GAMES, DYNAMICS & LEARNING

A. Giannou[1] T. Lianeas[1] P. Mertikopoulos[2] E. V. Vlatakis-Gkaragkounis[3]

[1]NTUA

[2]French National Center for Scientific Research (CNRS) & Criteo AI Lab

[3]Columbia University

ECE-NTUA – June 4, 2021

GAMES, DYNAMICS & LEARNING

4. LEARNING IN FINITE GAMES AND BANDITS, CONT'D

A. Giannou[1] T. Lianeas[1] P. Mertikopoulos[2] E. V. Vlatakis-Gkaragkounis[3]

[1]NTUA

[2]French National Center for Scientific Research (CNRS) & Criteo AI Lab

[3]Columbia University

ECE-NTUA – June 4, 2021

Outline

Overview

Online learning: Algorithms & guarantees

Overview

Learning in finite games

- **Frequencies** (pop. shares) $\leftrightsquigarrow$ **Choice probabilities** (mixed strategies)
- **Multi-agent** (game-theoretic) v. **online** ("playing against anything")
- **Dynamics** (last time) $\leftrightsquigarrow$ **Algorithms** (today)
- **Feedback:**
    - Full information
    - Pure/Noisy payoff vector
    - Bandit (only rewards)

- **Today:** Deep dives

Learning with a finite number of actions

Online decision-making with **mixed strategies**

---

repeat

At each epoch $t \geq 0$

Choose **mixed strategy** $x_t \in \mathcal{X} := \Delta(\mathcal{A})$

Encounter **payoff vector** $v_t \in \mathbb{R}^{\mathcal{A}}$ [depends on context]

Get **mean payoff** $u_t(x_t) = \langle v_t, x_t \rangle$

Receive **feedback** [depends on context]

until end

Key considerations

- **Time:** continuous or discrete?
- **Players:** ~~continuous or~~ discrete
- **Actions:** ~~continuous or~~ discrete
- **Payoffs:** determined by other players or "Nature"?
- **Feedback:** full info? payoff-based?

Online v. multi-agent learning

	Continuous Time	Discrete Time
Regret (Online)	$\mathcal{O}(1)$	$\mathcal{O}(\sqrt{T})$
Nash (Game-theoretic)	"Folk Theorem"	**TODAY**

Table: Recap of results so far

Recall:

- **Regret:** $\mathrm{Reg}(T) = \max_{x \in \mathcal{X}} \sum_{t=1}^{T} [u_t(x) - u_t(X_t)]$ [or integral]
- **Folk theorem:** Asymptotic stability $\iff$ Strict Nash equilibrium [cont. time]

Outline

Feedback

Feedback types (from best to worst):

- **Full information**: observe entire payoff vector $v_t \leftarrow v(X_t)$
- **Full, inexact information**: observe estimate V_t of v_t
- **Partial information / Bandit:** only chosen component $u_t(a_t) = v_{a_t,t}$

Feedback

Feedback types (from best to worst):

- **Full information**: observe entire payoff vector $v_t \leftarrow v(X_t)$
- **Full, inexact information**: observe estimate V_t of v_t
- **Partial information / Bandit:** only chosen component $u_t(a_t) = v_{a_t,t}$

Abstract feedback model

$$V_t = v_t + Z_t + b_t$$

where Z_t is *zero-mean* and b_t is the *bias* of V_t

Assumptions

- **Bias:** $\|b_t\| \leq B_t$ (a.s.)
- **Variance:** $\mathbb{E}[\|Z_t\|^2 \mid \mathcal{F}_t] \leq \sigma_t^2$ (a.s.)
- **Second moment:** $\mathbb{E}[\|V_t\|^2 \mid \mathcal{F}_t] \leq M_t^2$ (a.s.)

