The Annals of Applied Probability
2012, Vol. 22, No. 2, 457–476
DOI: 10.1214/10-AAP751

ON ERGODIC TWO-ARMED BANDITS

BY PIERRE TARRÈS[1] AND PIERRE VANDEKERKHOVE

CNRS, Université de Toulouse and Université Paris-Est

A device has two arms with *unknown deterministic payoffs* and the aim is to asymptotically identify the best one without spending too much time on the other. The Narendra algorithm offers a stochastic procedure to this end. We show under weak ergodic assumptions on these deterministic payoffs that the procedure eventually chooses the best arm (i.e., with greatest Cesaro limit) with probability one for appropriate step sequences of the algorithm. In the case of i.i.d. payoffs, this implies a "quenched" version of the "annealed" result of Lamberton, Pagès and Tarrès [*Ann. Appl. Probab.* **14** (2004) 1424–1454] by the law of iterated logarithm, thus generalizing it.

More precisely, if $(\eta_{\ell,i})_{i\in\mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$, $\ell \in \{A, B\}$, are the deterministic reward sequences we would get if we played at time i, we obtain infallibility with the same assumption on nonincreasing step sequences on the payoffs as in Lamberton, Pagès and Tarrès [*Ann. Appl. Probab.* **14** (2004) 1424–1454], replacing the i.i.d. assumption by the hypothesis that the empirical averages $\sum_{i=1}^{n} \eta_{A,i}/n$ and $\sum_{i=1}^{n} \eta_{B,i}/n$ converge, as n tends to infinity, respectively, to θ_A and θ_B, with rate at least $1/(\log n)^{1+\varepsilon}$, for some $\varepsilon > 0$.

We also show a fallibility result, that is, convergence with positive probability to the choice of the wrong arm, which implies the corresponding result of Lamberton, Pagès and Tarrès [*Ann. Appl. Probab.* **14** (2004) 1424–1454] in the i.i.d. case.

1. Introduction.

1.1. *General introduction.* The so-called two-armed bandit is a device with two arms, each one yielding an outcome in {0, 1} at each time step, irrespective of the strategy of the player, who faces the challenge of choosing the best one without losing too much time on the other.

The Narendra algorithm is a stochastic procedure devised to this end which was initially introduced by Norman [12] and Shapiro and Narendra [14] (see also [9, 10]) in the fields of mathematical psychology and learning automata. An application to optimal adaptive asset allocation in a financial context has been developed by Niang [11].

Received May 2009; revised November 2010.

[1]On leave from the University of Oxford. Supported in part by the Swiss National Foundation Grant 200021-1036251/1 and by a Leverhulme Prize.

MSC2010 subject classifications. Primary 62L20, 62L20; secondary 93C40, 91E40, 68T05, 91B32.

Key words and phrases. Convergence, ergodicity, stochastic algorithms, two-armed bandit.

Formally, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The Narendra two-armed bandit algorithm is defined as follows. At each time step $n \in \mathbb{N}$, we play source A (resp., source B) with probability X_n (resp., $1 - X_n$), where $X_0 = x \in (0, 1)$ is fixed and X_n is updated according to the following rule, for all $n \geq 0$:

$$X_{n+1} = \begin{cases} X_n + \gamma_{n+1}(1 - X_n), & \text{if } U_{n+1} = A \text{ and } \eta_{A,n+1} = 1, \\ (1 - \gamma_{n+1})X_n, & \text{if } U_{n+1} = B \text{ and } \eta_{B,n+1} = 1, \\ X_n, & \text{otherwise,} \end{cases} \tag{1}$$

where $(\gamma_n)_{n\geq 1}$ is a deterministic sequence taking values in $(0, 1)$, U_{n+1} is the random variable corresponding to the label of the arm played at time $n + 1$ and $\eta_{\ell,n+1}$ denotes the payoff, taking values in $\{0, 1\}$, of source $\ell \in \{A, B\}$ at time $n + 1$.

We assume without loss of generality that $U_{n+1} = A\mathbb{1}_{\{I_{n+1} \leq X_n\}} + B\mathbb{1}_{\{I_{n+1} > X_n\}}$, where $(I_n)_{n\geq 1}$ is a sequence of independent uniformly distributed random variables on $[0, 1]$.

The literature on this algorithm generally assumes that the sequences $(\eta_{A,n})_{n\geq 1}$ and $(\eta_{B,n})_{n\geq 1}$ are independent with Bernoulli distributions of parameters θ_A and θ_B, where $\theta_A > \theta_B$, the aim being to determine whether $(X_n)_{n\in\mathbb{N}}$ a.s. converges to 1 or not as n tends to infinity.

Notwithstanding the apparent simplicity of this stochastic procedure, the first criteria on a.s. convergence to "the good arm" under the above i.i.d. assumptions were only obtained thirty years after the original definition of this Narendra algorithm by Tarrès [15] and Lamberton, Pagès and Tarrès [6] in a more general framework. Recently Lamberton and Pagès established the corresponding rate of convergence [4] and proposed and studied a penalized version [5]. Note that a game theoretical question arising in the context of two-armed bandits was recently studied by Benaïm and Ben Arous [1] and Pagès [13].

Our work focuses on the understanding of the Narendra two-armed bandit algorithm under the assumption that the payoff sequences $(\eta_{\ell,n})_{n\geq 1}$, $\ell \in \{A, B\}$, are *unknown* and *deterministic*. Under the following condition (S) on the step sequence (required in [6] but without monotonicity) and weak ergodic assumption (E2) on the rate at which A must be asymptotically better than B, we show that X_n a.s. converges to 1. Heuristically, the result points out that, even with strongly dependent outcomes, X_n accumulates sufficient statistical information on the ergodic behavior of the two arms to induce a corresponding appropriate decision.

More precisely, let us introduce the following *step sequence* and *ergodic* assumptions.

Step sequence conditions. Let, for all $n \in \mathbb{N} \cup \{\infty\}$, $\Gamma_n = \sum_{k=1}^{n} \gamma_k$.
Let (S1) and (S2) be the following assumptions on the step sequence $(\gamma_n)_{n\in\mathbb{N}}$:

(S1) $(\gamma_n)_{n\geq 1}$ is nonincreasing and $\Gamma_\infty = \infty$;
(S2) $\gamma_n = O(\Gamma_n e^{-\theta_B \Gamma_n})$.

Let (S) be the set of conditions (S1) and (S2).

Ergodic conditions. Let (E) be the assumption that the ouputs of arms A and B satisfy

$$\text{(E)} \qquad \frac{1}{n}\sum_{k=1}^{n}\eta_{A,k} \underset{n\to\infty}{\longrightarrow} \theta_A \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{n}\eta_{B,k} \underset{n\to\infty}{\longrightarrow} \theta_B,$$

where θ_A, $\theta_B \in (0,1)$. The ergodic condition (E) means that the average payoff of arm A (resp., arm B) is θ_A (resp., θ_B) but does not assume anything on the corresponding rate of convergence. In order to introduce conditions on this rate, let us denote, for all $n \in \mathbb{N}$,

$$R_n := \max_{\ell\in\{A,B\}}\left|\sum_{i=1}^{n}(\eta_{\ell,i} - \theta_\ell)\right|.$$

Given a map $\phi : \mathbb{N} \longrightarrow \mathbb{R}_+$ and $\theta_A, \theta_B \in (0,1)$, let us denote by

(Eϕ) the assumption that $R_n/\phi(n) \underset{n\to\infty}{\longrightarrow} 0$.

Let (E1) and (E2) be condition (Eϕ), respectively, with the following assumption on ϕ:

(E1) ϕ is nondecreasing concave on $[k_0, \infty)$ for some $k_0 \in \mathbb{N}$ and

$$\sup_{n\in\mathbb{N}} \gamma_n \phi(n) < \infty.$$

(E2) $\phi(n) = \frac{n}{(\log(n+2))^{1+\varepsilon}}$ for some $\varepsilon > 0$.

Note that (E) corresponds to (Eϕ) with $\phi(n) = n$, $n \in \mathbb{N}$, under which (E1) holds, for instance, in the case of a step sequence $\gamma_n = c/(c+n)$, $c > 0$. Also, Lemma 1, proved in Section 2, implies that (S)–(E2) $\Longrightarrow$ (E1).

