Mini-Course: "Games, Dynamics and Learning"

A. Giannou T. Lianeas P. Mertikopoulos E. Vlatakis

School of Electrical and Computer Engineering,
National Technical University of Athens

04 June 2021

Follow the Regularized Leader

The algorithm of Follow the Regularized Leader is defined by the round-by-round recursive rule

$$
\begin{aligned}
X_{i,n} &= Q_i(Y_{i,n}) \\
Y_{i,n+1} &= Y_{i,n} + \gamma_n \hat{v}_{i,n}
\end{aligned}
\tag{FTRL}
$$

- $Q_i \colon \mathcal{Y}_i \to \mathcal{X}_i$ denotes the "choice map" of player $i \in \mathcal{N}$.
- $\gamma_n > 0$ is a "learning rate" parameter such that $\sum_n \gamma_n = \infty$.
- $\hat{v}_{i,n}$ is a "payoff signal" that provides an estimate for the mixed payoffs of player i at stage n.

Regularization

The second component of FTRL is the choice map

$$Q_i(y_i) = \arg\max_{x_i \in \mathcal{X}_i} \{\langle y_i, x_i \rangle - h_i(x_i)\}.$$

In the above, each player's *regularizer* $h_i \colon \mathcal{X}_i \to \mathbb{R}$ is defined as $h_i(x_i) = \sum_{\alpha_i \in \mathcal{A}_i} \theta_i(x_i)$ for some "kernel function" $\theta_i \colon [0,1] \to \mathbb{R}$ with the following properties:

(*i*) θ_i is *continuous* on $[0,1]$;

(*ii*) C^2-smooth on $(0,1]$; and

(*iii*) $\inf_{[0,1]} \theta_i'' > 0$.

Examples

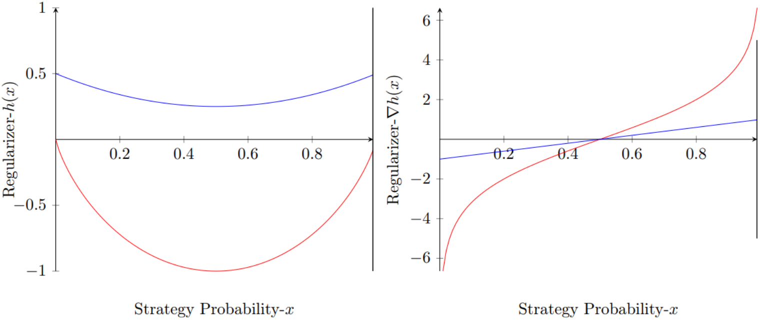
- Negative Shannon Entropy: $h(x) = \sum_i x_i \log(x_i)$
  - Exponential/Multiplicative Weight Updates

$$\Lambda_i(y) = \exp(y_i) / \sum_j \exp(y_j)$$

- Euclidean Regularizer: $h(x) = \sum_i {x_i}^2/2$
  - Euclidean Projection

$$\Pi(y) = \arg\min_{x \in \Delta} \|y - x\|^2$$

Dichotomy of regularizers



$$\begin{bmatrix} \text{steep} & h_1(x) & = & x\log(x) & + & (1-x)\log(1-x) \\ \text{non-steep} & h_2(x) & = & \frac{1}{2}x^2 & + & \frac{1}{2}(1-x)^2 \end{bmatrix}$$

The feedback model

We assume a "black-box" model for players' payoff vector of the form

$$\hat{v}_n = v(X_n) + Z_n \tag{1}$$

for some abstract error process $Z_n = (Z_{i,n})_{i \in \mathcal{N}}$.
We will further decompose Z_n as $Z_n = U_n + b_n$, where

- Random (zero-mean) error: $\mathbb{E}[U_n \,|\, \mathcal{F}_n] = 0$.
- Systematic error: $b_n = \mathbb{E}[Z_n \,|\, \mathcal{F}_n]$.

with $\mathcal{F}_n$ denoting the history of X_n up to stage n (inclusive).

