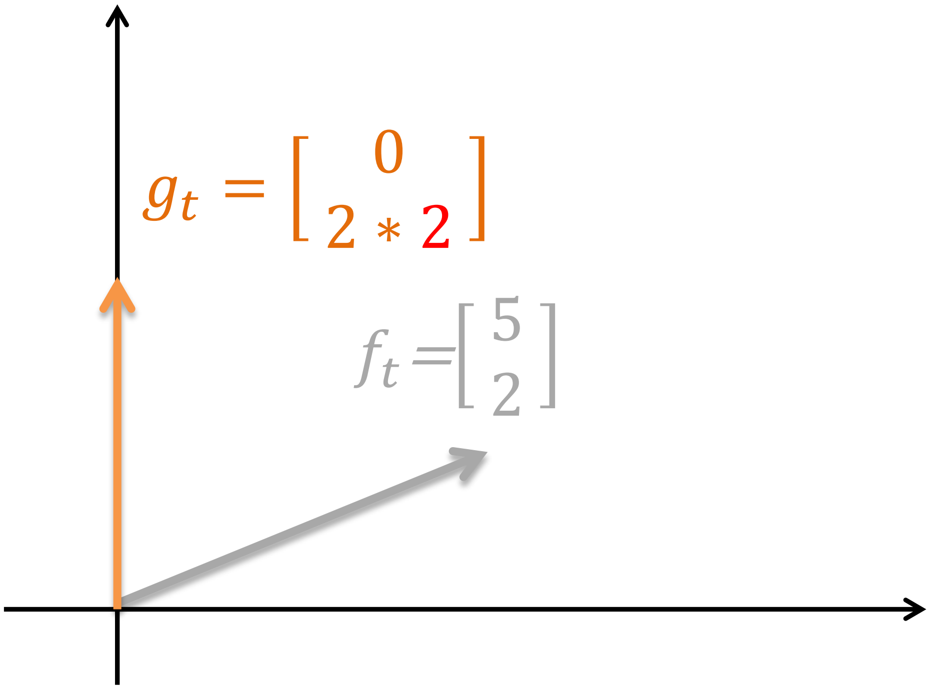
$$f_t = \begin{bmatrix}5\\2\end{bmatrix}$$

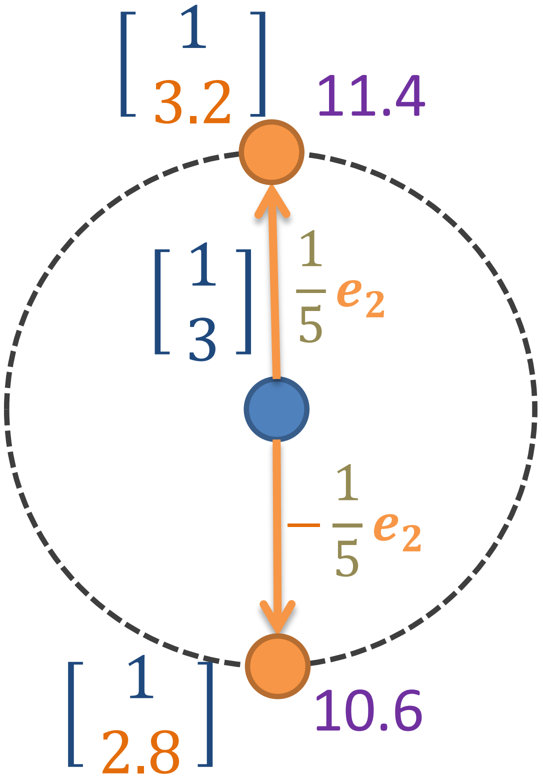
$$g_t = \begin{bmatrix}10\\0\end{bmatrix}$$