LEMMA 1. *If condition* (S) *holds*, *then*

$$\limsup_{n\to\infty} \frac{\gamma_n n}{\log n} \le \limsup_{n\to\infty} \frac{\Gamma_n}{\log n} \le 1/\theta_B.$$

Theorems 2 and 3 provide assumptions for convergence of the Narendra sequence $(X_n)_{n\ge 0}$ toward 0 or 1 as n tends to infinity, respectively, convergence toward 1 when $\theta_A > \theta_B$ (i.e., asymptotic choice of the "right arm").

THEOREM 2. *Under assumptions* (S1)–(E1), *the Narendra sequence* $(X_n)_{n\in\mathbb{N}}$ *converges* $\mathbb{P}_x$*-a.s. toward* 0 *or* 1 *as n tends to infinity.*

THEOREM 3. *Under assumptions* (S)–(E2) *and* $\theta_A > \theta_B$, *the Narendra sequence* $(X_n)_{n\in\mathbb{N}}$ *converges* $\mathbb{P}_x$*-a.s. toward* 1 *as n tends to infinity.*

Recall that the above conditions (E1) and (E2) are purely deterministic. If we let the sequences $(\eta_{A,i})_{i\in\mathbb{N}}$ and $(\eta_{B,i})_{i\in\mathbb{N}}$ be distributed as i.i.d. sequences with expectations θ_A and θ_B, then (E2) almost surely occurs as a consequence of the law of iterated logarithm. Assuming (S) and $\theta_A > \theta_B$, Theorem 3 implies that the algorithm $(X_n)_{n\in\mathbb{N}}$ almost surely converges to 1, which is a generalization of the corresponding infallibility Proposition 5 proved by Lamberton, Pagès and Tarrès in [6] for nonincreasing step sequences $(\gamma_n)_{n\in\mathbb{N}}$.

In practice, the Narendra algorithm is used in the context of performance assessment, or in applications either in automatic control or in financial mathematics and the i.i.d. assumption looks rather unrealistic since the performance depends in general on parameters that evolve slowly and randomly in time. The following framework provides a possible generalization.

Suppose that $(S_{\ell,i})_{i\in\mathbb{N}}$, $\ell \in \{A, B\}$, are ergodic stationary Markov chains taking values in a measurable space $(\mathbb{X}, \mathcal{X})$, with transition kernel Q_ℓ and stationary initial distribution π_ℓ. Let us consider a measurable event $\mathcal{C} \in \mathcal{X}$, and define sequences $(\eta_{\ell,i})_{i\in\mathbb{N}}$, for $\ell \in \{A, B\}$, as

$$\eta_{\ell,i} = \mathbb{1}_{\{S_{\ell,i}\in\mathcal{C}\}}, \qquad i \in \mathbb{N}. \tag{2}$$

These random sequences $(\eta_{\ell,i})_{i\in\mathbb{N}}$ are functions of the states of the Markov chains and satisfy, as a consequence, the ergodic condition (E), with

$$\theta_\ell = \pi_\ell(S_{\ell,0} \in \mathcal{C}).$$

The sequences $(S_{\ell,i})_{i\in\mathbb{N}}$, $\ell \in \{A, B\}$, represent the agents' outputs from which $(\eta_{\ell,i})_{i\in\mathbb{N}}$ extracts scores through target assessment. Note that, contrary to $(S_{\ell,i})_{i\in\mathbb{N}}$, $(\eta_{\ell,i})_{i\in\mathbb{N}}$ is not Markov in general.

Miao and Yang [8] establish under weak conditions (concerning mainly the transition kernels Q_ℓ) the law of iterated logarithm for additive functionals of Markov chains, thus providing the required ergodic rate of convergence (E2).

Let us now show a simple fallibility result that will also imply the corresponding result of [6] in the i.i.d. case.

THEOREM 4. *Assume* $\theta_A > \theta_B$ *and* $\sum_{n\geq 0}\prod_{k=1}^{n}(1 - \gamma_k\eta_{B,k}) < \infty$. *Then* $\mathbb{P}(\lim_{n\to\infty} X_n = 0) > 0$.

REMARK 1.1. In the case where $(\eta_{B,k})_{k\geq 0}$ is an i.i.d. sequence of random variables, then

$$\mathbb{E}_x\left(\sum_{n\geq 0}\prod_{k=1}^{n}(1 - \gamma_k\eta_{B,k})\right) = \sum_{n\geq 0}\prod_{k=1}^{n}(1 - \gamma_k\theta_B) < \infty$$

ensures that the third condition of Theorem 4 is fulfilled and, therefore, Theorem 4 implies the fallibility result Theorem 1(b) in [6].

REMARK 1.2. In the general (ergodic) case, if $\sum \gamma_n^2 < \infty$, $\sum \Gamma_n |\phi''(n)| < \infty$ and $\limsup \Gamma_n |\phi'(n)| < \infty$, then the proof of Lemma 10 implies that the conditions of Theorem 4 are equivalent to $\sum \exp(-\Gamma_n \theta_B) < \infty$ and $\theta_A > \theta_B$. These assumptions hold, for instance, if $\gamma_n = c/(c+n)$ and $\phi(n) = n/(\log(n+2))^{1+\varepsilon}$ for some $\varepsilon > 0$ and $c\theta_B > 1$ (see also the proof of Lemma 10).

PROOF OF THEOREM 4. Recall that $X_0 = x \in (0,1)$. Let A be the event

$$A := \{\forall k \geq 1, I_k \leq X_k\} = \left\{\forall n \geq 0, X_n = x \prod_{k=1}^{n} (1 - \gamma_k \eta_{B,k})\right\}.$$

Then

$$\mathbb{P}(A) = \prod_{n=1}^{\infty} \left(1 - x \prod_{k=1}^{n} (1 - \gamma_k \eta_{B,k})\right) > 0 \quad \iff \quad \sum_{n \geq 0} \prod_{k=1}^{n} (1 - \gamma_k \eta_{B,k}) < \infty,$$

and note that this last predicate, which is the second assumption of the theorem, obviously implies $\sum \gamma_n \eta_{B,n} = \infty$. Now, a.s. on A,

$$X_n \leq x \exp\left(-\sum_{k=1}^{n} \gamma_n \eta_{B,n}\right) \underset{n \to \infty}{\longrightarrow} 0,$$

which concludes the proof. $\square$

Notation. The letter C will denote a positive real constant that may change from one inequality to the other.

We write ϕ' and ϕ'' for the first- and second-order discrete derivatives of ϕ: for all $n \geq 1$,

$$\phi'(n) := \phi(n) - \phi(n-1) \quad \text{and} \quad \phi''(n) := \phi(n-1) + \phi(n+1) - 2\phi(n).$$

We let, for all $n \in \mathbb{N}$,

$$\alpha_n := R_n / \phi(n), \qquad \beta_n := \sup_{k \geq n} \alpha_k.$$

Note that, under assumption (Eϕ), $\alpha_n, \beta_n \underset{n \to \infty}{\longrightarrow} 0$.

Given two real sequences $(u_n)_{n \geq 0}$ and $(v_n)_{n \geq 0}$, we write

$$u_n = \square(v_n),$$

when, for all $n \geq 0$, $|u_n| \leq |v_n|$.

1.2. *Sketch of the proofs of Theorems 2 and 3.* Our first aim is to write down in Proposition 5 the evolution of $(X_n)_{n\geq 0}$ as a stochastic perturbation of the Cauchy–Euler procedure defined by

$$x_{n+1} = x_n + \gamma_{n+1} h(x_n), \tag{3}$$

where $h(x) := (\theta_A - \theta_B) f(x)$, with $f(x) := x(1-x)$.