Assumptions

We may then characterize the input signal $\hat{v}_n$ by means of the following statistics:

1. *Bias:* $\mathbb{E}[\|b_n\|_* \,|\, \mathcal{F}_n] \leq B_n$
2. *Variance:* $\mathbb{E}[\|U_n\|_*^2 \,|\, \mathcal{F}_n] \leq M_n^2$

In the above, B_n and M_n represent deterministic bounds on the bias and variance of the feedback signal $\hat{v}_n$.

Assumptions

For concreteness, we will also make the following blanket assumptions:

1. *Bias control:* $\lim_{n\to\infty} B_n = 0$ and $\sum_n \gamma_n B_n < \infty$.
2. *Variance control:* $\sum_n \gamma_n^2 M_n^2 < \infty$.
3. *Generic observation errors at equilibrium:* For every mixed Nash equilibrium x^* of Γ and for all $n = 0, 1, \ldots,$ there exists a player $i \in \mathcal{N}$ and strategies $a, b \in \operatorname{supp}(x_i^*)$ such that

$$\mathbb{P}(|\hat{v}_{ia,n} - \hat{v}_{ib,n}| \geq \beta | \mathcal{F}_n) > 0 \quad \text{for all sufficiently small } \beta > 0.$$

Examples

Model 1 - Oracle based feedback

- At each round n, every player $i \in \mathcal{N}$ picks an action $\alpha_{i,n} \in \mathcal{A}_i$ based on $X_{i,n} \in \mathcal{X}_i$.

Model 1 - Oracle based feedback

- At each round n, every player $i \in \mathcal{N}$ picks an action $\alpha_{i,n} \in \mathcal{A}_i$ based on $X_{i,n} \in \mathcal{X}_i$.
- An oracle reveals to each player the pure payoff vector $v_i(\alpha_n) \equiv (u_i(\alpha_i; \alpha_{-i,n}))_{\alpha_i \in \mathcal{A}_i}$.

Model 1 - Oracle based feedback

- At each round n, every player $i \in \mathcal{N}$ picks an action $\alpha_{i,n} \in \mathcal{A}_i$ based on $X_{i,n} \in \mathcal{X}_i$
- An oracle reveals to each player the pure payoff vector $v_i(\alpha_n) \equiv (u_i(\alpha_i; \alpha_{-i,n}))_{\alpha_i \in \mathcal{A}_i}$.
- Then the player's feedback signal is $\hat{v}_{i,n} = v_i(\alpha_n)$.

Model 1 - Oracle based feedback

Special case of our general model with

Model 1 - Oracle based feedback

Special case of our general model with

- Assumption for bias is trivial because $\mathbb{E}[\hat{v}_n \mid \mathcal{F}_n] = \mathbb{E}_{X_n}[v(\alpha_n)] = v(X_n)$, i.e., $b_n = 0$.

Model 1 - Oracle based feedback

Special case of our general model with

- Assumption for bias is trivial because $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = \mathbb{E}_{X_n}[v(\alpha_n)] = v(X_n)$, i.e., $b_n = 0$.
- Assumption for noise is satisfied as long as $\sum_n \gamma_n^2 < \infty$, since $\|U_n\|_* \leq 2 \max_X \|v(X)\|_*$.

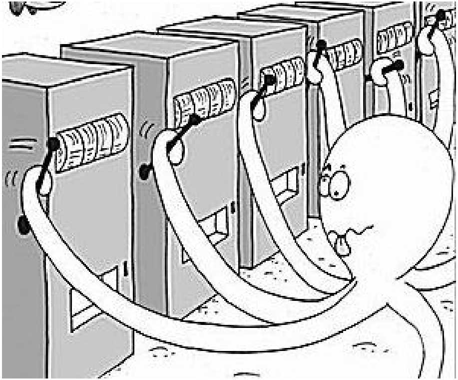
Model 1 - Oracle based feedback

Special case of our general model with

- (A1) is trivial because $\mathbb{E}[\hat{v}_n \,|\, \mathcal{F}_n] = \mathbb{E}_{X_n}[v(\alpha_n)] = v(X_n)$, i.e., $b_n = 0$.
- (A2) is satisfied as long as $\sum_n \gamma_n^2 < \infty$, since $\|U_n\|_* \leq 2 \max_X \|v(X)\|_*$.
- (A3) is an immediate consequence of genericity. Otherwise, the game should have pure Nash equilibria.

