No-regret algorithms and games

Sylvain Sorin

IMJ-PRG
Sorbonne Université
sylvain.sorin@imj-prg.fr

Workshop on Learning in Games
Institut de Mathématiques de Toulouse
July 1-3, 2024

No regret-algorithms appear in game dynamics, prediction of sequences and convex optimization.
We will describe the basic framework, the main tools and the fundamental results.

Part I deals with the initial discrete (in space) and random procedures.

Part II is concerned with continuous deterministic algorithms.

Part I

A. On-line algorithms and regret

1. No-regret properties
2. Approachability theory
3. Existence
4. Calibration.
5. Extensions

B) Application to finite games.

1. Unilateral procedures.
2. External regret and Hannan set.
3. Internal regret and correlated equilibria
4. Alternative approaches

Part II

1. On-line learning, vector field, convex optimization.
2. Closed field, continuous time
3.GD, MD and DA
4.Advances

The notion of regret appears in Hannan, 1957 [60], Blackwell, 1956 [16] in a game theoretical set-up.

Algorithms and properties are studied in this spirit in Foster and Vohra, 1993 [46], Fudenberg and Levine, 1995 [54], Foster and Vohra, 1999 [49], Hart and Mas-Colell, 2000 [62], Lehrer, 2003 [91] , Benaim, Hofbauer and Sorin, 2005 [19], Cesa-Bianchi and Lugosi, 2006 [36] ... among others.

This topic is analyzed in the following books:

Fudenberg and Levine (1998) *The Theory of Learning in Games*, MIT Press.
Young (2004) *Strategic Learning and Its Limits*, Oxford U. P.
Cesa-Bianchi and Lugosi (2006) *Prediction, Learning and Games*, Cambridge University Press.
Hart and Mas-Colell (2013) *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*, World Scientific Publishing.

and the connection with related notions of approachability and consistency is well presented in the survey:

Perchet (2014) *Approachability, regret and calibration: implications and equivalences*, [122]

Special Issue on Learning in Games in Honor of D. Blackwell (1999) *Games and Economic Behavior*, **29**.

Similar tools and properties occur in statistics and in the learning community:

Vvok, 1990 [159], Cover, 1991 [40], Littlestone and Warmuth, 1994 [94], Freund and Shapire, 1999 [53], Auer, Cesa-Bianchi, Freund and Shapire, 2002 [8], Cesa-Bianchi and Lugosi, 2003 [35], Stoltz and Lugosi, 2005 [150], Kalai and Vempala, 2005 [83], Blum and Mansour, 2007 [23], ...

Part I: No-regret (I)

Table of contents

A. On-line algorithms and regret

Consider an agent acting in discrete time and facing an unknown environment.
At each stage $n \in \mathbb{N}$
- choose k_n in a finite set K,
- then observe a reward vector $U_n \in \mathcal{U} = [-1, 1]^K$
- payoff is the k_n^{th} component: $\omega_n = U_n^{k_n}$.

Adversarial framework: no assumption on the reward process.
It is, at each stage n a function of the past history
$h_{n-1} = \{k_1, U_1, ..., k_{n-1}, U_{n-1}\} \in H_{n-1}$.

A strategy of the agent is a map σ from $H = \cup_{m=0}^{+\infty} H_m$ to $\Delta(K)$ (set of probabilities on K).

$\sigma(h_{n-1})$ is the "mixed move" at stage n.

1. No-regret properties

External regret

The external regret given $k \in K$ and $U \in \mathscr{U} \subset \mathbb{R}^K$ is the vector $R(k,U) \in \mathbb{R}^K$ defined by:

$$R(k,U)^\ell = U^\ell - U^k,\ \ell \in K.$$

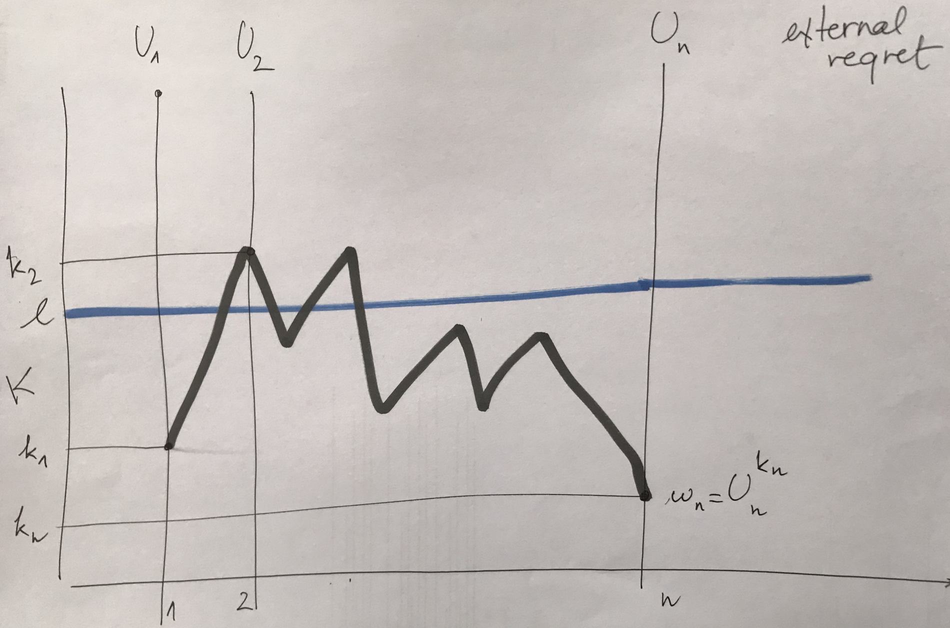
Evaluation of the procedure σ relies on the sequence of external regret vectors $\{R_n\}$ at each stage n with $R_n = R(k_n, U_n)$ thus :

$$R_n^\ell = U_n^\ell - \omega_n,\ \ell \in K.$$

The average external regret vector at stage n is $\overline{R}_n = \frac{1}{n}\sum_{m=1}^n R_m$, thus :

$$\overline{R}_n^\ell = \overline{U}_n^\ell - \overline{\omega}_n,\ \ell \in K.$$

This compares the realized average payoff to the average payoff corresponding to the choice of a constant component. See Hannan, 1957 [60], Fudenberg and Levine, 1995 [54], Foster and Vohra, 1999 [49], ...



U_1
U_2
U_n
external regret
k_2
ℓ
K
k_1
k_n
$\omega_n = U_n^{k_n}$
1
2
n

Definition

A strategy σ satisfies external consistency (or exhibits no external regret) if, for every process $\{U_m\} \in \mathscr{U}$:

$$\max_{k \in K} [\overline{R}_n^k]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

or, equivalently $\quad \sum_{m=1}^{n} (U_m^k - \omega_m) \leq o(n), \qquad \forall k \in K.$

The average payoff on the path is asymptotically greater than on any horizontal line.

Internal regret

The internal regret given (k,U) is the $K \times K$ matrix $S(k,U)$ with components:

$$S^{j\ell}(k,U) = (U^\ell - U^j)\, \mathbf{I}_{\{j=k\}}.$$

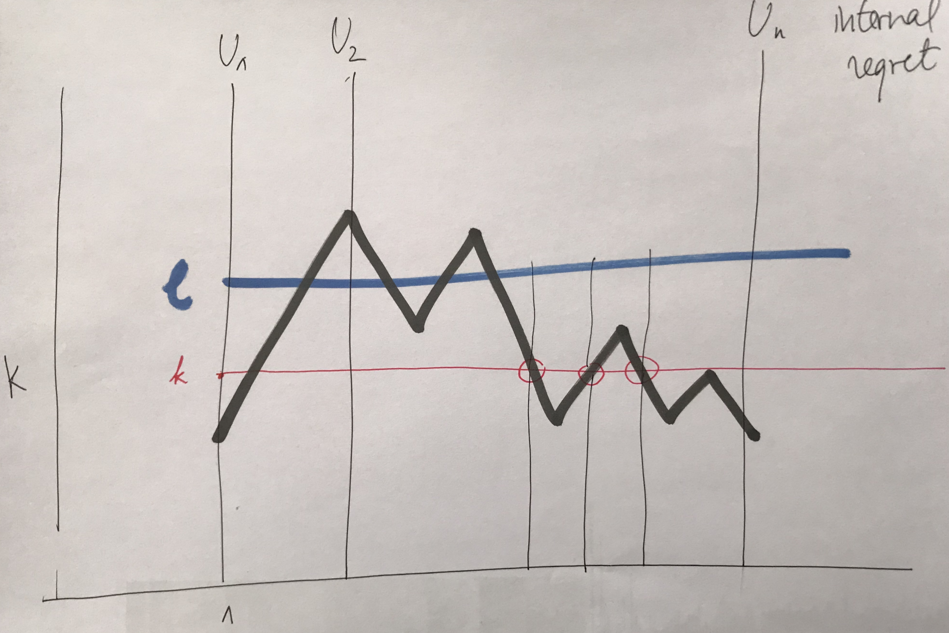
The evaluation at stage n is $S_n = S(k_n, U_n)$ hence defined by:

$$S_n^{k\ell} = \begin{cases} U_n^\ell - U_n^k & \textit{for } k = k_n \\ 0 & \textit{otherwise}. \end{cases}$$

Average internal regret matrix:

$$\overline{S}_n^{k\ell} = \frac{1}{n} \sum_{m=1, k_m=k}^{n} (U_m^\ell - U_m^k)$$

Comparison for each component k, of the average payoff obtained on the dates where k was played, to the average payoff for an alternative choice ℓ at these dates.
See Foster and Vohra (1999) [49], Fudenberg and Levine (1999) [56].



U_1
U_2
U_n
internal regret
ℓ
k
K
1

Definition

A strategy σ satisfies internal consistency (or exhibits no internal regret) if, for every process $\{U_m\} \in \mathscr{U}$ and every couple k, ℓ:

$$[\overline{S}_n^{k\ell}]^+ \longrightarrow 0 \text{ a.s., as } n \to +\infty$$

Comparison with horizontal lines on each section of the trajectory.

There are several proofs of existence of such strategies.
See Section B) 4.

We will use here approachability theory.

2. Approachability theory

All the results are due to Blackwell (1954) [21], (1956) [16].

Consider a two person game defined by A, a $I \times J$-matrix with entries in $\mathbb{R}^K$: $A_{ij} \in \mathbb{R}^K$ is the outcome if Player 1 plays $i \in I$ and player 2 plays $j \in J$.
The game is played in discrete time for infinitely many stages:
- at each stage $n = 1, 2, ...,$ after having observed the past history of vector payoffs $\{g_m = A_{i_m j_m}\}$ denoted $h_{n-1} = (g_1, ..., g_{n-1}) \in \mathscr{H}_{n-1} = (\mathbb{R}^K)^{n-1}$,
Player 1 chooses $x_n \in X = \Delta(I)$ (probabilities on I) and player 2 chooses $y_n \in Y = \Delta(J)$.
- then a couple $(i_n, j_n) \in I \times J$ is selected according to the product probability $x_n \otimes y_n$, and the game goes to stage $n+1$ with the history $h_n = (g_1, ..., g_n) \in \mathscr{H}_n$.

A strategy σ of Player 1 in the repeated game is a sequence $\sigma = (s_1, ..., s_n, ...)$ with $s_n : \mathscr{H}_{n-1} \to \Delta(I)$ for each n.
(Similarly for a strategy τ of player 2).

A couple (σ, τ) naturally defines a probability distribution $\mathbf{P}_{\sigma,\tau}$ over the set of plays $\mathscr{H}_\infty = (I \times J)^\infty$, endowed with the product σ-algebra, and $\mathbf{E}_{\sigma,\tau}$ is the associated expectation.

$\overline{g}_n = \frac{1}{n}[\sum_{m=1}^{n} g_m]$ is the average payoff up to stage n included.

$\|A\| = \max_{i \in I, j \in J, k \in K} |A_{ij}^k|$.

Definition

A set C in $\mathbb{R}^K$ is approachable by Player 1 if for any $\varepsilon > 0$ there exists a strategy σ and N such that, for any strategy τ of Player 2 and any $n \geq N$:

$$E_{\sigma,\tau}(d_n) \leq \varepsilon$$

and $d_n \to 0$, where d_n is the euclidean distance $d(\overline{g}_n, C)$.

A set C in $\mathbb{R}^K$ is excludable by Player 1 if for some $\delta > 0$, the set $C^c_\delta = \{z; d(z,C) \geq \delta\}$ is approachable by her.

From the definitions it is enough to consider closed sets C and even their intersection with the closed ball of radius $\|A\|$.

Given x in $X = \Delta(I)$, define the set

$$[xA] = co\ \{\sum_i x_i A_{ij};\ j \in J\} \subset \mathbb{R}^K.$$

If Player 1 uses x, the expected outcome will be in $[xA]$, whatever being the move of player 2.

B-sets and sufficient condition

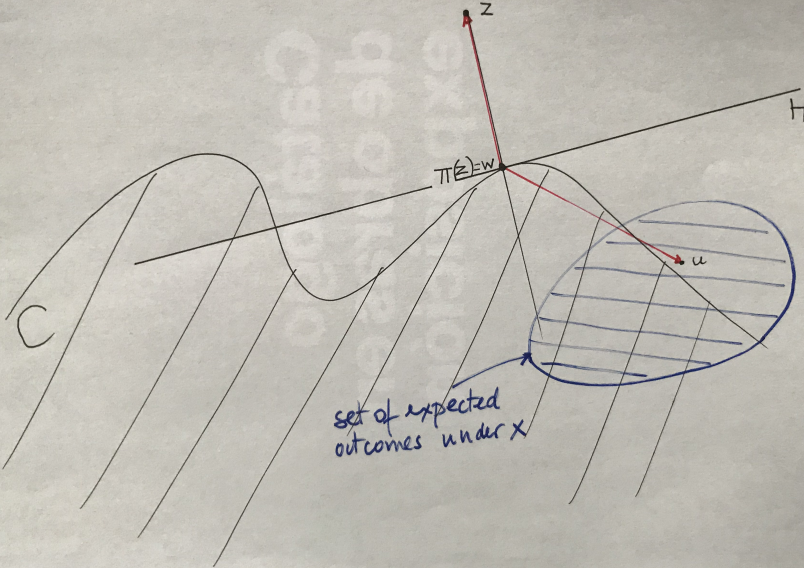
The first result is a sufficient condition for approachability based on the following notion:

Definition

A closed set C in $\mathbb{R}^K$ is a **B**-set for Player 1 if:
for any $z \notin C$, there exists a closest point $w = w(z)$ in C to z and a mixed move $x = x(z)$ in X, such that the hyperplane trough w orthogonal to the segment $[wz]$ separates z from $[xA]$. Explicitly:

$$\langle z - w, u - w \rangle \leq 0, \forall u \in [xA].$$

B-set $\langle z-w, u-w \rangle \leq 0$



Theorem

Let C be a **B***-set for Player 1.*
Then C is approachable by that player.
A strategy satisfying $\sigma(h_n) = x(\overline{g}_n)$, whenever $\overline{g}_n \notin C$, gives:

$$E_{\sigma\tau}(d_n) \;\leq\; \frac{2\|A\|}{\sqrt{n}}, \quad \forall\tau$$

and d_n converges $P_{\sigma\tau}$ a.s. to 0, more precisely:

$$P(\exists n \geq N; d_n^2 \geq \varepsilon) \leq \frac{8\|A\|^2}{\varepsilon N}.$$

Proof

Let Player 1 use a strategy σ as above.
Denote $w_n = w(\overline{g}_n)$ and $d_n^2 = \|\overline{g}_n - w_n\|^2$.

$$d_{n+1}^2 \leq \|\overline{g}_{n+1} - w_n\|^2 \tag{1}$$

implies:

$$d_{n+1}^2 \leq \|\overline{g}_{n+1} - \overline{g}_n\|^2 + \|\overline{g}_n - w_n\|^2 + 2\langle \overline{g}_{n+1} - \overline{g}_n, \overline{g}_n - w_n \rangle. \tag{2}$$

Decompose:

$$\begin{aligned} \langle \overline{g}_{n+1} - \overline{g}_n, \overline{g}_n - w_n \rangle &= \left(\frac{1}{n+1}\right)\langle g_{n+1} - \overline{g}_n, \overline{g}_n - w_n \rangle \\ &= \left(\frac{1}{n+1}\right)\left(\langle g_{n+1} - w_n, \overline{g}_n - w_n \rangle - \|\overline{g}_n - w_n\|^2\right) \end{aligned}$$

The property of $x(\overline{g}_n)$ implies that:

$$\langle E(g_{n+1}|h_n) - w_n, \overline{g}_n - w_n \rangle) \leq 0$$

since $E(g_{n+1}|h_n)$ belongs to $[x(\overline{g}_n)A]$.

