Multi-agent learning

Conditional Regret

Gerard Vreeswijk, Intelligent Systems Group, Computer Science Department, Faculty of Sciences, Utrecht University, The Netherlands.

Thursday 19th February, 2015

Conditional regret: example

Shapley's game:

$$
G = \begin{array}{c} \\ R \\ Y \\ B \end{array}
\begin{array}{c}
\begin{array}{ccc} R & Y & B \end{array} \\
\begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix}
\end{array}
$$

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds: 1 2 3 4 5 6 7 8 9 10

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:										

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y

Conditional regret: example

Shapley's game:

$$G = \begin{array}{cc} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:										

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c c} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:										

Conditional regret: example

Shapley's game:

$$G = \begin{array}{cc} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0

Conditional regret: example

Shapley's game:

$$G = \begin{array}{cc} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:										

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c c} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R
Hyp. payoff row:										

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R
Hyp. payoff row:	1	0	0	0	1	1	1	0	1	0

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R
Hyp. payoff row:	1	0	0	0	1	1	1	0	1	0

Better for row to play R in the three periods where he actually played Y?

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c c} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R
Hyp. payoff row:	1	0	0	0	1	1	1	0	1	0

Better for row to play R in the three periods where he actually played Y?

- Actual payoff Y: 3×0

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R
Hyp. payoff row:	1	0	0	0	1	1	1	0	1	0

Better for row to play R in the three periods where he actually played Y?

- Actual payoff Y: 3×0
- Hypothetical payoff R: 3×1.

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R
Hyp. payoff row:	1	0	0	0	1	1	1	0	1	0

Better for row to play R in the three periods where he actually played Y?

- Actual payoff Y: 3×0
- Hypothetical payoff R: 3×1.
- Average regret: $(3 - 0)/10$.

Conditional regret: example

Shapley's game:

$$G = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \\ 0,0 & 0,1 & 1,0 \end{pmatrix} \end{array}$$

Rounds:	1	2	3	4	5	6	7	8	9	10
Action col:	R	B	Y	Y	B	R	R	Y	R	Y
Action row:	R	R	B	B	B	Y	Y	R	Y	R
Payoff row:	1	0	0	0	1	0	0	0	0	0
Hyp. action row:	R	R	B	B	B	R	R	R	R	R
Hyp. payoff row:	1	0	0	0	1	1	1	0	1	0

Better for row to play R in the three periods where he actually played Y?

- Actual payoff Y: 3×0
- Hypothetical payoff R: 3×1.
- Average regret: $(3 - 0)/10$.

The complete conditional regret matrix is:

$$\mathbf{R} = \begin{array}{c} \\ R \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R & Y & B \end{array} \\ \begin{pmatrix} 0.0 & 0.1 & 0.0 \\ 0.3 & 0.0 & 0.0 \\ -0.1 & 0.1 & 0.0 \end{pmatrix} \end{array}$$

Row: original actions;
column: alternative actions.

Conditional regret matrix

The conditional regret matrix at time t is

$$R^t(\omega) =_{Def} \begin{pmatrix} (u(1,y^t) - u(1,y^t))e_1^t & \dots & (u(k,y^t) - u(1,y^t))e_1^t \\ \vdots & \ddots & \vdots \\ (u(1,y^t) - u(k,y^t))e_k^t & \dots & (u(k,y^t) - u(k,y^t))e_k^t \end{pmatrix}$$

Conditional regret matrix

The conditional regret matrix at time t is

$$R^t(\omega) =_{Def} \begin{pmatrix} (u(1,y^t) - u(1,y^t))e_1^t & \dots & (u(k,y^t) - u(1,y^t))e_1^t \\ \vdots & \ddots & \vdots \\ (u(1,y^t) - u(k,y^t))e_k^t & \dots & (u(k,y^t) - u(k,y^t))e_k^t \end{pmatrix}$$

The player's conditional regrets through time t are given by the average of the R^t's:

$$\bar{R}^t =_{Def} \frac{1}{t} \sum_{s=1}^{t} R^t.$$

Conditional regret matrix

The conditional regret matrix at time t is

$$R^t(\omega) =_{Def} \begin{pmatrix} (u(1,y^t) - u(1,y^t))e_1^t & \dots & (u(k,y^t) - u(1,y^t))e_1^t \\ \vdots & \ddots & \vdots \\ (u(1,y^t) - u(k,y^t))e_k^t & \dots & (u(k,y^t) - u(k,y^t))e_k^t \end{pmatrix}$$

The player's conditional regrets through time t are given by the average of the R^t's:

$$\bar{R}^t =_{Def} \frac{1}{t} \sum_{s=1}^{t} R^t.$$

- Both $\bar{R}^t$ and $\bar{R}^t$ have zeros on the diagonal.

Conditional regret matrix

The conditional regret matrix at time t is

$$R^t(\omega) =_{Def} \begin{pmatrix} (u(1,y^t) - u(1,y^t))e_1^t & \dots & (u(k,y^t) - u(1,y^t))e_1^t \\ \vdots & \ddots & \vdots \\ (u(1,y^t) - u(k,y^t))e_k^t & \dots & (u(k,y^t) - u(k,y^t))e_k^t \end{pmatrix}$$

The player's conditional regrets through time t are given by the average of the R^t's:

$$\bar{R}^t =_{Def} \frac{1}{t} \sum_{s=1}^{t} R^t.$$

- Both $\bar{R}^t$ and $\bar{R}^t$ have zeros on the diagonal.
- The average conditional regret vector at time t, $\bar{r}^t$, is just the sum of the columns of $\bar{R}^t$.

Blackwell's approachability theorem

Blackwell's approachability theorem

Blackwell's approachability theorem

- Blackwell's approachability theorem might be considered as the vector-valued form of the minmax-theorem.

Blackwell's approachability theorem

- Blackwell's approachability theorem might be considered as the vector-valued form of the minmax-theorem.
- Actions of row player: $x \in X$. Actions of nature: $y \in Y$.

Blackwell's approachability theorem

- Blackwell's approachability theorem might be considered as the vector-valued form of the minmax-theorem.
- Actions of row player: $x \in X$. Actions of nature: $y \in Y$.
- Vector-valued payoffs for row player: $v(x, y) \in \mathbb{R}^m$.

Blackwell's approachability theorem

- Blackwell's approachability theorem might be considered as the vector-valued form of the minmax-theorem.
- Actions of row player: $x \in X$. Actions of nature: $y \in Y$.
- Vector-valued payoffs for row player: $v(x, y) \in \mathbb{R}^m$.
- Let V be the vector-valued payoff matrix for the row player.

Blackwell's approachability theorem

- Blackwell's approachability theorem might be considered as the vector-valued form of the minmax-theorem.
- Actions of row player: $x \in X$. Actions of nature: $y \in Y$.
- Vector-valued payoffs for row player: $v(x, y) \in \mathbb{R}^m$.
- Let V be the vector-valued payoff matrix for the row player.
- Consider a set of, for row, desirable payoffs $C \subseteq \mathbb{R}^m$.

Blackwell's approachability theorem

- Blackwell's approachability theorem might be considered as the vector-valued form of the minmax-theorem.
- Actions of row player: $x \in X$. Actions of nature: $y \in Y$.
- Vector-valued payoffs for row player: $v(x, y) \in \mathbb{R}^m$.
- Let V be the vector-valued payoff matrix for the row player.
- Consider a set of, for row, desirable payoffs $C \subseteq \mathbb{R}^m$.
- If the row player has a strategy $\sigma : H \to \Delta(X)$ such that

$$\lim_{t\to\infty} d(\bar{v}^t, C) = 0 \quad \text{a.s.}$$

then C is said to be approachable.

When a half-space C is unapproachable

When a half-space C is unapproachable

- Let $\alpha \in \mathbb{R}^m$, $\alpha \neq 0$ and $C =_{Def} \{ v \in \mathbb{R}^m \mid \alpha \cdot v \leq 0 \}$.

When a half-space C is unapproachable

- Let $\alpha \in \mathbb{R}^m$, $\alpha \neq 0$ and $C =_{Def} \{ v \in \mathbb{R}^m \mid \alpha \cdot v \leq 0 \}$.
- Let

$$\mathbf{W} =_{Def} \begin{pmatrix} \alpha \cdot u(x_1, y_1) & \dots & \alpha \cdot u(x_1, y_l) \\ \vdots & \ddots & \vdots \\ \alpha \cdot u(x_k, y_1) & \dots & \alpha \cdot u(x_k, y_l) \end{pmatrix}.$$

When a half-space C is unapproachable

- Let $\alpha \in \mathbb{R}^m$, $\alpha \neq 0$ and $C =_{Def} \{ v \in \mathbb{R}^m \mid \alpha \cdot v \leq 0 \}$.
- Let
$$\mathbf{W} =_{Def} \begin{pmatrix} \alpha \cdot u(x_1, y_1) & \dots & \alpha \cdot u(x_1, y_l) \\ \vdots & \ddots & \vdots \\ \alpha \cdot u(x_k, y_1) & \dots & \alpha \cdot u(x_k, y_l) \end{pmatrix}.$$
- Suppose for some $q^N \in \Delta(Y)$ and some $\delta > 0$ we have $\mathbf{W}q^N \geq (\delta, \dots, \delta)$.