Model 2 - Payoff based feedback (Bandit)

Google Ads



Bandit Case

- At each round n, every player $i \in \mathcal{N}$ picks an action $\alpha_{i,n} \in \mathcal{A}_i$ based on $X_{i,n} \in \mathcal{X}_i$.

Bandit Case

- At each round n, every player $i \in \mathcal{N}$ picks an action $\alpha_{i,n} \in \mathcal{A}_i$ based on $X_{i,n} \in \mathcal{X}_i$.
- Players observe their realized payoffs $u_i(\alpha_{i,n}, \alpha_{-i,n})$

Bandit Case

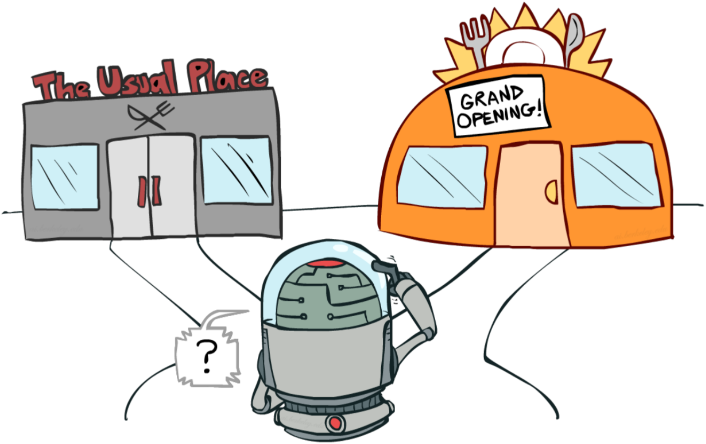
- At each round n, every player $i \in \mathcal{N}$ picks an action $\alpha_{i,n} \in \mathcal{A}_i$ based on $X_{i,n} \in \mathcal{X}_i$.
- Players observe their realized payoffs $u_i(\alpha_{i,n}, \alpha_{-i,n})$
- Players need to somehow estimate their payoffs!

Importance Weighted Estimator

$$\hat{v}_{ia,n} = \begin{cases} 0 & , \text{ if } a \neq a_{i,n} \\ \dfrac{u_i(a; a_{-i,n})}{x_{ia,n}} & , \text{ if } a = a_{i,n} \end{cases}$$

- Unbiased: $\mathbb{E}[\hat{v}_{i,n}] = v_i(X_n)$
- Unbounded Variance: $\mathbb{E}[\|\hat{v}_{i,n}\|_*^2 \,|\, \mathcal{F}_n] \sim \dfrac{1}{\min x_{ia,n}}$

Exploitation-Exploration



Let's leave our options open...

FTRL-exploration

- Idea: We do not limit from the beginning other options, we regularize the probabilities with an exploitation parameter that goes to zero in the infinity.

$$Y_{ia,n+1} = Y_{ia,n} + \gamma_n \hat{v}_{ia,n}$$
$$X_{i,n} = \arg\max_{X \in \Delta(\mathcal{A}_i)} \{\langle Y_{i,n}, X\rangle - h_i(X)\}$$
$$\hat{X}_{i,n} = (1-\epsilon_n) X_{i,n} + \frac{\epsilon_n}{A_i}$$

