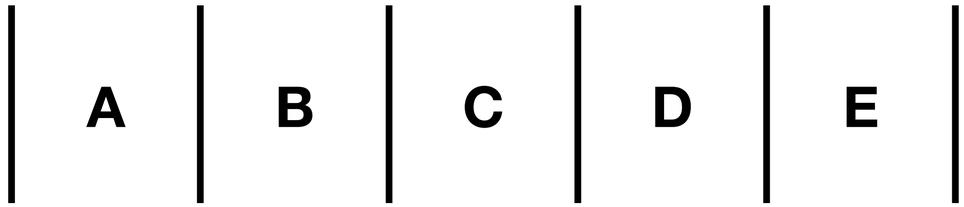
Time Efficient Swap Regret Minimization

Ashley Yu
Mentored by: Maxwell Fishelson

Online Learning

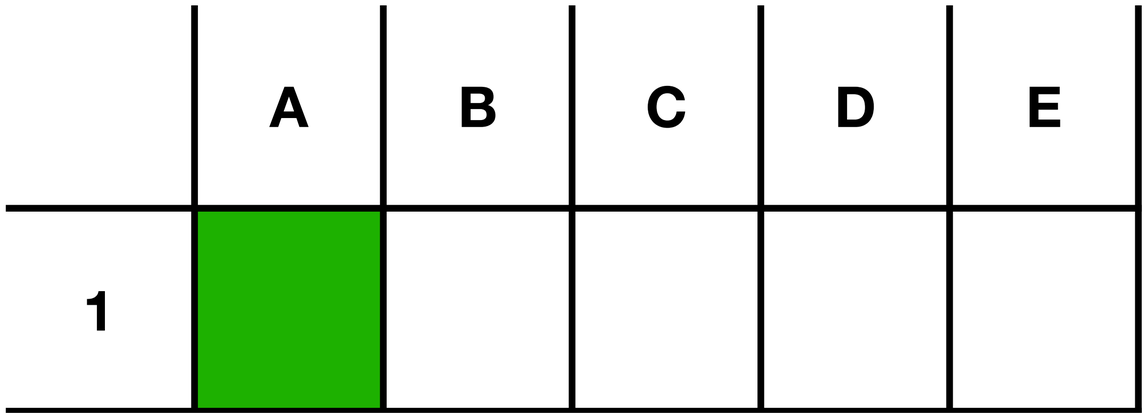
Online Learning



Online Learning

	A	B	C	D	E
1					

Online Learning



Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2					

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2					

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2	0.8	0.4	0	0.2	0.3

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2	0.8	0.4	0	0.2	0.3
3					

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2	0.8	0.4	0	0.2	0.3
3					

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2	0.8	0.4	0	0.2	0.3
3	0.1	0.2	0.3	0.4	0.5

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2	0.8	0.4	0	0.2	0.3
3	0.1	0.2	0.3	0.4	0.5

$$x^{(t)} \in \Delta(N)$$

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2	0.8	0.4	0	0.2	0.3
3	0.1	0.2	0.3	0.4	0.5