Examples

① Full Information

- Game-theoretic model: player i at time t observes

$$V_{i,t} = v_i(X_{i,t}; X_{-i,t})$$

(mixed of i; mixed strat of others)

Bias: $b_{i,t} = 0$ Noise: $Z_{i,t} = 0$

② Pure Payoff Information

At round t, each player selects $a_{i,t} \in \mathcal{A}_i$ based on $X_{i,t} \in \mathcal{X}_i$

Assume observed: $V_{i,t} = (v_{i a_i}(a_t))_{a_i \in \mathcal{A}_i}$

$\mathbb{E}[V_{i,t}] = v_i(X_{i,t}; X_{-i,t}) \Rightarrow$ Bias $b_{i,t} = 0$ Variance $\sigma_{i,t} = \mathcal{O}(1)$

Follow the regularized leader

Follow the regularized leader with abstract feedback

$$\begin{aligned} Y_{t+1} &= Y_t + V_t \\ X_{t+1} &= Q(\eta_{t+1} Y_{t+1}) \end{aligned} \qquad \text{(FTRL)}$$

constant η in the sequel

where η_t is a variable **learning rate** parameter

Follow the regularized leader

Follow the regularized leader with abstract feedback

$$
\begin{aligned}
Y_{t+1} &= Y_t + V_t \\
X_{t+1} &= Q(\eta_{t+1} Y_{t+1})
\end{aligned}
\qquad \text{(FTRL)}
$$

where η_t is a variable **learning rate** parameter

$Q(y) = \arg\max \{ \langle y, x \rangle - h(x) \}$

Technical: Will need Q Lipschitz continuous $\iff$ h is strongly convex

$$h(x') \geq h(x) + \langle \nabla h(x), x' - x \rangle + \frac{K}{2} \|x' - x\|^2$$

Follow the regularized leader

Follow the regularized leader with abstract feedback

$$\begin{aligned} Y_{t+1} &= Y_t + V_t \\ X_{t+1} &= Q(\eta_{t+1} Y_{t+1}) \end{aligned} \tag{FTRL}$$

where η_t is a variable **learning rate** parameter

Technical: Will need Q Lipschitz continuous $\iff$ h is strongly convex

$$h(x') \geq h(x) + \langle \nabla h(x), x' - x \rangle + \frac{K}{2} \|x' - x\|^2$$

Example: Multiplicative / Exponential Weights algorithm

$$\begin{aligned} Y_{t+1} &= Y_t + V_t \\ X_{t+1} &= \frac{(\exp(\eta_{t+1} Y_{a,t+1}))_{a \in \mathcal{A}}}{\sum_{a \in \mathcal{A}} \exp(\eta_{t+1} Y_{a,t+1})} \end{aligned} \tag{EW}$$

[Vovk, 1990; Littlestone and Warmuth, 1994; Auer et al., 1995; Freund and Schapire, 1999]

Regret guarantees of FTRL

Work as in continuous-time case

- Fenchel coupling

$$F_t = h(x) + g(Y_t) - \langle Y_t, x \rangle$$

where g is a potential function for Q, i.e.,

$$Q = \nabla g$$

- Discrete-time evolution

$$F_{t+1} \leq F_t + \gamma \langle V_t, X_t - x \rangle + \frac{\gamma^2}{2K} \| V_t \|_*^2$$

- Aggregate/Telescope:

$$\overline{\operatorname{Reg}}(T) = \mathcal{O}\left(\frac{\max h - \min h}{\gamma} + \sum_{t=1}^{T} B_t + \gamma \sum_{t=1}^{T} M_t^2 \right)$$

- Take $\gamma \propto 1/\sqrt{T}$: [Why?]

$$\overline{\operatorname{Reg}}(T) = \mathcal{O}\left(\sqrt{T} + \sum_{t=1}^{T} B_t + \frac{\sum_{t=1}^{T} M_t^2}{\sqrt{T}} \right)$$

Regret guarantees of FTRL

Theorem (Shalev-Shwartz and Singer, 2006; Shalev-Shwartz, 2011)

- Assume:
  - *feedback unbiased and bounded in mean square* ($B_t = 0$, $\sup_t M_t < M$)
  - $\gamma = (2/M)\sqrt{KH/T}$ *with* $H = \max h - \min h$
- Then: *FTRL enjoys the bound*

$$\overline{\mathrm{Reg}}(T) \leq 2M\sqrt{(H/K)T} = \mathcal{O}(\sqrt{T})$$

Observe:

- This bound is tight [Nesterov, 2004; Abernethy et al., 2008; Bubeck, 2015]
- Cannot achieve $\mathcal{O}(1)$ regret as in continuous time [Why?]
- How to do if T is unknown? [Exercise]

ANALYSIS

Fenchel Coupling: $F(x, y) = h(x) + g(y) - \langle y, x \rangle$ (x: mixed strat; y: score vector) ||| $\nabla g = Q$

Template Inequality: $F_{t+1} \leq F_t + \gamma \langle V_t, X_t - x \rangle + \frac{\gamma^2}{2K} \|V_t\|_*^2$

Proof: $F_{t+1} = F(x, Y_{t+1})$

$= h(x) + g(Y_{t+1}) - \langle Y_{t+1}, x \rangle$ [By FTRL]

$= h(x) + g(Y_t + \gamma V_t) - \langle Y_t + \gamma V_t, x \rangle$

$= h(x) + g(Y_t + \gamma V_t) - \langle Y_t, x \rangle - \gamma \langle V_t, x \rangle$

$= \underbrace{h(x) + g(Y_t) - \langle Y_t, x \rangle} + g(Y_t + \gamma V_t) - g(Y_t) - \gamma \langle V_t, x \rangle$

$= F_t + g(Y_t + V_t) - g(Y_t) - \gamma \langle V_t, x \rangle$

ANALYSIS

$\langle X_t, V_t \rangle$

Control: $g(Y_t + \gamma V_t) = g(Y_t) + \gamma \langle \nabla g(Y_t), V_t \rangle + \frac{\gamma^2}{2} V_t^T \,\mathrm{Hess}(g(\xi_t))\, V_t$

$\leq g(Y_t) + \gamma \langle V_t, X_t \rangle + \frac{\gamma^2}{2K} \|V_t\|_*^2$

* Proof not entirely rigorous, but can be made rigorous via cox arg.

** NB: $\|\cdot\|_*$ is the dual norm, i.e., $\|y\|_* = \max_{\|x\| \leq 1} |\langle y, x \rangle|$

Example: If $\|x\| = \sum_j |x_j| \rightsquigarrow \|y\|_* = \max_j |y_j|$

L^1 norm $\leftrightsquigarrow L^\infty$ norm

L^p norm $\leftrightsquigarrow L^q$ norm where $\frac{1}{p} + \frac{1}{q} = 1$

L^2 norm $\leftrightsquigarrow L^2$ norm

Analysis

Continue: $F_{t+1} = F_t + g(Y_t + V_t) - g(Y_t) - \gamma\langle V_t, x\rangle$

$\leq F_t + \gamma\langle V_t, X_t\rangle + \frac{\gamma^2}{2K}\|V_t\|_*^2 - \gamma\langle V_t, x\rangle$

$= F_t + \gamma\langle V_t, X_t - x\rangle + \frac{\gamma^2}{2K}\|V_t\|_*^2$

$V_t = v_t + Z_t + b_t$

$\Rightarrow \gamma\langle v_t, x - X_t\rangle \leq F_t - F_{t+1} + \gamma\langle Z_t, X_t - x\rangle$

$+ \gamma\langle b_t, X_t - x\rangle$

$+ \frac{\gamma^2}{2K}\|V_t\|_*^2$

$\mathcal{F}(X_1, \dots, X_t)$

$\Rightarrow \gamma\mathbb{E}[\langle v_t, x - X_t\rangle \mid \mathcal{F}_t] \leq F_t - \mathbb{E}[F_{t+1} \mid \mathcal{F}_t] + \gamma\langle \mathbb{E}[Z_t \mid \mathcal{F}_t], X_t - x^*\rangle$ (with $\mathbb{E}[Z_t \mid \mathcal{F}_t] = 0$)

