Almost sure convergence and asymptotical normality of a generalization of Kesten's stochastic approximation algorithm for multidimensional case

Pedro Cruz
pedrocruz@ua.pt
Universidade de Aveiro – Portugal

14 June, 2005

Abstract

It is shown the almost sure convergence and asymptotical normality of a generalization of Kesten's stochastic approximation algorithm for multidimensional case.

In this generalization, the step increases or decreases if the scalar product of two subsequente increments of the estimates is positive or negative.

This rule is intended to accelerate the entrance in the 'stochastic behaviour' when initial conditions cause the algorithm to behave in a 'deterministic fashion' for the starting iterations.

1 Introduction and problem statement

We consider the problem of finding the stationary point $x^* \in \mathbb{R}^n$ of a vector field $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ using the stochastic approximation algorithm

$$x_t = x_{t-1} - \gamma(s_{t-1})y_t, \quad t = 1, 2, \ldots \tag{1}$$

$$s_t = \left(s_{t-1} + \mathrm{u}(-y_t^T y_{t-1})\right)^+, \quad t = 2, 3, \ldots \tag{2}$$

where

- $y_t = \varphi(x_{t-1}) + \xi_t$, $y_t \in \mathbb{R}^n$ is the t^{th} measure of φ perturbated by the random vector $\xi_t \in \mathbb{R}^n$;
- $a^+ := \max\{a, 0\}$;
- u is a sigmoid function;
- The random vector $x_0 \in \mathbb{R}^n$, and the random variables s_0 and s_1 are initial problem conditions of the algorithm;

- $x_t \in \mathbb{R}^n$ is the t^{th} approximation to the stationary point $x^* \in \mathbb{R}^n$ of φ.

We suppose the following assumptions apply.

Assumptions B1

1. $\{x_0, \xi_1, \xi_2, \ldots,\}$ are mutually independent random vectors where vectors ξ_i are identically distributed with mean zero $\mathrm{E}\xi_t = 0$ and finite covariance matrix $S_\xi := \mathrm{E}\,\xi_t \xi_t^T$. We denote $\mathcal{F}_t$ the $\sigma-$algebra made by random vectors $\{x_0, \xi_1, \xi_2, \ldots, \xi_t\}$ and random variables s_0 and s_1. Assume s_0, s_1 are mutually independent random variables from $\{x_0, \xi_1, \xi_2, \ldots\}$.
2. There exists positive Ω such that for each open ball $I \subset \mathrm{B}(\Omega)$, $\mathrm{P}(\xi_t \in I) > 0$.
3. $\mathrm{E}|x_0| < \infty$.

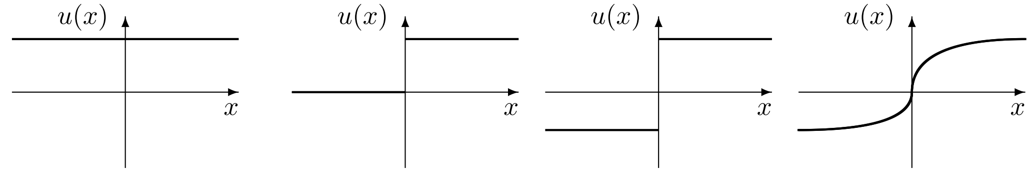
Assumptions B2

1. $\gamma(s)$ is a monotone decreasing function defined in $[0, +\infty)$ so $\gamma(0)$ will denote the maximum value of the step.
2. $\displaystyle\int_0^\infty \gamma(s) ds = \infty$.
3. $\displaystyle\int_0^\infty \gamma^2(s) ds < \infty$.