However, contrary to the case of i.i.d. payoff sequences $(\eta_{\ell,n})_{n\geq 0}$, $\ell \in \{A, B\}$, considered in [6], the perturbation of the scheme (3) under an ergodic assumption (E) does not only consist of a martingale, but also of an increment whose importance depends on ϕ, that is, on the rate of convergence of the mean payoffs to θ_A and θ_B. More precisely let, for all $n \geq 1$,

$$\wedge_n = \sum_{k=1}^{n} \gamma_k f(X_{k-1})\big(\eta_{A,k} - \eta_{B,k} - (\theta_A - \theta_B)\big)$$

with the convention that $\wedge_0 = 0$ and let $(M_n)_{n\geq 1}$ be an $(\mathcal{F}_n)_{n\geq 1}$-adapted martingale given by

$$M_n := \sum_{k=1}^{n} \gamma_k \varepsilon_k, \qquad M_0 := 0$$

with

$$\varepsilon_k := \eta_{A,k}(1 - X_{k-1})(\mathbb{1}_{U_k = A} - X_{k-1}) + \eta_{B,k} X_{k-1}\big((1 - X_{k-1}) - \mathbb{1}_{U_k = B}\big).$$

PROPOSITION 5. *For all $n \in \mathbb{N}$,*

$$X_n = x + M_n + \wedge_n + (\theta_A - \theta_B) \sum_{k=1}^{n} \gamma_k f(X_{k-1}).$$

PROOF. The updating rule (1) can be rewritten as

$$\begin{aligned} X_{n+1} &= X_n + \gamma_{n+1}\eta_{A,n+1}(1 - X_n)\mathbb{1}_{U_{n+1}=A} - \gamma_{n+1}\eta_{B,n+1} X_n \mathbb{1}_{U_{n+1}=B} \\ &= X_n + \gamma_{n+1}\eta_{A,n+1}(1 - X_n)(\mathbb{1}_{U_{n+1}=A} - X_n) \\ &\quad + \gamma_{n+1}\eta_{B,n+1} X_n\big((1 - X_n) - \mathbb{1}_{U_{n+1}=B}\big) \\ &\quad + \gamma_{n+1} f(X_n)(\eta_{A,n+1} - \eta_{B,n+1}). \end{aligned} \tag{4}$$

□

Note that Proposition 5 can be interpreted as the property that the noise is multiplicative in the sense that, for all n,

$$\gamma_{n+1}^{-1}(\Lambda_{n+1} - \Lambda_n) = f(X_n)\big(\eta_{A,k} - \eta_{B,k} - (\theta_A - \theta_B)\big)$$

is the product of a function of X_n and a function of $(\eta_{A,n+1}, \eta_{B,n+1})$ outcome of the two arms at time $n + 1$.

Let us now provide estimates of the evolution of $(\wedge_n)_{n\in\mathbb{N}}$, which will be necessary to the proof of Theorem 3; they will also imply Theorem 2 in passing. We note that Laruelle and Pagès [7] recently generalized the proof of this latter result as convergence of the ergodic dynamics toward an equilibrium point of the corresponding ODE under the assumption that the noise is multiplicative and a classical Lyapounov assumption, or more generally under a strong Lyapounov assumption, and technical conditions.

Our estimates of $\Lambda_n - \Lambda_m$ for large m and n are derived by discrete integration by parts. To this end, we need to get round the difficulty that the sequence $(\gamma_n f(X_{n-1}))_{n\in\mathbb{N}}$ is not monotonic in general.

Instead, let us define, for all $n \in \mathbb{N}$,

$$\Delta_n := \frac{\gamma_n}{\prod_{k=1}^n (1-\gamma_k)}, \qquad S_n := \frac{1}{\prod_{k=1}^n (1-\gamma_k)}$$

with the convention that $\Delta_0 = S_0 := 1$. Remark that $S_n \to \infty$ if and only if $\sum_{n\geq 1} \gamma_n = +\infty$.

Note that x/S_n is a trivial lower bound for X_n and that

$$\gamma_n = \frac{\Delta_n}{S_n} \qquad \text{with } S_n = \sum_{k=0}^n \Delta_k. \tag{5}$$

We first study the sequence $(\Psi_n)_{n\in\mathbb{N}}$ defined by

$$\Psi_n := \sum_{k=n+1}^{\infty} \frac{\gamma_k}{S_{k-1}} \big(\eta_{A,k} - \eta_{B,k} - (\theta_A - \theta_B)\big);$$

$(\Psi_n)_{n\geq 1}$ is well defined since, for all $\ell \in \{A, B\}$,

$$\sum_{k=2}^{\infty} \frac{\gamma_k}{S_{k-1}} |\eta_{\ell,k} - \theta_\ell| \leq \sum_{k=2}^{\infty} \frac{\gamma_k}{S_{k-1}} = \sum_{k=2}^{\infty} \left(\frac{1}{S_{k-1}} - \frac{1}{S_k}\right) = \frac{1}{S_1}$$

since under (S1) we have $S_n \underset{n\to\infty}{\longrightarrow} \infty$. Since $(\gamma_n/S_{n-1})_{n\in\mathbb{N}}$ is a nonincreasing sequence if $(\gamma_n)_{n\in\mathbb{N}}$ is itself nonincreasing [recall that $\gamma_n \in (0,1)$], we deduce Lemma 6 by an Abel transform, that is, discrete integration. Moreover, we observe that, for all $n \geq m \geq 0$, the evolution of $\wedge_\cdot$ between time steps m and n is given by

$$\wedge_n - \wedge_m = \sum_{k=m+1}^{n} S_{k-1} f(X_{k-1}) \frac{\gamma_k}{S_{k-1}} \big(\eta_{A,k} - \eta_{B,k} - (\theta_A - \theta_B)\big).$$

Now, $(S_k f(X_k))_{k\in\mathbb{N}}$ is a nondecreasing sequence. Indeed, for all $k \in \mathbb{N}$, $f(X_k) \geq (1-\gamma_k) f(X_{k-1})$ since f is concave and X_k is the barycentre of X_{k-1} and either 0 or 1, with weights $1-\gamma_k$ and γ_k, where $f(0) = f(1) = 0$. We rely on this monotonicity and apply an Abel transform again, which enables us to show Lemma 7.

LEMMA 6. *Assume that $(\gamma_n)_{n\in\mathbb{N}}$ is nonincreasing and that ϕ is nondecreasing concave on $[k_0, \infty)$ for some $k_0 \in \mathbb{N}$. Then, for all $n \geq k_0$,*

$$|\Psi_n| \leq \frac{2\beta_n}{S_{n-1}}[\phi'(n) + 2\gamma_n\phi(n)].$$

LEMMA 7. *Let, for all $n \in \mathbb{N}$,*

$$R'_n := \frac{2\sup_{k\geq n}\beta_k[\phi'(k) + 2\gamma_k\phi(k)]}{1-\gamma_n}.$$

Under the assumptions of Lemma 6 we have, for all $n \geq m \geq k_0$,

$$|\wedge_n - \wedge_m| \leq R'_m\left[\sum_{k=m+1}^{n}\gamma_k f(X_{k-1}) + 2f(X_n)\right].$$

Lemmas 6 and 7 are proved in Sections 3.2 and 3.3.

These results enable us to conclude the proof of Theorem 2. Indeed, by Proposition 5 and Lemma 7, for all $n \geq m \geq 0$,

$$\begin{aligned} X_n - X_m &= M_n - M_m + \wedge_n - \wedge_m + (\theta_A - \theta_B)\sum_{k=m+1}^{n}\gamma_k f(X_{k-1}) \\ &= M_n - M_m + \bigl(\theta_A - \theta_B + \square(R'_m)\bigr)\sum_{k=m+1}^{n}\gamma_k f(X_{k-1}) \qquad (6) \\ &\quad + 2\square(R'_m) f(X_n). \end{aligned}$$

We assume that (E1) and (S1) hold; thus, $R'_n \underset{n\to\infty}{\longrightarrow} 0$. Let us prove by contradiction that

$$\sum_{k=1}^{\infty}\gamma_k f(X_{k-1}) < \infty \qquad \text{a.s.} \tag{7}$$

holds. Indeed, let us assume the contrary; choose m such that $|R'_m| < |\theta_A - \theta_B|$. A.s. on $\{\sum_{k=1}^{\infty}\gamma_k f(X_{k-1}) = \infty\}$, using Chow's lemma (see, e.g., [3]) and $\mathbb{E}(\varepsilon_{k+1}^2|\mathcal{F}_k) \leq 2f(X_k)$, we deduce

$$M_n - M_m = o\left(\sum_{k=m+1}^{n}\gamma_k^2 f(X_{k-1})\right) = o\left(\sum_{k=m+1}^{n}\gamma_k f(X_{k-1})\right)$$

and, therefore, for all $n, m \in \mathbb{N}$,

$$X_n - X_m = \bigl(\theta_A - \theta_B + \square(R'_m) + o_{n\to\infty}(1)\bigr)\sum_{k=m+1}^{n}\gamma_k f(X_{k-1}) + O(1),$$

which is contradictory using $X_n \in [0, 1]$ for all $n \in \mathbb{N}$.