First Try

- In round t

$$u_t \in_R \left\{ \overset{e_1}{\begin{bmatrix}1\\0\end{bmatrix}}, \boxed{\overset{e_2}{\begin{bmatrix}0\\1\end{bmatrix}}} \right\}$$





$$f_t = \begin{bmatrix}5\\2\end{bmatrix}$$

$$g_t = \begin{bmatrix}0\\4\end{bmatrix}$$

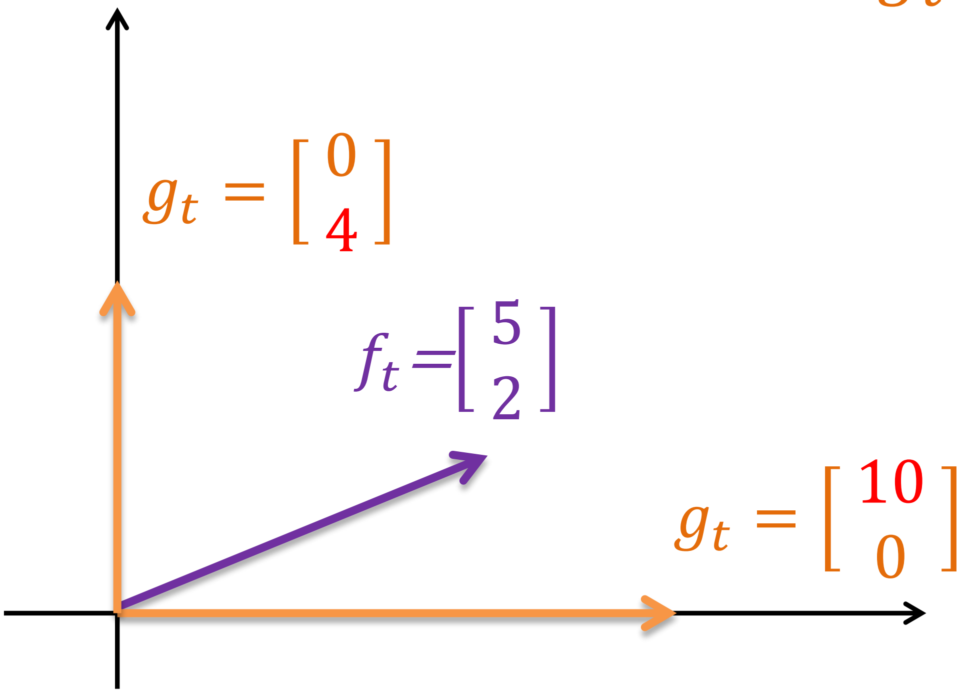
First Try

- In round t

$$u_t \in_R \left\{ \boxed{\overset{e_1}{\begin{bmatrix}1\\0\end{bmatrix}}} \boxed{\overset{e_2}{\begin{bmatrix}0\\1\end{bmatrix}}} \right\}$$