$$x^{(t)} \in \Delta(N)$$

$$u^{(t)} \in [0,1]^N$$

Online Learning

	A	B	C	D	E
1	0.1	0.3	0.9	0.5	0.7
2	0.8	0.4	0	0.2	0.3
3	0.1	0.2	0.3	0.4	0.5

$$x^{(t)} \in \Delta(N)$$

$$u^{(t)} \in [0,1]^N$$

$$\text{Total Reward} = \sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

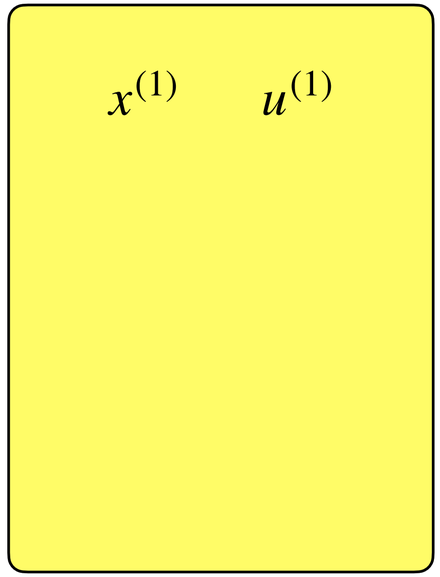
Regret

Regret

regret = reward obtained vs benchmark

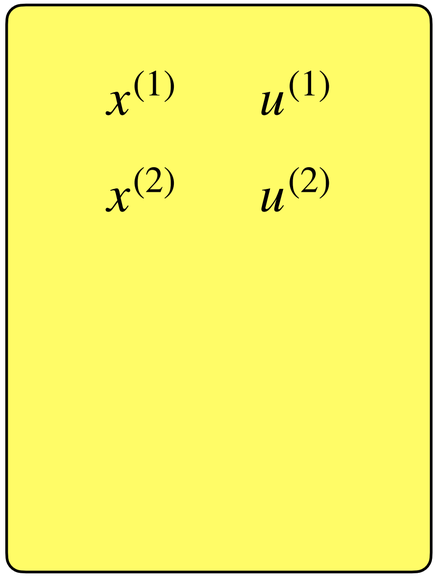
Regret

regret = reward obtained vs benchmark



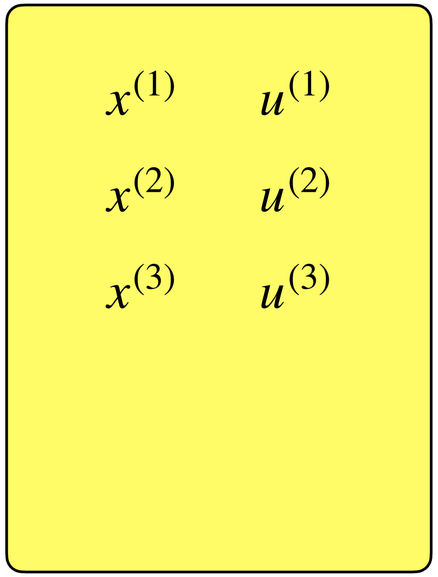
Regret

regret = reward obtained vs benchmark



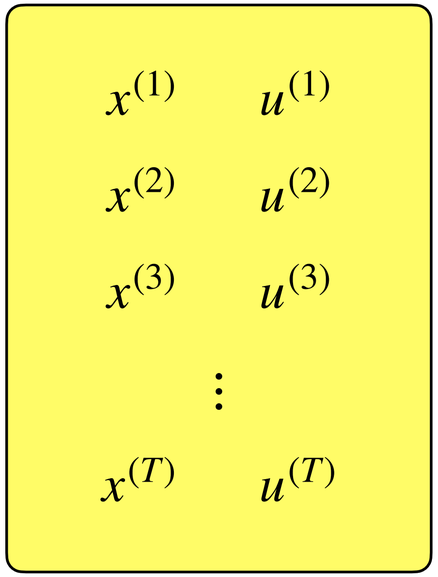
Regret

regret = reward obtained vs benchmark



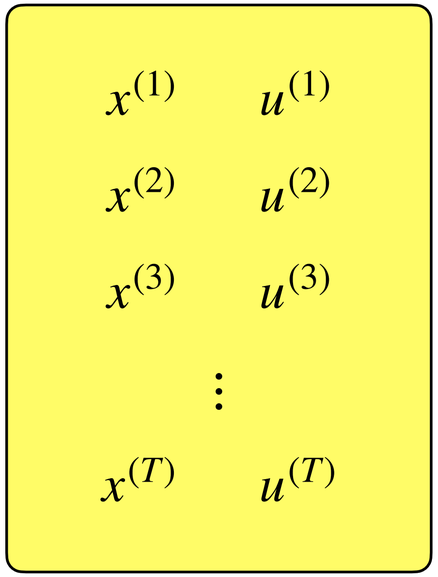
Regret

regret = reward obtained vs benchmark



Regret