When a half-space C is unapproachable

- Let $\alpha \in \mathbb{R}^m$, $\alpha \neq 0$ and $C =_{Def} \{ v \in \mathbb{R}^m \mid \alpha \cdot v \leq 0 \}$.
- Let
$$\mathbf{W} =_{Def} \begin{pmatrix} \alpha \cdot u(x_1, y_1) & \dots & \alpha \cdot u(x_1, y_l) \\ \vdots & \ddots & \vdots \\ \alpha \cdot u(x_k, y_1) & \dots & \alpha \cdot u(x_k, y_l) \end{pmatrix}.$$
- Suppose for some $q^N \in \Delta(Y)$ and some $\delta > 0$ we have $\mathbf{W}q^N \geq (\delta, \dots, \delta)$.
- Then immediately (by def. of expectation): $E[v \cdot \alpha] \geq \delta$.

When a half-space C is unapproachable

- Let $\alpha \in \mathbb{R}^m$, $\alpha \neq 0$ and $C =_{Def} \{ v \in \mathbb{R}^m \mid \alpha \cdot v \leq 0 \}$.
- Let

$$\mathbf{W} =_{Def} \begin{pmatrix} \alpha \cdot u(x_1, y_1) & \dots & \alpha \cdot u(x_1, y_l) \\ \vdots & \ddots & \vdots \\ \alpha \cdot u(x_k, y_1) & \dots & \alpha \cdot u(x_k, y_l) \end{pmatrix}.$$

- Suppose for some $q^N \in \Delta(Y)$ and some $\delta > 0$ we have $\mathbf{W} q^N \geq (\delta, \dots, \delta)$.
- Then immediately (by def. of expectation): $E[v \cdot \alpha] \geq \delta$.
- By the strong law of large numbers for dependent r.v.'s

$$\lim_{t \to \infty} \bar{v}^t \cdot \alpha \geq \delta \text{ a.s.}$$

When a half-space C is unapproachable

- Let $\alpha \in \mathbb{R}^m$, $\alpha \neq 0$ and $C =_{Def} \{ v \in \mathbb{R}^m \mid \alpha \cdot v \leq 0 \}$.
- Let

$$\mathbf{W} =_{Def} \begin{pmatrix} \alpha \cdot u(x_1, y_1) & \dots & \alpha \cdot u(x_1, y_l) \\ \vdots & \ddots & \vdots \\ \alpha \cdot u(x_k, y_1) & \dots & \alpha \cdot u(x_k, y_l) \end{pmatrix}.$$

- Suppose for some $q^N \in \Delta(Y)$ and some $\delta > 0$ we have $\mathbf{W} q^N \geq (\delta, \dots, \delta)$.
- Then immediately (by def. of expectation): $E[v \cdot \alpha] \geq \delta$.
- By the strong law of large numbers for dependent r.v.'s

$$\lim_{t \to \infty} \bar{v}^t \cdot \alpha \geq \delta \text{ a.s.}$$

- By definition of C we have $\lim_{t \to \infty} d(\bar{v}^t, C) \geq \delta$ a.s.

The converse also holds

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \ldots, \delta). \tag{1}$$

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \dots, \delta). \tag{1}$$

- Let $A =_{Def}$ the convex hull of $\mathbf{W}$'s columns.

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \dots, \delta). \tag{1}$$

- Let $A =_{Def}$ the convex hull of $\mathbf{W}$'s columns.
  By inequality (1), A does not meet the positive orthant.

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \ldots, \delta). \tag{1}$$

- Let $A =_{Def}$ the convex hull of $\mathbf{W}$'s columns.
  By inequality (1), A does not meet the positive orthant.
- Let $\mathbb{R}^m_{++}$ denote the strictly positive orthant. This set is also convex.

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \dots, \delta). \tag{1}$$

- Let $A =_{Def}$ the convex hull of $\mathbf{W}$'s columns.
  By inequality (1), A does not meet the positive orthant.
- Let $\mathbb{R}^m_{++}$ denote the strictly positive orthant. This set is also convex.
- Also, A and $\mathbb{R}^m_{++}$ are disjoint.

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \dots, \delta). \tag{1}$$

- Let $A =_{Def}$ the convex hull of $\mathbf{W}$'s columns.
  By inequality (1), A does not meet the positive orthant.
- Let $\mathbb{R}^m_{++}$ denote the strictly positive orthant. This set is also convex.
- Also, A and $\mathbb{R}^m_{++}$ are disjoint.
- From the separating hyperplane theorem it follows that there is a vector $q^* \in \mathbb{R}^m$ such that $q^*\mathbf{W} \leq 0$.

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \dots, \delta). \tag{1}$$

- Let $A =_{Def}$ the convex hull of **W**'s columns.
  By inequality (1), A does not meet the positive orthant.
- Let $\mathbb{R}^m_{++}$ denote the strictly positive orthant. This set is also convex.
- Also, A and $\mathbb{R}^m_{++}$ are disjoint.
- From the separating hyperplane theorem it follows that there is a vector $q^* \in \mathbb{R}^m$ such that $q^*\mathbf{W} \leq 0$.
- Without loss of generality, $q^* \in \Delta^m$.

The converse also holds

- Suppose there is no $q^N \in \Delta(Y)$ such that some $\delta > 0$ we have

$$\mathbf{W}q^N \geq (\delta, \dots, \delta). \tag{1}$$

- Let $A =_{Def}$ the convex hull of $\mathbf{W}$'s columns.
  By inequality (1), A does not meet the positive orthant.
- Let $\mathbb{R}^m_{++}$ denote the strictly positive orthant. This set is also convex.
- Also, A and $\mathbb{R}^m_{++}$ are disjoint.
- From the separating hyperplane theorem it follows that there is a vector $q^* \in \mathbb{R}^m$ such that $q^*\mathbf{W} \leq 0$.
- Without loss of generality, $q^* \in \Delta^m$.
- Claim: q^* ensures that row player approaches C.

q^* ensures that row player approaches C

q^* ensures that row player approaches C

- Claim: q^* ensures that row player approaches C.

q^* ensures that row player approaches C

- Claim: q^* ensures that row player approaches C.
- Let $v^1, v^2, \ldots$ be the payoffs that result when q^* is played by the row player.

q^* ensures that row player approaches C

- Claim: q^* ensures that row player approaches C.
- Let $v^1, v^2, \dots$ be the payoffs that result when q^* is played by the row player.
- The sequence of r.v.'s $w^1, w^2, \dots = \alpha \cdot v^1, \alpha \cdot v^2, \dots$ is bounded.

q^* ensures that row player approaches C

- Claim: q^* ensures that row player approaches C.
- Let $v^1, v^2, \dots$ be the payoffs that result when q^* is played by the row player.
- The sequence of r.v.'s $w^1, w^2, \dots = \alpha \cdot v^1, \alpha \cdot v^2, \dots$ is bounded.
- Because $q^* \mathbf{W} \leq 0$,
$$E[\, w^t \mid w^1, \dots, w^{t-1} \,] \leq 0.$$

q^* ensures that row player approaches C

- Claim: q^* ensures that row player approaches C.
- Let $v^1, v^2, \dots$ be the payoffs that result when q^* is played by the row player.
- The sequence of r.v.'s $w^1, w^2, \dots = \alpha \cdot v^1, \alpha \cdot v^2, \dots$ is bounded.
- Because $q^*\mathbf{W} \leq 0$,