- Unbiased: $\mathbb{E}[\hat{v}_{i,n}] = v_i(\hat{X}_n)$
- Bounded Variance: $\mathbb{E}[\|U_{i,n}\|_*^2 \,|\, \mathcal{F}_n] \sim \dfrac{1}{\min \hat{X}_{ia,n}} = \mathcal{O}(1/\epsilon_n)$
- Bias: $\|b_n\|_* = \|v(\hat{X}_n) - v(X_n)\|_* = O(\varepsilon_n)$

The Bandit Case

- (A1) is satisfied as long as $\varepsilon_n \to 0$ and $\sum_n \gamma_n \varepsilon_n < \infty$.
- (A2) is satisfied $\sum_n \gamma_n^2 \varepsilon_n^{-1} < \infty$.
- (A3) is an immediate consequence of genericity. Otherwise, the game should have pure Nash equilibria.

Asymptotic Stability

A point $x^* \in \mathcal{X}$ is said to be

1. *Stochastically stable* under (FTRL): If for all $\delta > 0$ and all neighborhoods $\mathcal{U}$ of x^* there exists open set of initial conditions $\mathcal{W}_0 \subseteq \mathcal{Y}$ such that

$$\mathbb{P}(X_n \in \mathcal{U} \text{ for all } n = 0, 1, \ldots) \geq 1 - \delta$$

   whenever $Y_0 \in \mathcal{W}_0$.

2. *Stochastically attracting* under (FTRL): If for all $\delta > 0$, there exists open set of initial conditions $\mathcal{W}_0 \subseteq \mathcal{Y}$ such that

$$\mathbb{P}(\lim_{n\to\infty} X_n = x^*) \geq 1 - \delta$$

   whenever $Y_0 \in \mathcal{W}_0$.

3. *Stochastically asymptotically stable* under (FTRL): if it is stochastically stable and attracting.

Main Results

Main Theorem. Suppose that Assumptions 1–3 hold. Then:
x^* is a strict Nash equilibrium $\iff$ x^* is stochastically asymptotically stable under (FTRL)

Main Results

Theorem

Let $x^* \in \mathcal{X}$ be a strict Nash equilibrium of Γ. If (FTRL) is run with inexact payoff feedback satisfying Assumptions 1 and 2, then x^* is stochastically asymptotically stable.

Theorem

Let x^* be a mixed Nash equilibrium of Γ. If (FTRL) is run with inexact payoff feedback satisfying assumption 3, then x^* is not stochastically asymptotically stable.

Proof techniques - Instability

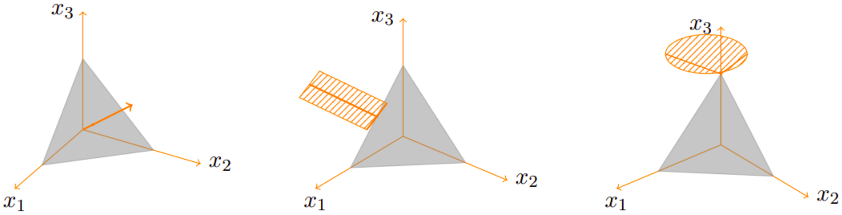


Figure: Polar cone

1. $x = Q(y) \Leftrightarrow y \in \partial h(x)$
2. $\partial h(x) = \nabla h(x) + \mathrm{PC}(x)$ for all $x \in \mathcal{X}$,
   where $PC(x) = \{y \in \mathcal{Y} : y_a \geq y_b$ for all $a, b \in \mathcal{A}\}$.