Taking conditional expectation with respect to the history h_n gives:

$$\mathsf{E}(d_{n+1}^2|h_n) \leq (1-\frac{2}{n+1})\ d_n^2 + (\frac{1}{n+1})^2 \mathsf{E}(\|g_{n+1}-\overline{g}_n\|^2|h_n). \qquad (3)$$

Since $\|g_{n+1}-\overline{g}_n\|^2 \leq 2\|g_{n+1}\|^2 + 2\|\overline{g}_n\|^2 \leq 4\|A\|^2$, we obtain:

$$\mathsf{E}(d_{n+1}^2) \leq (\frac{n-1}{n+1})\ \mathsf{E}(d_n^2) + (\frac{1}{n+1})^2\ 4\|A\|^2$$

and by induction:

$$\mathsf{E}(d_n^2) \leq \frac{4\|A\|^2}{n}.$$

This gives in particular the convergence in probability of d_n to 0.

Introduce the random variable:
$W_n = d_n^2 + \|A\|^2 \sum_{m=n+1}^{\infty} (\frac{1}{m^2} E(\|g_m - \overline{g}_m\|^2 | h_n)$.
(3) implies:

$$E(W_{n+1}|h_n) \leq W_n$$

thus W_n is a positive supermartingale hence converges $P_{\sigma\tau}$ a.s. to 0.
More precisely Doob's maximal inequality, see e.g. Neveu, 1972 [116], gives :

$$P(\exists n \geq N; d_n^2 \geq \varepsilon) \leq \frac{E(W_N)}{\varepsilon} \leq \frac{8\|A\|^2}{\varepsilon N}.$$

■

In particular one obtains:

Corollary

For any x in S, $[xA]$ is approachable by Player 1, with the constant strategy x.

It follows that a necessary condition for a set C to be approachable by Player 1 is that for any y in Y, $[Ay] \cap C \neq \emptyset$, otherwise C would be excludable by Player 2.

In fact this condition is also sufficient for convex sets.

Convex case

Theorem

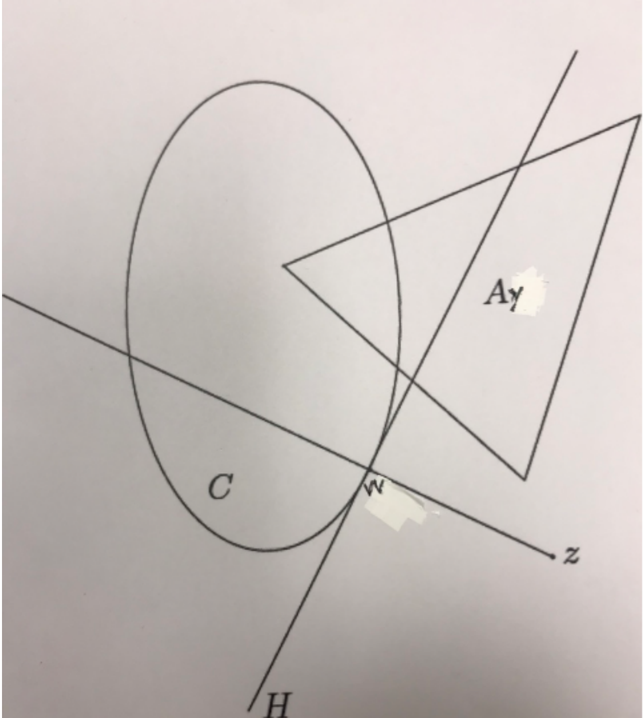
Assume C closed and convex in $\mathbb{R}^K$.
*C is a **B**-set for Player 1 iff*

$$(*) \qquad [Ay] \cap C \neq \emptyset, \qquad \forall y \in Y.$$

*In particular a set is approachable iff it is a **B**-set.*

Proof

By the previous Corollary, it is enough to prove that $(*)$ implies that C is a **B**-set.
The idea is to reduce by projection the problem to the one-dimentional case and to use the minmax theorem.



Ay
C
w
z
H

In fact, let $z \notin C$, $w = \Pi_C(z)$ the (orthogonal) projection on C, and consider the 2 person zero-sum game with real payoff matrix $B = \langle w - z, A \rangle$. Since $[Ay] \cap C \neq \emptyset$ for all $y \in Y$, this implies that its value is at least $\min_{c \in C} \langle w - z, c \rangle = \langle w - z, w \rangle$. Hence there exists an optimal strategy $x \in X$ of player 1 such that $\langle w - z, \sum_i x_i A_{ij} \rangle \ \geq \ \langle w - z, w \rangle$ for any $j \in J$, which shows that xA is on the opposite side of the hyperplane to z, and the result follows. ∎

Remarks

The first property is quite general: bounded random payoffs such that condition **B** makes sense.

For the convex case, only the existence of $\mathtt{val}\langle \alpha, A \rangle$ for any α in $\mathbb{R}^K$ is needed.

Extensions

1. In dimension 1, any set is either approachable or excludable. There exist sets that are neither approachable nor excludable.
Extension to random payoffs, uniformly bounded in L^2 (Blackwell, 1956).
2. Any set is either weakly approachable or weakly excludable (strategy adapted to the duration) [first link with differential games] (Vieille, 1992).
3. Any approachable set contains a **B**-set (Spinat, 2002).
4. Extension to infinite dimension (Lehrer, 2002).
5. General active states (Lehrer, 2003).
6. Idea of a potential (convex case)
Write

$$\langle x_{n+1} - \Pi_C(\bar{x}_n), \bar{x}_n - \Pi_C(\bar{x}_n)\rangle \leq 0, \tag{4}$$

as

$$\langle x_{n+1} - \bar{x}_n, \nabla P_C(\bar{x}_n)\rangle \leq -2\, P_C(\bar{x}_n), \tag{5}$$

with $P_C(x) = \|x - \Pi_C(x)\|^2$ and $\nabla P_C(x) = 2[x - \Pi_C(x)]$
7. Geometric condition and proximal normal (dual approach) (As Soulaimani, Quincampoix and Sorin, 2009)
8. Approachability and viability [second link with differential games] (As Soulaimani, Quincampoix and Sorin, 2009)

3. Existence

Existence External Consistency

We prove the existence of a strategy satisfying EC by showing that the negative orthant $D = \mathbb{R}^K_-$ is approachable by the sequence of regrets $\{R_n\}$ in the game on $K \times \mathscr{U}$ with payoff $R(k, U)$.

Average property 1

Lemma

$\forall x \in \Delta(K), \forall U \in \mathscr{U}$:

$$\langle x, \mathsf{E}_x[R(., U)]\rangle = 0.$$

Proof

One has:

$$\mathsf{E}_x[R(., U)] = \sum_{k \in K} x^k \, R(k, U) = \sum_{k \in K} xk(U - U^k \mathbf{1}) = U - \langle x, U\rangle \mathbf{1}$$

($\mathbf{1}$ is the K-vector of ones), thus $\langle x, \mathsf{E}_x[R(., U)]\rangle = 0$. ∎

Let $Z \in \mathbb{R}^K$ and define if $Z^+ \neq 0$, $x(Z)$ to be proportional to this vector.
Then the **B**-set condition is satisfied, in fact:

$$\langle Z - \Pi_D(Z), G - \Pi_D(Z) \rangle = 0, \quad \forall Z \in \mathbb{R}^K \tag{6}$$

where G is the expected outcome, given $x(Z)$.
Recall that $\Pi_D(Z) = Z^-$, $Z = Z^+ + Z^-$ and $\langle Z^-, Z^+ \rangle = 0$, $\forall Z \in \mathbb{R}^K$.
Now $\langle \Pi_D(Z),\, Z - \Pi_D(Z) \rangle = 0$ and using the previous Lemma:

$$\begin{aligned} \langle G,\, Z - \Pi_D(Z) \rangle &= \langle G,\, Z^+ \rangle \\ &\div \langle G,\, \sigma(Z) \rangle \\ &= \langle \mathsf{E}_x[R(.,U)],\, x \rangle, \quad \text{for } x = x(Z) \\ &= 0 \end{aligned}$$

Hence D is approachable so that $d(\bar{R}_n, \mathbb{R}^K_-)$ goes to 0.

Existence IC

Given a $K \times K$ real matrix A with nonnegative coefficients, let $Inv[A]$ be the non-empty set of invariant measures for A, namely vectors $\mu \in \Delta(K)$ satisfying:

$$\sum_{k \in K} \mu^k A^{k\ell} = \mu^\ell \sum_{k \in K} A^{\ell k} \qquad \forall \ell \in K.$$

(The existence follows from the existence of an invariant measure for a finite Markov chain - which is itself a consequence of the minmax theorem).

Average property 2

Lemma

Given $A \in \mathbb{R}_+^{K^2}$, *let* $\mu \in Inv[A]$ *then:*

$$\langle A, \mathsf{E}_\mu(S(.,U)) \rangle = 0, \quad \forall U \in \mathscr{U}\,.$$

Proof

$$\langle A, \mathsf{E}_\mu(S(.,U)) \rangle = \sum_{k,\ell} A^{k\ell} \mu^k (U^\ell - U^k)$$

and the coefficient of each U^ℓ is

$$\sum_{k \in K} \mu^k A^{k\ell} - \mu^\ell \sum_{k \in K} A^{\ell k} = 0$$

■

To prove the existence of a strategy satisfying internal consistency, we show that $\Delta = \mathbb{R}_-^{K \times K}$ is approachable by the sequence of internal regret $\{S_n\}$ in the game on $K \times \mathscr{U}$ with payoff $S(k,U)$.
Given $B \in \Delta$, define, if $B^+ \neq 0$, $x(B)$ to be an invariant measure of B^+.
One has, M being the expected regret matrix:

$$\langle M - \Pi_\Delta(B),\, B - \Pi_\Delta(B)\rangle = 0$$

since again $\langle \Pi_\Delta(B),\, B - \Pi_\Delta(B)\rangle = 0$ and using the Lemma above:

$$\begin{aligned}
\langle M,\, B - \Pi_\Delta(B)\rangle &= \langle M,\, B^+\rangle \\
&= \langle \mathsf{E}_\mu[S(.,U)],\, B^+\rangle, \quad \text{for } \mu = x(S) \in Inv[S^+] \\
&= 0
\end{aligned}$$

Then Δ is approachable hence $\max_{k,\ell}[\bar{S}_n^{k,\ell}]^+ \longrightarrow 0$.

Calibrating

Consider a sequence of random variables X_m with values in a finite set Ω (written as a basis of $\mathbb{R}^{\Omega}$).
Obviously any deterministic prediction algorithm ϕ_m - where the loss is measured by $\|X_m - \phi_m\|$ - will have a worst loss 1 and any random predictor a loss at least $1/2$ (take $X_m = 1$ iff $\phi_m(1) \leq 1/2$).
Introduce a predictor with values in a finite discretization V of $D = \Delta(\Omega)$.
"$\phi_m = v, v \in V$" means that the anticipated distribution of X is v, i.e. the predicted probability that $X_m = \omega$ (or $X_m^{\omega} = 1$) is v^{ω}.

Definition

ϕ is ε-calibrated if, for any $v \in V$:

$$\lim_{n \to +\infty} \frac{1}{n} \| \sum_{\{m \leq n, \phi_m = v\}} (X_m - v) \| \leq \varepsilon.$$

Dawid (1982), [41].

If the average number of times v is predicted does not vanish, the empirical average distribution of X_m on these dates is close to v.
More precisely let B_n^v the set of stages before n where v is announced, let N_n^v be its cardinal and $\bar{X}_n(v)$ the average of X_m (in the simplex $\Delta(\Omega)$) on these stages.
Then the condition writes:

$$\lim_{n \to +\infty} \frac{N_n^v}{n} \|\bar{X}_n(v) - v\| \leq \varepsilon, \qquad \forall v \in V.$$

From internal consistency to calibrating

Foster and Vohra (1997) [47].
Consider the online algorithm where the choice set of the forecaster is V and given X the reward is the vector with components:

$$U^v = \|X - v\|^2, \quad v \in V,$$

(where we use the L^2 norm).
Given an internal consistent procedure ϕ one obtains (the outcome is here a loss):

$$\frac{1}{n} \sum_{m \in B_n^v} (U_m^v - U_m^w) \leq o(n), \qquad \forall w \in V,$$

which is:

$$\frac{1}{n} \sum_{m \in B_n^v} (\|X_m - v\|^2 - \|X_m - w\|^2) \leq o(n), \qquad \forall w \in V,$$

hence implies:

$$\frac{N_n^v}{n}(\|\bar{X}_n(v) - v\|^2 - \|\bar{X}_n(v) - w\|^2) \leq o(n), \qquad \forall w \in V.$$

In particular by chosing w a closest point to $\bar{X}_n(v)$:

$$\frac{N_n^v}{n}(\|\bar{X}_n(v) - v\|^2) \leq \delta^2 + o(n)$$

where δ is the L^2 mesh of V, from which ε calibration follows.

From calibrating to approachability

Foster and Vohra (1997)
We use calibrating to prove approachability of convex sets.

Assume that the set C satisfies:

$$\forall y \in Y, \exists x \in X \qquad \text{such that} \qquad xAy \in C.$$

Consider a δ-grid of Y defined by $\{y_v, v \in V\}$.
A stage is of type v if player 1 predicts y_v and then plays a mixed move x_v such that $x_v A y_v \in C$.
By using a calibrated procedure, the average move of player 2 on the stages of type v will be δ close to y_v.
By a martingale argument the average payoff will then be ε close to $x_v A y_v$ for δ small enough and n large enough.
Finally the total average payoff is a convex combination of such amounts hence is close to C by convexity.

There is a huge literature on the relations between approachability, no-regret and calibrating.

See e.g.
Mannor and Stoltz, 2010 [98],
Abernethy, Bartlett and Hazan, 2011 [1],
Perchet, 2014 [122].

Extensions

Conditional expectation

Recall that the regret at stage n that the player wants to control is of the form:

$$\frac{1}{n}\sum_{m=1}^{n}[U_m^k - \omega_m], \quad k \in K$$

where $\omega_m = U_m^{k_m}$ is the (random) payoff at stage m.
Let $x_m \in \Delta(K)$ be the strategy of the player at stage m, then

$$\mathsf{E}(\omega_m | h_{m-1}) = \langle U_m, x_m \rangle$$

so that $\omega_m - \langle U_m, x_m \rangle$ is a bounded martingale difference.
Hoeffding-Azuma's [12], [72] concentration inequality for a process $\{Z_n\}$ of martingale differences with $|Z_n| \leq L$ gives that:

$$\mathsf{P}\{|\bar{Z}_n| \geq \varepsilon\} \leq 2\exp(-\frac{n\,\varepsilon^2}{2L^2})$$

Hence the average difference between the payoff and its conditional expectation is controlled.

Thus we will study conditions of the form:

$$\sum_{m=1}^{n} U_m^k - \langle U_m, x_m \rangle \leq o(n), \quad k \in K.$$

or equivalently, because of the linearity:

$$\sum_{m=1}^{n} \langle U_m, x \rangle - \langle U_m, x_m \rangle \leq o(n), \quad x \in \Delta(K).$$

and the requirement to be in the simplex will disappear.
Similarly the internal no-regret condition becomes:

$$\sum_{m=1}^{n} x_m^k [U_m^j - U_m^k] \leq o(n), \quad \forall k, j \in K.$$

Procedures in law

Assume that the actual move k_n is not observed and define a pseudo-process $\tilde{R}$ defined through the conditional expected regret:

$$R_n = U_n - \omega_n \mathbf{1}, \qquad \tilde{R}_n = U_n - \langle U_n, x_n \rangle \mathbf{1}.$$

More generally for Blackwell's theorem, replace the actual payoff g by its conditional expectation γ under x, note that the $\mathbf{B}$ condition still holds and the strategy $\tilde{\sigma}(h_n) = x(\bar{\gamma})$ will still imply that $\bar{\gamma}_n$ approaches C and the same is true for $\bar{g}_n$.