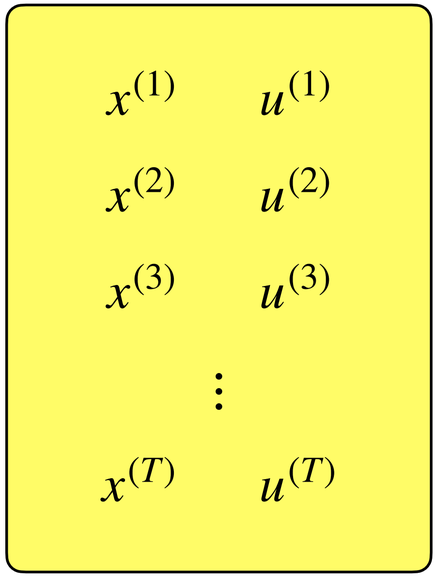
regret = reward obtained vs benchmark



$$\sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

Regret

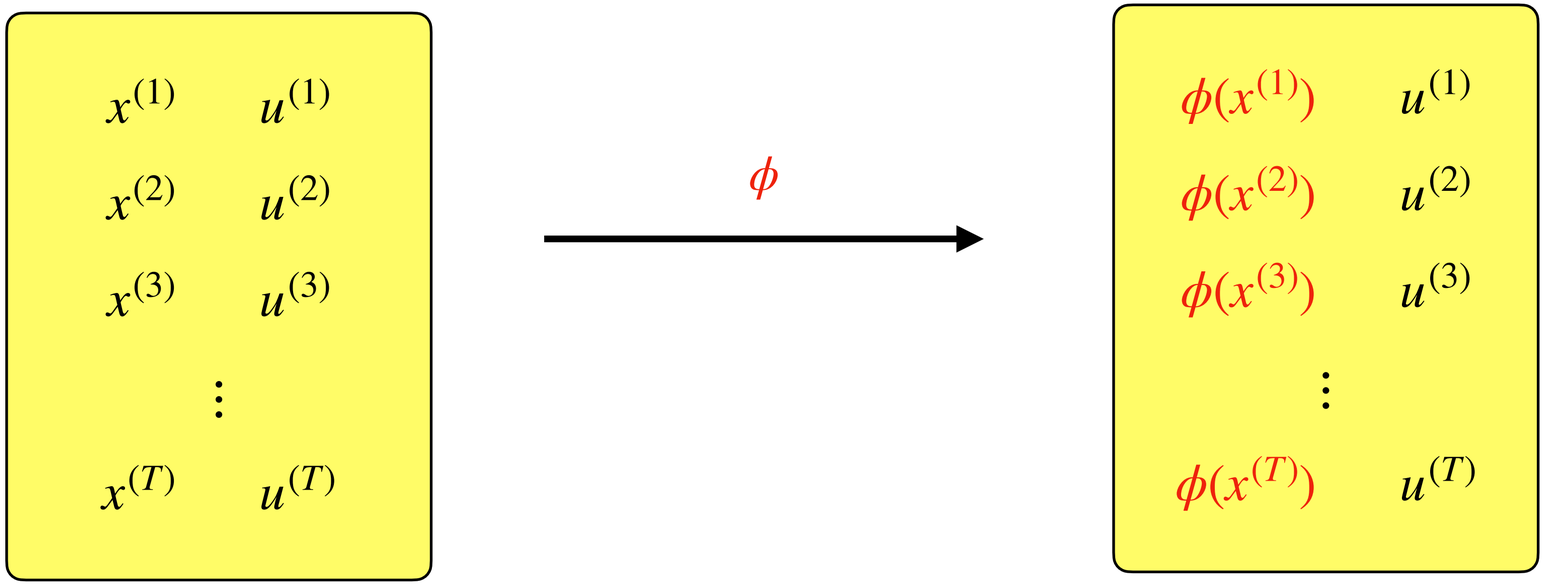
regret = reward obtained vs benchmark



$$\sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

Regret

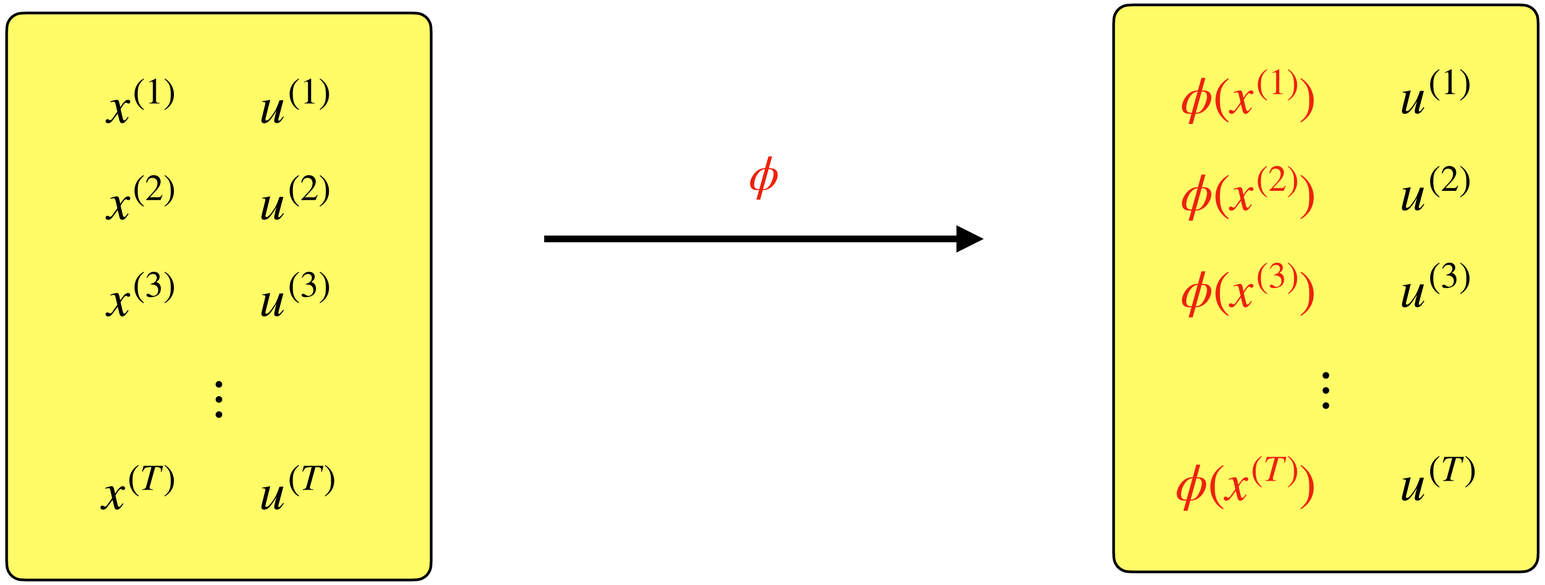
regret = reward obtained vs benchmark



$$\sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

Regret

regret = reward obtained vs benchmark

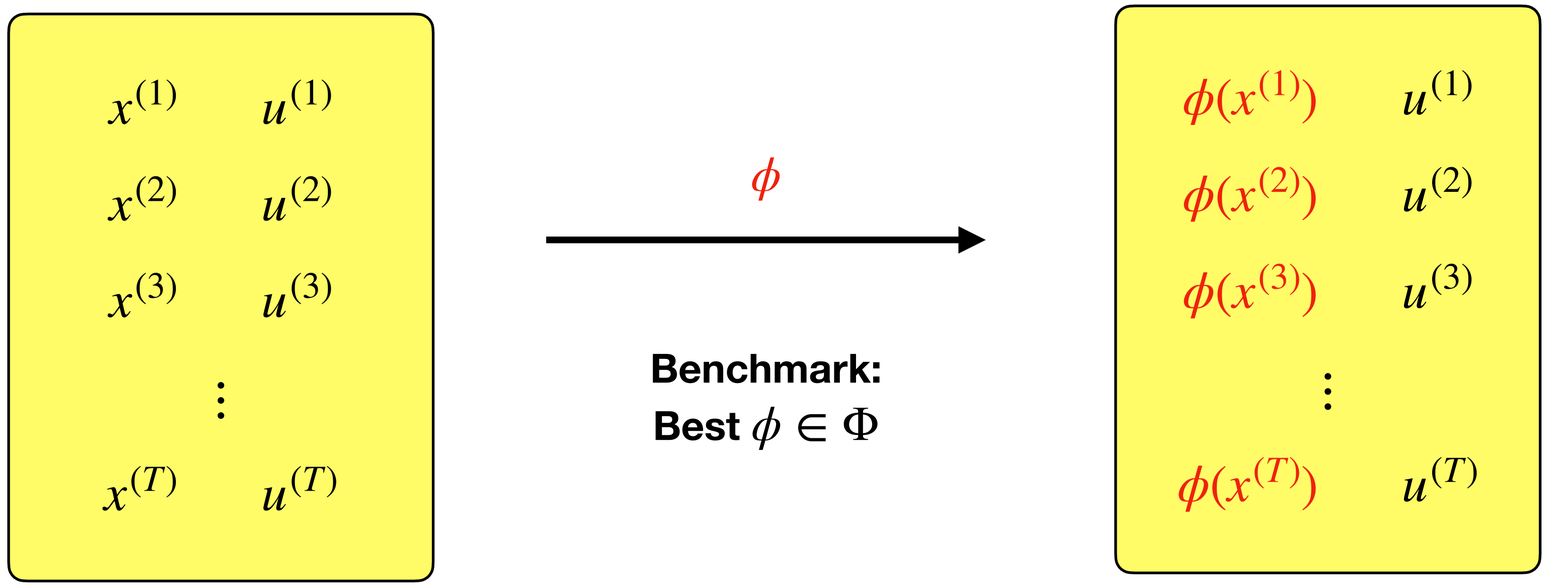


$$\sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

$$\sum_{t=1}^{T} \phi(x^{(t)}) \cdot u^{(t)}$$

Regret

regret = reward obtained vs benchmark

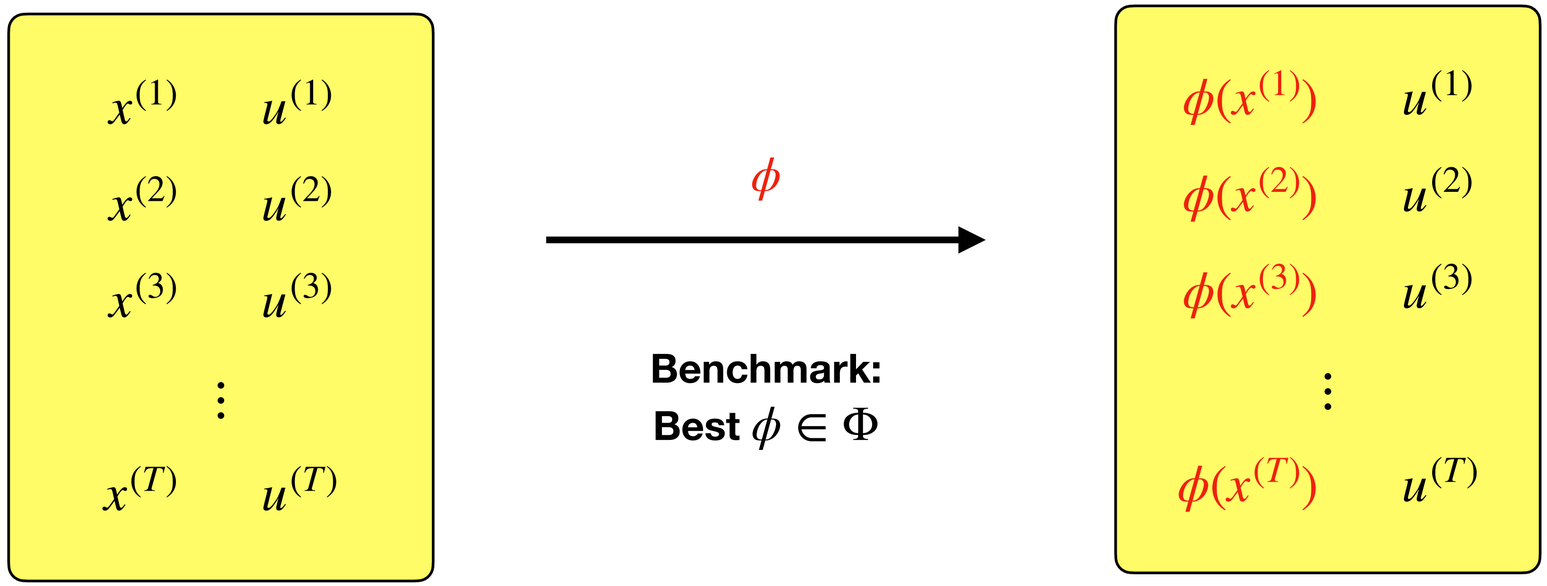


$$\sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