$$f_t = \begin{bmatrix}5\\2\end{bmatrix}$$

$$g_t = \begin{bmatrix}0\\4\end{bmatrix}$$



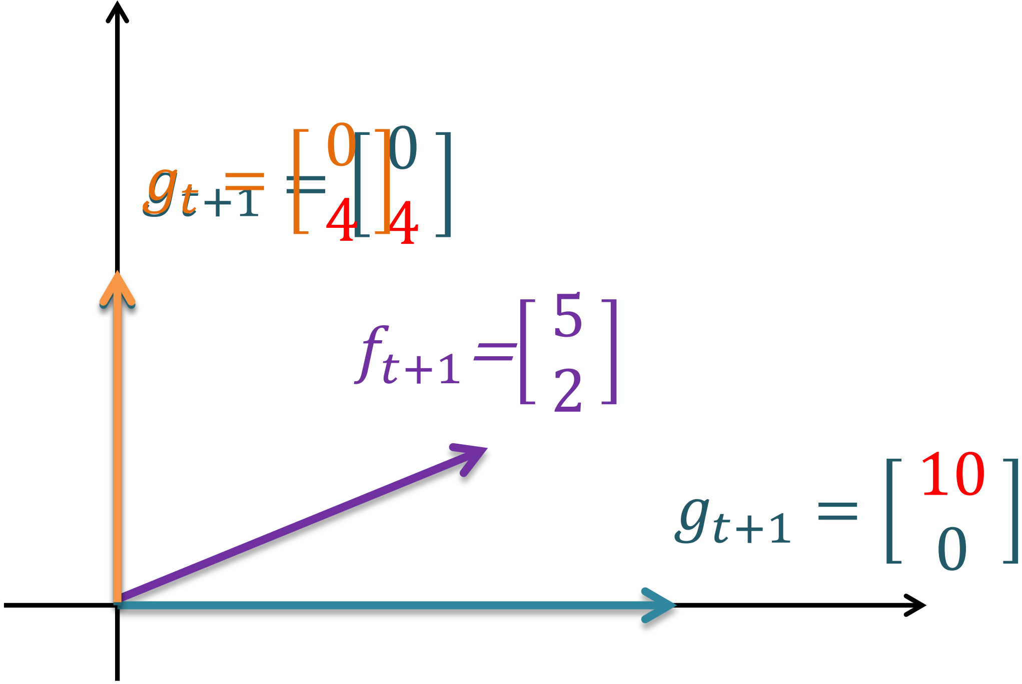
First Try

- In round t+1

$$u_{t+1} \in_R \left\{ \overset{e_1}{\boxed{\begin{bmatrix}1\\0\end{bmatrix}}} \overset{e_2}{\boxed{\begin{bmatrix}0\\1\end{bmatrix}}} \right\}$$

$$f_t = \begin{bmatrix}5\\2\end{bmatrix} \quad f_{t+1} = \begin{bmatrix}5\\2\end{bmatrix}$$

$$g_t = \begin{bmatrix}0\\4\end{bmatrix} \quad g_{t+1} = \begin{bmatrix}10\\0\end{bmatrix}$$