Then consistency holds both for the pseudo and the realized processes under $\tilde{\sigma}$.

Experts and generalized consistency

External consistency can be considered as a robustness property of σ facing a given finite family of "external" experts using procedures $\phi \in \Phi$:

$$\lim \frac{1}{n}[\sum_{m=1}^{n} \langle \phi_m - x_m, U_m \rangle]^+ = 0, \quad \forall \phi \in \Phi.$$

The typical case corresponds to a constant choice : $\phi = k$ and $\Phi = K$.

In general "k" will be the (random) move of expert k, that the player follows with probability x_m^k at stage m.

U_m^k is then the payoff to expert k at stage m.

Internal consistency corresponds to experts adjusting their behavior to the one of the agent.

From external to internal consistency

Stoltz and Lugosi (2005) [150]
Consider a family $\psi^{ij}, (i,j) \in K \times K$ of experts and θ an algorithm that satisfies external consistency with respect to this family.
Define σ inductively as follows.
Given some element $p \in \Delta(K)$, let $p(ij)$ be the vector obtained by adding p^i to the j^{th} component of p.
Let $q_{n+1}(p)$ be the distribution induced by θ at stage $n+1$ given the history h_n and the behavior $\psi^{ij}(h_n) = p(ij)$ of the experts.
Assume that the map $p \mapsto q_{n+1}(p)$ is continuous and let $\bar{p}_{n+1}$ be a fixed point which defines $\sigma(h_n) = x_{n+1}$.
The fact that σ is an incarnation of θ implies that it performs well facing any ψ^{ij} hence

$$[\sum_{m=0}^{n} \langle \psi_m^{ij} - x_m, U_m \rangle] \leq o(n), \qquad \forall i,j$$

which is

$$[\sum_{m=0}^{n} \langle \bar{p}(ij)_m - \bar{p}_m, U_m \rangle] \leq o(n), \qquad \forall i,j$$

hence

$$[\sum_{m=0}^{n} \bar{p}_m^i (U_m^j - U_m^i)] \leq o(n), \qquad \forall i,j$$

and this is the internal consistency condition.

Blum and Mansour (2007) [23]
Consider K parallel algorithms $\{\phi[k]\}$ having no external regret, that generates each a (row) vector $q[k] \in \Delta(K)$ then define σ by an invariant measure p satisfying

$$p = pq.$$

Given the outcome $U_m \in \mathbb{R}^K$, let $p^k U_m$ be the reward vector used for algorithm $\phi[k]$.
Expressing the fact that $\phi[k]$ satisfies no external regret gives, for all $j \in K$

$$[\sum_{m=0}^{n} p_m^k U_m^j - \langle q[k]_m, p_m^k U_m \rangle] \leq o(n)$$

Note that $\sum_k \langle q[k]_m, p_m^k U_m \rangle = \sum_k \langle p_m^k q[k]_m, U_m \rangle = \langle p_m, U_m \rangle$, hence by summing over k, for any function $M : K \mapsto K$, corresponding to a perturbation of σ with $j = M(k)$ the difference between the performances of σ_M and σ will satisfy as well

$$[\sum_{m=0}^{n} \sum_k p_m^k U_m^{M(k)} - \langle p_m, U_m \rangle] = [\sum_{m=0}^{n} \sum_k p_m^k (U_m^{M(k)} - U_m^k)] \leq o(n).$$

This is the internal consistency for “swap experts”.

Large range

Blum and Mansour (2007) [23]; Cesa-Bianchi and Lugosi (2006)[36]; Lehrer (2003) [91].
Consider an even larger set of experts that are allowed (in addition to be adapted to the past history) to choose their actions and to be active as a function of the choice of the predictor.
Explicitly every expert $s \in S$ (finite) is characterized, at stage m, conditional to the past, by :
- a choice function $f_m^s : K \to K$
- an activity function $\tau_m^s : K \to [0,1]$.

Given a predictor ϕ which prediction at stage m has a law p_m the regret facing s is :

$$r_m^s = \sum_k p_m^k \tau_m^s(k)[U_m^{f_m^s(k)} - U_m^k]$$

We assume that the functions f^s, τ^s are known by the predictor.

Then there exists a consistent procedure.

Bandit framework

This is the case where given the move k and the vector U the only information to the agent is the realization $\omega = U^k$ (the vector U is not announced).

Define the pseudo regret vector at each stage n by:

$$\hat{U}_n^k = \frac{\omega_n}{\sigma_n^k} \mathbf{I}_{\{k_n = k\}}$$

and note that it is an unbiased estimator of the true regret.

To keep the outcome bounded one may have to perturb the strategy and same asymptotic properties hold.

Auer, Cesa-Bianchi, Freund, Shapire (2002), [8].

Similarly the pseudo regret matrix is:

$$\hat{S}_n^{kj} = \frac{\sigma_n^k}{\sigma_n^j} \mathbf{1}_{\{k_n = j\}} - U^k \mathbf{1}_{\{k_n = k\}}$$

with expectation $\sigma_n^k [U^j - U^k]$.

For more advances, see Bubeck and Cesa-Bianchi (2012), chapter 5, [33].

Table of contents

1. Unilateral procedures

Let $\mathscr{G}$ a finite game in strategic form.
Finitely many players labeled $i = 1, 2, \ldots, I$.
S^i: finite set of actions of player i, $S = \prod_i S^i$, and $Z = \Delta(S)$ is the set of probabilities on S (correlated moves).
Consider repeated interaction in discrete time where at each stage the players observe the actions of their opponents.

We want evaluate the joint impact on the play of a prescribed behavior of the players.

Since we will study the procedure from the view point of player 1 it is convenient to set $S^1 = K, X = \Delta(K)$ (mixed moves of player 1), $L = \prod_{i \neq 1} S^i$, and $Y = \Delta(L)$ (correlated moves of player 1's opponents) hence $Z = \Delta(K \times L)$.
$F : S \to \mathbb{R}$ denotes the payoff function of player 1.

Player 1 faces the on-line problem, corresponding to the repeated game, where at stage m if ℓ_m is the profile of moves of his opponents, the outcome vector is is $U_m = \{F(k, \ell_m\}_{k \in K}\}$.

2. External regret and Hannan's set

Let $r(z)$ denote the K-dimensional vector at z in Z, defined by:

$$r^j(z) = \sum_{(k,\ell)\in K\times L} z(k,\ell)[F(j,\ell) - F(k,\ell)]$$

Player 1 compares his payoff using an alternative move j to his actual payoff at z, assuming the other players' behavior fixed.

Definition

H^1 (for Hannan's set), Hannan (1956) [60] is the set of correlated moves in Z satisfying the external no-deviation condition for player 1. Formally:

$$H^1 = \{z \in Z : \sum_{(k,\ell)\in K\times L} z(j,\ell)[F(k,\ell) - F(j,\ell)] \leq 0, \forall j \in K\}$$

$$= \{z \in Z : r(z) \in D = \mathbb{R}^K_-\}.$$

The empirical average distribution of correlated moves on a play is $z_n \in Z$ with:

$$z_n^{k\ell} = \frac{1}{n} \sum_{m=1}^{n} \mathbf{1}_{(k_m, \ell_m) = (k, \ell)}$$

Proposition

If Player 1 follows some external consistent procedure, the empirical average distribution of moves converges to the Hannan set H^1 (in the sense that all accumulation points will be in H^1).

Proof
The external no regret property is:

$$\frac{1}{n}\sum_{m=1}^{n}[U_m^j - U_m^{k_m}] \leq o(n) \qquad \forall j \in K$$

which is:

$$\frac{1}{n}\sum_{m=1}^{n}F(j,\ell_m) - \frac{1}{n}\sum_{m=1}^{n}F(k_m,\ell_m) \leq o(n) \qquad \forall j \in K$$

and this expression is:

$$r^j(z_n) = \sum_{k,\ell} z_n(k,\ell)[\,F(j,\ell) - F(k,\ell)] \leq o(n) \qquad \forall j \in K$$

so that the accumulation points of z_n are in H^1. ∎

Alternative proof

We consider an auxiliary game with vector payoffs in $\mathbb{R}^M$, where the dimension is $M = L+1$, and the payoff $g(k,\ell) = (F(k,\ell),\ell)$ is the couple of the current payoff in the original game and of the opponent(s) profile.
D_1 is the convex set:

$$D_1 = \{(u,\theta) \in \mathbb{R} \times \Delta(L); u \geq \max_{k \in K} F(k,\theta)\}.$$

Proposition

D_1 is approachable.

Proof

The proof that D_1 is approachable is that it is not excludable: namely, for any $\theta \in \Delta(L)$, there is some $k \in K$ such that $g(k,\theta) \in D_1$. ∎

This obviously implies the non emptiness of H^1 since by approachability $d(\bar{g}_n, D_1)$ goes to 0 hence also

$[\max_{k \in K} r^k z_n]^+$.

One defines similarly H^i for each player and $H = \cap_{i \in I} H^i$ which is the global Hannan's set.

Proposition

If each player follows some external consistent procedure, the empirical distribution of moves converges a.s. to the Hannan set H.

Note that no coordination is required.

3. Internal regret and correlated equilibria

Given $z \in Z$, introduce the comparison matrix

$$c^{(k,j)}(z) = \sum_{\ell \in L} z(k,\ell)[F(j,\ell) - F(k,\ell)], \quad j,k \in K$$

This corresponds to the change in the payoff of Player 1 at z when replacing move k by j.

Definition

C^1 is the set of correlated moves satisfying the internal no deviation property for player 1, namely:

$$C^1 = \{z \in Z; \sum_{\ell \in L} z(k,\ell)[F(j,\ell) - F(k,\ell)] \leq 0, \forall j,k \in K\}.$$

$$= \{z \in Z; c^{(k,j)}(z) \leq 0, \forall j,k \in K\}.$$

It is obviously a subset of H^1 since

$$\sum_{j \in K} c^{(j,k)}(z) = r^k(z).$$

Proposition

If Player 1 follows some internal no-regret procedure, the empirical distribution of moves converges to the set C^1.

Proof

The internal no regret property is:

$$\frac{1}{n}\sum_{m=1, k_m=k}^{n}(U_m^j - U_m^{k_m}) \leq o(n) \qquad \forall j,k \in K$$

thus:

$$\frac{1}{n}\sum_{m=1, k_m=k}^{n}(F(j,\ell_m) - F(k_m,\ell_m) \leq o(n) \qquad \forall j,k \in K$$

and this expression is:

$$c^{(k,j)}(z_n) = \sum_{k,\ell} z_n(k,\ell)[\, F(j,\ell) - F(k,\ell)] \leq o(n) \qquad \forall j,k \in K$$

so that the accumulation points of z_n are in C^1. ∎

Recall that the set of correlated equilibrium distributions (Aumann, 1974, [9]) of the game $\mathscr{G}$ with payoff functions $\{F^i\}_{i\in I}$ is defined by:

$$C = \{z \in Z; \sum_{\ell \in S^{-i}} z(k,\ell)[F^i(j,\ell) - F^i(k,\ell)] \leq 0, \quad \forall j,k \in S^i, \forall i \in I\}.$$

Hence one has :

Proposition

The intersection over all $i \in I$ of the sets C^i is the set of correlated equilibrium distributions of the game.

Thus we obtain:

Theorem

If each player follows some internal consistency procedure, the empirical average distribution of moves converges to the set of correlated equilibria.

Note that this provides an alternative proof of existence of correlated equilibrium through the existence of internally consistent procedures.

Alternative joint procedure

Hart and Mas Colell (2000) [62]
The procedure is defined (for player 1) by x^1_{n+1} being a function of his average regret, his last move $s^1_n = j$ and some large parameter L, as follows:

$$x^1_{n+1}(k) = \overline{R}^+_n(j,k)/L, \quad k \neq j \qquad ; \qquad x^1_{n+1} = 1 - \sum_{k \neq j} x^1_{n+1}(k).$$

Theorem

If all players use the above procedure, the empirical distribution of moves converge to the set of correlated equilibria.

From calibrating to correlated equilibrium

Foster and Vohra (1997) [47]

Consider the case where Player 1 is forecasting the behavior (a profile in L) of his opponents.

Given a precision level δ, Player 1 is predicting points in a δ-grid $\{p[v], v \in V\}$ of $\Delta(L)$ and then plays a (pure) best reply to his forecast.

It is thus clear that if the forecast is calibrated the empirical distribution of the moves of the opponents, will converge to the forecast, on each event of the form $\{m; p_m = p[v]\}$, hence eventually the action chosen by Player 1, k, will be close to a best reply to the frequency near $p[v]$.

When looking at the average empirical distribution z, the conditional distribution $z|k$ of z given k, will correspond to a convex combination of distributions $p[v]$ to which k is best reply, hence k will again be an (approximate) best reply to $z|k$: hence z is (approximately) in C^1.

If all players use calibrated strategies the empirical average frequency converges to C.

Note that this is in the spirit of the dual approachability property and was used to obtain approachability of convex sets.

4. Alternative approaches

Smooth fictitious play

This procedure is based only on the previous observations of the outcome vectors and not on the moves of the predictor, hence the regret cannot be used, Fudenberg and Levine (1995) [54].

Recall that fictitious play corresponds to

$$x^i_{m+1} \in BR^i(\bar{x}^{-i}_m)$$

where BR^i is the best reply correspondence. Even for 2 player matrix games this procedure does not have global convergence properties, Shapley (1964), [142].

We consider here a unilateral version so player 1 faces a sequence of vectors $U_m = F(., \ell_m)$ and plays a perturbation of a best reply to the average $\bar{U}_n$.

Definition

A smooth perturbation of the payoff on $\Delta(K) \times U \in \mathscr{U}$ is a map $V^{\varepsilon}(x,U) = \langle x, U\rangle - \varepsilon\rho(x), \quad 0 < \varepsilon < \varepsilon_0$, such that:

(i) $\rho : X \to \mathbb{R}$ is a $\mathscr{C}^1$ function with $\|\rho\| \leq 1$,

(ii) $\text{argmax}_{x\in X} V^{\varepsilon}(.,U)$ reduces to one point and defines a continuous map $\mathbf{br}^{\varepsilon} : \mathscr{U} \to X$, called a smooth best reply function,

(iii) Let

$$W^{\varepsilon}(U) = \max_x V^{\varepsilon}(x,U) = V^{\varepsilon}(\mathbf{br}^{\varepsilon}(U),U).$$

then:

$$DW^{\varepsilon}(U) = \mathbf{br}^{\varepsilon}(U).$$

(which is a version of the enveloppe theorem)

A typical example is obtained via the entropy function on $X = \Delta(K)$:

$$\rho(x) = \sum_k x^k \log x^k. \tag{7}$$

which leads to:

$$[\mathbf{br}^{\varepsilon}(U)]^k = \frac{\exp(U^k/\varepsilon)}{\sum_{j \in K} \exp(U^j/\varepsilon)}. \tag{8}$$

Definition

A smooth fictitious play strategy σ^ε is defined by

$$\sigma^\varepsilon(h_n) = \mathbf{br}^\varepsilon(\overline{U}_n).$$

The corresponding discrete dynamics written in the spaces of both vectors and payoffs is:

$$\overline{U}_{n+1} - \overline{U}_n = \frac{1}{n+1}[U_{n+1} - \overline{U}_n]. \tag{9}$$

$$\overline{\omega}_{n+1} - \overline{\omega}_n = \frac{1}{n+1}[\omega_{n+1} - \overline{\omega}_n]. \tag{10}$$

with

$$\mathsf{E}(\omega_{n+1}|h_n) = \langle \mathbf{br}^\varepsilon(\overline{U}_n), U_{n+1}\rangle. \tag{11}$$

External consistency is as usual:

$$\frac{1}{n}\sum_{m=1}^{n} U_m^k - \omega_m \le o(n), \qquad \forall k \in K$$

or

$$\langle x, \bar{U}_n\rangle - \bar{\omega}_n \le o(n), \qquad \forall x \in \Delta(K)$$

Fudenberg and Levine (1995) [54]

Theorem

For any $\eta > 0$, there exists $\bar{\varepsilon}$ such that for $\varepsilon \leq \bar{\varepsilon}$, SFP$(\varepsilon)$ is η-consistent.