$$\sum_{t=1}^{T} \phi(x^{(t)}) \cdot u^{(t)}$$

Regret

regret = reward obtained vs benchmark



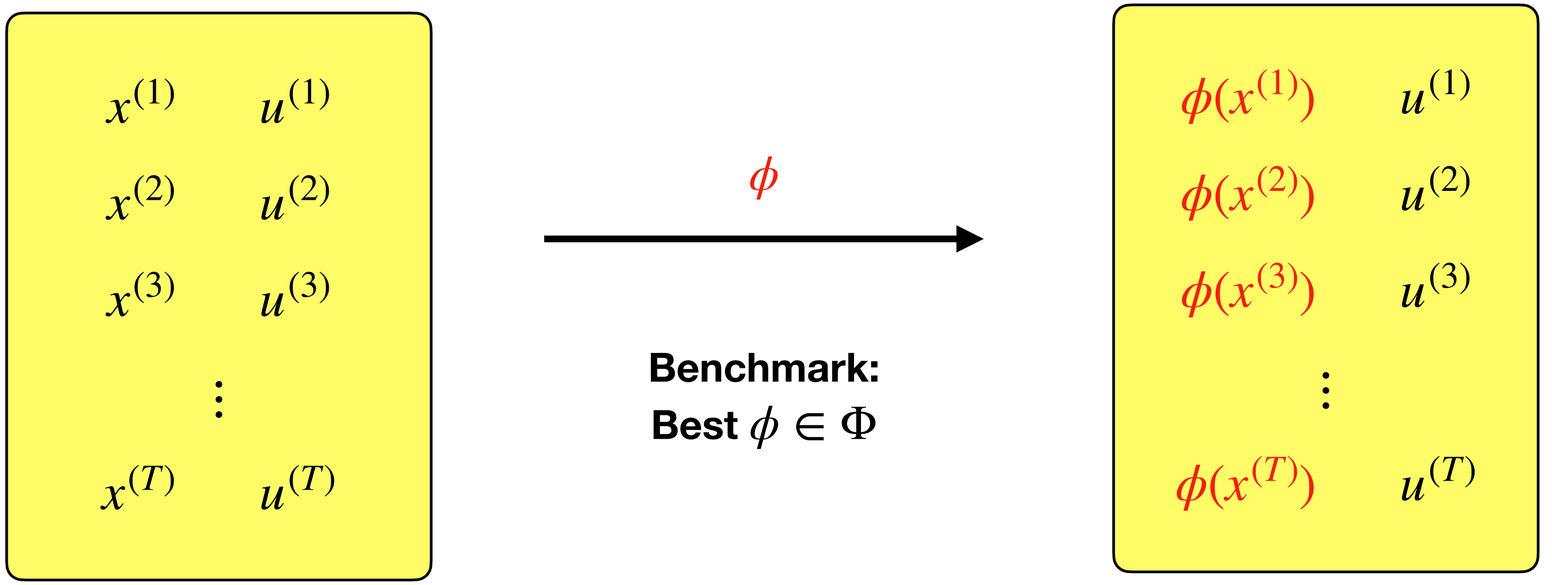
$$\sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

$$\sum_{t=1}^{T} \phi(x^{(t)}) \cdot u^{(t)}$$

$$\text{Regret}\left(x^{(1:T)}, u^{(1:T)}\right) = \max_{\phi \in \Phi} \sum_{t=1}^{T} \phi(x^{(t)}) \cdot u^{(t)} - \sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

Regret

regret = reward obtained vs benchmark



$$\sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

$$\sum_{t=1}^{T} \phi(x^{(t)}) \cdot u^{(t)}$$

$$\text{Regret}\left(x^{(1:T)}, u^{(1:T)}\right) = \max_{\phi \in \Phi} \sum_{t=1}^{T} \phi(x^{(t)}) \cdot u^{(t)} - \sum_{t=1}^{T} x^{(t)} \cdot u^{(t)}$$