$g_{t+1} \not\approx g_t$
even $f_{t+1} = f_t$

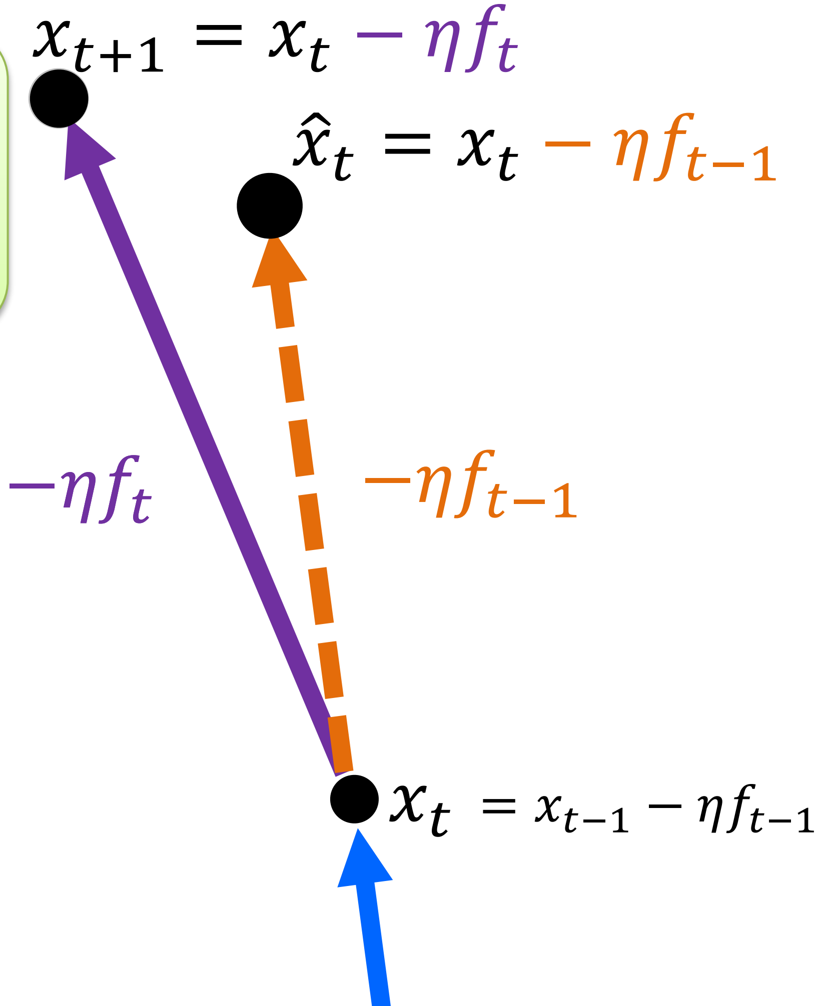
MAIN ALGORITHM

Our Algorithm Idea

Full information Algorithm’s idea

Regret

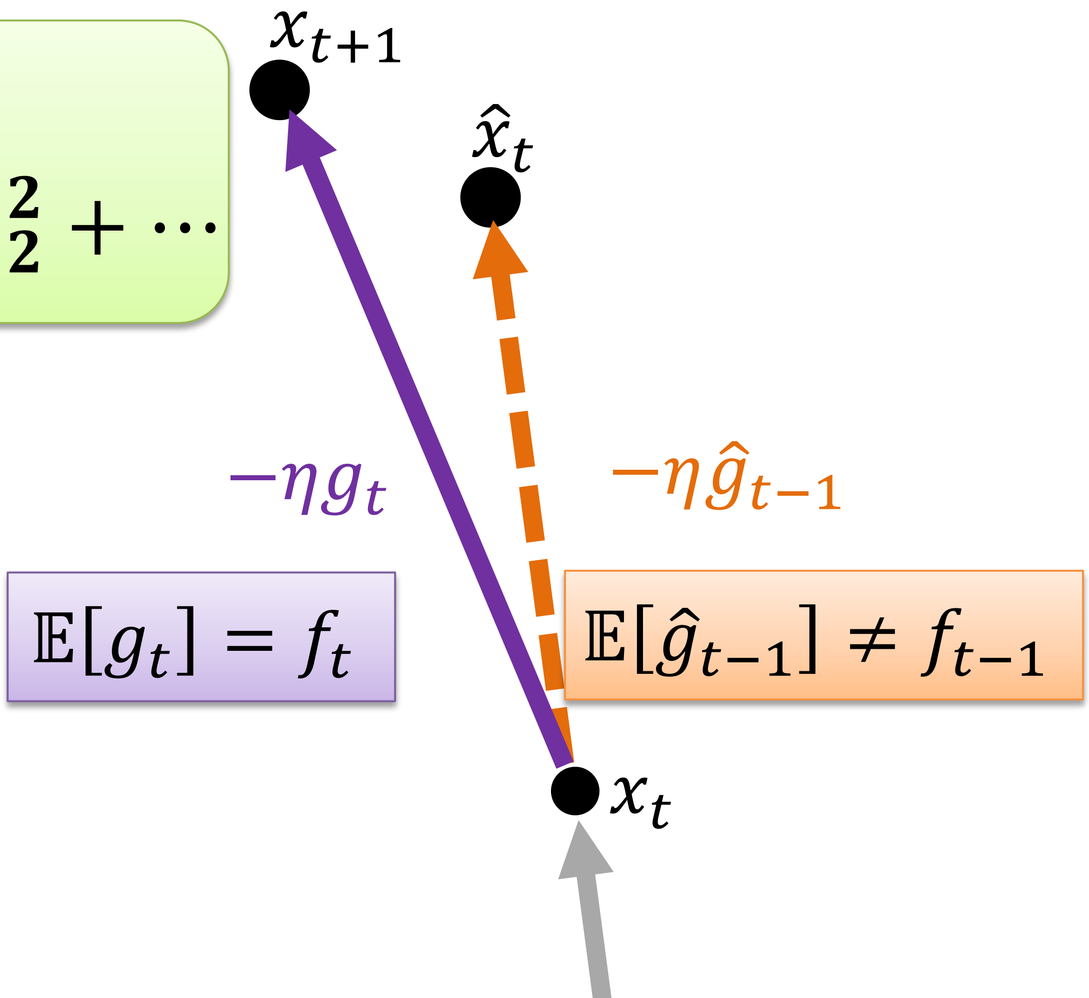
$$= \sum_{t=1}^{T} \eta \| \boxed{f_t} - \boxed{f_{t-1}} \|_2^2 + \cdots$$