Hence, $\mathbb{P}_x$-almost surely, $(X_n)_{n\geq 0}$ is a Cauchy sequence and, therefore, converges to a limit random variable $X_\infty \in [0, 1]$. Now (7) implies that $f(X_\infty) = 0$, since $\Gamma_\infty = \infty$ and, therefore, $X_\infty = 0$ or 1 a.s.

The proof of Theorem 3 itself has two parts. The first one consists in showing a "brake phenomenon," that is, that $(X_n)_{n\geq 0}$ cannot in any case decrease too rapidly to 0 as n goes to infinity. We already observed that, trivially, X_n is lower bounded by x/S_n. A better lower bound can easily be obtained; let us define, for all $n \in \mathbb{N}$,

$$S_n^B := \frac{1}{\prod_{k=1}^n (1 - \gamma_k \mathbb{1}_{\{I_k > X_{k-1}, \eta_{B,k}=1\}})} \qquad \text{with initial condition } S_0^B = 0$$

and, for all $n \geq 1$,

$$\Delta_n^B := \gamma_n S_n^B, \qquad Y_n^B := S_n^B X_n.$$

Note that, as a consequence of the definition of the Narendra algorithm (1), for all $n \geq 0$,

$$Y_{n+1}^B = \begin{cases} Y_n^B + \Delta_{n+1}^B (1 - X_n), & \text{if } U_{n+1} = A \text{ and } \eta_{A,n+1} = 1, \\ Y_n^B, & \text{otherwise.} \end{cases} \tag{8}$$

Roughly speaking, S_n^B is the product S_n restricted to playing and winning with B; x/S_n^B is straightforwardly a lower bound of X_n. Proposition 8, proved in Section 4.1, further claims that, for any $C > 0$, $C \log S_n^B / S_n^B$ is an asymptotic lower bound of X_n a.s. on $\{X_\infty = 0\}$.

PROPOSITION 8. *Under assumptions* (S) *and* (E2),

$$\Big\{\lim_{n\to\infty} X_n = 0\Big\} \subseteq \Big\{\limsup_{n\to\infty} \frac{X_n}{\log S_n^B / S_n^B} = \infty\Big\}, \qquad \mathbb{P}_x\text{-}a.s.$$

The second part of the proof of Theorem 3 assumes $\theta_A > \theta_B$ and is given in Section 4.2. Recall that, by Theorem 2, X_n converges a.s. to 0 or 1 [using the remark that (S)–(E2) implies (E1), see the remark before the statement of Lemma 1] so that we only need to show that $\mathbb{P}(\lim X_n = 0) = 0$.

We study $(X_n)_{n\geq 0}$ as a perturbed Cauchy–Euler scheme and prove by Doob's inequality that, starting from $C \log S_n^B / S_n^B$ for sufficiently large $C > 0$, X_n remains bounded away from 0 with lower bounded probability, which enables us to conclude that $X_\infty \neq 0$ a.s.

2. Deterministic estimates on the step sequence. We first recall below the two following preliminary remarks in [6] that (S2) implies on one hand that $\sum_{n=1}^\infty \gamma_n^2 < \infty$ and, on the other hand, that $\Gamma_n - \log S_n$ converges as n goes to infinity.

Then we prove Lemma 1 that (S) implies explicit asymptotic upper bounds on $(\gamma_n)_{n\in\mathbb{N}}$ and $(\Gamma_n)_{n\in\mathbb{N}}$.

PRELIMINARY REMARK 1. Assumption (S2) implies $\sum_{n=1}^{\infty} \gamma_n^2 < \infty$ since, for all $n \in \mathbb{N}$,

$$\sum_{k=1}^{n} \gamma_k^2 \le C \sum_{k=1}^{n} (\Gamma_k - \Gamma_{k-1}) \Gamma_k e^{-\theta_B \Gamma_k}$$
$$\le C \int_0^{\Gamma_n} u e^{-\theta_B u} \, du \le C \int_0^{+\infty} u e^{-\theta_B u} \, du < \infty$$

using that $u \mapsto u e^{-\theta_B u}$ is nonincreasing for $u > \theta_B^{-1}$.

PRELIMINARY REMARK 2. The partial sums S_n and Γ_n satisfy for every $n \ge 1$,

$$\log S_n - \sum_{k=1}^{n} \frac{\gamma_k^2}{1 - \gamma_k} \le \Gamma_n \le \log S_n. \tag{9}$$

This follows from the easy comparisons

$$\Gamma_n = \sum_{k=1}^{n} \frac{\Delta_k}{S_k} \begin{cases} \le \displaystyle\int_1^{S_n} \frac{du}{u} = \log S_n, \\ = \displaystyle\sum_{k=1}^{n} \frac{S_{k-1}}{S_k} \int_{S_{k-1}}^{S_k} \frac{du}{S_{k-1}} \ge \sum_{k=1}^{n} (1 - \gamma_k) \int_{S_{k-1}}^{S_k} \frac{du}{u}, \\ \ge \log S_n - \displaystyle\sum_{k=1}^{n} \frac{\gamma_k^2}{1 - \gamma_k}. \end{cases}$$

PROOF OF LEMMA 1. The first inequality is elementary, since $\Gamma_n \ge n\gamma_n$, using that $(\gamma_n)_{n \ge 1}$ is a nonincreasing sequence by (S1). By assumption (S2), for some $C > 0$, for all $n \in \mathbb{N}$,

$$C \ge \frac{\gamma_n e^{\theta_B \Gamma_n}}{\Gamma_n}.$$

Using that $u \mapsto e^{\theta_B u}/u$ is increasing on $[1/\theta_B, \infty)$ we obtain that, for sufficiently large $n_0 \in \mathbb{N}$,

$$C(n - n_0) \ge \int_{\Gamma_{n_0}}^{\Gamma_n} \frac{e^{\theta_B x}}{x} \, dx \underset{n \to \infty}{\sim} \frac{e^{\theta_B \Gamma_n}}{\theta_B \Gamma_n}. \tag{10}$$

Trivially, $\log(e^{\theta_B \Gamma_n}/\theta_B \Gamma_n) \sim_{n \to \infty} \theta_B \Gamma_n$, so that (10) proves the second inequality. □

3. Abel transforms.

3.1. *Preliminary estimates.* Lemmas 9 and 10 estimate the error in replacing the payoffs $\eta_{\ell,k}$ by their "average success rate" θ_ℓ in a sum weighted by a decreasing sequence $(\xi_n)_{n\in\mathbb{N}}$, by the use of Abel transforms, that is, discrete integrations by parts. More precisely let, for all $n \in \mathbb{N}$ and $\ell \in \{A, B\}$,

$$\Phi^{\ell}_{n,\xi} = \sum_{k=1}^{n} \xi_k(\eta_{\ell,k} - \theta_\ell)$$

be the corresponding deviation. Lemma 9 upper bounds $|\Phi^{\ell}_{n,\xi} - \Phi^{\ell}_{m,\xi}|$ for all $n \geq m$, whereas Lemma 10 shows that $\Phi_{n,\xi}$ converges to a finite value under certain assumptions, which are fulfilled, for instance, when $\xi := \gamma$ and (S)–(E2) hold.

Lemma 9 is the main tool in the proof of Lemmas 6 and 7 and the second part of Lemma 10 will be useful in the proof of Proposition 8 providing "brake phenonemon" bounds.