ϵ-Regret: $\leq \epsilon T$

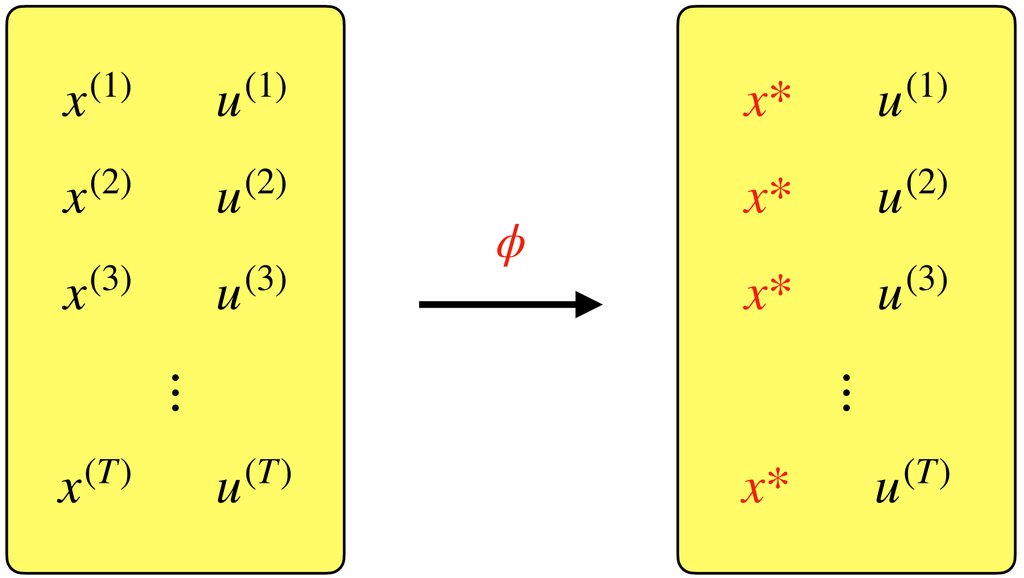
External Regret

External Regret

Φ = all constant maps

External Regret

Φ = all constant maps



Swap Regret

Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$

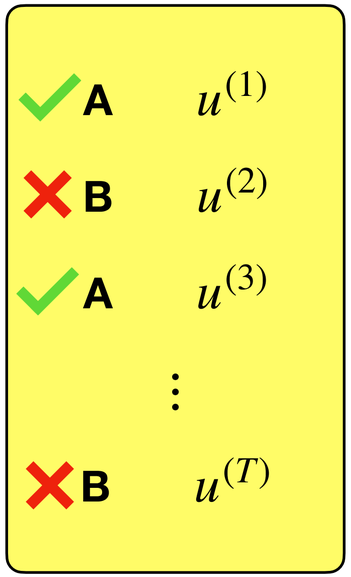
Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$

A	$u^{(1)}$
B	$u^{(2)}$
A	$u^{(3)}$
	$\vdots$
B	$u^{(T)}$

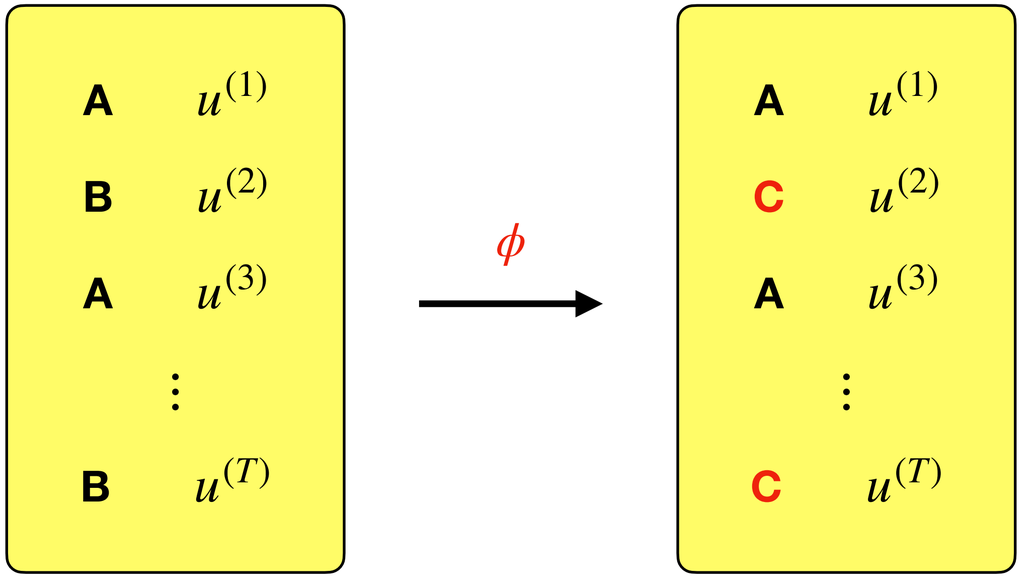
Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$



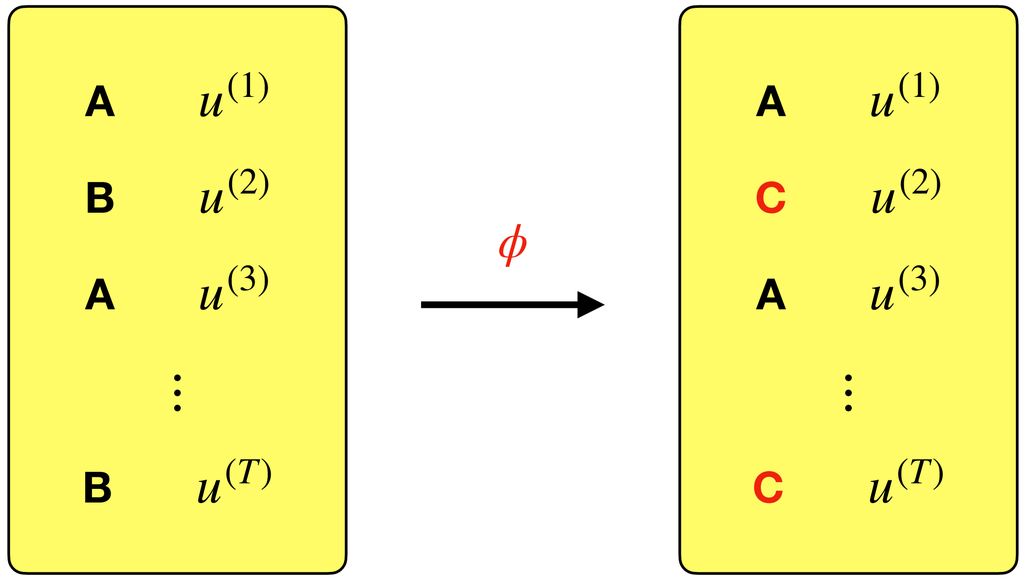
Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$



Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$

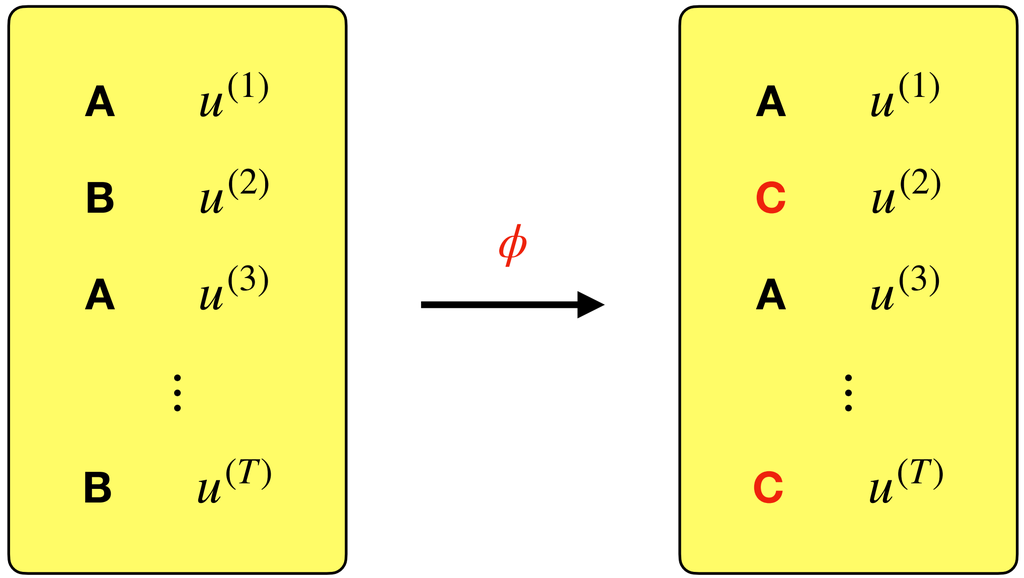


ϕ

A	A
B	C
C	C
⋮	⋮

Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$



ϕ

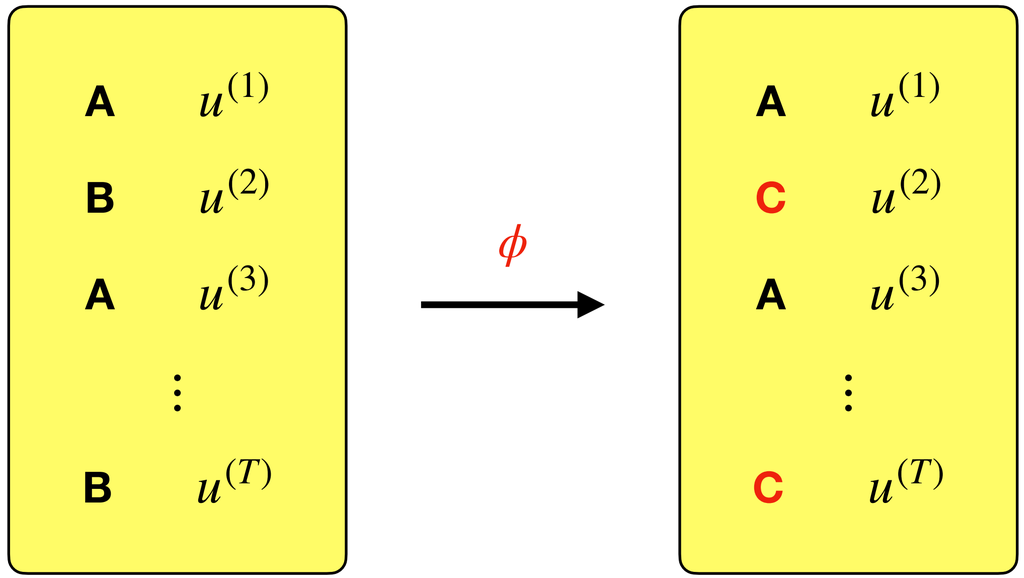
A	A
B	C
C	C
⋮	⋮

ϕ

A	D
B	D
C	D
⋮	⋮

Swap Regret

Φ = **all maps** $[N] \to [N]$



ϕ

A	A
B	C
C	C
⋮	⋮

ϕ

A	D
B	D
C	D
⋮	⋮

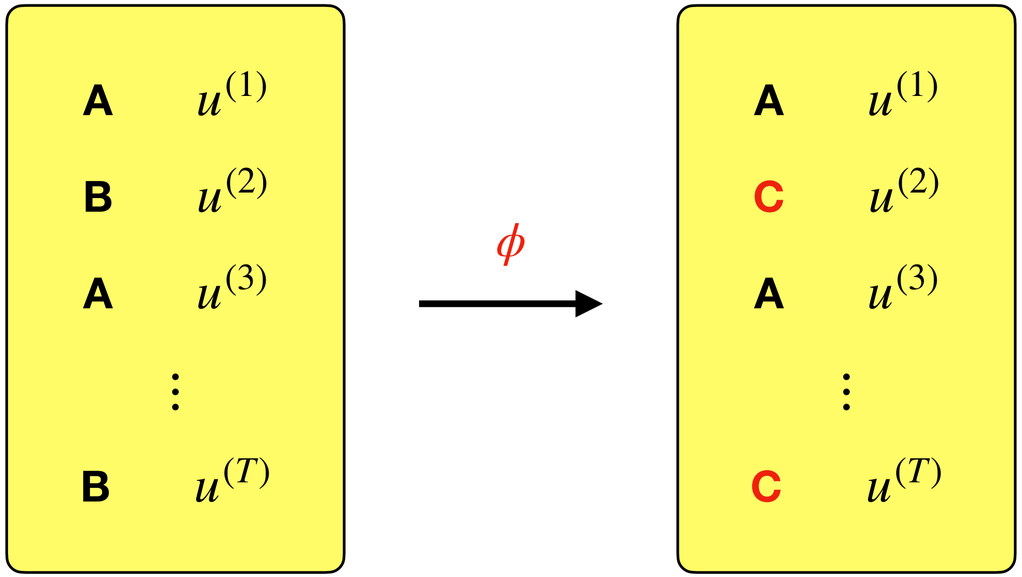
ϕ

A	X
B	Y
C	Z
⋮	⋮

Swap Regret

Φ = **all maps** $[N] \to [N]$

Φ = **all linear maps** $\Delta(N) \to \Delta(N)$



ϕ

A	A
B	C
C	C
⋮	⋮

ϕ

A	D
B	D
C	D
⋮	⋮

ϕ

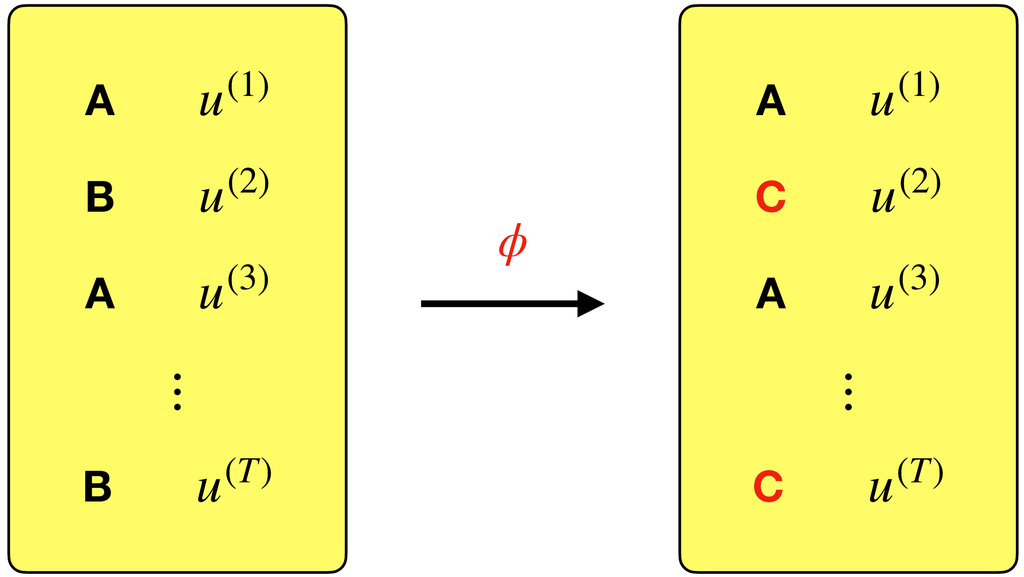
A	X
B	Y
C	Z
⋮	⋮

Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$

$\Phi =$ **all linear maps** $\Delta(N) \to \Delta(N)$

$\Phi =$ **ALL maps** $\Delta(N) \to \Delta(N)$



ϕ

A	A
B	C
C	C
⋮	⋮

ϕ

A	D
B	D
C	D
⋮	⋮

ϕ

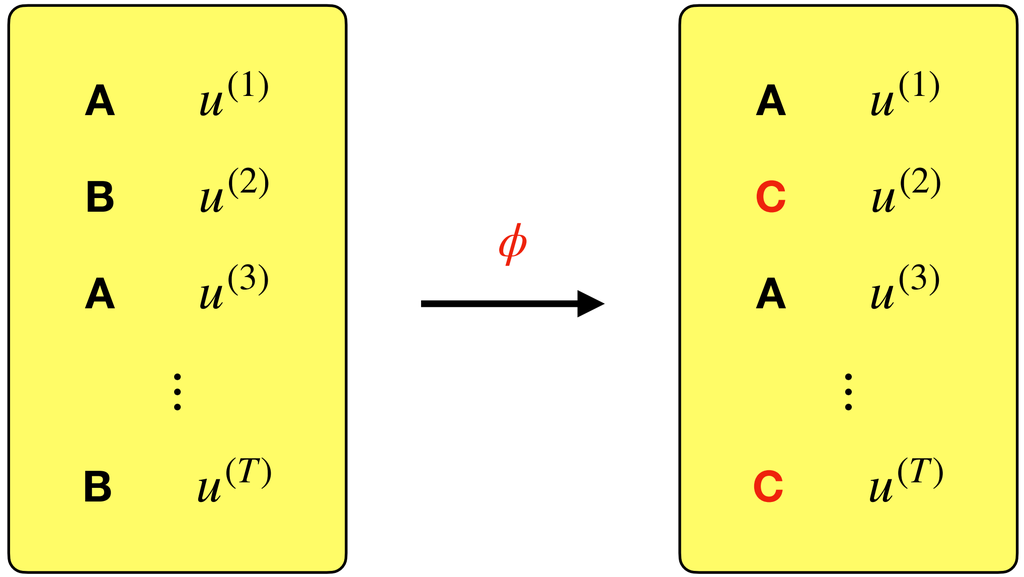
A	X
B	Y
C	Z
⋮	⋮

Swap Regret

$\Phi =$ **all maps** $[N] \to [N]$

$\Phi =$ **all linear maps** $\Delta(N) \to \Delta(N)$