Proof techniques - Instability

Lemma (Informal)
Let $X_{i,n}$ be the sequence of play in (FTRL) i.e., $X_{i,n} = Q(Y_{i,n}) \in \mathcal{X}_i$ of player $i \in \mathcal{N}$; and for some round $n \geq 0$ let $a, b \in \operatorname{supp}(X_{i,n})$ be two pure strategies of player $i \in \mathcal{N}$. Then it holds:

$$(\theta_i'(X_{ia,n+1}) - \theta_i'(X_{ia,n})) - (\theta_i'(X_{ib,n+1}) - \theta_i'(X_{ib,n})) = \gamma_n(\hat{v}_{ia,n} - \hat{v}_{ib,n})$$

Proof techniques - Instability

Lemma (Informal)

Let $X_{i,n}$ be the sequence of play in (FTRL) i.e., $X_{i,n} = Q(Y_{i,n}) \in \mathcal{X}_i$ of player $i \in \mathcal{N}$; and for some round $n \geq 0$ let $a, b \in \text{supp}(X_{i,n})$ be two pure strategies of player $i \in \mathcal{N}$. Then it holds:

$$(\theta_i'(X_{ia,n+1}) - \theta_i'(X_{ia,n})) - (\theta_i'(X_{ib,n+1}) - \theta_i'(X_{ib,n})) = \gamma_n(\hat{v}_{ia,n} - \hat{v}_{ib,n})$$

- Assume ad absurdum that a mixed Nash equilibrium x^* is stochastically asymptotically stable. Since x^* is mixed, there exist $a, b \in \text{supp}(x^*)$.

Proof techniques - Instability

Lemma (Informal)
Let $X_{i,n}$ be the sequence of play in (FTRL) i.e., $X_{i,n} = Q(Y_{i,n}) \in \mathcal{X}_i$ of player $i \in \mathcal{N}$; and for some round $n \geq 0$ let $a, b \in \operatorname{supp}(X_{i,n})$ be two pure strategies of player $i \in \mathcal{N}$. Then it holds:

$$(\theta_i'(X_{ia,n+1}) - \theta_i'(X_{ia,n})) - (\theta_i'(X_{ib,n+1}) - \theta_i'(X_{ib,n})) = \gamma_n(\hat{v}_{ia,n} - \hat{v}_{ib,n})$$

- Assume ad absurdum that a mixed Nash equilibrium x^* is stochastically asymptotically stable. Since x^* is mixed, there exist $a, b \in \operatorname{supp}(x^*)$.
- The stochastic stability implies that for all $\varepsilon, \delta > 0$ if X_0 belongs to an initial neighborhood $\mathcal{U}_\varepsilon$, then $\|X_n - x^*\| < \varepsilon$ for all $n \geq 0$, with probability at least $1 - \delta$.

Proof techniques - Instability

- By the triangle inequality for two consecutive instances of the sequence of play $X_{i,n}, X_{i,n+1}$ for any player $i \in \mathcal{N}$ it holds:

  $|X_{ia,n+1} - X_{ia,n}| + |X_{ib,n+1} - X_{ib,n}| < \mathcal{O}(\varepsilon)$ with probability $1-\delta$

Proof techniques - Instability

- By the triangle inequality for two consecutive instances of the sequence of play $X_{i,n}, X_{i,n+1}$ for any player $i \in \mathcal{N}$ it holds:

  $$|X_{ia,n+1} - X_{ia,n}| + |X_{ib,n+1} - X_{ib,n}| < \mathcal{O}(\varepsilon) \text{ with probability } 1-\delta$$

- Consider ε sufficiently small, such that the probabilities of the strategies that belong to the support of the equilibrium are bounded away from 0, for all the points of the neighborhood. Since θ_i is continuously differentiable in $(0, 1]$, the differences decribed in the lemma above are bounded from $\mathcal{O}(\varepsilon)$.

Proof techniques - Instability

- If the sequence of play X_n is contained to an $\varepsilon-$neighborhood of x^*, then the difference of the feedback, for any player $i \in \mathcal{N}$, to two strategies of the equilibrium is $\mathcal{O}(\varepsilon/\gamma_n)$ with probability at least $1 - \delta$:

$$\mathbb{P}(|\hat{v}_{ia,n} - \hat{v}_{ib,n}| = \mathcal{O}(\varepsilon/\gamma_n) \mid \mathcal{F}_n) \geq 1 - \delta$$