A similar result holds for conditional smooth fictitious play generating an internal no-regret procedure, Fudenberg and Levine(1999), [56].

Stochastic approximation

We summarize here results from Benaïm, Hofbauer and Sorin (2005) [19], (2006) [20], following the approach for ODE by Benaïm (1996) [15], (1999) [16], Benaïm and Hirsch (1996).

Consider a random discrete process defined on a compact subset of $\mathbb{R}^K$ and satisfying the differential inclusion :

$$Y_n - Y_{n-1} \in a_n[T(Y_{n-1}) + W_n]$$

where
i) T is an u.s.c. correspondence with compact convex values
ii) $a_n \geq 0$, $\sum_n a_n = +\infty$, $\sum_n a_n^2 < +\infty$
iii) $E(W_n | Y_1, ..., Y_{n-1}) = 0$.

Theorem

The set of accumulation points of $\{Y_n\}$ is almost surely a compact set, invariant and attractor free for the dynamical system defined by the differential inclusion:

$$\dot{Y} \in T(Y).$$

One says that the process $\{Y_n\}$ is a Discrete Stochastic Approximation of the differential inclusion.

A typical application is the case where:

$$Y_n - Y_{n-1} = a_n Z_n$$

with Z_n random, satisfying $E[Z_n|Y_1,...,Y_{n-1}] \in \mathbf{T}(Y_{n-1})$ where one writes:

$$Y_n - Y_{n-1} = a_n[E[Z_n|Y_1,...,Y_{n-1}] + Z_n - E[Z_n|Y_1,...,Y_{n-1}])]$$

$$\in a_n[\mathbf{T}(Y_{n-1}) + W_n]$$

i)Back to SFP

Lemma
The process $(\overline{U}_n, \overline{\omega}_n)$ is a Discrete Stochastic Approximation for the differential inclusion:

$$(\dot{\mathbf{u}}, \dot{\omega}) \in \{(U, \langle \mathbf{br}^{\varepsilon}(\mathbf{u}), U\rangle) - (\mathbf{u}, \omega); U \in \mathscr{U}\}. \tag{12}$$

The main property of the continuous dynamics is given by:
Theorem
The set $\{(u,\omega) \in \mathscr{U} \times \mathbb{R} : W^{\varepsilon}(u) - \omega \leq \varepsilon\}$ is a global attracting set for the continuous dynamics.
In particular, for any $\eta > 0$, there exists $\bar{\varepsilon}$ such that for $\varepsilon \leq \bar{\varepsilon}$, $\limsup_{t\to\infty} W^{\varepsilon}(\mathbf{u}(t)) - \omega(t) \leq \eta$ (i.e. continuous SFP(ε) satisfies η-consistency).
Proof
Let $q(t) = W^{\varepsilon}(\mathbf{u}(t)) - \omega(t)$.
Taking time derivative one obtains:

$$\begin{aligned}
\dot{q}(t) &= DW^{\varepsilon}(\mathbf{u}(t)).\dot{\mathbf{u}}(t) - \dot{\omega}(t) \\
&= \langle \mathbf{br}^{\varepsilon}(\mathbf{u}(t)), \dot{\mathbf{u}}(t)\rangle - \dot{\omega}(t) \\
&= \langle \mathbf{br}^{\varepsilon}(\mathbf{u}(t)), U - \mathbf{u}(t)\rangle - (\langle \mathbf{br}^{\varepsilon}(\mathbf{u}(t)), U\rangle - \omega(t)) \\
&= -(\langle \mathbf{br}^{\varepsilon}(\mathbf{u}(t)), \mathbf{u}(t)\rangle - \omega(t)) \\
&\leq -q(t) + \varepsilon.
\end{aligned}$$

so that $q(t) \leq \varepsilon + Me^{-t}$ for some constant M and the result follows. ∎
Theorem
For any $\eta > 0$, there exists $\bar{\varepsilon}$ such that for $\varepsilon \leq \bar{\varepsilon}$, SFP$(\varepsilon)$ is η- consistent.
Proof
The assertion follows from the previous result and the DSA property. ∎

A similar result holds for conditional smooth fictitious play generating an internal no-regret procedure.
Recent advances: Benaim and Faure (2013) [17] obtain consistency with vanishing perturbation $\varepsilon_n = n^{-a}, a < 1$.

Process non longer autonomous.

ii) Approachability

In the framework of Blackwell's theorem (convex case) one has:

$$\bar{g}_{n+1} - \bar{g}_n = \frac{1}{n}[g_{n+1} - \bar{g}_n]$$

with

$$\langle \bar{g}_n - \Pi_C(\bar{g}_n), E(g_{n+1}|h_n) - \Pi_C(\bar{g}_n)\rangle \leq 0$$

which is a DSA of

$$\dot{z} \in N(z) - z$$

with $N(z) \subset \{v; \langle z - \Pi_C(z), v - \Pi_C(z)\rangle \leq 0\}$.
Let $P(z) = d^2(z, C) = \|z - \Pi_C(z)\|^2$ and $Q(t) = P(z_t)$.
One has $\nabla P(z) = 2[z - \Pi_C(z)]$. Then:

$$\dot{Q}(t) = \langle \nabla P(z_t), \dot{z}_t\rangle \in 2\langle z_t - \Pi_C(z_t), v_t - \Pi_C(z_t) + \Pi_C(z_t) - z_t\rangle$$

$$\leq -2Q(t).$$

So that $Q(t)$ decreases exponentially to 0 hence C is a global attractor and the limit points of the DSA $\{\bar{g}_n\}$ are in C. The same proof works for any $\mathscr{C}^1$ function P non negative with $P^{-1}(0) = C$ and satisfying for some $a > 0$:

$$N(z) \subset \{v; \langle \nabla P(z), v - z\rangle \leq -aP(z)\}$$

Finally if one has, when $z \notin C$

$$N(z) \subset \{v; \langle \nabla P(z), v - z\rangle < 0\}$$

P is a strong Lyapounov function related to C for the differential inclusion and any bounded DSA of the differential inclusion

converges to C.

iii) Regret In this framework the same function can be used to define the distance and the strategy.

Definition

P is a potential function for $D = \mathbb{R}^K_-$ if
(i) P is $\mathscr{C}^1$ from $\mathbb{R}^K$ to $\mathbb{R}^+$
(ii) $P(w) = 0$ iff $w \in D$
(iii) $\nabla P(w) \in \mathbb{R}^K_+$
(iv) $\langle \nabla P(w), w \rangle > 0, \forall w \notin D$.

Compare Hart and Mas Colell (2001, 2003) [63], [64].

Example: $P(w) = \sum_k([w^k]^+)^2 = d(w,D)^2$.

1. External regret

Given a potential P for $D = \mathbb{R}^K_-$, the P-regret-based discrete procedure for player 1 is defined by

$$\sigma(h_n) \div \nabla P(\overline{R}_n) \qquad \text{if} \quad \overline{R}_n \notin D \tag{13}$$

and arbitrarily otherwise.

Discrete dynamics associated to the average regret:

$$\overline{R}_{n+1} - \overline{R}_n = \frac{1}{n+1}(R_{n+1} - \overline{R}_n)$$

By the choice of σ

$$\langle \nabla P(\overline{R}_n), \mathsf{E}(R_{n+1}|h_n) \rangle = 0.$$

(recall $\langle x, \mathsf{E}_x(R(., U)) \rangle = 0.$)
The continuous time version is expressed by the following differential inclusion in $\mathbb{R}^m$:

$$\dot{\mathbf{w}} \in N(\mathbf{w}) - \mathbf{w} \tag{14}$$

where N is a correspondence that satisfies

$$\langle \nabla P(\mathbf{w}), \mathbf{N}(\mathbf{w}) \rangle = \mathbf{0}.$$

Theorem

The potential P is a Lyapounov function associated to $D = \mathbb{R}^K_-$. Hence, D contains a global attractor attractor.

Proof
For any solution $\mathbf{w}$, if $\mathbf{w}(t) \notin D$ then

$$\frac{d}{dt}P(\mathbf{w}(t)) = \langle \nabla P(\mathbf{w}(t)), \dot{\mathbf{w}}(t) \rangle$$

$$\in \langle \nabla P(\mathbf{w}(t)), N(\mathbf{w}(t)) - \mathbf{w}(t) \rangle = -\langle \nabla P(\mathbf{w}(t)), \mathbf{w}(t) \rangle < 0$$

∎

Corollary

Any P-regret-based discrete dynamics satisfies internal consistency.

Proof
$D = \mathbb{R}^K_-$ contains an attractor whose basin of attraction contains the range $\mathscr{R}$ of R and the discrete process for $\bar{R}_n$ is a bounded DSA. ∎

2. Internal regret

Given a potential Q for $M = \mathbb{R}_{-}^{K^2}$, a Q-regret-based discrete procedure for player 1 is a strategy σ satisfying

$$\sigma(h_n) \in Inv[\nabla Q(\bar{S}_n)] \qquad \text{if} \qquad \bar{S}_n \notin M \tag{15}$$

and arbitrarily otherwise.
The discrete process of internal regret matrices is:

$$\bar{S}_{n+1} - \bar{S}_n = \frac{1}{n+1}[S_{n+1} - \bar{S}_n]. \tag{16}$$

with the property:

$$\langle \nabla Q(\bar{S}_n), \mathsf{E}(S_{n+1}|h_n)\rangle = 0.$$

(Recall $\langle A, \mathsf{E}_{\mu}(S(.,U))\rangle = 0$.)
Corresponding continuous procedure with $w \in \mathbb{R}^{K^2}$

$$\dot{\mathbf{w}}(t) \in N(\mathbf{w}(t)) - \mathbf{w}(t) \tag{17}$$

and

$$\langle \nabla Q(w), N(w)\rangle = 0.$$

The previous continuous time process satisfy:

$$\mathbf{w}^{+}_{k\ell}(t) \to_{t\to\infty} 0.$$

Corollary
The discrete process (16) satisfy:

$$[\bar{S}^{k\ell}_n]^{+} \to_{t\to\infty} 0 \qquad a.s.$$

hence conditional consistency (internal no regret) holds.

Multiplicative Weight Algorithm: discrete and continuous time

Evolution of a single population with K types modelized through a symmetric 2 person game with $K \times K$ payoff (fitness) matrix A

A_{ij} is the payoff of "i" facing "j".

x_t^k : frequency of type k at time t.

Replicator equation on the simplex $\Delta(K)$ of $\mathbb{R}^K$

$$\dot{x}_t^k = x_t^k \left(e^k A x_t - x_t A x_t \right), \quad k \in K \qquad (RD) \tag{18}$$

Taylor and Jonker (1978) [155]

Replicator dynamics for I populations

$$\dot{x}_t^{ip} = x_t^{ip}[F^i(e^{ip}, x_t^{-i}) - F^i(x_t^i, x_t^{-i})], \qquad p \in S^i, i \in I$$

natural interpretation: $x_t^i = \{x_t^{ip}, p \in S^i\}$, is a mixed strategy of player i.

The model is in the framework of an I-person game but we consider the dynamics for one player, without hypotheses on the behavior of the others.

Hence, from the point of view of this player, he is facing a (measurable) vector outcome process $\{U_t, t \geq 0\}$, with values in the cube $\mathscr{U} = [-1, 1]^K$ where K is his move's set.

U_t^k is the payoff at time t if k is the choice at that time.

The $\mathscr{U}$-replicator process (RP) is specified by the following equation on $\Delta(K)$:

$$\dot{x}_t^k = x_t^k[U_t^k - \langle x_t, U_t \rangle], \qquad k \in K. \tag{19}$$

The logit map L from $\mathbb{R}^K$ to $\Delta(K)$ is defined by:

$$L^k(V) = \frac{\exp V^k}{\sum_j \exp V^j}. \tag{20}$$

Recall that L satisfies

Proposition

$$L(V) = argmax_{\Delta(K)}[\ \langle V, x\rangle - \rho(x)].$$

The following procedure has been introduced in discrete time in the framework of on-line algorithms under the name “multiplicative weight algorithm", Vovk, 1990 [159], Littlestone and Warmuth (1994) [94].

$$x_{n+1} = L(\sum_{m=1}^{n} U_m)$$

Define the continuous exponential weight process (CEW), Sorin (2009) [144] on $\Delta(K)$ by:

$$x_t = L(\int_0^t U_s ds).$$

(CEW) provides an explicit solution of (RP), Rustichini (1999) [135], Hofbauer, Sorin and Viossat (2009) [80]

Proposition

(CEW) *satisfies* (RP).

Hofbauer, Sorin and Viossat (2009) [80]

Proposition

(CEW) or (RP) satisfies external consistency.

Proof

By integrating:

$$\frac{\dot{x}_t^k}{x_t^k} = [U_t^k - \langle x_t, U_t \rangle], \qquad k \in K. \tag{21}$$

one obtains, on the support of x_0:

$$\int_0^t [U_s^k - \langle x_s, U_s \rangle] ds = \int_0^t \frac{\dot{x}_s^k}{x_s^k} ds = \log(\frac{x_t^k}{x_0^k}) \leq -\log x_0^k.$$

∎

Back to a game framework this implies that if player 1 follows (RP) the set of accumulation points of the corrrelated distribution induced by the empirical process of moves will belong to his Hannan set H^1.

The example due to Viossat (2007) [157] of a game where the limit set for the replicator dynamics is disjoint from the unique correlated equilibrium shows that (RP) does not satisfy internal consistency.

To obtain similar results in discrete time one can starting from a discrete process construct a continuous time interpolation then use an adapted consistent procedure. It remains to describe a discretization and to evaluate the error terms.
This was done in Sorin (2009) [144] for the entropy on the simplex.
A much more general result will be described in section 3.
The result implies that for

$$x_{m+1} = L(a \sum_{s=1}^{m} U_s)$$

the n stage regret satisfies

$$\sum_{s=1}^{n} U_s^k - \langle x_s, U_s \rangle \leq 0(\sqrt{n})$$

which match the initial result of Auer et alii (1995) (2022) [8] using the discrete MWA.

Hence the general picture is (ρ is the entropy function)

$argmax\langle\sum_{m=1}^{n} U_m, x\rangle$ or $argmax\langle\bar{U}_n, x\rangle$ gives fictitious play
$argmax\langle\bar{U}_n, x\rangle - \varepsilon\rho(x)$ gives approximate consistency
$argmax\langle\bar{U}_n, x\rangle - (1/\sqrt{n})\rho(x)$ corresponds to a vanishing perturbation.

In continuous time one can take $argmax\langle\bar{U}_t, x\rangle - (1/t)\rho(x)$

No convergence to Nash

There is no uncoupled deterministic smooth dynamic that converges to Nash equilibrium in all finite 2-person games: Hart and Mas-Colell (2003).
Similarly there are no learning process with finite memory such that the stage behavior will converge to Nash equilibrium: Hart and Mas-Colell (2005).
Similar results were obtained for MAD dynamics, Hofbauer and Swinkels (1995)
see also Foster and Young (2001) On the impossibility of predicting.

Young (2002) On the limits to rational learning .

Part II: No-regret (II)

General framework

V normed vector space, finite dimensional,
dual V^* and duality map $\langle .|.\rangle$,
$X \subset V$ compact convex.

Consider algorithms that associate to a trajectory of parameters $\{u_t \in V^*, t \geq 0\}$ in the dual space, a process of actions/controls $\{x_t \in X, t \geq 0\}$ in the primal space, where x_t depends on $\{(x_s, u_s), 0 \leq s < t\}$.

$$R_t(y) = \int_0^t \langle u_s | y - x_s \rangle ds, \qquad t \geq 0, \quad y \in X \tag{22}$$

or in discrete time:

$$R_n(y) = \sum_{m=1}^{n} \langle u_m | y - x_m \rangle, \qquad y \in X. \tag{23}$$

The procedure satisfies the no-regret property if:

$$R_t(y) \leq o(t), \qquad \forall y \in X, \tag{24}$$

or

$$R_n(y) \leq o(n), \qquad \forall y \in X. \tag{25}$$

A) We compare the performance of the algorithms in terms of regret under three (increasing) assumptions:
(I) general case: $\{u_t\}$ is a bounded measurable process in V^*,
(II) closed form: $u_t = \phi(x_t)$ for a continuous vector field $\phi : X \to V^*$,
(III) convex gradient: $u_t = -\nabla f(x_t), f : X \to \mathbb{R}$, $\mathscr{C}^1$ convex function (with similar properties in discrete time).