$+ \gamma \operatorname{diam}(\mathcal{X}) \cdot B_t$

$+ \frac{\gamma^2}{2K} M_t^2$

Analysis

$$\gamma \mathbb{E}[\langle v_t, x - X_t\rangle] \leq \gamma \mathbb{E}[F_t - F_{t+1}] + \gamma \operatorname{diam}(\mathcal{X}) B_t + \frac{\gamma^2}{2K} M_t^2$$

$$\boxed{\overline{\operatorname{Reg}}(T) \leq F_1/\gamma + \operatorname{diam}(\mathcal{X}) \sum_{t=1}^T B_t + \frac{\gamma}{2K} \sum_{t=1}^T M_t^2}$$

- If $B_t = 0$, $M_t \leq M$

$$\hookrightarrow \overline{\operatorname{Reg}}(T) \leq \frac{\max h - \min h}{\gamma} + \frac{\gamma}{2K} M^2 T$$

Take $\gamma = 1/\sqrt{T} \rightsquigarrow \boxed{\overline{\operatorname{Reg}}(T) = \mathcal{O}(\sqrt{T})}$

Regret guarantees of FTRL

Theorem (Shalev-Shwartz and Singer, 2006; Shalev-Shwartz, 2011)

- Assume:
  - *feedback unbiased and bounded in mean square* ($B_t = 0$, $\sup_t M_t < M$)
  - $\gamma = (2/M)\sqrt{KH/T}$ *with* $H = \max h - \min h$
- Then: *FTRL enjoys the bound*

$$\overline{\mathrm{Reg}}(T) \leq 2M\sqrt{(H/K)T} = \mathcal{O}(\sqrt{T})$$

Observe:

- This bound is tight [Nesterov, 2004; Abernethy et al., 2008; Bubeck, 2015]
- Cannot achieve $\mathcal{O}(1)$ regret as in continuous time [Why?]
- How to do if T is unknown? [Exercise]

Which regularizer to pick?

- Assume perfect info, $v_{a,t} \in [0,1]$ [for simplicity]

Which regularizer to pick?

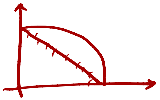
- Assume perfect info, $v_{a,t} \in [0,1]$

$\|v_t\|^2 = \sum_{i=1}^{n} v_{a,t}^2 \leq n = M^2$ [for simplicity]

- Euclidean regularization
  - L^2-norm bound $M = |\mathcal{A}|^{1/2}$
  - Strong convexity modulus $K = 1$; $H \leq 1/2$
  - Optimal tuning gives

$$\overline{\mathrm{Reg}}(T) \leq 2M\sqrt{(H/K)T} = \mathcal{O}(\sqrt{T})$$

$H = \max h - \min h$



$$\overline{\mathrm{Reg}}(T) \leq 2\sqrt{|\mathcal{A}| \cdot T}$$

Which regularizer to pick?

- Assume perfect info, $v_{a,t} \in [0,1]$ [for simplicity]

$$\overline{\mathrm{Reg}}(T) \le 2M\sqrt{(H/K)T} = \mathcal{O}(\sqrt{T})$$

- Euclidean regularization
  - L^2-norm bound $M = |\mathcal{A}|^{1/2}$
  - Strong convexity modulus $K = 1$; $H \le 1/2$
  - Optimal tuning gives

$$\overline{\mathrm{Reg}}(T) \le 2\sqrt{|\mathcal{A}| \cdot T}$$

- Entropic regularization / Exponential weights
  - L^∞-norm bound $M = 1$ $\|v_t\|_\infty = \max_a |v_{a,t}| \le 1 = M$
  - Strong convexity modulus $K = 1$; $H = \log|\mathcal{A}|$
  - Optimal tuning gives

$$\overline{\mathrm{Reg}}(T) \le 2\sqrt{\log|\mathcal{A}| \cdot T}$$

Which regularizer to pick?