Proof techniques - Instability

- If the sequence of play X_n is contained to an $\varepsilon-$neighborhood of x^*, then the difference of the feedback, for any player $i \in \mathcal{N}$, to two strategies of the equilibrium is $\mathcal{O}(\varepsilon/\gamma_n)$ with probability at least $1-\delta$:

$$\mathbb{P}(|\hat{v}_{ia,n} - \hat{v}_{ib,n}| = \mathcal{O}(\varepsilon/\gamma_n) \mid \mathcal{F}_n) \geq 1-\delta$$

- From assumption 3 for a fixed round n and some player $i \in \mathcal{N}$, there exist $\beta, \pi > 0$ such that: $\mathbb{P}(|\hat{v}_{ia,n} - \hat{v}_{ib,n}| \geq \beta \mid \mathcal{F}_n) = \pi > 0$.

Proof techniques - Instability

- If the sequence of play X_n is contained to an $\varepsilon-$neighborhood of x^*, then the difference of the feedback, for any player $i \in \mathcal{N}$, to two strategies of the equilibrium is $\mathcal{O}(\varepsilon/\gamma_n)$ with probability at least $1-\delta$:

$$\mathbb{P}(|\hat{v}_{ia,n} - \hat{v}_{ib,n}| = \mathcal{O}(\varepsilon/\gamma_n) \mid \mathcal{F}_n) \geq 1 - \delta$$

- From assumption 3 for a fixed round n and some player $i \in \mathcal{N}$, there exist $\beta, \pi > 0$ such that: $\mathbb{P}(|\hat{v}_{ia,n} - \hat{v}_{ib,n}| \geq \beta \mid \mathcal{F}_n) = \pi > 0$.
- Thus by choosing $\varepsilon = \mathcal{O}(\beta\gamma_n)$ and $\delta = \pi/2$, we obtain a contradiction and our proof is complete.

Nash equilibria - reminder

A point x^* is a *Nash equilibrium* of Γ if

$$u_i(x^*) \geq u_i(x_i; x^*_{-i}) \quad \text{for all } x_i \in \mathcal{X}_i \text{ and all } i \in \mathcal{N}. \qquad \text{(NE)}$$

We call support of x^* the set: $\text{supp}(x_i^*) = \{\alpha_i \in \mathcal{A}_i : x^*_{i\alpha_i} > 0\}$.
Equivalently, Nash equilibria can be characterized by means of the variational inequality

$$v_{i\alpha_i^*}(x^*) \geq v_{i\alpha_i}(x^*) \quad \text{for all } \alpha_i^* \in \text{supp}(x_i^*) \text{ and all } \alpha_i \in \mathcal{A}_i,\ i \in \mathcal{N}.$$

Proof techniques - Stability

- Let $x^* = (\alpha_1^*, \ldots, \alpha_N^*) \in \mathcal{A}$ be a strict Nash equilibrium. Then for every $\varepsilon \in (0, 1)$, there exist constants $M_{i,\varepsilon}$ and the corresponding score-dominant open sets for each player $i \in \mathcal{N}$ such that: $\prod_{i \in \mathcal{N}} Q_i(\mathcal{W}_i(M_{i,\varepsilon})) \subseteq \mathcal{U}_\varepsilon$, where $\mathcal{U}_\varepsilon = \{x \in \mathcal{X} : x_{i\alpha_i^*} > 1 - \varepsilon \text{ for all } i \in \mathcal{N}\}$ and

$$\mathcal{W}_i(M_i) = \{Y_i : Y_{i\alpha_i^*} - Y_{i\alpha_i} > M_i \text{ for all } \alpha_i \neq \alpha_i^*, \alpha_i \in \mathcal{A}_i\}$$
$$\text{for each player } i \in \mathcal{N}$$

Proof techniques - Stability

- Fix a confidence level $\delta > 0$, focus on one player $i \in \mathcal{N}$ and drop the index i for simplicity; consider a neighborhood $\mathcal{U}$ of x^* that can be described as the one above and for which $u_\alpha(X) - u_{\alpha^*}(X) \leq -c$ for some $c > 0$, for all $\alpha \neq \alpha^*$, $\alpha \in \mathcal{A}_i$ and all $X \in \mathcal{U}$.
- We will prove by induction that there exists an open set of initial conditions $\mathcal{W}_0$, such that whenever $Y_0 \in \mathcal{W}_0$ then $Y_n \in \mathcal{W}$ for all $n = 0, 1, \ldots$.