B) We consider three different procedures:
a) Projected dynamics (PD),
b) Mirror descent (MD),
c) Dual averaging (DA).

C) We analyze the relations between the continuous and discrete time processes, in particular in terms of speed of convergence to 0 of the average regret.

D) We also study the convergence of the trajectories of $\{x_t\}$ or $\{x_n\}$ (in classes (II) and (III)).

Framework (I) corresponds to the usual model of on-line learning where the agent observes $\{u_s, s < t\}$ and chooses x_t.

The next two frameworks (II) and (III), describe more specific cases where the parameter u_t is a function of the action x_t.

Famework (II), *closed form*, is relevant for game dynamics and variational inequalities.

Consider a strategic game $\Gamma(\phi)$ with a finite set of players I, where the equilibrium set E is given by the solutions $x \in X$ of the following variational inequalities:

$$\langle \phi^i(x) | x^i - y^i \rangle \geq 0, \qquad \forall y^i \in X^i, \forall i \in I.$$

Here $X^i \subset V^i$ is the strategy set of player $i \in I$, $X = \prod_i X^i$, and $\phi^i : X \to V^{i*}$ is her evaluation function.

Examples include:
- finite games (with mixed extension): ϕ^i is the vector payoff VG^i.
- continuous games with payoff G^i, $\mathscr{C}^1$ and concave wrt x^i, $\forall i \in I$ then ϕ^i is the gradient of G^i w.r.t. x^i.
- population games (Wardrop equilibria), X^i is the simplex $\Delta(S^i)$ and ϕ^i corresponds to the outcome function $F^i : S^i \times X \longrightarrow \mathbb{R}$.

For each player i, the reference process is $u_t^i = \phi^i(x_t)$ which, as a function of x_t, is determined by the behavior of all players. Hence the overall global dynamics of $\{x_t\}$ is generated by a family of unilateral procedures since for each i, x_t^i depends on (u^i, x^i) only.
In particular for each player i, the knowledge of $\phi^j, j \neq i$, is not assumed.

Thus for each player individually the situation is like *general case* (I), while the private parameters of the players are linked via x_t.

We will analyze the consequences on the process $\{x_t\}$, assuming only that each player uses a procedure satisfying the no-regret condition.

Obviously the (global) algorithm associated to the global parameter $\phi = \{\phi^i\}$ will also share the no-regret property since:

$$\int_0^t \langle \phi^i(x_s) | x^i - x_s^i \rangle ds \leq o(t), \qquad \forall x^i \in X^i, \qquad \forall i \in I,$$

implies:

$$\int_0^t \langle \phi(x_s) | x - x_s \rangle ds \leq o(t), \qquad \forall x \in X.$$

But in addition it is decentralized in the sense that x^i depends upon ϕ^i only.

Framework (III) covers the case of convex optimization where the parameter, after the choice x_t, is the gradient of the (unknown) convex function and $u_t = -\nabla f(x_t)$.

The research in this area is extremely active and very diverse; it links basic optimization algorithms, Polyak, 1987 [127], Nemirovski and Yudin, 1983 [110], Nesterov, 2004 [113], to on-line procedures, see e.g. Zinkevich, 2003 [165].

Recent books and lecture notes include:

Bubeck S. (2011) *Introduction to online optimization*, Lecture Notes.
Bubeck S. (2015) Convex optimization: Algorithms and complexity, *Fondations and Trends in Machine Learning*, **8**, 231-357.
Hazan E. (2011)The convex optimization approach to regret minimization, *Optimization for machine learning*, S. Sra, S. Nowozin, S. Wright eds, MIT Press, 287-303.
Hazan E. (2015) Introduction to Online Convex Optimization, *Fondations and Trends in Optimization*, **2**, 157-325.
Hazan E. (2019) Optimization for Machine Learning , https://arxiv.org/pdf/1909.03550.pdf.
Rakhlin A. (2009) *Lecture notes on on-line learning*.
Shalev-Shwartz S. (2012) Online Learning and Online Convex Optimization, *Foundations and Trends in Machine Learning*, **4**, 107-194.

Related algorithms have also been developped in Operations Research (transportation, networks), see e.g. Dupuis and Nagurney, 1993 [44], Nagurney and Zhang, 1996 [107], Smith, 1984 [143].

Note that each community (learning, game theory, optimization) has its own terminology and point of view.

2. Closed field, continuous time

Definitions and notations

We describe here some relations with variational inequalities when the parameter process has a *closed form*: $u = \phi(x)$.
$NE(\phi)$ is the set of (internal) solutions, in X, of the variational inequality:

$$\langle \phi(x) | y - x \rangle \leq 0, \qquad \forall y \in X. \tag{26}$$

a) If ϕ is the evaluation function in a game $\Gamma(\phi)$, $NE(\phi)$ corresponds to the set of equilibria.

b) The minimization of a $\mathscr{C}^1$ convex function f on X corresponds to the variational inequality (26) with $\phi = -\nabla f$.
This case presents two properties:
ϕ is dissipative,
ϕ is a gradient.
The general definitions are as follows.

Definition

$\phi : X \to V^*$ is dissipative if it satisfies:

$$\langle \phi(x) - \phi(y) | x - y \rangle \leq 0, \qquad \forall x, y \in X. \tag{27}$$

A game $\Gamma(\phi)$ is dissipative if ϕ is dissipative.
This notion is related to the monotonicity requirement in Rosen (1965) [134].
The terminology is "stable" in Hofbauer and Sandholm (2009) [76], "contractive" in Sandholm (2015) [139] and "dissipative" in Sorin and Wan, 2016 [149].

$SE(\phi)$ is the set of (external) solutions, in X, of the variational inequality:

$$\langle \phi(y) | y - x \rangle \leq 0, \qquad \forall y \in X. \tag{28}$$

Note that $SE(\phi)$ is convex.

Recall, see Minty, 1967 [103], that if ϕ is dissipative, then :

$$NE(\phi) \subset SE(\phi) \neq \emptyset$$

and if ϕ is continuous the reverse inclusion is satisfied:

$$SE(\phi) \subset NE(\phi) \neq \emptyset.$$

If $NE(\phi) = SE(\phi)$ we will also use the notation $E(\phi) = E$ for this set.

Fundamental example: 0-sum game

If $F : X = X^1 \times X^2 \to \mathbb{R}$ is $\mathscr{C}^1$ and concave/convex, the vector field $\phi = (\nabla^1 F, -\nabla^2 F)$ is dissipative, Rockafellar (1970) [133]. The elements of $NE(\phi) = SE(\phi) = E$ are optimal strategies of the associated 0-sum game.

We define a potential for a vector field, see e.g. Sorin and Wan (2016) [149].

Definition

A real function W of class $\mathscr{C}^1$ on $X = \prod_i X^i$, is a potential for ϕ if there exist strictly positive functions μ^i on X, $i \in I$, such that:

$$\left\langle \nabla^i W(x) - \mu^i(x)\phi^i(x), y^i - x^i \right\rangle = 0, \quad \forall x \in X, \forall y^i \in X^i, \forall i \in I. \tag{29}$$

where $\nabla^i W$ denotes the gradient w.r.t. x^i.
A game $\Gamma(\phi)$ corresponding to such ϕ is a potential game.

Alternative previous definitions include:
Monderer and Shapley [105] for finite games,
Sandholm [137] for population games.

The following result is classical, see e.g. Sandholm (2001) [137].

Proposition

Let ϕ be a vector field with potential Φ.
1. Every local maximum of Φ belongs to $NE(\phi)$.
2. If Φ is concave on X, then any element in $NE(\phi)$ is a global maximum of Φ on X.

Results

Assume that the procedure satisfies the no-regret property:

$$R_t(y) \leq o(t), \qquad \forall y \in X,$$

where:

$$R_t(y) = \int_0^t \langle \phi(x_s) | y - x_s \rangle ds, \qquad t \geq 0, y \in X.$$

A first property deals with convergent trajectories $\{x_t\}$.

Lemma

If ϕ is continuous and $x_s \to x$, then $x \in NE(\phi)$.

Proof:

Since $R_t(y) = \int_0^t \langle \phi(x_s) | y - x_s \rangle ds$:

$$\frac{R_t(y)}{t} \to \langle \phi(x) | y - x \rangle, \qquad \forall y \in X. \tag{30}$$

and $R_t(y) \leq o(t)$ implies $x \in NE(\phi)$. ■

In particular, if x is a stationary point for the discrete or continuous time procedure, then $x \in NE(\phi)$.

Define the time average trajectories :

$$\bar{x}_t = \frac{1}{t}\int_0^t x_s ds \qquad \text{and} \qquad \bar{x}_n = \frac{1}{n}\sum_{m=1}^{n} x_m.$$

Lemma

If ϕ is dissipative, the accumulation points of $\{\bar{x}_t\}$ or $\{\bar{x}_n\}$ are in $SE(\phi)$.

Proof:

$$\frac{R_t(y)}{t} = \frac{1}{t}\int_0^t \langle \phi(x_s) | y - x_s \rangle \geq \frac{1}{t}\int_0^t \langle \phi(y) | y - x_s \rangle = \langle \phi(y) | y - \bar{x}_t \rangle.$$

Hence under the no-regret condition any accumulation point $\hat{x}$ of $\{\bar{x}_t\}$ will satisfy $\langle \phi(y) | y - \hat{x} \rangle \leq 0$. ∎

This result implies the non-emptiness of $SE(\phi)$ for dissipative ϕ. In particular the minmax theorem (in the $\mathscr{C}^1$ case) follows from the existence of no-regret procedures.

Class (III): convex gradient.

Since $u_t = -\nabla f(x_t)$ with f $\mathscr{C}^1$ convex, this corresponds to a specific case of dissipative and continuous vector field ϕ, hence: $SE(\phi) = NE(\phi) = E = argmin_X f$.
Use the basic convexity property:

$$\langle \nabla f(x_t) | y - x_t \rangle \leq f(y) - f(x_t)$$

to obtain with $u_t = -\nabla f(x_t)$ in the definition of the regret $R_t(y)$:

$$\int_0^t [f(x_s) - f(y)]\, ds \leq \int_0^t \langle -\nabla f(x_s) | y - x_s \rangle ds = R_t(y)$$

which implies by Jensen's inequality:

$$f(\bar{x}_t) - f(y) \leq \frac{1}{t} \int_0^t [f(x_s) - f(y)]\, ds \leq \frac{R_t(y)}{t}. \tag{31}$$

In particular one obtains:

Lemma

i) The accumulation points of $\{\bar{x}_t\}$ *or* $\{\bar{x}_n\}$ *belong to* E.
ii) If $t \mapsto f(x_t)$ *(resp.* $n \mapsto f(x_n)$*) is decreasing, the accumulation points of* $\{x_t\}$ *or* $\{x_n\}$ *belong to* E.

Continuous time

A very useful tool is available in this set-up.

Level functions

Definition

$P:\mathbb{R}^+\times X\to\mathbb{R}^+$ is a level function (for $\{u_t,x_t\}$) if:

$$\langle u_t, x_t - y\rangle \geq \frac{d}{dt}P(t;y), \quad \forall t\in\mathbb{R}^+, \forall y\in X. \tag{32}$$

Proposition

$R_t(y)=\int_0^t\langle u_s|y-x_s\rangle ds \leq P(0;y)-P(t;y)$ *is bounded.*

(1) no-regret property: Rate of convergence $1/t$*.*

(2) Class (II): Assume $y^*\in SE(\phi)$*, then* $P(t;y^*)$ *is decreasing:*

$$\frac{d}{dt}P(t;y^*)\leq\langle\phi(x_t),x_t-y^*\rangle\leq 0.$$

Positive correlation

Given a dynamics $\dot{x}_t = D(x_t)$, f decreases on trajectories if:

$$\frac{d}{dt} f(x_t) = \langle \nabla f(x_t) | \dot{x}_t \rangle \leq 0.$$

The analogous property for a vector field ϕ is:

$$\langle \phi(x_t) | \dot{x}_t \rangle \geq 0.$$

In the framework of games, a similar condition was described in discrete time as Myopic Adjustment Dynamics, Swinkels (1993) [154] : if $x^i_{n+1} \neq x^i_n$ then $G^i(x^i_{n+1}, x^{-i}_n) > G^i(x^i_n, x^{-i}_n)$.

The corresponding notion in continuous time is positive correlation, (between the dynamics and the vector field), Sandholm (2010) [138]:

$$\dot{x}^i_t \neq 0 \Longrightarrow \langle \phi^i(x_t), \dot{x}^i_t \rangle > 0.$$

Proposition

Consider a vector field ϕ with potential Φ.
If the dynamics satisfies positive correlation, then Φ is a strict Lyapunov function.
All ω-limit points are rest points.

This result is proved by Sandholm (2001) [137] for his version of potential population game, see extensions in Benaim, Hofbauer and Sorin (2005) [19].
A similar property for fictitious play in discrete time is established in Monderer and Shapley (1996) [105].

We will show that this property holds for the three dynamics defined below.

3. GD, MD and DA

We first describe in this subsection three procedures in continuous time that satisfy the no-regret property. Their discrete time counterparts will be analyzed in the next subsection.

As usual, discrete time dynamics are easier to describe but their mathematical properties are more difficult to establish. This explain why we choose to start with the continuous time versions.

We now introduce and study three dynamics:

- Projected dynamics (PD),
- Mirror descent (MD),
- Dual averaging (DA).

Continuous time

In each case we first define the dynamics, then control the values of the regret by exhibiting a level function and finally study the trajectories for class (II) and (III).

Hilbertian framework: Projected Dynamics

V Hilbert, $X \subset V$, convex closed.

Dynamics

analogous to projected gradient descent (Levitin and Polyak, 1966) and defined, as projected dynamics (PD), by $x_t \in X$ with:

$$\langle u_t - \dot{x}_t, y - x_t \rangle \leq 0, \forall y \in X. \tag{33}$$

which is:

$$\dot{x}_t = \Pi_{T_X(x_t)}(u_t). \tag{34}$$

since $T_C(x)$ is a cône (where Π_C is the projection on the closed convex set C and $T_C(x)$ is the tangent cône to C at x).

Values

Proposition

$$V(t;y) = \frac{1}{2}\|x_t - y\|^2, \quad y \in X, \tag{35}$$

is a level function.

Proof:
(33) gives:

$$\langle u_t, y - x_t \rangle \leq \langle \dot{x}_t, y - x_t \rangle = -\frac{d}{dt} V(t;y).$$

■

Trajectories

Proposition

Assume ϕ dissipative and $E \neq \emptyset$.
$\{\bar{x}_t\}$ converges weakly to a point in E.

Proof:

- $\{\bar{x}_t\}$ is bounded hence has weak accumulation points.
- The weak limit points of $\{\bar{x}_t\}$ are in E
- $\|x_t - y^*\|$ converges when $y^* \in E$

Hence by Opial's lemma [118], $\bar{x}_t$ converges weakly to a point in E. ∎

Lemma

Positive correlation holds.

Proof:

$$\langle \phi(x_t), \dot{x}_t \rangle = \|\dot{x}_t\|^2$$

since $\langle u_t - \dot{x}_t, \dot{x}_t \rangle = 0$ by (33) and Moreau's decomposition, Moreau, 1965 [106]. ■

Consider class (III): $u_t = -\nabla f(x_t)$.

Proposition

$f(x_t)$ *is decreasing and converges to* $f^* = \min_X f$ *at speed* $1/t$
Assume $E \neq \emptyset$*.* $\{x_t\}$ *weakly converges to a point in* E*.*

Proof:

Weak accumulation points of $\{x_t\}$ are in E.
Then Opial's lemma applies. ■

Mirror descent

Continuous version of “Mirror descent algorithm”,
Nemirovski and Yudin (1983), [110], Beck and Teboulle (2003). [14]

Dynamics H strictly convex, $\mathscr{C}^2$,

X, compact, convex $\subset dom\ H$.
The continuous time process mirror descent (MD) satisfies, $x_t \in X$ and:

$$\langle u_t - \frac{d}{dt}\nabla H(x_t) | y - x_t\rangle \leq 0, \forall y \in X. \tag{36}$$

The previous analysis corresponds to the case: $H(x) = \frac{1}{2}\|x\|^2$.