LEMMA 9. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a positive real-valued nonincreasing sequence. Assume ϕ is nondecreasing on $[k_0, \infty)$ for some $k_0 \in \mathbb{N}$, then, for all $n \geq m \geq k_0$,*

$$|\Phi^{\ell}_{n,\xi} - \Phi^{\ell}_{m,\xi}| \leq \beta_m\left(\sum_{k=m+1}^{n} \xi_k\phi'(k) + 2\xi_m\phi(m)\right).$$

PROOF. Let, for all $n \in \mathbb{N}$ and $\ell \in \{A, B\}$, $\kappa^{\ell}_n := \sum_{k=1}^{n}(\eta_{\ell,k} - \theta_\ell)$. If $n \geq m \geq k_0$, then

$$\begin{aligned}
\Phi^{\ell}_{n,\xi} - \Phi^{\ell}_{m,\xi} &= \sum_{k=m+1}^{n} \xi_k(\eta_{\ell,k} - \theta_\ell)\\
&= \sum_{k=m+1}^{n} \xi_k(\kappa^{\ell}_k - \kappa^{\ell}_{k-1}) = \sum_{k=m+1}^{n} \xi_k\kappa^{\ell}_k - \sum_{k=m}^{n-1} \xi_{k+1}\kappa^{\ell}_k \qquad (11)\\
&= \sum_{k=m}^{n-1}(\xi_k - \xi_{k+1})\kappa^{\ell}_k + \xi_n\kappa^{\ell}_n - \xi_m\kappa^{\ell}_m.
\end{aligned}$$

Now, using that $(\xi_n)_{n\geq 0}$ is nonincreasing,

$$\begin{aligned}
&\left|\sum_{k=m}^{n-1}(\xi_k - \xi_{k+1})\kappa^{\ell}_k\right|\\
&\quad\leq \sum_{k=m}^{n-1}(\xi_k - \xi_{k+1})R_k = \sum_{k=m}^{n-1}(\xi_k - \xi_{k+1})\alpha_k\phi(k)\\
&\quad\leq \beta_m \sum_{k=m}^{n-1}(\xi_k - \xi_{k+1})\phi(k) = \beta_m\left(\sum_{k=m}^{n-1} \xi_k\phi(k) - \sum_{k=m+1}^{n} \xi_k\phi(k-1)\right) \qquad (12)\\
&\quad= \beta_m\left(\sum_{k=m+1}^{n} \xi_k\big(\phi(k) - \phi(k-1)\big) + \xi_m\phi(m) - \xi_n\phi(n)\right).
\end{aligned}$$

In summary, (11) and (12) imply

$$|\Phi^{\ell}_{n,\xi} - \Phi^{\ell}_{m,\xi}| \le \beta_m \left(\sum_{k=m+1}^{n} \xi_k \big(\phi(k) - \phi(k-1)\big) + 2\xi_m \phi(m) \right)$$
$$= \beta_m \left(\sum_{k=m+1}^{n} \xi_k \phi'(k) + 2\xi_m \phi(m) \right). \qquad \square$$

REMARK 3.1. Under assumption (E2), that is, when $\phi(k) := k(\log(k+2))^{-(1+\varepsilon)}$ for some $\varepsilon > 0$, then

$$\phi'(k) \le \frac{1}{(\log(k+1))^{1+\varepsilon}}, \qquad k \in \mathbb{N}.$$

Indeed, for all $x \in \mathbb{R}^+$,

$$\left(\frac{d\phi}{dx}\right)(x) = \frac{1}{(\log(x+2))^{1+\varepsilon}} - \frac{(1+\varepsilon)x}{(x+2)(\log(x+2))^{2+\varepsilon}}$$

and

$$\phi'(k) \le \sup_{x \in [k-1,k]} \left(\frac{d\phi}{dx}\right)(x).$$

LEMMA 10. *Given a positive real-valued nondecreasing sequence* $(\xi_n)_{n\in\mathbb{N}}$, *let, for all* $n \in \mathbb{N}$, $\Xi_n := \sum_{k=1}^{n} \xi_k$. *If* ϕ *is nondecreasing on* $[k_0, \infty)$ *for some* $k_0 \in \mathbb{N}$, $\sum_{k=1}^{\infty} \Xi_k |\phi''(k)| < \infty$ *and* $\limsup_{n\in\mathbb{N}} \Xi_n |\phi'(n)| = 0$ *then, for all* $\ell \in \{A, B\}$, $(\Phi^{\ell}_{n,\xi})_{n\in\mathbb{N}}$ *converges to a finite real value as n goes to infinity.*

In particular, under assumptions (S) *and* (E2), *for all* $\ell \in \{A, B\}$, $(\Phi^{\ell}_{n,\gamma})_{n\in\mathbb{N}}$ *and* $(\Phi^{\ell}_{n,\gamma/\Gamma})_{n\in\mathbb{N}}$ [*where* $\gamma = (\gamma_n)_{n\in\mathbb{N}}$ *and* $\gamma/\Gamma = (\gamma_n/\Gamma_n)_{n\in\mathbb{N}}$] *converge to a finite real value as n goes to infinity.*

PROOF. For all $m, n \ge k_0$ with $n \ge m$, Lemma 9 implies

$$|\Phi^{\ell}_{n,\xi} - \Phi^{\ell}_{m,\xi}| \le \beta_m \left(\sum_{k=m+1}^{n} \xi_k \phi'(k) + 2\xi_m \phi(m) \right).$$

But

$$\sum_{k=m+1}^{n} \xi_k \phi'(k) = \sum_{k=m+1}^{n} (\Xi_k - \Xi_{k-1})\phi'(k) = \sum_{k=m+1}^{n} \Xi_k \phi'(k) - \sum_{k=m}^{n-1} \Xi_k \phi'(k+1)$$
$$= \sum_{k=m}^{n-1} \Xi_k \big(\phi'(k) - \phi'(k+1)\big) - \Xi_m \phi'(m) + \Xi_n \phi'(n)$$
$$= -\sum_{k=m}^{n-1} \Xi_k \phi''(k) - \Xi_m \phi'(m) + \Xi_n \phi'(n).$$

Let us now prove the convergence of $(\Phi^{\ell}_{n,\gamma})_{n\in\mathbb{N}}$ under assumptions (S)–(E2). Then $\Gamma_n = O(\log n)$ by Lemma 1 and $\phi'(n) = o(\frac{1}{\log n})$ (see Remark 3.1) so that $\Gamma_n \phi'(n) \underset{n\to\infty}{\longrightarrow} 0$. Now, there exist λ, $\mu \in (0, 1)$ such that

$$\begin{aligned}
|\phi''(k)| &= \big|\big(\phi(k+1) - \phi(k)\big) - \big(\phi(k) - \phi(k-1)\big)\big| \\
&= \left|\frac{d\phi}{dx}(k+\mu) - \frac{d\phi}{dx}(k-\lambda)\right| \\
&\le 2 \sup_{x\in[k-1,k+1]} \left|\left(\frac{d^2\phi}{dx^2}\right)\right|
\end{aligned}$$

and

$$\begin{aligned}
\left(\frac{d^2\phi}{dx^2}\right)(x) &= \frac{1+\varepsilon}{(x+2)(\log(x+2))^{2+\varepsilon}} \\
&\quad \times \left[-2 + \frac{x}{x+2}\left(1 + \frac{2+\varepsilon}{\log(x+2)}\right)\right] \\
&= O\left(\frac{1}{x(\log(x+2))^{2+\varepsilon}}\right), \qquad x \in \mathbb{R}^+ \setminus \{0\},
\end{aligned}$$

so that $\sum \Gamma_k |\phi''(k)| < \infty$ and the assumptions of the first statement are fulfilled. The convergence of $(\Phi^{\ell}_{n,\gamma/\Gamma})_{n\in\mathbb{N}}$ follows similarly, since $\gamma_n/\Gamma_n = O(\gamma_n)$. $\square$

3.2. *Proof of Lemma* 6. Recall that $\Psi_\infty = 0$ [see the first paragraph after the definition of $(\Psi_n)_{n\in\mathbb{N}}$, Section 1.2]. Hence, using Lemma 9,

$$\begin{aligned}
|\Psi_n| &= \left|\sum_{k=n+1}^{\infty} \frac{\gamma_k}{S_{k-1}}\big(\eta_{A,k} - \eta_{B,k} - (\theta_A - \theta_B)\big)\right| \\
&\le 2\beta_n \left\{\sum_{k=n+1}^{\infty} \frac{\gamma_k}{S_{k-1}}\phi'(k) + 2\frac{\gamma_n}{S_{n-1}}\phi(n)\right\} \qquad (13) \\
&\le 2\beta_n \left\{\phi'(n) \sum_{k=n+1}^{\infty} \frac{\gamma_k}{S_{k-1}} + 2\frac{\gamma_n}{S_{n-1}}\phi(n)\right\},
\end{aligned}$$

where we use the concavity of ϕ in the last inequality.