- Notice that whenever $X \in \mathcal{U}$, the payoffs belong to the set $\mathcal{W} = Q^{-1}(\mathcal{U})$. Furthermore, the payoff differences $Y_\alpha - Y_{\alpha^*}$ between every pure strategy $\alpha \in \mathcal{A}_i$, $\alpha \neq \alpha^*$ and the strategy of the equilibrium α^* can be expressed as

$$\begin{aligned} Y_{\alpha,n+1} - Y_{\alpha^*,n+1} = & Y_{\alpha,0} - Y_{\alpha^*,0} + \sum_{k=0}^{n} \gamma_k (u_\alpha(X_k) - u_{\alpha^*}(X_k)) \\ & + \sum_{k=0}^{n} \gamma_k Noise_k + \sum_{k=0}^{n} \gamma_k Bias_k \end{aligned}$$

- Notice that whenever $X \in \mathcal{U}$, the payoffs belong to the set $\mathcal{W} = Q^{-1}(\mathcal{U})$. Furthermore, the payoff differences $Y_\alpha - Y_{\alpha^*}$ between every pure strategy $\alpha \in \mathcal{A}_i$, $\alpha \neq \alpha^*$ and the strategy of the equilibrium α^* can be expressed as

$$
\begin{aligned}
Y_{\alpha,n+1} - Y_{\alpha^*,n+1} = & Y_{\alpha,0} - Y_{\alpha^*,0} + \sum_{k=0}^{n} \gamma_k (u_\alpha(X_k) - u_{\alpha^*}(X_k)) \\
& + \sum_{k=0}^{n} \gamma_k Noise_k + \sum_{k=0}^{n} \gamma_k Bias_k
\end{aligned}
$$

- Using martingale limit theory we control the terms $\sum_{k=0}^{n} \gamma_k Noise_k$, $\sum_{k=0}^{n} \gamma_k Bias_k$ as to be less than $\varepsilon_1 = \sqrt{2 \sum_{k=0}^{\infty} \gamma_k^2 M_k^2 / \delta}$, $\varepsilon_2 = 2 \sum_{k=0}^{\infty} \gamma_k B_k / \delta$ equivalently with probability at least $1 - \delta$.

- Let $R_n = \sum_{k=0}^{n} \gamma_k (U_{\alpha,k} - U_{\alpha^*,k})$, which is a martingale.
- Consider the event $D_{n,\varepsilon_1} = \{\sup_{0 \le k \le n} R_k \ge \varepsilon_1\}$, then

$$\mathbb{P}(D_{n,\varepsilon_1}) \le \frac{\mathbb{E}[R_n^2]}{{\varepsilon_1}^2} \le \frac{2\sum_{k=0}^{n} \gamma_k^2 M_k^2}{\varepsilon_1^2}$$

- Notice that

$$\begin{aligned}\mathbb{E}[R_n^2] &= \sum_{k=0}^{n} \gamma_k^2 \, \mathbb{E}[|U_{\alpha,k} - U_{\alpha^*,k}|^2] \le 2 \sum_{k=0}^{n} \gamma_k^2 \, \mathbb{E}[\|U_k\|_*^2] \\ &= 2 \sum_{k=0}^{n} \gamma_k^2 \, \mathbb{E}[\mathbb{E}[\|U_k\|_*^2 \mid \mathcal{F}_k]] \le 2 \sum_{k=0}^{n} \gamma_k^2 M_k^2\end{aligned}$$

and $\mathbb{E}[U_{\alpha,k} U_{b,l}] = \mathbb{E}[\mathbb{E}[U_{\alpha,k} U_{b,l} \mid \mathcal{F}_{k \vee l}]] = 0$ for all $k \ne l$ and a, b be either of the pure strategy α and the strategy of the equilibrium α^*, due to the noise being zero-mean.