- Assume perfect info, $v_{a,t} \in [0,1]$ [for simplicity]
- Euclidean regularization
  - L^2-norm bound $M = |\mathcal{A}|^{1/2}$
  - Strong convexity modulus $K = 1$; $H \leq 1/2$
  - Optimal tuning gives

$$\overline{\operatorname{Reg}}(T) \leq 2\sqrt{|\mathcal{A}| \cdot T}$$

- Entropic regularization / Exponential weights
  - L^∞-norm bound $M = 1$
  - Strong convexity modulus $K = 1$; $H = \log|\mathcal{A}|$
  - Optimal tuning gives

$$\overline{\operatorname{Reg}}(T) \leq 2\sqrt{\log|\mathcal{A}| \cdot T}$$

- Huge reduction in dimensionality!

Learning with bandit feedback

The bandit / partial info case:

- Play action $a_t \in \mathcal{A}$ according to mixed strategy $X_t \in \mathcal{X}$
- Receive payoff $u_t(a_t) = v_{a_t,t} \in [0,1]$

Learning with bandit feedback

The bandit / partial info case:

- Play action $a_t \in \mathcal{A}$ according to mixed strategy $X_t \in \mathcal{X}$
- Receive payoff $u_t(a_t) = v_{a_t,t} \in [0,1]$

- Importance weighted estimator:

$$V_{a,t} = \frac{\mathbb{1}(a_t = a)}{\mathbb{P}(a_t = a)} u_t(a_t) = \begin{cases} 0 & \text{if } a \neq a_t \\ \dfrac{u_t(a_t)}{X_{a,t}} & \text{if } a = a_t \end{cases}$$

Learning with bandit feedback

The bandit / partial info case:

- Play action $a_t \in \mathcal{A}$ according to mixed strategy $X_t \in \mathcal{X}$
- Receive payoff $u_t(a_t) = v_{a_t,t} \in [0,1]$

- Importance weighted estimator:

$$V_{a,t} = \frac{\mathbb{1}(a_t = a)}{\mathbb{P}(a_t = a)} u_t(a_t) = \begin{cases} 0 & \text{if } a \neq a_t \\ \dfrac{u_t(a_t)}{X_{a,t}} & \text{if } a = a_t \end{cases}$$

✓ Unbiased estimator [Verify this]

$$\Rightarrow \mathbb{E}[V_{a,t}] = v_{a,t}. \quad \text{Indeed: } \mathbb{E}[V_{a,t}] = \sum_{a' \in \mathcal{A}} V_{a,t} \, \mathbb{P}(a_t = a') = \frac{u_t(a)}{X_{a,t}} \mathbb{P}(a = a_t) + \sum_{a' \neq a} 0 \cdot \mathbb{P}(a_t = a')$$

Learning with bandit feedback

The bandit / partial info case:

- Play action $a_t \in \mathcal{A}$ according to mixed strategy $X_t \in \mathcal{X}$
- Receive payoff $u_t(a_t) = v_{a_t,t} \in [0,1]$

- Importance weighted estimator:

$$V_{a,t} = \frac{\mathbb{1}(a_t = a)}{\mathbb{P}(a_t = a)} u_t(a_t) = \begin{cases} 0 & \text{if } a \neq a_t \\ \dfrac{u_t(a_t)}{X_{a,t}} & \text{if } a = a_t \end{cases}$$

✓ Unbiased estimator [Verify this]

✗ Not bounded in mean square! [$X_{a,t}$ can become arbitrarily small]

Possible outlets

Two approaches:

1. **Adjust the algorithm:** [valid for all regularizers]

   - Reduce variance by increasing exploration

     $$X_t \leftarrow (1-\varepsilon_t)Q(Y_t) + \varepsilon_t \text{unif}$$

   - Still unbiased; variance = $\mathcal{O}(1/\varepsilon_t)$
   - Adaptation of FTRL bounds yields

     $$\overline{\text{Reg}}(T) = \mathcal{O}(T^{2/3})$$ $\rightsquigarrow \Theta(T^{1/6})$ gap

     suboptimal relative to full / noisy information bound

ANALYSIS

$X_t = (1-\varepsilon_t) Q(Y_t) + \varepsilon_t \,\mathrm{unif}$

$Q(Y_t) = \frac{1}{1-\varepsilon_t} X_t - \frac{\varepsilon_t}{1-\varepsilon_t} \mathrm{unif}$