Now

$$\sum_{k=n+1}^{\infty} \frac{\gamma_k}{S_k} = \sum_{k=n+1}^{\infty} \frac{\Delta_k}{S_k^2} = \sum_{k=n+1}^{\infty} \frac{S_k - S_{k-1}}{S_k^2} \le \frac{1}{S_n},$$

so that inequality (13) implies the result.

3.3. *Proof of Lemma 7*. Note that

$$
\begin{aligned}
\wedge_n - \wedge_m &= \sum_{k=m+1}^{n} S_{k-1} f(X_{k-1}) \frac{\gamma_k}{S_{k-1}} \bigl(\eta_{A,k} - \eta_{B,k} - (\theta_A - \theta_B)\bigr) \\
&= \sum_{k=m+1}^{n} S_{k-1} f(X_{k-1}) (\Psi_{k-1} - \Psi_k) \\
&= \sum_{k=m+1}^{n} \Psi_k \bigl(S_k f(X_k) - S_{k-1} f(X_{k-1})\bigr) \\
&\quad + \Psi_m S_m f(X_m) - \Psi_n S_n f(X_n).
\end{aligned} \tag{14}
$$

Recall that $(S_k f(X_k))_{k \in \mathbb{N}}$ is a nondecreasing sequence (see last paragraph before the statements of Lemmas 6 and 7) so that (14) implies, together with Lemma 6, that, for all $n \geq m \geq k_0$,

$$
\begin{aligned}
|\wedge_n - \wedge_m| &\leq R'_m \Biggl[\sum_{k=m+1}^{n} \frac{S_k f(X_k) - S_{k-1} f(X_{k-1})}{S_k} + f(X_m) + f(X_n) \Biggr] \\
&= R'_m \Biggl[\sum_{k=m+1}^{n} [f(X_k) - f(X_{k-1}) + \gamma_k f(X_{k-1})] + f(X_m) + f(X_n) \Biggr] \\
&= R'_m \Biggl[\sum_{k=m+1}^{n} \gamma_k f(X_{k-1}) + 2 f(X_n) \Biggr].
\end{aligned}
$$

4. Proof of Theorem 3.

4.1. *Brake phenomenon bound*: *Proof of Proposition 8*. Assume that (S) and (E2) hold. Let

$$
\mathcal{A} := \Bigl\{ \limsup_{n \to \infty} \frac{Y_n^B}{\log S_n^B} < \infty \Bigr\} \cap \Bigl\{ \lim_{n \to \infty} X_n = 0 \Bigr\}.
$$

In order to prove Proposition 8, that is, that $\mathbb{P}(\mathcal{A}) = 0$, we first upper bound S_n^B in Lemma 11. Then we show that $Y_n^B \underset{n \to \infty}{\longrightarrow} \infty$ a.s. on $\mathcal{A}$ in Lemma 12 so that, for every $\lambda > 0$, $X_n > \lambda / S_n^B$ for large $n \in \mathbb{N}$. Both lemmas are shown in Section 4.1.1; we finally conclude in Section 4.1.2 that $\mathcal{A}$ almost surely does not occur.

4.1.1. *Brake phenomenon*: *Preliminary estimates*.

LEMMA 11. *Under assumptions* (S)–(E2), *there exists* $L > 0$ *such that*, *for all* $n \in \mathbb{N}$, $S_n^B \leq L e^{\theta_B \Gamma_n}$ *a.s.*

PROOF. Recall that (S) implies $\sum \gamma_n^2 < \infty$ (see Preliminary Remark 1, Section 2, or Lemma 1), so that there exists $K > 0$ such that, for all $n \in \mathbb{N}$,

$$S_n^B \le K \exp\left(\sum_{k=1}^{n} \gamma_k \mathbb{1}_{\{I_k > X_k, \eta_{B,k}=1\}}\right) \qquad \text{a.s.}$$

Now observe that

$$\begin{aligned}
\sum_{k=1}^{n} \gamma_k \mathbb{1}_{\{I_k > X_k, \eta_{B,k}=1\}} &= \theta_B \Gamma_n + \sum_{k=1}^{n} \gamma_k (\eta_{B,k} - \theta_B) - \sum_{k=1}^{n} \gamma_k \eta_{B,k} \mathbb{1}_{\{I_k \le X_k\}} \\
&= \theta_B \Gamma_n + \Phi_{n,\gamma}^B - \sum_{k=1}^{n} \gamma_k \eta_{B,k} \mathbb{1}_{\{I_k \le X_k\}},
\end{aligned} \tag{15}$$

which enables us to conclude since $\Phi_{n,\gamma}^B$ converges to a finite value by Lemma 10. $\square$

LEMMA 12. *Under assumptions* (S)–(E2), $\mathcal{A} \subseteq \{\limsup_{n\to\infty} Y_n^B = \infty\}$, $\mathbb{P}_x$*-a.s.*

PROOF. There exist L, $L' > 0$ such that, for all $n \in \mathbb{N}$,

$$\frac{\gamma_{n+1} S_n^B}{\Gamma_{n+1}} \le \frac{\gamma_n S_n^B}{\Gamma_n} \le L' e^{-\theta_B \Gamma_n} S_n^B \le L L', \tag{16}$$

where we use (S2) in the first inequality and Lemma 11 in the last one.

Now

$$\begin{aligned}
\left\{\limsup_{n\to\infty} Y_n^B = \infty\right\} &= \left\{\sum_{k=1}^{\infty} (Y_{k+1}^B - Y_k^B) = \infty\right\} \\
&\supseteq \left\{\sum_{k=1}^{\infty} \frac{Y_{k+1}^B - Y_k^B}{\Gamma_k} = \infty\right\} \\
&= \left\{\sum_{k=1}^{\infty} \frac{\Delta_{k+1}^B (1 - X_k)}{\Gamma_k} \mathbb{1}_{\{U_{k+1}=A\}} \eta_{A,k+1} = \infty\right\} \\
&\supseteq \mathcal{A} \cap \left\{\sum_{k=1}^{\infty} \frac{\gamma_k S_{k-1}^B}{\Gamma_k} \mathbb{1}_{\{U_k=A\}} \eta_{A,k} = \infty\right\} \\
&= \mathcal{A} \cap \left\{\sum_{k=1}^{\infty} \frac{\gamma_k S_{k-1}^B X_{k-1}}{\Gamma_k} \eta_{A,k} = \infty\right\} \\
&\supseteq \mathcal{A} \cap \left\{\sum_{k=1}^{\infty} \frac{\gamma_k}{\Gamma_k} \eta_{A,k} = \infty\right\}.
\end{aligned}$$

We use $X_n \underset{n\to\infty}{\longrightarrow} 0$ a.s. on $\mathcal{A}$ (and $\gamma_n \to 0$) in the second inclusion, whereas, in the third equality, we apply conditional Borel–Cantelli lemma (see, e.g., [2], Theorem 2.7.33), which claims, given a filtration $\mathbb{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ and an $\mathbb{F}$-adapted bounded real sequence $(\xi_n)_{n\geq 0}$ (i.e., $\exists M > 0$ s.t. $\xi_n \leq M$ a.s.), that

$$\left\{\sum_{n\in\mathbb{N}} \xi_n = \infty\right\} = \left\{\sum_{n\in\mathbb{N}} \mathbb{E}(\xi_n|\mathcal{F}_{n-1}) = \infty\right\}.$$

Here $\xi_n := \gamma_n S_{n-1}^B \mathbb{1}_{\{U_n=A\}}\eta_{A,n}/\Gamma_n$ is bounded, using (16). The last inclusion makes use of $S_n^B X_n \geq x$ for all $n \in \mathbb{N}$.