- Let $\Gamma_1 = 2\sum_{k=0}^{\infty} \gamma_k^2 M_k^2$ and choose $\varepsilon_1 = \sqrt{2\Gamma_1/\delta}$.
- The event $D_{\varepsilon_1} = \cup_{n=0}^{\infty} D_{\varepsilon_1,n}$ will happen with probability at most $\delta/2$.

- Notice that

$$|\sum_{k=0}^{n} \gamma_k (b_{\alpha,k} - b_{\alpha^*,k})| \leq \sum_{k=0}^{n} \gamma_k |b_{\alpha,k} - b_{\alpha^*,k}| \leq 2 \sum_{k=0}^{n} \gamma_k \|b_k\|_*$$

- Let $S_n = 2 \sum_{k=0}^{n} \gamma_k \|b_k\|_*$, which is a submartingale.
- If $E_{n,\varepsilon_2} = \{\sup_{0\leq k \leq n} S_k \geq \varepsilon_2\}$ then it holds

$$\mathbb{P}(E_{n,\varepsilon_1}) \leq \frac{\mathbb{E}[S_n]}{\varepsilon_2} = \frac{2 \sum_{k=0}^{n} \gamma_k \, \mathbb{E}[\mathbb{E}[\|b_k\|_* \,|\, \mathcal{F}_k]]}{\varepsilon_2}$$
$$\leq \frac{2 \sum_{k=0}^{n} \gamma_k B_k}{\varepsilon_2}$$

- Let $\Gamma_2 = 2\sum_{k=0}^{\infty} \gamma_k B_k$ and choose $\varepsilon_2 = 2\Gamma_2/\delta$.
- Then the event $E_{\varepsilon_2} = \cup_{n=0}^{\infty} E_{n,\varepsilon_2}$ will occur with probability at most $\delta/2$.

- Choose $M_0 > M + \varepsilon_1 + \varepsilon_2$ and let $\mathcal{W}_0 = \{Y : Y_\alpha < -M_0 \text{ for all } \alpha \neq \alpha^*\}$. If $Y_0 \in \mathcal{W}_0$ then with probability at least $1 - \delta$ we prove that $Y_n \in M$ for all $n = 1, 2, \dots$ and thus the equilibrium is stochastically stable.

- Choose $M_0 > M + \varepsilon_1 + \varepsilon_2$ and let $\mathcal{W}_0 = \{Y : Y_\alpha < -M_0 \text{ for all } \alpha \neq \alpha^*\}$. If $Y_0 \in \mathcal{W}_0$ then with probability at least $1 - \delta$ we prove that $Y_n \in M$ for all $n = 1, 2, \dots$ and thus the equilibrium is stochastically stable.
- Since with probability at least $1 - \delta$ the sequence remains in the neighborhood $\mathcal{U}$ we have

$$Y_{\alpha,n+1} - Y_{\alpha^*,n+1} \leq -c \sum_{k=0}^{n} \gamma_k + \varepsilon_1 + \varepsilon_2 \qquad (2)$$

which implies that the score differences go to $-\infty$, thus all the strategies except for the strategy of the equilibrium become dominated.As a result the point is stochastically asymptotically stable.

Permitted parameters

The above conditions for the method's learning rate and exploration parameters can be achieved by using schedules of the form

- $\gamma_n \propto 1/n^p$
- $\varepsilon_n \propto 1/n^q$

with $p + q > 1$ and $2p - q > 1$. A popular choice is $p = 2/3 + \delta$ and $q = 1/3 + \delta$ for some arbitrarily small $\delta > 0$ – or $\delta = 0$ and including an extra logarithmic factor.