$$E[\, w^t \mid w^1, \dots, w^{t-1} \,] \leq 0.$$

- By the strong law of large numbers for dependent r.v.'s

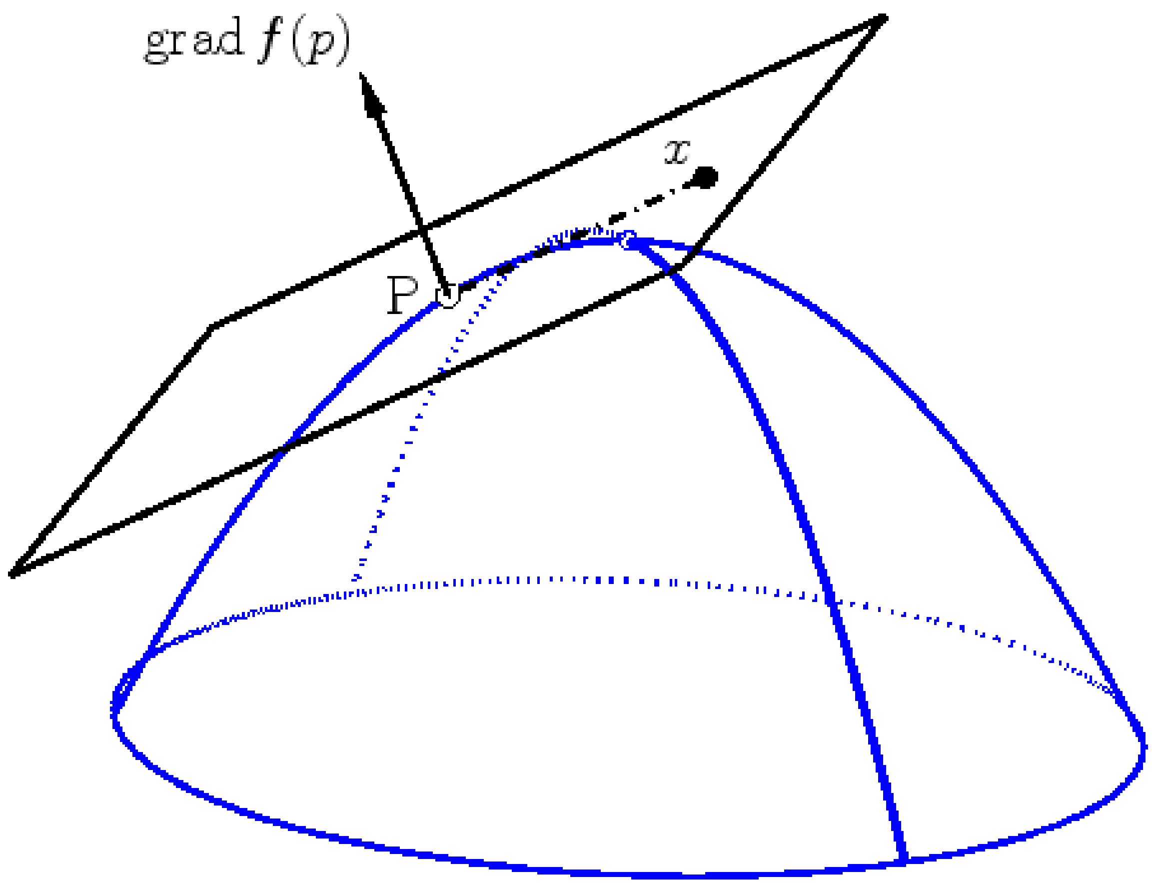
$$\limsup_{t \to \infty} \bar{w}^t \leq 0 \quad \text{a.s.}$$

q^* ensures that row player approaches C

- Claim: q^* ensures that row player approaches C.
- Let $v^1, v^2, \dots$ be the payoffs that result when q^* is played by the row player.
- The sequence of r.v.'s $w^1, w^2, \dots = \alpha \cdot v^1, \alpha \cdot v^2, \dots$ is bounded.
- Because $q^* \mathbf{W} \leq 0$,
$$E[\, w^t \mid w^1, \dots, w^{t-1} \,] \leq 0.$$
- By the strong law of large numbers for dependent r.v.'s
$$\limsup_{t \to \infty} \bar{w}^t \leq 0 \quad \text{a.s.}$$
- Therefore $\lim_{t\to\infty} d(\bar{v}^t, C) = 0$ a.s.

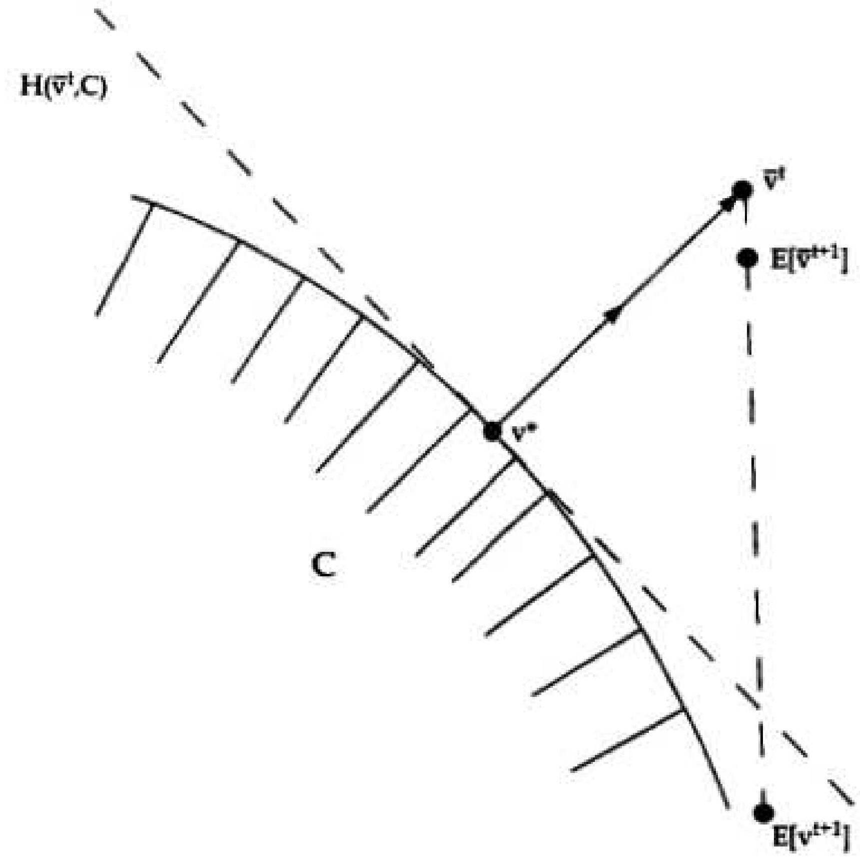
Tangent plane

Blackwell's theorem make use of the concept of tangent plane:



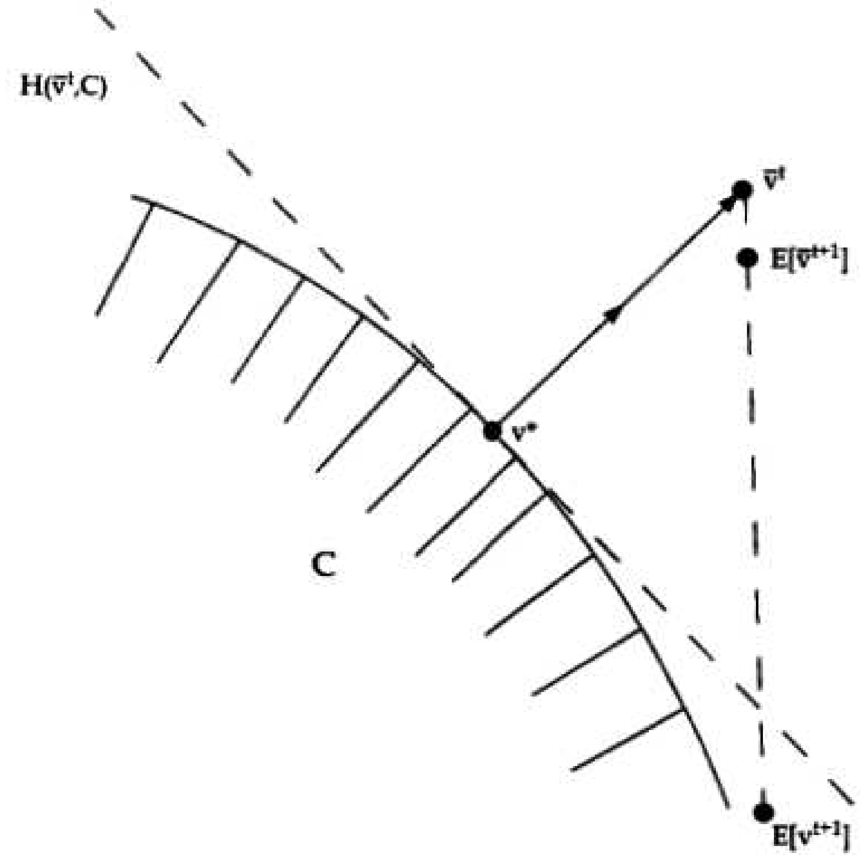
Blackwell's theorem

Theorem (Blackwell, 1956) Let G be a finite two-player game with payoffs in $\mathbb{R}^m$. A closed non-empty convex set C in $\mathbb{R}^m$ is approachable by the row player if and only if every tangent half-space containing C is approachable.