Control: $g(Y_t + \gamma V_t) = g(Y_t) + \gamma \langle \nabla g(Y_t), V_t \rangle$

$\frac{\gamma}{1-\varepsilon} \langle V_t, X_t \rangle - \frac{\varepsilon}{1-\varepsilon} \langle V_t, \mathrm{unif} \rangle$

"trembling hand"

ε-FTRL: $\overline{\mathrm{Reg}}(T) \leq \frac{F_1}{\gamma} + O(\varepsilon T) + O(\gamma T/\varepsilon)$

FTRL: $\overline{\mathrm{Reg}}(T) \leq F_1/\gamma + \mathrm{diam}(\mathcal{X}) \sum_{t=1}^{T} B_t + \frac{\gamma}{2K} \sum_{t=1}^{T} M_t^2$

Possible outlets

Two approaches:

1. **Adjust the algorithm:** [valid for all regularizers]
   - Reduce variance by increasing exploration
   $$X_t \leftarrow (1-\varepsilon_t)Q(Y_t) + \varepsilon_t \mathrm{unif}$$
   - Still unbiased; variance = $\mathcal{O}(1/\varepsilon_t)$
   - Adaptation of FTRL bounds yields
   $$\overline{\mathrm{Reg}}(T) = \mathcal{O}(T^{2/3})$$

2. **Adjust the analysis:** [only valid for (EW)]
   - Derive refined bound on KL divergence [not clear for other FTRL]
   - Refined bound suitable for variance growth up to $\mathcal{O}(1/\min|X_{a,t}|)$
   - Almost tight bound:
   $$\overline{\mathrm{Reg}}(T) = \mathcal{O}(\sqrt{|\mathcal{A}|\log|\mathcal{A}|\cdot T})$$

→ Auer et al. 1995 "Gambling in a rigged casino"

[Log factor can be shaved off, cf. Audibert and Bubeck, 2010]

Recap

Quick recap:

- In general, no-regret learning does not converge to equilibrium ✗
- Multi-agent FTRL echoes replicator properties
- Discrete-time analysis $\rightsquigarrow$ next lecture
- **Regret guarantees:** $\mathcal{O}(1)$ in continuous time, $\mathcal{O}(\sqrt{T})$ in discrete [both tight]
- Non-Euclidean regularization can be **very** beneficial [EW algo]
- Bandit framework much harder, but still possible to achieve $\mathcal{O}(\sqrt{T})$ ✓

References I

J. Abernethy, P. L. Bartlett, A. Rakhlin, and A. Tewari. Optimal strategies and minimax lower bounds for online convex games. In *COLT '08: Proceedings of the 21st Annual Conference on Learning Theory*, 2008.

S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: A meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

J.-Y. Audibert and S. Bubeck. Regret bounds and minimax policies under partial monitoring. *Journal of Machine Learning Research*, 11:2635–2686, 2010.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, 1995.

S. Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3-4):231–358, 2015.

S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

Y. Freund and R. E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*, volume 2 of *Economic learning and social evolution*. MIT Press, Cambridge, MA, 1998.

References II

D. Fudenberg and J. Tirole. *Game Theory*. The MIT Press, 1991.

J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.

T. Lattimore and C. Szepesvári. *Bandit Algorithms*. Cambridge University Press, Cambridge, UK, 2020.

N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108 (2):212–261, 1994.

Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Number 87 in Applied Optimization. Kluwer Academic Publishers, 2004.

N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.

W. H. Sandholm. *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge, MA, 2010.

S. Shalev-Shwartz. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.

S. Shalev-Shwartz and Y. Singer. Convex repeated games and Fenchel duality. In *NIPS' 06: Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pages 1265–1272. MIT Press, 2006.

S. Sorin. Exponential weight algorithm in continuous time. *Mathematical Programming*, 116(1):513–528, 2009.

References III

V. G. Vovk. Aggregating strategies. In *COLT '90: Proceedings of the 3rd Workshop on Computational Learning Theory*, pages 371–383, 1990.

J. W. Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, MA, 1995.