Algorithm’s Idea

Regret

$$= \sum_{t=1}^{T} \eta \|g_t - \hat{g}_{t-1}\|_2^2 + \cdots$$



$\mathbb{E}[g_t] = f_t$

$\mathbb{E}[\hat{g}_{t-1}] \neq f_{t-1}$

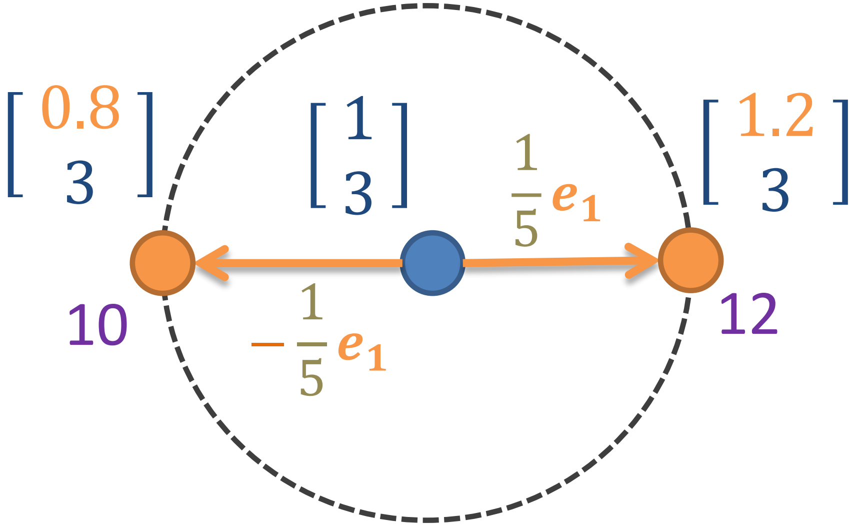
Algorithm

$$f_t \quad \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

- In round t

Know $\hat{g}_{t-1} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$

$$u_t \in_R \left\{ \begin{array}{cc} e_1 & e_2 \\ \boxed{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} , & \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{array} \right\}$$



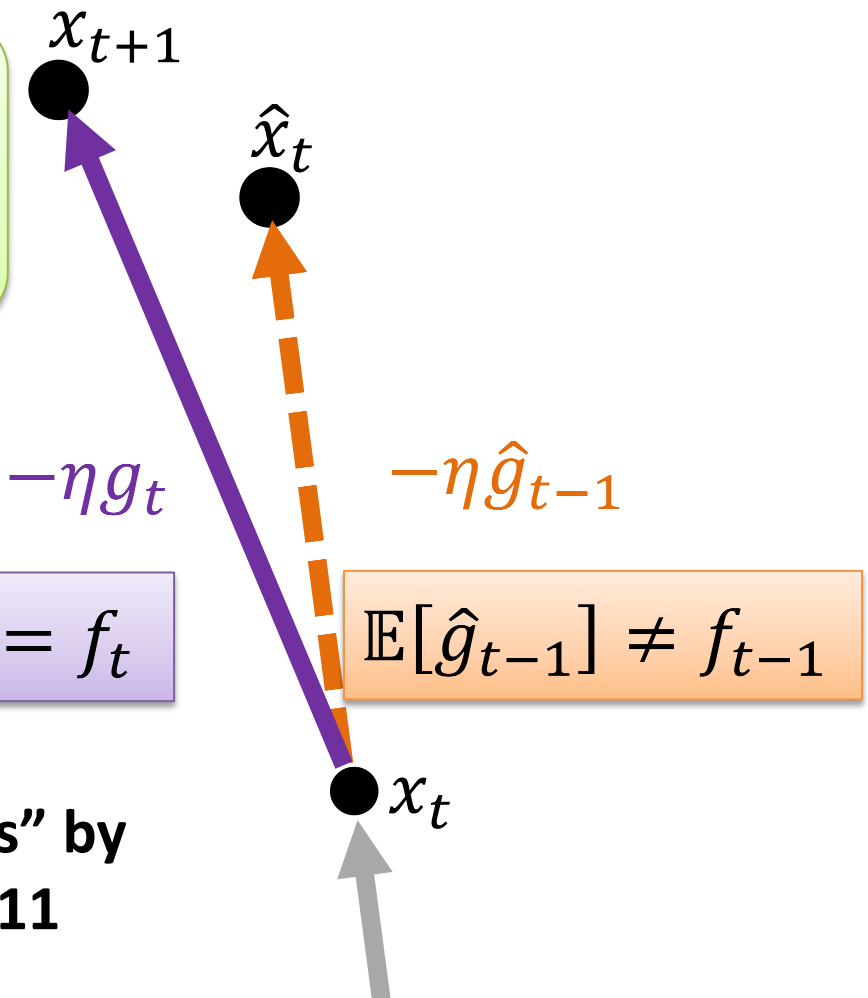
$$12 - 10 = \left(\frac{1}{5} + \frac{1}{5}\right) * 5$$