Blackwell's theorem

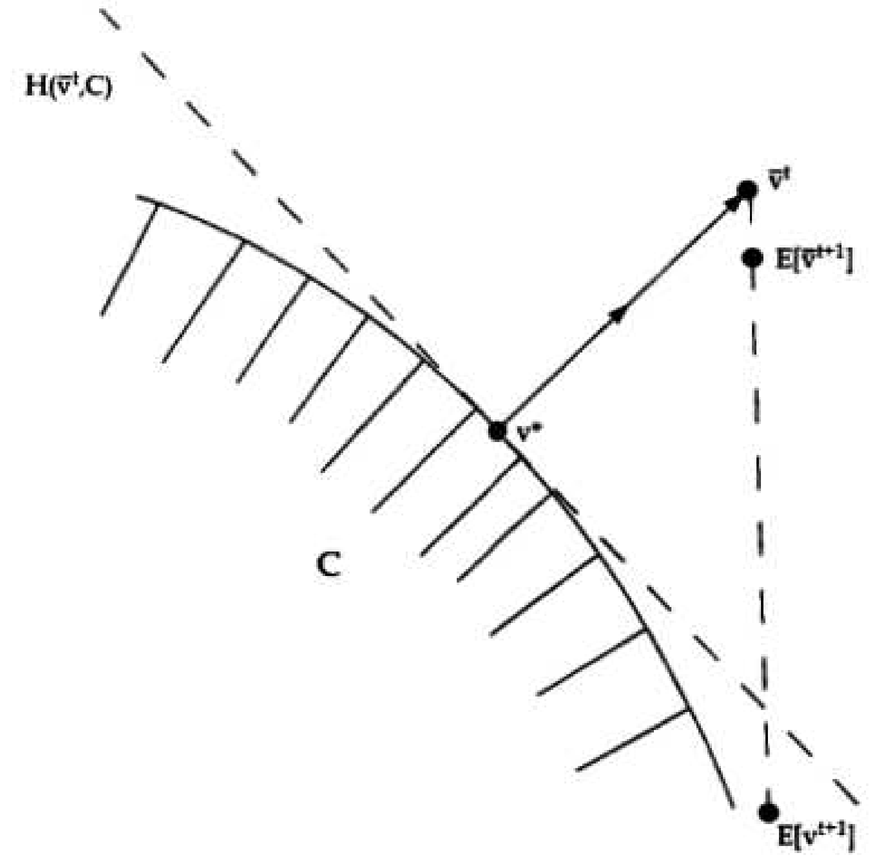
Theorem (Blackwell, 1956) Let G be a finite two-player game with payoffs in $\mathbb{R}^m$. A closed non-empty convex set C in $\mathbb{R}^m$ is approachable by the row player if and only if every tangent half-space containing C is approachable.



- If $\bar{v}^t \in C$ then play anything.

Blackwell's theorem

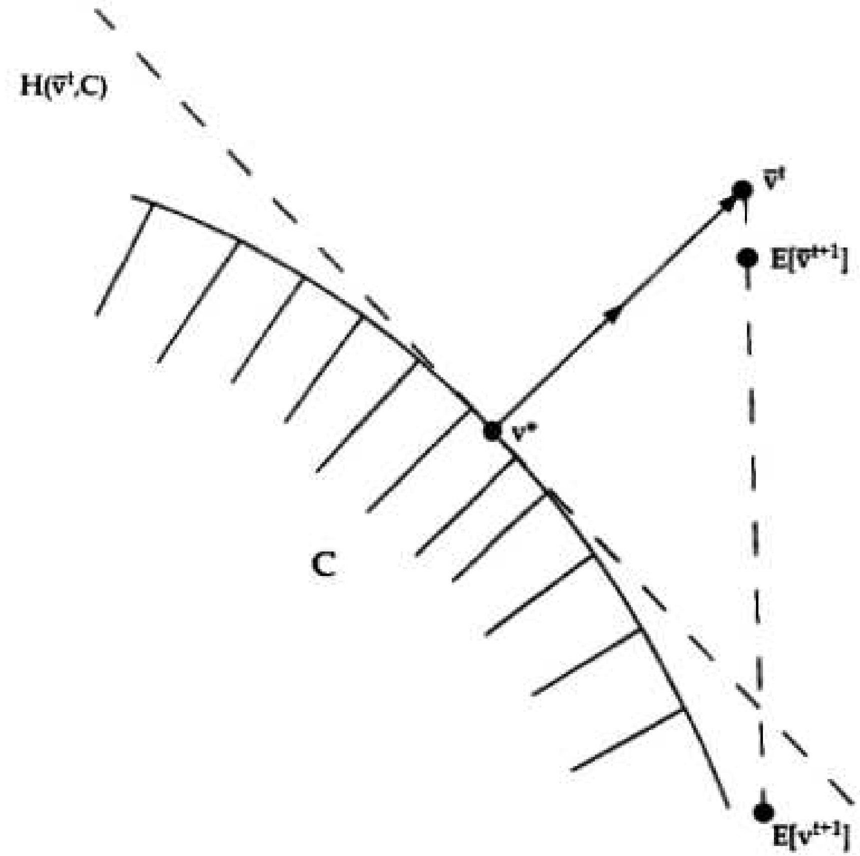
Theorem (Blackwell, 1956) Let G be a finite two-player game with payoffs in $\mathbb{R}^m$. A closed non-empty convex set C in $\mathbb{R}^m$ is approachable by the row player if and only if every tangent half-space containing C is approachable.



- If $\bar{v}^t \in C$ then play anything.
- If $\bar{v}^t \notin C$ then play a q on the rows

Blackwell's theorem

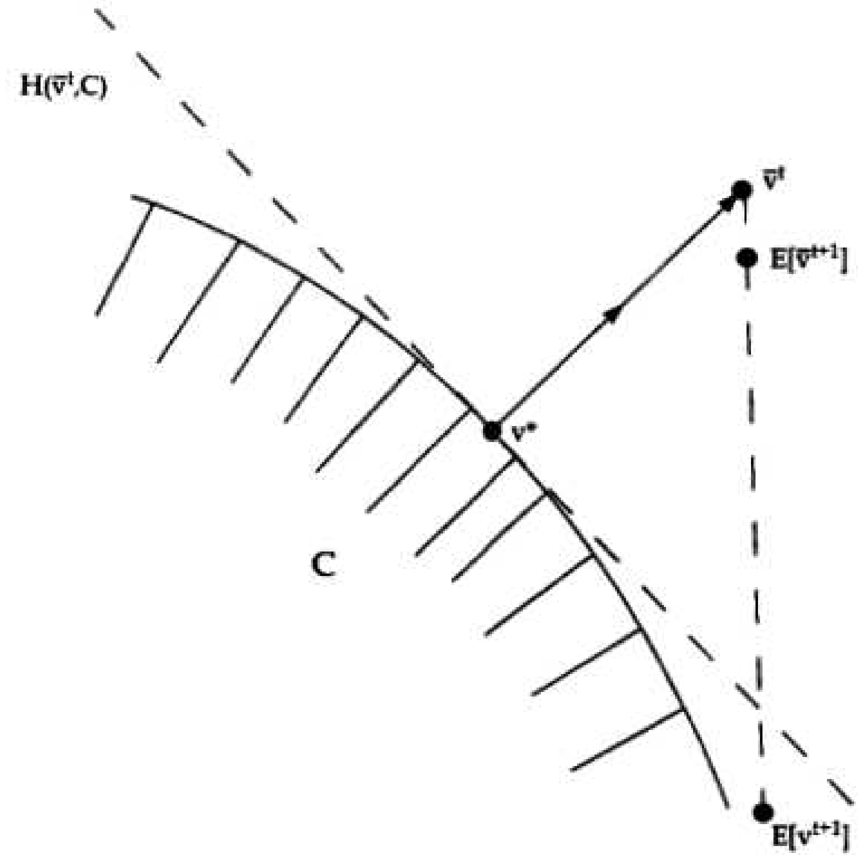
Theorem (Blackwell, 1956) Let G be a finite two-player game with payoffs in $\mathbb{R}^m$. A closed non-empty convex set C in $\mathbb{R}^m$ is approachable by the row player if and only if every tangent half-space containing C is approachable.



- If $\bar{v}^t \in C$ then play anything.
- If $\bar{v}^t \notin C$ then play a q on the rows such that the expected payoff next period lies in the half-space

Blackwell's theorem

Theorem (Blackwell, 1956) Let G be a finite two-player game with payoffs in $\mathbb{R}^m$. A closed non-empty convex set C in $\mathbb{R}^m$ is approachable by the row player if and only if every tangent half-space containing C is approachable.