$\Phi =$ **ALL maps** $\Delta(N) \to \Delta(N)$

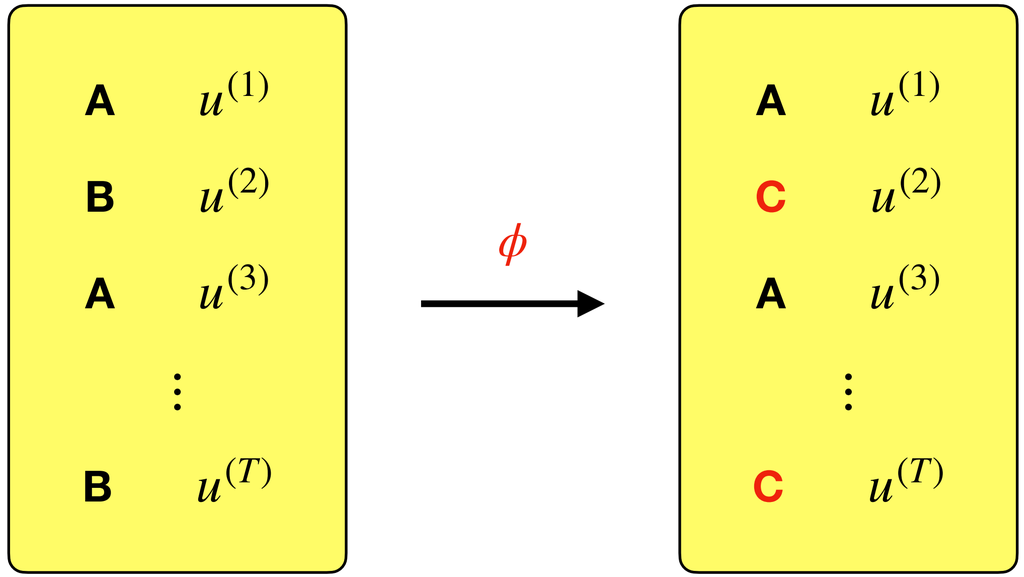


Swap Regret

Φ = **all maps** $[N] \to [N]$

Φ = **all linear maps** $\Delta(N) \to \Delta(N)$

Φ = **ALL maps** $\Delta(N) \to \Delta(N)$



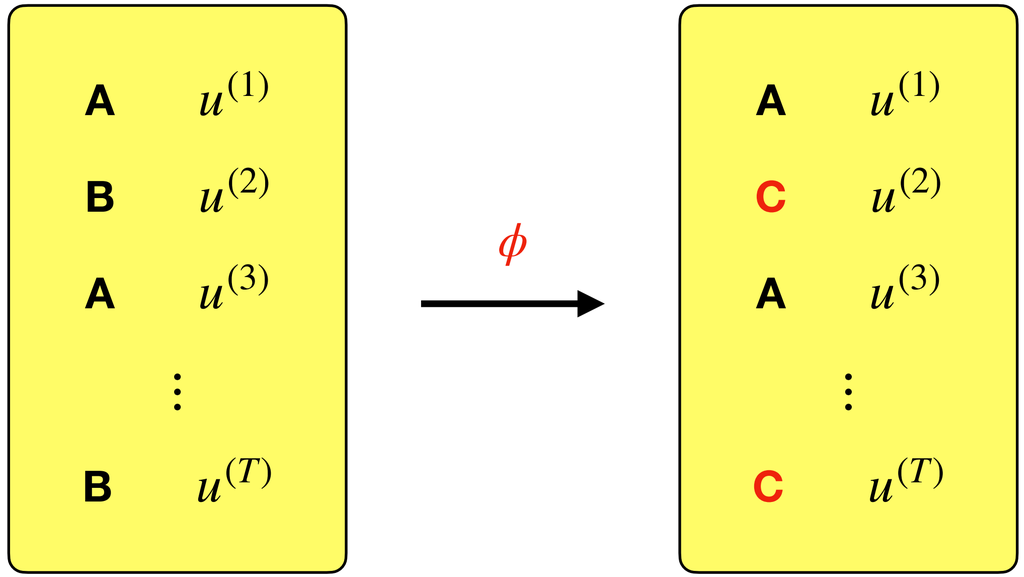
Why?

Swap Regret

Φ = **all maps** $[N] \to [N]$

Φ = **all linear maps** $\Delta(N) \to \Delta(N)$

Φ = **ALL maps** $\Delta(N) \to \Delta(N)$



Why?

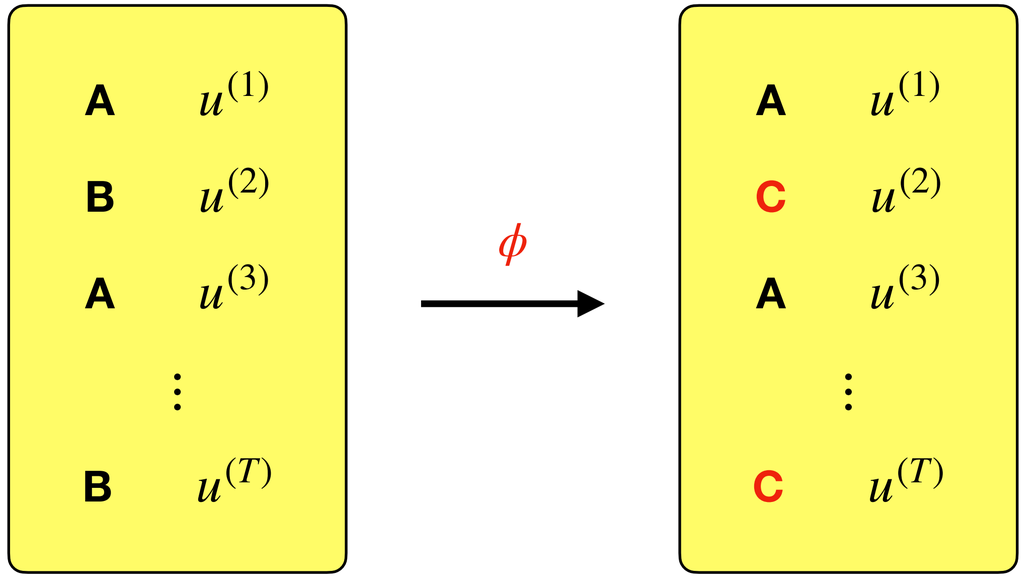
- Unexploitable

Swap Regret

Φ = **all maps** $[N] \to [N]$

Φ = **all linear maps** $\Delta(N) \to \Delta(N)$

Φ = **ALL maps** $\Delta(N) \to \Delta(N)$



Why?

- Unexploitable
- Games!

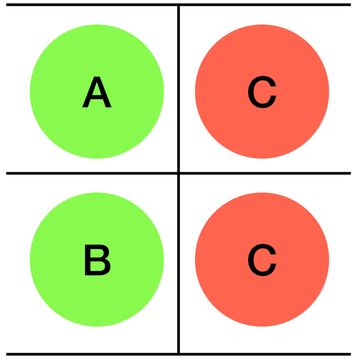
Learning in Games

ϵ external regret over T rounds → average strategies are ϵ coarse-correlated-equilibrium

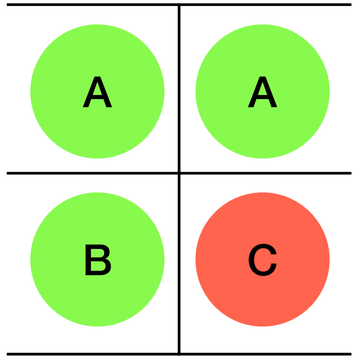
ϵ swap regret over T rounds → average strategies are ϵ correlated-equilibrium

Game Equilibrium: Strategy profile where no player wants to deviate

Coarse Deviations



Swap Deviations



CCE: No player wants to make a coarse deviation

So why not Swap Regret?

So why not Swap Regret?

Multiplicative Weight Updates: ϵ-External-Regret for $T = \Omega\left(\frac{\log N}{\epsilon^2}\right)$

So why not Swap Regret?

Multiplicative Weight Updates: ϵ-External-Regret for $T = \Omega\left(\frac{\log N}{\epsilon^2}\right)$

Blum-Mansour MWU [BM07]**:** ϵ-Swap-Regret for $T = \Omega\left(\frac{N \log N}{\epsilon^2}\right)$

So why not Swap Regret?

Multiplicative Weight Updates: ϵ-External-Regret for $T = \Omega\left(\frac{\log N}{\epsilon^2}\right)$

Blum-Mansour MWU [BM07]**:** ϵ-Swap-Regret for $T = \Omega\left(\frac{N \log N}{\epsilon^2}\right)$

Question: can we improve this?

So why not Swap Regret?

Blum Mansour task the learner with computing the stationary distribution of a Markov chain at each time step.

Instead, we propose estimating the stationary distribution through sampling