Now $\sum \gamma_k \eta_{A,k}/\Gamma_k = \infty$ a.s. on $\mathcal{A}$, since, on one hand,

$$\sum_{k=1}^{\infty} \frac{\gamma_k}{\Gamma_k} \geq \sum_{k=1}^{\infty} \frac{\Gamma_{k+1} - \Gamma_k}{\Gamma_k} \geq \int_{\Gamma_1}^{\infty} \frac{dx}{x}$$

and, on the other hand,

$$\Phi_{n,\gamma/\Gamma}^A := \sum_{k=1}^{n} \frac{\gamma_k}{\Gamma_k}(\eta_{A,k} - \theta_A)$$

converges (deterministically) to a finite value by Lemma 10. $\square$

4.1.2. *Proof of Proposition* 8. We assume that on the contrary $\mathbb{P}(\mathcal{A}) > 0$ and reach a contradiction by proving that $\limsup_{n\to\infty} Y_n^B/\log(S_n^B) = \infty$ a.s. on $\mathcal{A}$. Note that

$$Y_n^B = \sum_{k=0}^{n-1} \Delta_{k+1}^B \mathbb{1}_{\{I_{k+1}\leq X_k\}}\eta_{A,k+1}(1 - X_k) + x$$

and let, for all $\lambda > 0$,

$$Z_n^{B,\lambda} := \sum_{k=0}^{n-1} \gamma_{k+1} S_k^B \mathbb{1}_{\{I_{k+1}\leq \lambda/S_k^B\}}\eta_{A,k+1},$$

$$\tilde{Z}_n^{B,\lambda} := \sum_{k=0}^{n-1} \gamma_{k+1} S_k^B \min\left(1, \frac{\lambda}{S_k^B}\right)\eta_{A,k+1}.$$

Almost surely on $\mathcal{A}$, $\limsup_{n\to\infty} Y_n^B = \infty$ by Lemma 12 and $\lim_{n\to\infty} X_n = \lim_{n\to\infty} \gamma_n = 0$, so that, for all $\lambda > 0$

$$\limsup_{n\to\infty} \frac{Y_n^B}{\log(S_n^B)} \geq \limsup_{n\to\infty} \frac{Z_n^{B,\lambda}}{\log(S_n^B)} \qquad \text{a.s.}$$

Fix $\lambda > 0$. To show that the right-hand side of this last inequality is infinite a.s. on $\mathcal{A}$, we aim to estimate $\mathbb{E}(Z_n^{B,\lambda}) = \mathbb{E}(\tilde{Z}_n^{B,\lambda})$ and to upper bound $\mathbb{E}((Z_n^{B,\lambda} -$

$\tilde{Z}_n^{B,\lambda})^2)$. In order to yield the latter we first observe that there exists $M > 0$ such that, for all $k \in \mathbb{N}$, $\gamma_{k+1} S_k^B \le \Delta_k^B \le M\Gamma_k$, by inequality (16).

Now

$$\begin{aligned}
&\mathbb{E}\big((Z_n^{B,\lambda} - \tilde{Z}_n^{B,\lambda})^2\big) \\
&\quad = \mathbb{E}\left(\sum_{k=0}^{n-1} (\gamma_{k+1} S_k^B)^2 \min\left(1, \frac{\lambda}{S_k^B}\right)\left(1 - \min\left(1, \frac{\lambda}{S_k^B}\right)\right)\eta_{A,k+1}\right) \qquad (17)\\
&\quad \le M\Gamma_n \mathbb{E}\left(\sum_{k=0}^{n-1} \gamma_{k+1} S_k^B \min\left(1, \frac{\lambda}{S_k^B}\right)\eta_{A,k+1}\right) = M\Gamma_n \mathbb{E}(Z_n^{B,\lambda}).
\end{aligned}$$

On the other hand, for all $M > 0$ and $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{E}(Z_n^{B,\lambda}) &= \mathbb{E}\left(\sum_{k=0}^{n-1} \gamma_{k+1} S_k^B \min\left(1, \frac{\lambda}{S_k^B}\right)\eta_{A,k+1}\right) \\
&\ge \lambda(1-\varepsilon)\mathbb{P}(\mathcal{A}) \sum_{k=k_0(\varepsilon,\lambda)}^{n-1} \gamma_{k+1}\eta_{A,k+1},
\end{aligned}$$

where we use that $S_n^B = Y_n^B / X_n \to \infty$ a.s. on $\mathcal{A}$, $k_0(\varepsilon, \lambda)$ being a constant depending on ε and λ. Now $\Phi_{n,\gamma}^A = \sum_{k=0}^{n-1} \gamma_{k+1}\eta_{A,k+1} - \Gamma_n \theta_A$ converges by Lemma 10, so that we obtain

$$\lambda\theta_A \ge \limsup_{n\to\infty} \frac{\mathbb{E}(Z_n^{B,\lambda})}{\Gamma_n} \ge \liminf_{n\to\infty} \frac{\mathbb{E}(Z_n^{B,\lambda})}{\Gamma_n} \ge \lambda\mathbb{P}(\mathcal{A})\theta_A.$$

Fix $\rho \in (0, 1)$ and let

$$B_{n,\lambda} := \{|Z_n^{B,\lambda} - \tilde{Z}_n^{B,\lambda}| \le \rho\mathbb{E}(Z_n^{B,\lambda})\}.$$

By (17) and Chebyshev's inequality,

$$\mathbb{P}(B_{n,\lambda}^c) \le \frac{M\Gamma_n}{\rho^2\mathbb{E}(Z_n^{B,\lambda})}.$$

Therefore, for all $\lambda > 0$, if we let $\mathcal{C}_\lambda := \mathcal{A} \cap \limsup_{n\to\infty} B_{n,\lambda}$,

$$\mathbb{P}(\mathcal{C}_\lambda) \ge \limsup_{n\to\infty} \mathbb{P}(\mathcal{A} \cap B_{n,\lambda}) \ge \mathbb{P}(\mathcal{A}) - \frac{M}{\lambda\rho^2\theta_A\mathbb{P}(\mathcal{A})} > 0,$$

if we choose λ such that $\lambda > M\theta_A^{-1}(\rho\mathbb{P}(\mathcal{A}))^{-2}$.

Now, almost surely on $\mathcal{C}_\lambda \subseteq \mathcal{A}$, $\tilde{Z}_n^{B,\lambda}/\Gamma_n \xrightarrow[n\to\infty]{} \lambda\theta_A$ (since $S_n^B \xrightarrow[n\to\infty]{} \infty$; see above), so that

$$\limsup_{n\to\infty} \frac{Y_n^B}{\log S_n^B} \ge \frac{\lambda(1-\rho)\theta_A}{\theta_B}$$

using that $\limsup_{n\to\infty} \log S_n^B / \Gamma_n \le \theta_B$ by Lemma 11.

Therefore,

$$\mathbb{P}\left(\left\{\limsup_{n\to\infty} \frac{Y_n^B}{\log S_n^B} = \infty\right\} \cap \mathcal{A}\right) \ge \mathbb{P}\Big(\limsup_{\lambda\in\mathbb{N},\lambda\to\infty} \mathcal{C}_\lambda\Big)$$
$$\ge \limsup_{\lambda\in\mathbb{N},\lambda\to\infty} \mathbb{P}(\mathcal{C}_\lambda) \ge \mathbb{P}(\mathcal{A}),$$

which enables us to conclude.

4.2. *Conclusion of the proof of Theorem* 3. Let, for all $n \ge 0$, $T_n^B := e^{\theta_B \Gamma_n}$. It follows from Proposition 8 that

$$\limsup_{n\to\infty} \frac{X_n}{\log T_n^B / T_n^B} = \infty \qquad \text{a.s. on } X_\infty = 0$$

using that $\limsup_{n\to\infty} S_n^B / T_n^B < \infty$ by Lemma 11.