- If $\bar{v}^t \in C$ then play anything.
- If $\bar{v}^t \notin C$ then play a q on the rows such that the expected payoff next period lies in the half-space orthogonal to $\bar{v}^t - v^*$ containing C.

Eliminating regret

(Result left hanging in Ch. 2)

Eliminating regret

Eliminating regret

- Let $X =_{Def} \{1, 2, \ldots, k\}$.

Eliminating regret

- Let $X =_{Def} \{1, 2, \ldots, k\}$.
- Let
$$r(x, y) =_{Def} [\, u(1, y) - u(x, y)\,, \ldots,\, u(k, y) - u(x, y)\,].$$

We now have a vector-valued game.

Eliminating regret

- Let $X =_{Def} \{1, 2, \ldots, k\}$.
- Let
$$r(x, y) =_{Def} [\ u(1, y) - u(x, y)\ , \ldots ,\ u(k, y) - u(x, y)\ ].$$
We now have a vector-valued game.
- Let $r^t =_{Def} r(x_t, y_t)$.

Eliminating regret

- Let $X =_{Def} \{1, 2, \ldots, k\}$.
- Let
$$r(x, y) =_{Def} [\ u(1, y) - u(x, y)\ , \ldots,\ u(k, y) - u(x, y)\ ].$$
We now have a vector-valued game.
- Let $r^t =_{Def} r(x_t, y_t)$.
- Let $\bar{r}^t$ be the average of the vector-valued payoffs $r^1, \ldots, r^t$.

Eliminating regret

- Let $X =_{Def} \{1, 2, \dots, k\}$.
- Let
$$r(x, y) =_{Def} [\ u(1, y) - u(x, y)\ , \dots ,\ u(k, y) - u(x, y)\ ].$$
  We now have a vector-valued game.
- Let $r^t =_{Def} r(x_t, y_t)$.
- Let $\bar{r}^t$ be the average of the vector-valued payoffs $r^1, \dots, r^t$.
- The goal is to let $\bar{r}^t \leq 0$ as t becomes large.

Eliminating regret

- Let $X =_{Def} \{1, 2, \dots, k\}$.
- Let
$$r(x, y) =_{Def} [\ u(1, y) - u(x, y)\ , \dots ,\ u(k, y) - u(x, y)\ ].$$
We now have a vector-valued game.
- Let $r^t =_{Def} r(x_t, y_t)$.
- Let $\bar{r}^t$ be the average of the vector-valued payoffs $r^1, \dots, r^t$.
- The goal is to let $\bar{r}^t \leq 0$ as t becomes large.
- Hence, the goal is to approach the convex set $C = \mathbb{R}^k_-$.

Eliminating regret

- Let $X =_{Def} \{1, 2, \ldots, k\}$.
- Let
$$r(x, y) =_{Def} [\, u(1, y) - u(x, y) \,, \ldots, \, u(k, y) - u(x, y) \,].$$
We now have a vector-valued game.
- Let $r^t =_{Def} r(x_t, y_t)$.
- Let $\bar{r}^t$ be the average of the vector-valued payoffs $r^1, \ldots, r^t$.
- The goal is to let $\bar{r}^t \leq 0$ as t becomes large.
- Hence, the goal is to approach the convex set $C = \mathbb{R}^k_-$.
- Suppose $\bar{r}^t \notin C$. Then the closest point in C is $\bar{r}^t_-$.

Eliminating regret

- Let $X =_{Def} \{1, 2, \dots, k\}$.
- Let
$$r(x, y) =_{Def} [\ u(1, y) - u(x, y)\ , \dots ,\ u(k, y) - u(x, y)\ ].$$
We now have a vector-valued game.
- Let $r^t =_{Def} r(x_t, y_t)$.
- Let $\bar{r}^t$ be the average of the vector-valued payoffs $r^1, \dots, r^t$.
- The goal is to let $\bar{r}^t \leq 0$ as t becomes large.
- Hence, the goal is to approach the convex set $C = \mathbb{R}^k_-$.
- Suppose $\bar{r}^t \notin C$. Then the closest point in C is $\bar{r}^t_-$.
- Blackwell: randomise play such that $E[r^{t+1}] \perp (\bar{r}^t - \bar{r}^t_-)$, so $E[r^{t+1}] \perp \bar{r}^t_+$.

Eliminating regret

Let $\mathbf{A}$ be

$$\begin{pmatrix} u(1, y^{t+1}) - u(1, y^{t+1}) & u(2, y^{t+1}) - u(1, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(1, y^{t+1}) \\ u(1, y^{t+1}) - u(2, y^{t+1}) & u(2, y^{t+1}) - u(2, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(2, y^{t+1}) \\ \vdots & \vdots & \ddots & \vdots \\ u(1, y^{t+1}) - u(k, y^{t+1}) & u(2, y^{t+1}) - u(k, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(k, y^{t+1}) \end{pmatrix}$$

Eliminating regret

Let $\mathbf{A}$ be

$$\begin{pmatrix} u(1, y^{t+1}) - u(1, y^{t+1}) & u(2, y^{t+1}) - u(1, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(1, y^{t+1}) \\ u(1, y^{t+1}) - u(2, y^{t+1}) & u(2, y^{t+1}) - u(2, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(2, y^{t+1}) \\ \vdots & \vdots & \ddots & \vdots \\ u(1, y^{t+1}) - u(k, y^{t+1}) & u(2, y^{t+1}) - u(k, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(k, y^{t+1}) \end{pmatrix}$$

- If q is played, $E[r^{t+1}]$ can be written as $q\mathbf{A}$.

Eliminating regret

Let $\mathbf{A}$ be

$$\begin{pmatrix} u(1,y^{t+1}) - u(1,y^{t+1}) & u(2,y^{t+1}) - u(1,y^{t+1}) & \dots & u(\mathsf{k},y^{t+1}) - u(1,y^{t+1}) \\ u(1,y^{t+1}) - u(2,y^{t+1}) & u(2,y^{t+1}) - u(2,y^{t+1}) & \dots & u(\mathsf{k},y^{t+1}) - u(2,y^{t+1}) \\ \vdots & \vdots & \ddots & \vdots \\ u(1,y^{t+1}) - u(k,y^{t+1}) & u(2,y^{t+1}) - u(k,y^{t+1}) & \dots & u(\mathsf{k},y^{t+1}) - u(k,y^{t+1}) \end{pmatrix}$$

- If q is played, $E[r^{t+1}]$ can be written as $q\mathbf{A}$.
- To apply Blackwell's theorem, we must ensure $q\mathbf{A} \perp \bar{r}^t_+$, i.e., $q\mathbf{A}\bar{r}^t_+ = 0$.

Eliminating regret

Let $\mathbf{A}$ be

$$\begin{pmatrix} u(1,y^{t+1}) - u(1,y^{t+1}) & u(2,y^{t+1}) - u(1,y^{t+1}) & \dots & u(\mathsf{k},y^{t+1}) - u(1,y^{t+1}) \\ u(1,y^{t+1}) - u(2,y^{t+1}) & u(2,y^{t+1}) - u(2,y^{t+1}) & \dots & u(\mathsf{k},y^{t+1}) - u(2,y^{t+1}) \\ \vdots & \vdots & \ddots & \vdots \\ u(1,y^{t+1}) - u(k,y^{t+1}) & u(2,y^{t+1}) - u(k,y^{t+1}) & \dots & u(\mathsf{k},y^{t+1}) - u(k,y^{t+1}) \end{pmatrix}$$

- If q is played, $E[r^{t+1}]$ can be written as $q\mathbf{A}$.
- To apply Blackwell's theorem, we must ensure $q\mathbf{A} \perp \bar{r}^t_+$, i.e., $q\mathbf{A}\bar{r}^t_+ = 0$.
- Since A is skew-symmetric ($\mathbf{A} = -\mathbf{A}^T$), we have

$$q\mathbf{A}q^T = 0.$$

Eliminating regret

Let $\mathbf{A}$ be

$$\begin{pmatrix} u(1, y^{t+1}) - u(1, y^{t+1}) & u(2, y^{t+1}) - u(1, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(1, y^{t+1}) \\ u(1, y^{t+1}) - u(2, y^{t+1}) & u(2, y^{t+1}) - u(2, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(2, y^{t+1}) \\ \vdots & \vdots & \ddots & \vdots \\ u(1, y^{t+1}) - u(k, y^{t+1}) & u(2, y^{t+1}) - u(k, y^{t+1}) & \dots & u(\mathsf{k}, y^{t+1}) - u(k, y^{t+1}) \end{pmatrix}$$

- If q is played, $E[r^{t+1}]$ can be written as $q\mathbf{A}$.
- To apply Blackwell's theorem, we must ensure $q\mathbf{A} \perp \bar{r}^t_+$, i.e., $q\mathbf{A}\bar{r}^t_+ = 0$.
- Since A is skew-symmetric ($\mathbf{A} = -\mathbf{A}^T$), we have

$$q\mathbf{A}q^T = 0.$$

So take $q^T \sim \bar{r}^t_+$, i.e., take q^T proportional to $\bar{r}^t_+$.