Values

Bregman distance associated to H:

$$D_H(y,x) = H(y) - H(x) - \langle \nabla H(x) | y - x \rangle \; (\geq 0).$$

$$\frac{d}{dt} D_H(y,x_t) = \langle -\frac{d}{dt} \nabla H(x_t) | y - x_t \rangle, \tag{37}$$

so that (36) implies:

$$\langle u_t | y - x_t \rangle \leq -\frac{d}{dt} D_H(y,x_t).$$

Proposition

$P(t;y) = D_H(y,x_t)$ *is a level function.*

Trajectories

The use of special functions H adapted to X allows to produce a trajectory that remains in $int X$ hence to get rid of the normal cône .
This leads to:

$$\frac{d}{dt}\nabla H(x_t) = u_t \tag{38}$$

$$\dot{x}_t = \nabla^2 H(x_t)^{-1} u_t. \tag{39}$$

which corresponds to a Riemannian metric, see Bolte and Teboulle, 2003 [25], Alvarez, Bolte and Brahic, 2004 [2], Mertikopoulos and Sandholm, 2018 [101].

Lemma

Positive correlation holds.

Proof :

$$\langle \phi(x_t) | \dot{x}_t \rangle = \langle \phi(x_t) | \nabla^2 H(x_t)^{-1} \phi(x_t) \rangle \geq 0.$$

Consider now class (III).

By Lemma 33 the accumulation points of $\{x_t\}$ are in E.

To prove convergence one introduces the following :

Hypothesis [H1]: if $z^k \to y^* \in S$ then $D_H(y^*, z^k) \to 0$.
For example H is L-smooth (see e.g. Nesterov, 2004 [113] Section 1.2.2.) and then:

$$0 \leq D_H(x,y) \leq \frac{L}{2}\|x-y\|^2.$$

Hypothesis [H2]: if $D_H(y^*, z^k) \to 0, y^* \in S$ then $z^k \to y^*$.
For example H is β-strongly convex (see e.g. Nesterov, 2004 [113] Section 2.1.3.) and then:

$$D_H(x,y) \geq \frac{\beta}{2}\|x-y\|^2.$$

Proposition

Consider class (III). If H is smooth and strongly convex, $\{x_t\}$ converges weakly to some $x^* \in E$.

Proof:

Let x^* be an accumulation point of $\{x_t\}$. Then $x^* \in E$ by Lemma 33 and thus $D_H(x^*, x_t)$ is decreasing. Since this sequence is decreasing to 0 on a subsequence $x_{t_k} \to x^*$ by [H1], it is decreasing to 0, hence by [H2] $x_t \to x^*$. ∎

Dual averaging

Continuous version of dual averaging Nesterov, 2009 [114].
We follow Kwon and Mertikopoulos, 2017 [89].

Dynamics

Assume h bounded strictly convex s.c.i. with $dom\, h = X \subset V$ convex compact.
Let $h^*(w) = \sup_{x \in V} \langle w|x \rangle - h(x)$ be the Fenchel conjugate. h^* is differentiable.
Introduce :

$$U_t = \int_0^t u_s ds$$

and let the dual averaging (DA) dynamics be defined by:

$$x_t = argmax\{\langle U_t|x \rangle - h(x); x \in V\} = argmax\{\langle U_t|x \rangle - h(x); x \in X\}.$$

The dynamics can be written as:

$$x_t = \nabla h^*(U_t) \in X \qquad (40)$$

Values

Consider the Fenchel coupling:

$$W(t;y) = h^*(U_t) + h(y) - \langle U_t | y \rangle \qquad (\geq 0). \tag{41}$$

$$\frac{d}{dt} h^*(U_t) = \langle u_t | \nabla h^*(U_t) \rangle = \langle u_t | x_t \rangle \tag{42}$$

thus:

$$\frac{d}{dt} W(t;y) = \langle u_t | x_t - y \rangle$$

Proposition

W is a level function.

There is an important literature on continuous time dynamics with similar features, see e.g. :
- in convex optimization: Attouch and Teboulle, 2004 [3], Attouch, Bolte, Redont and Teboulle, 2004 [4], Auslender and Teboulle, 2006 [10], 2009 [11]... Teboulle, 2018 [156],
- in game theory: Hofbauer and Sandholm, 2009 [76], Coucheney, Gaujal and Mertikopoulos, 2015 [39], Mertikopoulos and Sandholm, 2016 [100], Mertikopoulos and Sandholm (2018) [101], Mertikopoulos and Zhou (2019) [102]
...

Discrete time: general case

We consider now discrete time algorithms.

Remark that the dynamics depends on an additional parameter, the step size.

Hilbertian framework: Projected Dynamics

Dynamics

Levitin and Polyak (1966) [93], Polyak (1987) [127], gradient projection method:

$$\begin{aligned} x_{m+1} &= argmin_X\{-\langle u_m, x\rangle + \frac{1}{2\eta_m}\|x - x_m\|^2\}, \\ &= argmax_X\{\langle u_m, x\rangle - \frac{1}{2\eta_m}\|x - x_m\|^2\}, \end{aligned} \quad (43)$$

with η_m decreasing, which corresponds to:

$$x_{m+1} = \Pi_X[x_m + \eta_m u_m], \quad (44)$$

or with variational characterization:

$$\langle x_m + \eta_m u_m - x_{m+1}, y - x_{m+1}\rangle \leq 0, \forall y \in X. \quad (45)$$

Values

Let $m(X)$ be the diameter of X. Assume $\|u_m\| = \|u_m\|_* \le M$.

Proposition

$$R_n(x) \le \frac{1}{2\eta_n} m(X)^2 + \frac{M^2}{2} \sum_{m=1}^{n} \eta_m$$

hence with $\eta_n = 1/\sqrt{n}$:

$$R_n(x) \le O(\sqrt{n}).$$

Trajectories

Lemma

For $x^* \in SE(\phi)$, $\|x_m - x^*\|$ converges if $\eta_n \in \ell^2$.

Proposition

If $\eta_n \in \ell^2$ and g is dissipative, $\{\bar{x}_n\}$ converges to a point in $SE(\phi)$.

Mirror descent

Assumption:
H, L-*strongly convex for some norm* $\|.\|$ *on* $V = \mathbb{R}^n$.
$\|u_n\|_* \leq M$.

Dynamics

Nemirovski and Yudin (1983), [110], Beck and Teboulle (2003) [14].
The mirror descent algorithm is given by :

$$x_{m+1} = argmin_X\{-\langle u_m | \, x\rangle + \frac{1}{\eta_m} D_H(x, x_m)\}, \tag{46}$$

Variational formulation:

$$\langle \nabla H(x_m) + \eta_m u_m - \nabla H(x_{m+1}) | x - x_{m+1}\rangle \leq 0, \forall x \in X. \tag{47}$$

Values

Proposition

$$R_n(x) \le \frac{D_H(x,x_1)}{\eta} + n\eta\frac{M^2}{2L}.$$

Then $\eta = 1/\sqrt{n}$ and $R_n(x) \le O(\sqrt{n})$.

Same property with $\eta_n = 1/\sqrt{n}$ via double trick.

Trajectories

Lemma

For $x^* \in SE(\phi)$, $D_H(x^*,x_n)$ converges if $\{\eta_n\} \in \ell^2$.

Dual averaging

Assumptions:
a) h is a l.s.c. function from V to $\mathbb{R} \cup \{+\infty\}$, L-strongly convex for some norm $\|.\|$ on $V = \mathbb{R}^n$, with $dom\, h = X$.
*b) $\|u_m\|_* \leq M, \forall n \in$ **N**.*

Dynamics

Dual averaging, Nesterov (2009).
Let $U_m = \sum_{k=1}^{m} u_k$
The algorithm is again given by a maximization property:

$$\begin{aligned} x_{m+1} &= argmin_V\{-\langle U_m|x\rangle + (1/\eta_m)h(x)\}, \\ &= argmax_X\{\langle U_m|x\rangle - (1/\eta_m)h(x)\} \end{aligned} \quad (48)$$

which is:

$$x_{m+1} = \nabla h^*(\eta_m U_m).$$

where $\{\eta_m\}$ is decreasing.

Values

Xiao (2010) [162] or discrete approximation of (40), Kwon and Mertikopoulos (2017). [89]:

Proposition

$$R_n(x) = \sum_{m=1}^{n} \langle u_m | x - x_m \rangle \leq \frac{r_X(h)}{\eta_n} + \frac{\sum_{m=1}^{n} \eta_{m-1} \|u_m\|_*^2}{2L}. \qquad (49)$$

Assume: $\|u_m\|_* \leq M$.
Convergence rate $O(\sqrt{n})$ with time varying parameters $\eta_m = 1/\sqrt{m}$.

Comments on the discrete dynamics framework

1) The three algorithms achieve the same bound $O(1/\sqrt{n})$ for the speed of convergence of the average regret, which is optimal already in class (III), Nesterov, 2004 [113] , using time varying step sizes $\eta_n = 1/\sqrt{n}$.

2) More precise properties concerning the trajectories are available only in the (PD) set-up. The results are similar to the ones in the continuous case, Section 3.2, if $\eta_n \in \ell^2$. (Compare to the analysis in Peypouquet and Sorin, 2010 [126] for dynamics induced by maximal monotone operators in discrete and continuous time.)

3) For vector fields ϕ with potential W one does not have the property $W(x_n)$ decreasing.

This section deals with class (III) *convex gradient*, where in addition f satisfies some regularity properties.
Recall that f is β smooth if:

$$|f(y)-f(x)-\langle \nabla f(x)|y-x\rangle| \leq \frac{\beta}{2}\|x-y\|^2. \tag{50}$$

Equivalently, ∇f is β-Lipschitz.
Hilbertian framework: Projected Dynamics
Assumption: f is β smooth.
Same procedure with constant steps:

$$x_{m+1} = \Pi_X(x_m - \eta \nabla f(x_m)).$$

The analysis in this section is standard, see e.g. Nesterov, 2004 [113].
Take $\eta = 1/\beta$ and define $v_n = \beta(x_{n+1} - x_n)$.
The main tool is the following:
Descent lemma

$$f(x_{n+1}) - f(y) \leq \langle v_n, y - x_n\rangle - \frac{1}{2\beta}\|v_n\|^2.$$

In particular $f(x_n)$ decreasing and $\{\|v_n\|\} \in \ell^2$.
Values

$$n[f(x_{n+1}) - f(y)] \leq R_n(y) - \frac{1}{2\beta}\|\sum_{m=1}^{n}\|v_m\|^2 = \frac{\beta}{2}\|y - x_1\|^2.$$

Hence convergence rate of the order $\frac{1}{n}$ with constant step size.
Trajectories
Lemma Let $y^* \in E$. Then $\|x_n - y^*\|$ decreases.

Proposition $\{x_n\}$ converges to a point in E.

Mirror descent

The dynamics is still:

$$\langle \nabla H(x_n) - \lambda \nabla f(x_n) - \nabla H(x_{n+1}) | x - x_{n+1} \rangle \leq 0, \forall x \in X.$$

We follow Bauschke, Bolte and Teboulle, 2017 [13]
H and f are $\mathscr{C}^1$
Hypothesis [A]:
there exists $L > 0$ such that:

$$L\, D_H - D_f \geq 0$$

(preorder: $L\,H - f$ convex, Nguyen, 2017 [117])
If H is strongly convex and f is smooth, [A] holds.

Values

One has, by [A]:

$$f(x) \leq f(y) + \langle \nabla f(z) | x - y \rangle + L D_h(x,z) - D_f(y,z)$$

(the last term is ≤ 0 when f is convex).
Take $2\lambda L = 1$

Proposition

Assume H convex.
1) $f(x_n)$ is decreasing.
2) $\sum D_H(x_{n+1}, x_n) < +\infty$.
3) Assume f convex, lower bounded.

$$f(x_n) - f(y) \leq \frac{2L}{n} D_H(y, x_1)$$

Recent result: Bui and Combettes, 2020 [34] Theorem 3.9:
variable metrics H_n allow to reach $f(x_n) - f^* = o(1/n)$.

Trajectories

Propositon Assume f convex.
1) $y^* \in E$ implies $D_H(y^*, x_n)$ decreases.
2) Assume:
$[H1] : x^k \to x^* \in E \Rightarrow D_H(x^*, x^k) \to 0$
$[H2] : x^* \in E, D_H(x^*, x^k) \to 0 \Rightarrow x^k \to x^*$

Then $\{x_n\}$ converges to a point in E.

Dual averaging

We follow Lu, Freund and Nesterov (2018)
Dual averaging with constant step size under Hypothesis $[A]$:
$L\,h - f$ convex
f convex and $\mathscr{C}^1$
$h : V \to \mathbb{R} \cup \{+\infty\}$ l.s.c. with $dom\, h = X$.

$$x_{m+1} = argmax_X \{\langle U_m | x\rangle - L\,h(x)\} \tag{51}$$

with $u_k = -\nabla f(x_k)$.
Values
Proposition
f convex, lower bounded.

$$f(\bar{x}_n) - f(y) \leq \frac{L}{n} h(y), \quad \forall y \in X.$$

Comments on the regular case

1) In the three cases (PD), (MD) and (DA) the speed of convergence of the values is $O(1/n)$ and the algorithms use a constant step parameter.
2) Using (PD) with f smooth implies $f(x_n)$ decreasing and the convergence of $\{x_n\}$.
3) The approach in Section 5.2 shows that similar results can be obtained using (MD) without assuming f with Lipschitz gradient if the regularization function H is adapted to f : condition $([A]$.

4) Analogous results for the values are much simpler to obtain in the (DA) framework. However the properties concern the value at

the average $f(\bar{x}_n)$ and no result is available on the trajectories.

Concluding remarks

For the three dynamics (PG), (MD) and (DA) the following 1), 2) and 3) holds:
1) In continuous time the speed of convergence of the average regret to 0, of the order $O(1/t)$ is not better in the general gradient convex case than in on-line learning.
2) In discrete time the speed of convergence of the average regret to 0, of the order $O(1/\sqrt{n})$ is not better in the general gradient convex case than in on-line learning.
3) Adding a regularity hypothesis on the convex function does not change the convergence rate in continuous time but allow a better convergence in discrete time from $O(1/\sqrt{n})$ to $O(1/n)$.
4) A similar phenomena appears with the so-called acceleration procedures following Nesterov, 1983 [112].
In the continuous time case a second order ODE leads to a speed of convergence $f(x_t) - f(x^*) \leq O(\frac{1}{t^2})$ with no further hypothesis on f, see Su, Boyd and Candes, 2014 [152], 2016 [153], Krichene, Bayen and Bartlett, 2015 [87], 2016 [88], Wibisono, Wilson and Jordan, 2016 [161], Attouch and Peypouquet, 2016 [5], Attouch, Chbani, Peypouquet and Redont, 2018 [6]...
To obtain a similar property in discrete time, namely $f(x_n) - f(x^*) \leq O(\frac{1}{n^2})$ one has to assume f smooth.
The same remark apply to the (weak) convergence of the trajectory, where the smooth hypothesis on f is needed in discrete time and not in continuous time, Chambolle and Dossal, 2015[38], Attouch, Chbani, Peypouquet, Redont 2018 [6]...
5) Concerning the link between discrete and continuous time dynamics, there are no direct results of the form: no-regret property in continuous time imply no-regret property in discrete time but analogy of the tools used and ad-hoc choice of the stage step size, see Sorin, 2009 [144], Kwon and Mertikopoulos, 2017 [89] and the Lyapounov functions in Krichene, Bayen and Bartlett, 2015 [87], 2016 [88], Wibisono, Wilson and Jordan, 2016 [161].
6) The Hilbert framework for (PD) allows to obtain convergence results on the trajectories.
The two other algorithms are more flexible and can achieve better explicit speed of convergence of the values by choosing an adequate norm, adapted to the problem, see the discussion in Bauschke, Bolte and Teboulle, 2017 [13].
For (MD), specific regularization functions H can also lead to convergence of the trajectories.

(DA) is much simpler to implement due to its integral formulation. However no convergence properties of the trajectories are in general available.