Given $l \in \mathbb{N}$, let us estimate $\mathbb{P}(X_\infty = 0 | \mathcal{F}_l)$. Using identity (6) and the assumption $\theta_A > \theta_B$, there exists $n_0 \in \mathbb{N}$ deterministic such that, for all $n \ge m \ge n_0$,

$$\begin{aligned} X_n - X_m &= M_n - M_m + \big(\theta_A - \theta_B + \Box(R'_m)\big) \sum_{k=m+1}^{n} \gamma_k f(X_{k-1}) \\ &\quad + 2\Box(R'_m) f(X_n) \\ &\ge M_n - M_m - X_n, \end{aligned}$$

so that

$$2X_n \ge X_m + M_n - M_m. \tag{18}$$

Let $(N_n)_{n\ge l}$ be the $(\mathcal{F}_n)_{n\ge l}$ adapted martingale given by

$$N_n := \sum_{i=l+1}^{n} \gamma_i \mathbb{1}_{\{X_{i-1}\le X_l\}} \varepsilon_i, \qquad N_l := 0;$$

recall that $(\varepsilon_i)_{i\in\mathbb{N}}$ was defined before the statement of Proposition 5.

Let n_0 be sufficiently large, so that $\gamma_{n_0} \le 1/2$; then, for all $n \ge n_0$, $X_{n+1} > X_n/2$. Thus, for all $n \ge l \ge n_0$, inequality (18) implies

$$2X_n \ge X_m + N_n - N_m \ge X_l/2 + N_n - N_m, \tag{19}$$

where $m := \max\{l \le i \le n : X_i > X_l/2\}$; indeed, if $m < n$ then, for all $m \le k \le n-1$, $X_{k+1} \le X_l/2$, hence, $X_k \le X_l$; (19) also trivially holds in the case $n = m$. Hence, if $x^- := \max(-x, 0)$ denotes the negative part of x, then

$$(2X_\infty - X_l/2)^- \le \sup_{m,n\ge l} |N_n - N_m| \le 2 \sup_{n\ge l} |N_n - N_l|.$$

Therefore, by Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}(X_\infty = 0|\mathcal{F}_l) &\le \frac{4\mathbb{E}[[(2X_\infty - X_l/2)^-]^2|\mathcal{F}_l]}{X_l^2} \\ &\le 16\frac{\mathbb{E}[\sup_{n\ge l}(N_n - N_l)^2|\mathcal{F}_l]}{X_l^2}.\end{aligned} \tag{20}$$

Now observe that, for all $k \in \mathbb{N}$, $\mathbb{E}(\varepsilon_{k+1}^2|\mathcal{F}_k) \le f(X_k) \le X_k$, so that Doob's inequality implies

$$\begin{aligned}\mathbb{E}\Big[\sup_{n\ge l}(N_n - N_l)^2|\mathcal{F}_l\Big] &\le 4\mathbb{E}\left(\sum_{n=l+1}^{\infty}\gamma_n^2 \mathbb{1}_{\{X_{n-1}\le X_l\}} f(X_{n-1})\right) \\ &\le 4X_l \sum_{n=l+1}^{\infty}\gamma_n^2.\end{aligned} \tag{21}$$

Let us upper bound $\sum_{i=n+1}^{\infty}\gamma_i^2$ in terms of T_n. For sufficiently large $k \in \mathbb{N}$,

$$T_{k+1}^B - T_k^B = e^{\theta_B\Gamma_{k+1}}(1 - e^{-\theta_B\gamma_{k+1}}) \ge \frac{T_{k+1}^B\theta_B\gamma_{k+1}}{2}$$

and, on the other hand, by assumption (S),

$$\gamma_k \le C\Gamma_k e^{-\theta_B\Gamma_k} = \frac{C\log(T_k^B)}{\theta_B T_k^B}.$$

Hence, if $l \in \mathbb{N}$ was assumed sufficiently large,

$$\sum_{n=l+1}^{\infty}\gamma_n^2 \le C\sum_{n=l+1}^{\infty}(T_n^B - T_{n-1}^B)\frac{\log T_n^B}{(T_n^B)^2} \le C\int_{T_l^B}^{\infty}\frac{\log t}{t^2}\,dt \le 2C\frac{\log T_l^B}{T_l^B}. \tag{22}$$

In summary, it follows from identities (20)–(22) that

$$\mathbb{P}(X_\infty = 0|\mathcal{F}_l) \le C\frac{\log T_l^B}{X_l T_l^B}.$$

Now the bounded martingale $\mathbb{P}(X_\infty = 0|\mathcal{F}_l)$ converges, as l goes to infinity, to

$$\mathbb{1}_{\{X_\infty=0\}} \le C\liminf_{l\to\infty}\frac{\log T_l^B}{X_l T_l^B} = 0 \qquad \text{a.s.}$$

so that $\mathbb{P}(X_\infty = 0) = 0$.

Acknowledgments. We are grateful to the referee for very helpful comments.

REFERENCES

[1] BENAÏM, M. and BEN AROUS, G. (2003). A two armed type bandit problem. *Internat. J. Game Theory* **32** 3–16.

[2] DACUNHA-CASTELLE, D. and DUFLO, M. (1986). *Probability and Statistics. Vol. II*. Springer, New York. MR0837655

[3] DUFLO, M. (1996). *Algorithmes Stochastiques*. *Mathématiques and Applications* (*Berlin*) [*Mathematics and Applications*] **23**. Springer, Berlin. MR1612815

[4] LAMBERTON, D. and PAGÈS, G. (2008). How fast is the bandit? *Stoch. Anal. Appl.* **26** 603–623. MR2401408

[5] LAMBERTON, D. and PAGÈS, G. (2008). A penalized bandit algorithm. *Electron. J. Probab.* **13** 341–373. MR2386736

[6] LAMBERTON, D., PAGÈS, G. and TARRÈS, P. (2004). When can the two-armed bandit algorithm be trusted? *Ann. Appl. Probab.* **14** 1424–1454. MR2071429

[7] LARUELLE, S. and PAGÈS, G. (2010). Stochastic approximation with averaging innovation. Preprint.

[8] MIAO, Y. and YANG, G. (2008). The law of the iterated logarithm for additive functionals of Markov chains. *Statist. Probab. Lett.* **78** 265–270. MR2387547

[9] NARENDRA, K. S. and THATHACHAR, M. A. L. (1974). Learning automata—a survey. *IEEE Trans. Systems Man Cybernet.* **SMC-4** 323–334. MR0469583

[10] NARENDRA, K. S. and THATHACHAR, M. A. L. (1989). *Learning Automata—An Introduction*. Prentice Hall, Englewood Cliffs, NJ.

[11] NIANG, M. (1999). Algorithme de Narendra et application à l'allocation d'actifs. Rapport de stage de DEA, Olympia Capital Management et Univ. Marne-la-vallée (France).

[12] NORMAN, M. F. (1968). On the linear model with two absorbing barriers. *J. Math. Psych.* **5** 225–241. MR0226954

[13] PAGÈS, G. (2005). A two-armed bandit type problem revisited. *ESAIM Probab. Stat.* **9** 277–282 (electronic). MR2174870

[14] SHAPIRO, I. J. and NARENDRA, K. S. (1969). Use of stochastic automata for parameter self-optimization with multi-modal perfomance criteria. *IEEE Trans. Syst. Sci. Cybern.* **SSC-5** 352–360.

[15] TARRÈS, P. (2001). Algorithmes stochastiques et marches aléatoires renforcées. Thèse de l'ENS Cachan (France).

INSTITUT DE MATHÉMATIQUES
CNRS, UNIVERSITÉ DE TOULOUSE
118 ROUTE DE NARBONNE
31062 TOULOUSE CEDEX 9
FRANCE
E-MAIL: tarres@math.univ-toulouse.fr

UNIVERSITÉ PARIS-EST
MARNE-LA-VALLÉE, LAMA
5 BOULEVARD DESCARTES
CHAMPS-SUR-MARNE
77454 MARNE-LA-VALLÉE CEDEX 2
FRANCE
E-MAIL: pierre.vandek@univ-mlv.fr