Eliminating conditional regret

Eliminating conditional regret

Eliminating conditional regret

- The goal is to suppress $\overline{\mathbf{R}}^t$ as t becomes large: $\limsup_{t\to\infty} \overline{\mathbf{R}}^t \leq 0$.

Eliminating conditional regret

- The goal is to suppress $\overline{\mathbf{R}}^t$ as t becomes large: $\limsup_{t\to\infty} \overline{\mathbf{R}}^t \leq 0$.
- So, the goal is to approach
  $C =_{Def} \{\text{all } k \times k\text{-matrices with non-positive entries}\}$.

Eliminating conditional regret

- The goal is to suppress $\overline{\mathbf{R}}^t$ as t becomes large: $\limsup_{t\to\infty} \overline{\mathbf{R}}^t \leq 0$.
- So, the goal is to approach
  $C =_{Def} \{\text{all } k \times k\text{-matrices with non-positive entries}\}$.
- The closest point in C to $\overline{\mathbf{R}}^t$ is the matrix $\overline{\mathbf{R}}^t_-$.

Eliminating conditional regret

- The goal is to suppress $\overline{\mathbf{R}}^t$ as t becomes large: $\limsup_{t\to\infty} \overline{\mathbf{R}}^t \leq 0$.
- So, the goal is to approach
  $C =_{Def}$ {all $k \times k$-matrices with non-positive entries}.
- The closest point in C to $\overline{\mathbf{R}}^t$ is the matrix $\overline{\mathbf{R}}^t_-$.
- Blackwell $\rightarrow$ find a randomised play $q \in \Delta^k$ that makes the expected incremental conditional regret matrix orthogonal to $\overline{\mathbf{R}}^t - \overline{\mathbf{R}}^t_- = \overline{\mathbf{R}}^t_+$.

Eliminating conditional regret

- The goal is to suppress $\overline{\mathbf{R}}^t$ as t becomes large: $\limsup_{t\to\infty} \overline{\mathbf{R}}^t \leq 0$.
- So, the goal is to approach
  $C =_{Def} \{\text{all } k \times k\text{-matrices with non-positive entries}\}$.
- The closest point in C to $\overline{\mathbf{R}}^t$ is the matrix $\overline{\mathbf{R}}^t_-$.
- Blackwell $\to$ find a randomised play $q \in \Delta^k$ that makes the expected incremental conditional regret matrix orthogonal to $\overline{\mathbf{R}}^t - \overline{\mathbf{R}}^t_- = \overline{\mathbf{R}}^t_+$.
- Write

$$\mathbf{P} =_{Def} \begin{pmatrix} u(1, y^{t+1}) & \dots & u(1, y^{t+1}) \\ \vdots & \ddots & \vdots \\ u(k, y^{t+1}) & \dots & u(k, y^{t+1}) \end{pmatrix}, \mathbf{A} =_{Def} \mathbf{P} - \mathbf{P}^T, \mathbf{Q} =_{Def} \begin{pmatrix} q_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & q_k \end{pmatrix}.$$

Eliminating conditional regret

- The goal is to suppress $\overline{\mathbf{R}}^t$ as t becomes large: $\limsup_{t\to\infty} \overline{\mathbf{R}}^t \leq 0$.
- So, the goal is to approach
  $C =_{Def} \{\text{all } k \times k\text{-matrices with non-positive entries}\}$.
- The closest point in C to $\overline{\mathbf{R}}^t$ is the matrix $\overline{\mathbf{R}}^t_-$.
- Blackwell $\rightarrow$ find a randomised play $q \in \Delta^k$ that makes the expected incremental conditional regret matrix orthogonal to $\overline{\mathbf{R}}^t - \overline{\mathbf{R}}^t_- = \overline{\mathbf{R}}^t_+$.
- Write

$$\mathbf{P} =_{Def} \begin{pmatrix} u(1, y^{t+1}) & \dots & u(1, y^{t+1}) \\ \vdots & \ddots & \vdots \\ u(k, y^{t+1}) & \dots & u(k, y^{t+1}) \end{pmatrix}, \mathbf{A} =_{Def} \mathbf{P} - \mathbf{P}^T, \mathbf{Q} =_{Def} \begin{pmatrix} q_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & q_k \end{pmatrix}.$$

- Then the expected incremental conditional regret matrix is $\mathbf{QA} = \mathbf{Q}(\mathbf{P} - \mathbf{P}^T)$.

Eliminating conditional regret

Eliminating conditional regret

- The expected incremental conditional regret matrix is $\mathbf{QA} = \mathbf{Q}(\mathbf{P} - \mathbf{P}^T)$.

Eliminating conditional regret

- The expected incremental conditional regret matrix is $\mathbf{QA} = \mathbf{Q}(\mathbf{P} - \mathbf{P}^T)$.
- Blackwell's theorem tells us to choose $\mathbf{Q}$ such that

$$(\mathbf{QA}) \cdot \overline{\mathbf{R}}_+ = 0.$$

Here, "$\cdot$" is the dot-product of two matrices in $\mathbb{R}^k \times \mathbb{R}^k$, not their matrix product.

Eliminating conditional regret

- The expected incremental conditional regret matrix is $\mathbf{QA} = \mathbf{Q}(\mathbf{P} - \mathbf{P}^T)$.
- Blackwell's theorem tells us to choose $\mathbf{Q}$ such that

$$(\mathbf{QA}) \cdot \overline{\mathbf{R}}_+ = 0.$$

  Here, "$\cdot$" is the dot-product of two matrices in $\mathbb{R}^k \times \mathbb{R}^k$, not their matrix product.
- Since $\mathbf{Q}$ and $\mathbf{A}$ commute (check), we have

$$(\mathbf{QA}) \cdot \overline{\mathbf{R}}_+ = \mathbf{A} \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) = (\mathbf{P} - \mathbf{P}^T) \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) = \mathbf{P} \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) - \mathbf{P}^T \cdot (\mathbf{Q}\overline{\mathbf{R}}_+).$$

Eliminating conditional regret

- The expected incremental conditional regret matrix is $\mathbf{QA} = \mathbf{Q}(\mathbf{P} - \mathbf{P}^T)$.
- Blackwell's theorem tells us to choose $\mathbf{Q}$ such that

$$(\mathbf{QA}) \cdot \overline{\mathbf{R}}_+ = 0.$$

  Here, "$\cdot$" is the dot-product of two matrices in $\mathbb{R}^k \times \mathbb{R}^k$, not their matrix product.

- Since $\mathbf{Q}$ and $\mathbf{A}$ commute (check), we have

$$(\mathbf{QA}) \cdot \overline{\mathbf{R}}_+ = \mathbf{A} \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) = (\mathbf{P} - \mathbf{P}^T) \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) = \mathbf{P} \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) - \mathbf{P}^T \cdot (\mathbf{Q}\overline{\mathbf{R}}_+).$$

- The problem boils down to ensure that, for every $1 \leq j \leq k$:

$$\sum_h q_h (\bar{r}_{hj})_+ \; - \; q_j \sum_h (\bar{r}_{jh})_+ \; = \; 0. \tag{2}$$

Eliminating conditional regret

- The expected incremental conditional regret matrix is $\mathbf{QA} = \mathbf{Q}(\mathbf{P} - \mathbf{P}^T)$.
- Blackwell's theorem tells us to choose $\mathbf{Q}$ such that

$$(\mathbf{QA}) \cdot \overline{\mathbf{R}}_+ = 0.$$

  Here, "$\cdot$" is the dot-product of two matrices in $\mathbb{R}^k \times \mathbb{R}^k$, not their matrix product.
- Since $\mathbf{Q}$ and $\mathbf{A}$ commute (check), we have

$$(\mathbf{QA}) \cdot \overline{\mathbf{R}}_+ = \mathbf{A} \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) = (\mathbf{P} - \mathbf{P}^T) \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) = \mathbf{P} \cdot (\mathbf{Q}\overline{\mathbf{R}}_+) - \mathbf{P}^T \cdot (\mathbf{Q}\overline{\mathbf{R}}_+).$$

- The problem boils down to ensure that, for every $1 \leq j \leq k$:

$$\sum_h q_h (\bar{r}_{hj})_+ \; - \; q_j \sum_h (\bar{r}_{jh})_+ \; = \; 0. \tag{2}$$

- Every q satisfying this equation is called a left-invariant for $\overline{\mathbf{R}}_+$.