$$\hat{g}_t = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$$

Algorithm's Idea

Regret

$$= \sum_{t=1}^{T} \eta \|g_t - \hat{g}_{t-1}\|_2^2 + \cdots$$



$\mathbb{E}[g_t] = f_t$

$\mathbb{E}[\hat{g}_{t-1}] \neq f_{t-1}$

"Better Algorithms for Benign Bandits" by Elad Hazan and Satyen Kale, JMLR 2011

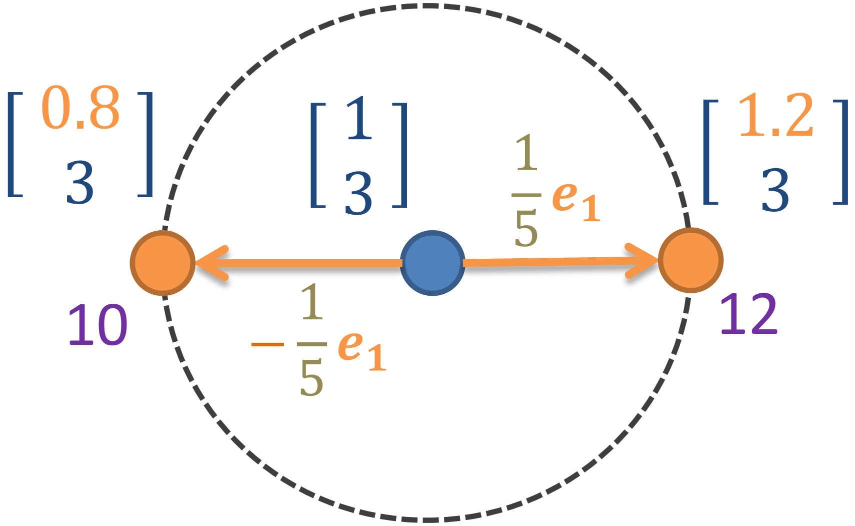
$$f_t \quad \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

Algorithm

- In round t

Know $\hat{g}_{t-1} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$

$$u_t \in_R \left\{ \begin{array}{cc} e_1 & e_2 \\ \boxed{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} & \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{array} \right\}$$