Advances

Popov - trajectories

We follow Popov (1980) [128], see also Korpelevich (1976) [86]. y_0 and x_0 are given (in fact only y_0 and $\phi(x_0)$ are used). The general on-line dynamics is as follows:

$$\begin{aligned} y_{t+1} &= \Pi_Z[y_t + \alpha_t] \\ x_{t+1} &= \Pi_Z[y_{t+1} + \alpha_t] \end{aligned} \tag{52}$$

and we consider the closed case with constant step size, $\alpha_t = \eta\phi(x_t)$.

Two levels: action and memory.

The action x_t determines the parameter α_t.

Then one updates the memory y_t through α_t.

The new memory y_{t+1} and the parameter α_t defines the new action x_{t+1}.

(PG)

Theorem (Popov, 1980)

Finite dimensional, euclidean.
Assume:
ϕ L *Lipschitz and* $3\eta L < 1$.
ϕ *dissipative.*
$E \neq \emptyset$ *(no compactness on* X*).*
Then $\|x_t - y_t\|$ *tends to 0 and* $\{x_t\}$ *converges to a point in* E.

(MD)

Theorem

Finite dimensional
Assume:
ϕ L *Lipschitz and* $3\eta L < 1$.
ϕ *dissipative*
$E \neq \emptyset$.
$x_k \to x$ *implies* $D_H(y, x_k) \to D_H(y, x)$.
Then $\|x_t - y_t\|$ *tends to 0 and* $\{x_t\}$ *converges to a point in* E.

Two levels procedures and optimism

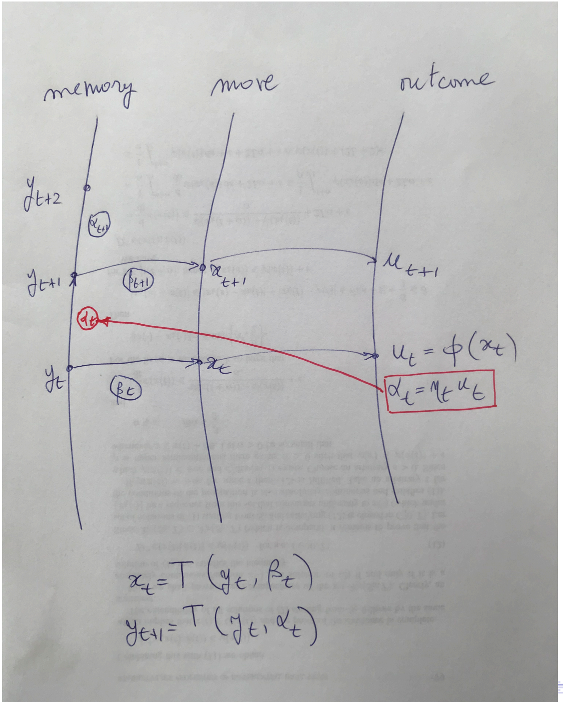
We follow Rakhlin and Sridharan (2013) [130].
The general on-line algorithm is as follows (MD):
$y_{t+1} = T(y_t; \alpha_t)$

$$\langle \nabla H(y_t) + \alpha_t - \nabla H(y_{t+1}) | z - y_{t+1} \rangle \leq 0, \forall z \in X$$

updating of the memory using the parameter α_t and
$x_{t+1} = T(y_{t+1}; \beta_{t+1})$

$$\langle \nabla H(y_{t+1}) + \beta_{t+1} - \nabla H(x_{t+1}) | z - x_{t+1} \rangle \leq 0, \forall z \in X$$

updating of the move using an anticipation β_{t+1}.
Popov's algorithm corresponds to the closed form:
$\alpha_t = \eta \phi(x_t) = \beta_{t+1}$


memory
move
outcome
y_{t+2}
α_{t+1}
y_{t+1}
β_{t+1}
x_{t+1}
u_{t+1}
α_t
$u_t = \phi(x_t)$
y_t
x_t
β_t
$\alpha_t = \eta_t u_t$
$x_t = T(y_t, \beta_t)$
$y_{t+1} = T(y_t, \alpha_t)$

Theorem (Rakhlin Sridharan, 2013)

H 1-strongly convex.
Take $\alpha_t = u_t/\sqrt{t}$ and $\beta_t = v_t/\sqrt{t}$

$$R_T(x) = \sum_{t=1}^{T} \langle u_t | x - x_t \rangle \leq O(\sqrt{T}).$$

More precisely:

$$\langle \alpha_t | x - x_t \rangle \leq (1/2)\|\alpha_t - \beta_t\|_*^2 + D_H(x, y_t) - D_H(x, y_{t+1}) - D_H(x_t, y_t). \tag{53}$$

same order of convergence than discrete time (MD) with variable step size.

Popov - evaluation

Framework of RS with datas Popov

$\alpha_t = \eta\phi(x_t)$.

We follow Hsieh, Iutzeler, Malick and Mertikopoulos (2019) [82]

Theorem

Assume ϕ L Lipschitz and $\eta L \leq 1/2$:

$$\begin{aligned} \eta\langle\phi(x_t)|z - x_t\rangle &\leq (1/2)(\eta L)^2[\|x_{t-1} - x_{t-2}\|^2 - \|x_t - x_{t-1}\|^2] \\ &\quad + D_H(z, y_t) - D_H(z, y_{t+1}) \end{aligned}$$

Assume in addition ϕ dissipative:

$$T\langle\phi(z)|z - \bar{x}_T\rangle \leq R_T(z) = \sum_{t=1}^{T}\langle\phi(x_t)|z - x_t\rangle \leq K.$$

New directions

Last iterate convergence

Mixed two levels procedure:
(MD) for the action, (DA) for the memory
different step sizes

Adaptive step size, Hsieh, Iutzeler, Malick and Mertikopoulos (2021)

Open problems

Higher recall et parallel memory updating

Link with continuous time and acceleration

MISSING
Imperfect monitoring
Stochastic aspects
Hypothesis testing
Weak calibration
Complexity and speed of convergence

(and lot of other topics are missing ... but no regret !!!)

Open problems

Higher recall et parallel memory updating

Link with continuous time and acceleration

MISSING
Imperfect monitoring
Stochastic aspects
Hypothesis testing
Weak calibration
Complexity and speed of convergence

(and lot of other topics are missing ... but no regret !!!)

Open problems

Higher recall et parallel memory updating

Link with continuous time and acceleration

MISSING

Imperfect monitoring
Stochastic aspects
Hypothesis testing
Weak calibration
Complexity and speed of convergence

(and lot of other topics are missing ... but no regret !!!)

Open problems

Higher recall et parallel memory updating

Link with continuous time and acceleration

MISSING

Imperfect monitoring
Stochastic aspects
Hypothesis testing
Weak calibration
Complexity and speed of convergence

(and lot of other topics are missing ... but no regret !!!)

Bibliography I

ABERNETHY J., P. L. BARTLETT AND E. HAZAN (2011) Blackwell approachability and no-regret learning are equivalent, *JMLR Workshop and Conference Proceedings*, **19**, 27-46.

ALVAREZ F., J. BOLTE AND O. BRAHIC (2004) Hessian Riemannian gradient flows in convex programming, *SIAM Journal on Control and Optimization*, **43**, 477-501.

ATTOUCH H. AND M. TEBOULLE (2004) Regularized Lotka-Volterra dynamical system as continuous proximal-like method in optimization, *JOTA*, **121**, 541-570.

ATTOUCH H., BOLTE J., REDONT P. AND M. TEBOULLE (2004) Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization, *Optimization*, **53**, 435-454.

Bibliography II

- ATTOUCH H. AND J. PEYPOUQUET (2016) The rate of convergence of Nesterov's accelerated forward backward method is actually faster than $1/k^2$, *SIAM Journal on Optimization*, **26**, 1824-1834.
- ATTOUCH H., Z. CHBANI, J. PEYPOUQUET AND P. REDONT (2018) Fast convergence of inertial dynamics and algorithms with asymptotic vanishing damping, *Mathematical Programming*, **168**, 123-175.
- AUER, P., N. CESA-BIANCHI, Y. FREUND AND R.E. SCHAPIRE (1995) Gambling in a rigged casino: the adversarial multi-armed bandit problem, *Proceedings of the 36 th Annual Symposium on Foundations of Computer Science*, 322-331.

Bibliography III

- AUER P., CESA-BIANCHI N., FREUND Y. AND R.E. SHAPIRE (2002) The nonstochastic multiarmed bandit problem, *SIAM J. Comput.*, **32**, 48-77.
- AUMANN R.J. (1974) Subjectivity and correlation in randomized strategies, *Journal of Mathematical Economics*, **1**, 67-96.
- AUSLENDER A. AND M. TEBOULLE (2006) Interior gradient and proximal methods for convex and conic optimization, *SIAM J Optimization*, **16**, 697-725.
- AUSLENDER A. AND M. TEBOULLE (2009) Projected subgradient methods with non Euclidean distances for non-differentiable convex minimization and variational inequalities, *Mathematical Programming*, **120**, 27-48.
- AZUMA K. (1967) Weighted sum of certain dependent random variables, *Tohoku Math. J.*, **68**, 357-367.

Bibliography IV

- BAUSCHKE H., BOLTE J. AND M. TEBOULLE (2017) A descent Lemma beyond Lipschitz gradient continuty: first-order methods revisited and applications, *Mathematics of Operations Research*, **42**, 330-348.
- BECK A. AND M. TEBOULLE (2003) Mirror descent and nonlinear projected subgradient methods for convex optimization, *Oper. Res. Lett.*, **31**, 167-175.
- M. BENAÏM (1996) A dynamical system approach to stochastic approximations, *SIAM Journal on Control and Optimization*, **34**, 437-472.
- BENAIM M., (1999) *Dynamics of Stochastic Algorithms* , Séminaire de Probabilités, XXIII, Azéma J. and alii eds, LNM **1709**, Springer, 36-72.

Bibliography V

- BENAIM M. AND M. FAURE (2013) Consistency of vanishingly smooth fictitious play, *Mathematics of Operations Research*, **38**, 437-450.
- M. BENAÏM AND M.W. HIRSCH (1996) Asymptotic pseudotrajectories and chain recurrent flows, with applications, *J. Dynam. Differential Equations*, **8**, 141-176.
- BENAIM M., J. HOFBAUER AND S. SORIN (2005) Stochastic approximations and differential inclusions, *SIAM J. Opt. and Control*, **44**, 328-348.
- BENAIM M., J. HOFBAUER AND S. SORIN (2006) Stochastic approximations and differential inclusions. Part II: applications, *Mathematics of Operations Research*, **31**, 673-695.

Bibliography VI

- BLACKWELL D. (1954) Controlled random walks, *Proceedings of the International Congress of Mathematicians, 1954, Amsterdam*, Erven P. Noordhoff N.V., North-Holland, **III**, 336-338.
- BLACKWELL D. (1956) An analog of the minimax theorem for vector payoffs, *Pacific Journal of Mathematics*, **6**, 1-8.
- BLUM A. AND Y. MANSOUR (2007) From external to internal regret, *Journal of Machine Learning Reserach*, **8**, 1307-1324.
- BOLTE J. (2003) Continuous gradient projection method in Hilbert spaces, *JOTA*, **119**, 235-259.
- BOLTE J. AND M. TEBOULLE (2003) Barrier operators and associated gradient-like dynamical systems for constrained minimization problems, *SIAM Journal on Control and Optimization*, **42**, 1266-1292.

Bibliography VII

BOT R. I., E. R. CSETNEK, AND D. K. NGUYEN (2023), Fast OGDA method in continuous and discrete time, *Foundations of Computational Mathematics*, https://doi.org/10.1007/s10208-023-09636-5.

BRÉZIS H. (1973) *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North Holland.

BROWN G.W. (1949) Some notes on computation of games solutions, *RAND Report*, P-78, The RAND Corporation.

BROWN G. W.(1951) Iterative solution of games by fictitious play, in Koopmans T.C. (ed.) *Activity Analysis of Production and Allocation* , Wiley, 374-376.

Bibliography VIII

- BROWN G.W. AND J. VON NEUMANN, (1950) *Solutions of games by differential equations*, *Contributions to the Theory of Games, I*, Kuhn H. and A.W. Tucker eds., Princeton University Press, Annals of Mathematical Studies, **24**, 73-79.
- BUBECK S. (2011) *Introduction to online optimization*, Lecture Notes.
- BUBECK S. (2015) Convex optimization: Algorithms and complexity, *Fondations and Trends in Machine Learning*, **8**, 231-357.
- BUBECK S. AND N. CESA-BIANCHI (2012) Regret analysis of stochastic and nonstochastic multi-armed bandit problems, *Fondations and Trends in Machine Learning*, **5**, 1-122.

Bibliography IX

- BUI M.N. AND P.L. COMBETTES (2021) Bregman forward-backward operator splitting, *Set-Valued and Variational Analysis*, **29**, 583-603.
- CESA-BIANCHI N. AND G. LUGOSI (2003) Potential-based algorithms in on-line prediction and game theory, *Machine Learning*, **51**, 239-261.
- CESA-BIANCHI N. AND G. LUGOSI (2006) *Prediction, Learning and Games*, Cambridge University Press.
- CESA-BIANCHI N., G. LUGOSI AND G. STOLTZ (2006) Regret minimization under partial monitoring, *Mathematics of Operations Research*, **31**, 562-580.
- CHAMBOLLE A. AND C. DOSSAL (2015) On the convergence of the iterates of FISTA, *Journal of Optimization Theory and Applications*, **166**, 968-982.

Bibliography X

- COUCHENEY P., B. GAUJAL AND P. MERTIKOPOULOS (2015) Penalty-regulated dynamics and robust learning procedures in games, *Mathematics of Operations Research*, **40**, 611-633.
- COVER T. (1991) Universal Portfolios, *Math. Finance*, **1**, 1-29.
- DAWID A. (1982) The well-calibrated Bayesian, *Journal of the American Statistical Association*, **77**, 605-613.
- DAFERMOS S.C. (1980) Traffic equilibrium and variational inequalities, *Transportation Sci.*, **14** , 42–54.
- DUCHI J., E. HAZAN AND Y. SINGER (2011) Adaptive Subgradient Methods for Online Learning and Stochastic Optimization, *Journal of Machine Learning Research*, **12**, 2121-2159.

Bibliography XI

- DUPUIS P. AND A. NAGURNEY (1993) Dynamical systems and variational inequalities, *Ann. Oper. Res.*, **44**, 9–42.
- FACCHINEI F. AND J. PANG (2007) *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer.
- FOSTER D. AND R. VOHRA (1993) A randomization rule for selecting forecasts, *Operations Research*, **41**, 704-707.
- FOSTER D. AND R. VOHRA (1997) Calibrated learning and correlated equilibria, *Games and Economic Behavior*, **21**, 40-55.
- FOSTER D. AND R. VOHRA (1998) Asymptotic calibration, *Biometrika*, **85**, 379-390.

Bibliography XII

- FOSTER D. AND R. VOHRA (1999) Regret in the on-line decision problem, *Games and Economic Behavior*, **29**, 7-35.
- FOSTER D. AND P. YOUNG, (1998) On the nonconvergence of fictitons play in coordination games, *Games and Economic Behavior*, **25**, 79-96.
- FOSTER D. AND P. YOUNG, (2003) Learning, hypothesis testing, and Nash equilibria, *Games and Economic Behavior*, **45**, 73-96.
- FOSTER D. AND P. YOUNG, (2005) Regret testing, *Games and Economic Behavior*, **25**, 79-96.
- FREUND Y. AND R.E. SCHAPIRE (1999) Adaptive game playing using multiplicative weights, *Games and Economic Behavior*, **29**, 79-103.

Bibliography XIII

- Fudenberg D. and D. K. Levine (1995) Consistency and cautious fictitious play, *Journal of Economic Dynamics and Control*, **19**, 1065-1089.
- Fudenberg D. and D. K. Levine (1998) *The Theory of Learning in Games*, MIT Press.
- Fudenberg D. and D. K. Levine (1999) Conditional universal consistency, *Games and Economic Behavior*, **29**, 104-130.
- Gaunersdorfer A. and J. Hofbauer (1995) Fictitious play, Shapley polygons and the replicator equation, *Games and Economic Behavior*, **11**, 279-303.
- Gilboa I. and A. Matsui (1991) Social stability and equilibrium, *Econometrica*, **59**, 859-867.