Claim: every zero-diagonal matrix possesses a left-invaria

Claim: every zero-diagonal matrix possesses a left-invaria

- Every $k \times k$ non-negative matrix $\mathbf{M}$ with a zero diagonal possesses a left-invariant vector.

Claim: every zero-diagonal matrix possesses a left-invaria

- Every $k \times k$ non-negative matrix $\mathbf{M}$ with a zero diagonal possesses a left-invariant vector.

- Define $\mathbf{N} =_{Def}$

$$\begin{pmatrix} -\sum_j m_{1j} & m_{12} & \dots & m_{1k} \\ m_{21} & -\sum_j m_{2j} & \dots & m_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & \dots & m_{2k} & -\sum_j m_{kj} \end{pmatrix}$$

Claim: every zero-diagonal matrix possesses a left-invari

- Every $k \times k$ non-negative matrix $\mathbf{M}$ with a zero diagonal possesses a left-invariant vector.

- Define $\mathbf{N} =_{Def}$

$$\begin{pmatrix} -\sum_j m_{1j} & m_{12} & \dots & m_{1k} \\ m_{21} & -\sum_j m_{2j} & \dots & m_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & \dots & m_{2k} & -\sum_j m_{kj} \end{pmatrix}$$

- To solve (2) we need to find a q such that $q\mathbf{N} = 0$, $q \geq 0$, $|q| = 1$.

Claim: every zero-diagonal matrix possesses a left-invaria

- Every $k \times k$ non-negative matrix $\mathbf{M}$ with a zero diagonal possesses a left-invariant vector.
- Define $\mathbf{N} =_{Def}$

$$\begin{pmatrix} -\sum_j m_{1j} & m_{12} & \dots & m_{1k} \\ m_{21} & -\sum_j m_{2j} & \dots & m_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & \dots & m_{2k} & -\sum_j m_{kj} \end{pmatrix}$$

- To solve (2) we need to find a q such that $q\mathbf{N} = 0$, $q \geq 0$, $|q| = 1$.
- Let β be as least as large as the absolute value of the most negative of the diagonal elements in $\mathbf{M}$. Then

$$\mathbf{I} + \frac{1}{\beta}\mathbf{N}$$

is non-negative and row stochastic.

Hence, it has at least one fixed point q^*.

Claim: every zero-diagonal matrix possesses a left-invaria

- Every $k \times k$ non-negative matrix $\mathbf{M}$ with a zero diagonal possesses a left-invariant vector.
- Define $\mathbf{N} =_{Def}$

$$\begin{pmatrix} -\sum_j m_{1j} & m_{12} & \dots & m_{1k} \\ m_{21} & -\sum_j m_{2j} & \dots & m_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & \dots & m_{2k} & -\sum_j m_{kj} \end{pmatrix}$$

- To solve (2) we need to find a q such that $q\mathbf{N} = 0$, $q \geq 0$, $|q| = 1$.
- Let β be as least as large as the absolute value of the most negative of the diagonal elements in $\mathbf{M}$. Then

$$\mathbf{I} + \frac{1}{\beta}\mathbf{N}$$

is non-negative and row stochastic.

Hence, it has at least one fixed point q^*.

- Any such q^* has the property that $q\mathbf{N} = 0$, so q^* is the desired left-invariant for $\mathbf{M}$.

Eliminating conditional regret

Theorem (Foster and Vohra, 1999). *Given a finite game* G*, if in each period a player plays a distribution* q *that satisfies*

$$\sum_h q_h(\bar{r}_{hj})_+ \; - \; q_j \sum_h (\bar{r}_{jh})_+ \; = \; 0,$$

and $\overline{\boldsymbol{R}}^t$ *is his conditional regret matrix through* t *then* $\overline{\boldsymbol{R}}^t \rightsquigarrow 0$ *almost surely, independently of the behaviour of the opponents.*

Eliminating conditional regret

Theorem (Foster and Vohra, 1999). *Given a finite game G, if in each period a player plays a distribution q that satisfies*

$$\sum_h q_h (\bar{r}_{hj})_+ \; - \; q_j \sum_h (\bar{r}_{jh})_+ \; = \; 0,$$

and $\overline{\boldsymbol{R}}^t$ is his conditional regret matrix through t then $\overline{\boldsymbol{R}}^t \rightsquigarrow 0$ almost surely, independently of the behaviour of the opponents.

Corollary. When every player uses the Foster-Vohra algorithm to suppress conditional regret, the empirical joint frequency of play converges almost surely to the set of correlated equilibria.

Example

In an earlier example, the conditional regret matrix after 10 rounds was:[1]

$$\mathbf{R} = \begin{array}{c c} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 0.0 & 0.1 & 0.0 \\ 0.3 & 0.0 & 0.0 \\ -0.1 & 0.1 & 0.0 \end{pmatrix} \end{array}$$

So,

$$\mathbf{N} = \begin{array}{c c} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} -0.1 & 0.1 & 0.0 \\ 0.3 & -0.3 & 0.0 \\ -0.1 & 0.1 & 0.0 \end{pmatrix} \end{array}$$

[1]There is a typo in SLaiL, third row.

We need to find a $q \geq 0$ such that $q\mathbf{N} = 0$, $|q| = 1$. If we set $\beta = 0.3$, then q is the fixed-point of

$$\mathbf{I} + \frac{1}{\beta}\mathbf{N} = \begin{array}{c c} & \begin{array}{ccc} R & Y & B \end{array} \\ \begin{array}{c} R \\ Y \\ B \end{array} & \begin{pmatrix} 2/3 & 1/3 & 0 \\ 1 & 0 & 0 \\ 0 & 1/3 & 2/3 \end{pmatrix} \end{array}$$

with $|q| = 1$.
By the theory of Markov chains such a fixed point exists. In this case, it is

$$q^* = (1/3, 1/3, 1/3).$$

Incremental conditional regret matching

Incremental conditional regret matching

- A disadvantage of the Foster-Vohra algorithm is that every period a new system of linear equations must be solved.

Incremental conditional regret matching

- A disadvantage of the Foster-Vohra algorithm is that every period a new system of linear equations must be solved.
- What we would like is a learning rule such as

$$q^{t+1} = q^t + \lambda\delta^t$$

Incremental conditional regret matching

- A disadvantage of the Foster-Vohra algorithm is that every period a new system of linear equations must be solved.
- What we would like is a learning rule such as

$$q^{t+1} = q^t + \lambda \delta^t$$

where λ is a learning parameter

Incremental conditional regret matching

- A disadvantage of the Foster-Vohra algorithm is that every period a new system of linear equations must be solved.
- What we would like is a learning rule such as

$$q^{t+1} = q^t + \lambda\delta^t$$

where λ is a learning parameter, and δ^t is an adaptation in the right direction.

Incremental conditional regret matching

- A disadvantage of the Foster-Vohra algorithm is that every period a new system of linear equations must be solved.
- What we would like is a learning rule such as

$$q^{t+1} = q^t + \lambda \delta^t$$

where λ is a learning parameter, and δ^t is an adaptation in the right direction.

- It turns out that the following works:

$$\delta^t_j =_{Def}$$

Incremental conditional regret matching

- A disadvantage of the Foster-Vohra algorithm is that every period a new system of linear equations must be solved.
- What we would like is a learning rule such as

$$q^{t+1} = q^t + \lambda \delta^t$$

where λ is a learning parameter, and δ^t is an adaptation in the right direction.

- It turns out that the following works:

$$\delta_j^t =_{Def} \underbrace{\sum_h q_h^t (\bar{r}_{hj}^t)_+}_{\text{The expected positive regrets from playing } q^t \text{ instead of } j.} -$$

Incremental conditional regret matching

- A disadvantage of the Foster-Vohra algorithm is that every period a new system of linear equations must be solved.
- What we would like is a learning rule such as

$$q^{t+1} = q^t + \lambda \delta^t$$

where λ is a learning parameter, and δ^t is an adaptation in the right direction.

- It turns out that the following works:

$$\delta_j^t =_{Def} \underbrace{\sum_h q_h^t (\bar{r}_{hj}^t)_+}_{\text{The expected positive regrets from playing } q^t \text{ instead of } j.} - \underbrace{q_j^t \sum_h (\bar{r}_{jh}^t)_+}_{\text{A normalisation that makes the } \delta_i \text{ sum to zero.}}$$

Incremental conditional regret matching

Theorem. If, in a finite game G, a player uses incremental conditional regret matching with a sufficiently small learning parameter, his conditional regrets become non-positive almost surely, independently of the behaviour of the participants.