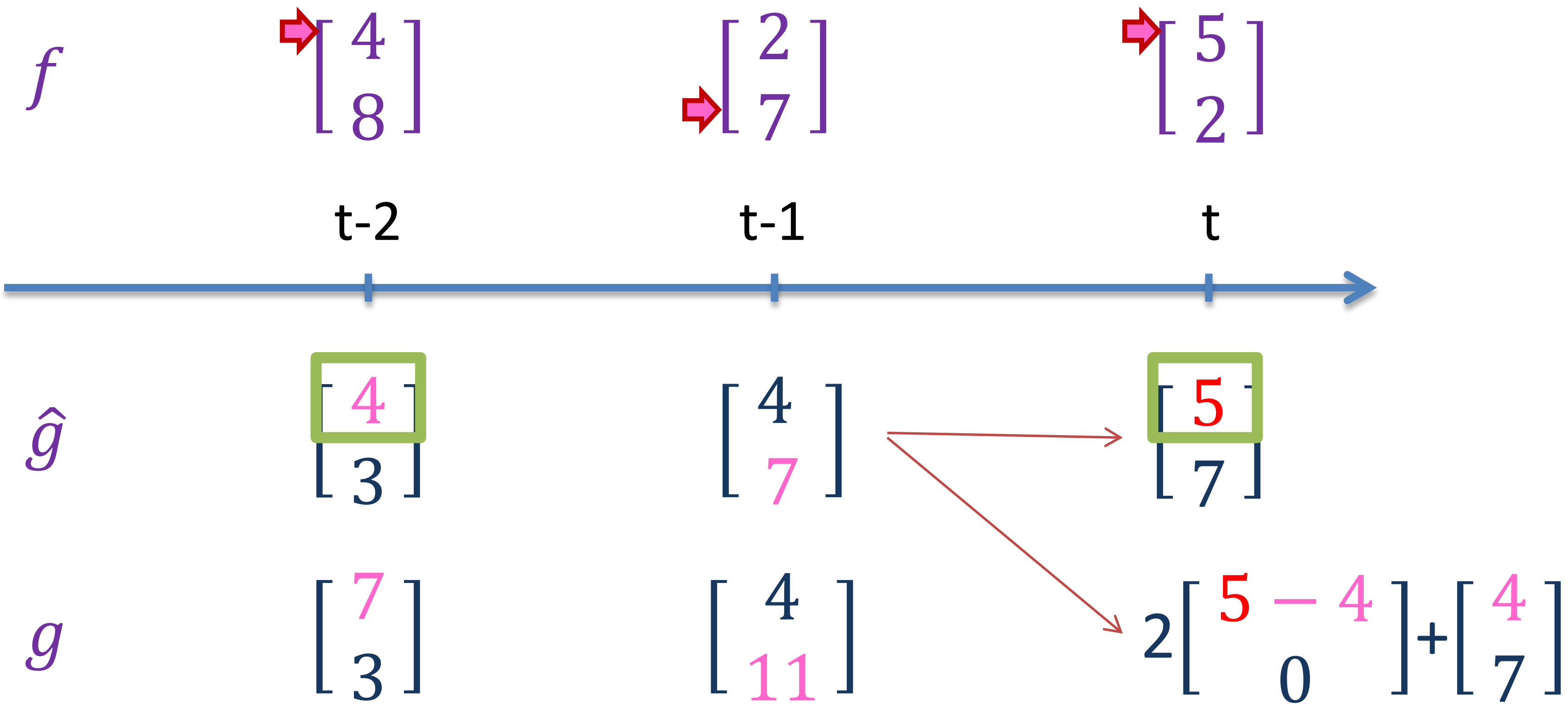


$$g_t = h_t + \hat{g}_{t-1}$$

$$= 2\begin{bmatrix} 5-4 \\ 0 \end{bmatrix} + \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

$$\mathbb{E}[g_t] = f_t$$

Analysis



f $\begin{bmatrix}4\\8\end{bmatrix}$ $\begin{bmatrix}2\\7\end{bmatrix}$ $\begin{bmatrix}5\\2\end{bmatrix}$

t-2 t-1 t

$\hat{g}$ $\begin{bmatrix}\boxed{4}\\3\end{bmatrix}$ $\begin{bmatrix}4\\7\end{bmatrix}$ $\begin{bmatrix}\boxed{5}\\7\end{bmatrix}$

g $\begin{bmatrix}7\\3\end{bmatrix}$ $\begin{bmatrix}4\\11\end{bmatrix}$ $2\begin{bmatrix}5-4\\0\end{bmatrix}+\begin{bmatrix}4\\7\end{bmatrix}$

$$\|g_t - \hat{g}_{t-1}\|_2 = 2|5-4| = 2\left\|f_{t,1} - f_{t-2,1}\right\|_2$$

Results

$$\sum_{t=1}^{T} \|f_t - f_{t-1}\|^2$$

Problems	T rounds		Deviation D	
	Full	Two-point Bandit	Full	Two-point Bandit
Online Linear Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL+12]	$O(\sqrt{D})$
Online Convex Optimization	$O(\sqrt{T})$ [Zin03]	$O(\sqrt{T})$ [ADX10]	$O(\sqrt{D})$ [CYL+12]	$O(\sqrt{D})$
Online Strongly Convex Optimization	$O(\log T)$ [HAK07]	$O(\log T)$ [ADX10]	$O(\log D)$ [CYL+12]	$O(\log D)$

Thank you !!