Bibliography XIV

- GUL F., D. PEARCE AND E. STACCHETTI, *A bound on the proportion of pure strategy equilibria in generic games*, Mathematics of Operations Research, **18**, 548-552, 1993.
- HANNAN J. (1957) Approximation to Bayes risk in repeated plays, *Contributions to the Theory of Games, III*, Drescher M., A.W. Tucker and P. Wolfe eds., Princeton University Press, 97-139.
- HART S. (2005) Adaptive heuristics, *Econometrica*, **73**, 1401-1430.
- HART S. AND A. MAS-COLELL, (2000) A simple adaptive procedure leading to correlated equilibrium, *Econometrica*, **68**, 1127-1150.
- HART S. AND A. MAS-COLELL (2001) A general class of adaptive strategies, *Journal of Economic Theory*, **98**, 26-54.

Bibliography XV

- HART S. AND A. MAS-COLELL (2003) Regret-based continuous time dynamics, *Games and Economic Behavior*, **45**, 375-394.
- HART S. AND A. MAS-COLELL (2003) Uncoupled dynamics do not lead to Nash equilibrium, *Am. Econ. Rev.*, **93**, 1830-1836.
- S. HART AND A. MAS-COLELL (2006) Stochastic uncoupled dynamics and Nash equilibrium, *Games and Economic Behavior*, **57**, 286-303.
- HART S. AND A. MAS COLELL (2013) *Simple Adaptive Strategies: From Regret-Matching to Uncoupled Dynamics*, World Scientific Publishing.
- HART S. AND D. SCHMEIDLER (1989) Existence of correlated equilibria, *Mathematics of Operations Research*, **14**, 18-25.

Bibliography XVI

- HAZAN E. (2011)The convex optimization approach to regret minimization, *Optimization for machine learning*, S. Sra, S. Nowozin, S. Wright eds, MIT Press, 287-303.
- HAZAN E. (2015) Introduction to Online Convex Optimization, *Fondations and Trends in Optimization*, **2**, 157-325; (2023) https://arxiv.org/pdf/1909.0527.pdf.
- HAZAN E. (2019) Optimization for Machine Learning , https://arxiv.org/pdf/1909.03550.pdf.
- HOEFFDING W. (1963) Probability inequalities for sum of bounded random variables, *J. Amer. Statist. Assoc.*, **58**, 13-30.
- HOFBAUER J.(2000) From Nash and Brown to Maynard Smith: Equilibria, dynamics and ESS, *Selection*, **1**, 81–88.

Bibliography XVII

- HOFBAUER J. (2011) Deterministic evolutionary game dynamics, in *Evolutionary Game Dynamics*, K. Sigmund ed., Proceedings of Symposia in Applied Mathematics, **69**, A.M.S., 61-79.
- HOFBAUER J. AND W. H. SANDHOLM (2002) On the global convergence of stochastic fictitious play, *Econometrica*, **70**, 2265-2294.
- HOFBAUER J. AND W. H. SANDHOLM (2009) Stable games and their dynamics, *Journal of Economic Theory*, **144**, 1710-1725.
- HOFBAUER J. AND K. SIGMUND, (1998) *Evolutionary Games and Population Dynamics*, Cambridge U.P.
- HOFBAUER J. AND K. SIGMUND, (2003) Evolutionary Games Dynamics, *Bulletin A.M.S.*, **40**, 479-519.

Bibliography XVIII

- HOFBAUER J. AND S. SORIN, (2006) Best response dynamics for continuous zero-sum games, *Discrete and Continuous Dynamical Systems -Series B*, **6**, 215-224.
- HOFBAUER J., S. SORIN AND Y. VIOSSAT (2009) Time average replicator and best reply dynamics, *Mathematics of Operations Research*, **34**, 263-269.
- HOFBAUER J AND J. SWINKELS (1996) A universal Shapley example. Preprint
- HSIEH Y-G, F. IUTZELER , J.MALICK AND P. MERTIKOPOULOS (2019) On the convergence of single-call stochastic extra-gradient methods. In 33 Annual Conference on Neural Information Processing Systems.
- KALAI A. AND S. VEMPALA (2005) Efficient algorithms for online decision problems, *Journal of Computer and System Sciences*, **71**, 291-307.

Bibliography XIX

- KINDERLEHRER D. AND G. STAMPACCHIA (1980) *An Introduction to Variational Inequalities and their Applications*, Academic Press.
- KINGMA D.P. AND J. L. BA (2015) Adam: a Method for Stochastic Optimization, *Interna- tional Conference on Learning Representations*, 1-13.
- KORPELEVICH G. (1976) The extragradient method for finding saddle points and other problems, *Ekonomika i Matematicheskie Metody*, **12**, 747-756 (English version, Matekon, 35-49).
- KRICHENE W., A. BAYEN, AND P. BARTLETT (2015) Accelerated mirror descent in continuous and discrete time, *NIPS 2015*.

Bibliography XX

- KRICHENE W., A. BAYEN, AND P. BARTLETT (2016) Adaptive Averaging in Accelerated Descent Dynamics, *NIPS 2016*.
- KWON J. AND P. MERTIKOPOULOS (2017) A continuous time approach to on-line optimization, *Journal of Dynamics and Games*, **4**, 125-148.
- KWON J. AND V. PERCHET (2017) Online learning and Blackwell approachability with partial monitoring: optimal convergence rates, *Proceedings of Machine Learning Research (AISTATS 2017)*, **54**, 604-613.
- LEHRER E. (2003) A wide range no-regret theorem, *Games and Economic Behavior*, **42**, 101-115.
- LEHRER E. AND E. SOLAN (2016) A general internal regret-free strategy, *Dynamic Games and Applications*, **6**, 112-138.

Bibliography XXI

- LEVITIN E.S. AND B.T. POLYAK (1966) Constrained minimization methods, *USSR Computational Mathematics and Mathematical Physics*, **6**, 1-50.
- LITTLESTONE N. AND M.K. WARMUTH (1994) The weighted majority algorithm, *Information and Computation*, **108**, 212-261.
- LUGOSI G., S. MANNOR AND G. STOLTZ (2008) Strategies for prediction under imperfect monitoring, *Mathematics of Operations Research*, **33**, 513-528.
- LU H., FREUND AND Y. NESTEROV (2018) Relatively smooth convex optimization by first-order methods, and applications, Siam J Opt, 28, 333-354.

Bibliography XXII

- MANNOR S., V. PERCHET AND G. STOLTZ (2014) Set-valued approachability and online learning with partial monitoring, *Journal of Machine Learning Research*, **15**, 3247-3295.
- MANNOR S. AND G. STOLTZ (2010) A geometric proof of calibration, *Mathematics of Operations Research*, **35**, 721-727.
- MERTENS, J.-F, S. SORIN AND S. ZAMIR (2015), *Repeated Games*, Cambridge University Press.
- MERTIKOPOULOS P. AND W. H. SANDHOLM (2016) Learning in games via reinforcement and regularization, *Mathematics of Operations Research*, **41**, 1297-1324.
- MERTIKOPOULOS P. AND W. H. SANDHOLM (2018) Riemannian game dynamics, *Journal of Economic Theory*, **177**, 315-364.

Bibliography XXIII

- Mertikopoulos P. and Z. Zhou (2019) Learning in games with continuous action sets and unknown payoff functions, *Mathematical Programming*, **173**, 465-507.
- Minty G. J.(1967) On the generalization of a direct method of the calculus of variations, *Bulletin A.M.S.*, **73**, 315–321.
- Monderer D. and L.S. Shapley (1996) Fictitious play property for games with identical interest, *Journal of Economic Theory*, **68**, 258-265.
- Monderer D. and L.S. Shapley (1996) Potential games, *Games and Economic Behavior*, **14**, 124-143.
- Moreau J.J. (1965) Proximité et dualité dans un espace Hilbertien, *Bull. Soc. Math. France*, **93**, 273–299.

Bibliography XXIV

- NAGURNEY A. AND D. ZHANG (1996) *Projected Dynamical Systems and Variational Inequalities with Applications* , Kluwer.
- NASH J. (1950) Equilibrium points in n-person games, *Proceedings of the National Academy of Sciences*, **36**, 48–49.
- NASH J. (1951) Non-cooperative games, *Annals of Mathematics*, **54**, 286-295.
- NEMIROVSKI A. AND D. YUDIN (1983) *Problem Complexity and Method Efficiency in Optimization*, Wiley.
- NEMIROVSKI A. (2004) Prox-method with rate of convergence O(1/t) for variatioal inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *Siam J Opt*, **15**, 229-251.

Bibliography XXV

- NESTEROV Y. (1983) A method of solving a convex programming problem with convergence rate o(1/k2), *Soviet Mathematics Doklady*, **27**, 372-376.
- NESTEROV Y. (2004) *Introductory Lectures on Convex Optimization*, Kluwer.
- NESTEROV Y. (2009) Primal-dual subgradient methods for convex problems, *Mathematical Programming*, **120**, 221-259.
- NESTEROV Y. (2018) *Lectures on Convex Optimization*, Springer.
- NEVEU J. (1972) *Martingales à Temps Discret*, Masson.
- NGUYEN Q.V. (2017) Forward-Backward splitting with Bregman distances, *Vietnam J. Math.*, **45**, 519-539.

Bibliography XXVI

- OPIAL Z. (1967) Weak Convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.*, **73**, 591-597.
- PERCHET V. (2009) Calibration and internal no-regret with random signals, *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, LNAI **5809**, 68-82.
- PERCHET V. (2011a) Approachability of convex sets in games with partial monitoring, *Journal of Optimization Theory and Applications*, **149**, 665-677.
- PERCHET V. (2011b) Internal regret with partial monitoring calibration-based optimal algorithms, *Journal of Machine Learning Research*, **12**, 1893-1921.

Bibliography XXVII

PERCHET V. (2014) Approachability, regret and calibration: implications and equivalences, *Journal of Dynamics and Games*, **1**, 181-254.

PERCHET V. (2015) Exponential weight approachability, applications to calibration and regret minimization, *Dynamic Games and Applications*, **5**, 136-153.

PERCHET V. AND M. QUINCAMPOIX (2014) On a unified framework for approachability with full or partial monitoring, *Mathematics of Operations Research*, **40**, 596-610.

PERCHET V. AND M. QUINCAMPOIX (2019) A differential game on Wasserstein space, application to weak approachability with partial monitoring, *Journal of Dynamics and Games*, **6**, 65-85.

Bibliography XXVIII

- PEYPOUQUET J. AND S. SORIN (2010) Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time, *Journal of Convex Analysis*, **17**, 1113-1163.
- POLYAK B. (1987) *Introduction to optimization*, Optimization software.
- POPOV L. D. (1980) A modification of the Arrow-Hurwicz method for search of saddle points, *Mathematical notes of the Academy of Sciences of the USSR*, **28**, 845-848.
- RAKHLIN A. (2009) *Lecture notes on On Line Learning*.
- RAKHLIN A. AND K. SRIDHARAN (2013) Online learning with predictable sequences, *COLT 13: Proceedings of the 26th Annual Conference on Learning Theory JMLR*, 1-27.

Bibliography XXIX

- ROBINSON J. (1951) An iterative method of solving a game, *Annals of Mathematics*, **54**, 296-301.
- ROCKAFELLAR R.T. (1970) *Convex Analysis*, Princeton University Press.
- ROCKAFELLAR R.T. (1970) Monotone operators associated with saddle-functions and minmax problems, *Nonlinear Functional Analysis, F. Browder, ed.*, Proceedings of Symposia in Pure Math, **18**, AMS , 241-250.
- ROSEN J.B. (1965) Existence and uniqueness of equilibrium points for concave N-person games, *Econometrica*, **33**, 520-534.
- RUSTICHINI A. (1999) Minimizing regret: the general case, *Games and Economic Behavior*, **29**, 224-243.

Bibliography XXX

- RUSTICHINI A. (1999) Optimal properties of stimulus-response learning models, *Games and Economic Behavior*, **29**, 244-273.
- SANDHOLM W. H. (2001) Potential games with continuous player sets, *Journal of Economic Theory*, **97**, 81-108.
- SANDHOLM W. H. (2010) *Population Games and Evolutionary Dynamics*, MIT Press.
- SANDHOLM W. H. (2015) Population games and deterministic evolutionary dynamics, H. P. Young, S. Zamir, eds *Handbook of Game Theory IV*, Elsevier, 703-778.
- SHALEV-SHWARTZ S. (2011) Online learning and online convex optimization, *Foundations and Trends in Machine Learning*, **4**, 107-194.

Bibliography XXXI

- SHAHSHAHANI S. (1979) A new mathematical framework for the study of linkage and selection, *Memoirs of the American Mathematical Society*, 211.
- SHAPLEY L.S. (1964) Some topics in two-person games, in *Advances in Game Theory*, M. Dresher, L.S. Shapley and A. W. Tucker (eds.), Annals of Mathematical Studies, **52**, Princeton University Press, 1-28.
- SMITH M.J. (1984) A descent algorithm for solving monotone variational inequalities and monotone complementarity problems, *J. Optim. Theory Appl.*, **44**, 485–496.
- SORIN S. (2009) Exponential weight algorithm in continuous time, *Mathematical Programming*, Ser. B **116**, 513-528.

Bibliography XXXII

- Sorin S. (2011) On some global and unilateral adaptive dynamics, In *Evolutionary Game Dynamics*, K. Sigmund (ed.), Proceedings of Symposia in Applied Mathematics, 81-109.
- Sorin S. (2020) Replicator dynamics: Old and new, *Journal of Dynamics and Games*, **7**, 365-385.
- Sorin S. (2023) Continuous time algorithms in optimization and game theory, *Dynamic Games and Applications*, **13**, 3-24.
- Sorin S. (2024) No-regret algorithms in on-line learning, games and convex optimization, *Mathematical Programming*, **203**, 645-686.
- Sorin S. and C. Wan (2016) Finite composite games: equilibria and dynamics, *Journal of Dynamics and Games*, **3**, 101-120.

Bibliography XXXIII

- STOLTZ G. AND G. LUGOSI (2005) Internal Regret in On-Line Portfolio Selection, *Machine Learning*, **59**, 125-159.
- STOLTZ G. AND G. LUGOSI (2007) Learning correlated equilibria in games with compact sets of strategies, *Games and Economic Behavior*, **59**, 187-208.
- SU W., S. BOYD, AND E. CANDES (2014) A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights, *NIPS 2014*.
- SU W., S. BOYD, AND E. CANDES (2016) A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights, *Journal of Machine Learning Research*, **17**, 1-43.

Bibliography XXXIV

- SWINKELS J.M. (1983) Adjustment dynamics and rational play in games, *Games and Economic Behavior*, **5**, 455–484.
- TAYLOR P.B. AND L.B. JONKER (1978) Evolutionary stable strategies and game dynamics, *Math. Biosci.*, **40**, 145–156.
- TEBOULLE M. (2018) A simplified view of first order methods for optimization, *Math Programming*, **170**, 67-96.
- VIOSSAT Y. (2007) The replicator dynamics does not lead to correlated equilibria, *Games and Economic Behavior*, **59**, 397-407
- VIOSSAT Y. AND A. ZAPECHELNYUK (2013) No-regret dynamics and fictitious play, *Journal of Economic Theory*, **148**, 825-842.

Bibliography XXXV

- VOVK V. (1990) Aggregating strategies, *Proceedings of the 3rd Annual Conference on Computational Learning Theory*, 371-383.
- Vovk, V. (1995) A game of prediction with expert advice, *COLT 95: Proceedings of the 8thAnnual Conference on Computational Learning Theory*, 51- 60
- WIBISONO A., A. C. WILSON, AND M. I. JORDAN (2016) A variational perspective on accelerated methods in optimization, *PNAS* , **113**, 7351-7358.
- XIAO L. (2010) Dual averaging methods for regularized stochastic learning and online optimization, *Journal of Machine Learning Research*, **11**, 2543-2596.
- YOUNG P. (2004) *Strategic Learning and Its Limits*, Oxford U. P.

Bibliography XXXVI

ZHANG D. AND A. NAGURNEY (1997) Formulation, stability, and computation of traffic network equilibria as projected dynamical systems, *J. Optim. Theory Appl.*, **93**, 417–444.

ZINKEVICH M. (2003) Online convex programming and generalized infinitesimal gradient ascent, *Proceedings of the Twentieth International Conference on Machine Learning*, Washington, 928-935.