Incremental conditional regret matching

Theorem. If, in a finite game G, a player uses incremental conditional regret matching with a sufficiently small learning parameter, his conditional regrets become non-positive almost surely, independently of the behaviour of the participants.

Proof. Based on the standard iterative procedure for find an invariant distribution of a finite Markov chain.

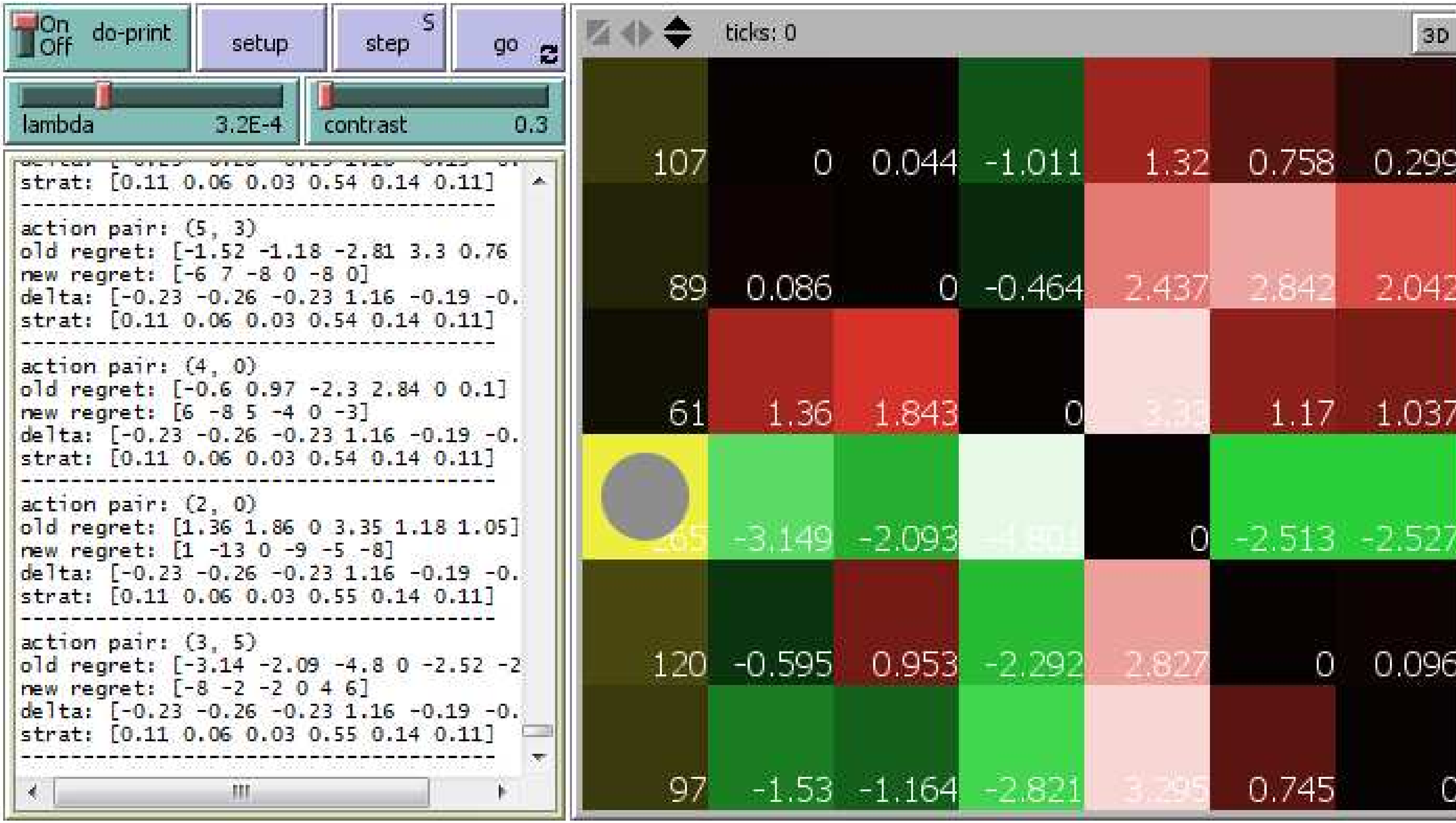
Incremental conditional regret matching

Theorem. If, in a finite game G, a player uses incremental conditional regret matching with a sufficiently small learning parameter, his conditional regrets become non-positive almost surely, independently of the behaviour of the participants.

Proof. Based on the standard iterative procedure for find an invariant distribution of a finite Markov chain.

Corollary. If, in a finite game G, all players use incremental conditional regret matching with a sufficiently small learning parameter, the empirical joint frequency of play converges almost surely to the set of correlated equilibria.

Demo



On
Off
do-print
setup
step
S
go
lambda 3.2E-4
contrast 0.3
strat: [0.11 0.06 0.03 0.54 0.14 0.11]
action pair: (5, 3)
old regret: [-1.52 -1.18 -2.81 3.3 0.76
new regret: [-6 7 -8 0 -8 0]
delta: [-0.23 -0.26 -0.23 1.16 -0.19 -0.
strat: [0.11 0.06 0.03 0.54 0.14 0.11]
action pair: (4, 0)
old regret: [-0.6 0.97 -2.3 2.84 0 0.1]
new regret: [6 -8 5 -4 0 -3]
delta: [-0.23 -0.26 -0.23 1.16 -0.19 -0.
strat: [0.11 0.06 0.03 0.54 0.14 0.11]
action pair: (2, 0)
old regret: [1.36 1.86 0 3.35 1.18 1.05]
new regret: [1 -13 0 -9 -5 -8]
delta: [-0.23 -0.26 -0.23 1.16 -0.19 -0.
strat: [0.11 0.06 0.03 0.55 0.14 0.11]
action pair: (3, 5)
old regret: [-3.14 -2.09 -4.8 0 -2.52 -2
new regret: [-8 -2 -2 0 4 6]
delta: [-0.23 -0.26 -0.23 1.16 -0.19 -0.
strat: [0.11 0.06 0.03 0.55 0.14 0.11]
ticks: 0
3D
107 0 0.044 -1.011 1.32 0.758 0.299
89 0.086 0 -0.464 2.437 2.842 2.042
61 1.36 1.843 0 1.17 1.037
-3.149 -2.093 0 -2.513 -2.527
120 -0.595 0.953 -2.292 2.827 0 0.096
97 -1.53 -1.164 -2.821 0.745 0

Refinements of learning rules

Conditional regret matching with ϵ-experimentation

Refinements of learning rules

Conditional regret matching with ϵ-experimentation

- Allow players to experiment ϵ% of the time.

Refinements of learning rules

Conditional regret matching with ϵ-experimentation

- Allow players to experiment ϵ% of the time. Then the conditional regrets in experimenting can be used as a proxy for the conditional regrets over all periods.

Refinements of learning rules

Conditional regret matching with ϵ-experimentation

- Allow players to experiment ϵ% of the time. Then the conditional regrets in experimenting can be used as a proxy for the conditional regrets over all periods.

- In this way, δ^t can be made to depend solely on realised past payoffs.

Conditional regret matching with inertia

Refinements of learning rules

Conditional regret matching with ϵ-experimentation

- Allow players to experiment ϵ% of the time. Then the conditional regrets in experimenting can be used as a proxy for the conditional regrets over all periods.

- In this way, δ^t can be made to depend solely on realised past payoffs.

Conditional regret matching with inertia

- Play the same action as in the previous period with a high probability. Otherwise, randomise play by q, where

$$q \sim \text{the non-negative regrets of the last action.}$$

Refinements of learning rules

Conditional regret matching with ϵ-experimentation

- Allow players to experiment ϵ% of the time. Then the conditional regrets in experimenting can be used as a proxy for the conditional regrets over all periods.

- In this way, δ^t can be made to depend solely on realised past payoffs.

Conditional regret matching with inertia

- Play the same action as in the previous period with a high probability. Otherwise, randomise play by q, where

$$q \sim \text{the non-negative regrets of the last action.}$$

- The effect is that change in behaviour over periods is discontinuous.

Refinements of learning rules

Conditional regret matching with ϵ-experimentation

- Allow players to experiment $\epsilon\%$ of the time. Then the conditional regrets in experimenting can be used as a proxy for the conditional regrets over all periods.
- In this way, δ^t can be made to depend solely on realised past payoffs.

Conditional regret matching with inertia

- Play the same action as in the previous period with a high probability. Otherwise, randomise play by q, where

$$q \sim \text{the non-negative regrets of the last action.}$$

- The effect is that change in behaviour over periods is discontinuous.
- All players need to use this method to eliminate conditional regret.