Assumptions B3

1. There exists a continuous function $V(x) : \mathbb{R}^n \to \mathbb{R}^+$ such that
   (a) $V(x^*) = 0$;
   (b) $\nabla^2 V(x) \le M$ for each x, $M > 0$ (the largest eigenvalue of $\nabla^2 V(x)$ is less than M);
   (c) $\varphi(x)^T \nabla V(x) > 0$ for each $x \ne x^*$;
   (d) For each $\gamma^* < \gamma(0)$ and for each z_0, the sequence
   $$z_t = z_{t-1} - \gamma^* \varphi(z_{t-1})$$
   converges deterministically for the stationary point x^* and verify that $\{V(z_t), t = 1, 2, \ldots\}$ is a monotonous decreasing sequence.
2. There exists positive R and β_0 such that
   $$\varphi(x)^T \nabla V(x) \ge \frac{1}{2}\gamma(0) \cdot (\varphi(x)^T M \varphi(x) + \mathrm{tr}(S_\xi M)) + \beta_0$$
   for $|x - x^*| \ge R$. This condition limits the maximum step $\gamma(0)$ and guarantees $\inf_{x \ne x^*} |\varphi(x)| > 0$.

Assumptions B4



(a) Case of Robbins-Monroe algorithm [8]. (b) Case of Kesten algorithm [2]. (c) Case similiar to Plakhov-Almeida algorithm [5]. (d) Some generic case.

Figure 1: Examples of function u.

1. u is a monotone, increasing and bounded function $\mathbb{R} \to \mathbb{R}$, for which

$$\mathrm{u}_+ = \lim_{x\to+\infty} \mathrm{u}(x) > 0 \text{ e } \mathrm{u}_- = \lim_{x\to-\infty} \mathrm{u}(x).$$

2. Denote $\mathrm{E}_\omega = \mathrm{E}[\mathrm{u}(X^{(\omega)})]$ where

$$X^{(\omega)} = \inf_{\substack{|\varphi_1|\le\omega \\ |\varphi_2|\le\omega}} \left[-(\xi_1+\varphi_1)^T(\xi_2+\varphi_2)\right].$$

   Define $\mathrm{E}_0 := \lim_{\omega\to0^+} \mathrm{E}_\omega$. Constant E_0 must be positive.

   Figure 1 shows possible example for function u where cases for known algorithms are included.

Comment 1 *Suppose we are observing the process (1), (2) starting in $t_0 > 1$. This new process, with initial conditions x_{t_0}, s_{t_0}, s_{t_0+1} and the random sequence $\xi_{t_0}, \xi_{t_0+1}, \ldots$ also satisfies conditions. Lemma 4, for example, makes use of this comment.*

Comment 2 *If* u *or the distribution of ξ_t are continuous, then* $\mathrm{E}_0 = \mathrm{E}[\mathrm{u}(-\xi_1^T\xi_2)]$. *More, if* u *is continuous and verifies* $\mathrm{u}(x) > -\mathrm{u}(-x)$ *when $x \neq 0$, then B4.2 is valid for any distribution of ξ_t with non zero variance.*

Comment 3 *We use the following notation for φ and V: φ' denotes a matrix, ∇V a vector and $\nabla^2 V$ a matrix.*

Theorem 1 *Suppose Assumptions B1 to B4 are verified. Then, almost surely,* $\lim_{t\to\infty} x_t = x^*$.

Assumptions for asymptotical normality are all assumptions for *almost sure* convergence and three more assumptions: Assumptions B3.3, B3.4 e B4.3.

Assumption B3.3 All eigenvalues of $\frac{I}{2} - (1/\mathrm{E}_0)\varphi'(x^*)$ are negative, where I is the identity matrix.

Assumption B3.4 Assume Taylor decomposition for φ,

$$\frac{|\varphi(x) - \varphi'(x^*)\,(x - x^*)|}{|x - x^*|} = \mathrm{O}((1), \text{ when } x \to x^*\,. \tag{3}$$

Comment 4 *From this assumption it follows*

$$\sup |\varphi(x)|/|x - x^*| < \infty \tag{4}$$

because

$$\frac{|\varphi(x) - \varphi'(x^*)\,(x - x^*)|}{|x - x^*|} \geq \frac{|\varphi(x)|}{|x - x^*|} - |\varphi'(x^*)|$$

and so

$$\begin{aligned} |o(1)| &\geq \frac{|\varphi(x)|}{|x - x^*|} - |\varphi'(x^*)| \\ \frac{|\varphi(x)|}{|x - x^*|} &\leq |\varphi'(x^*)| - |o(1)| < \infty \end{aligned}$$

Assumption B4.3 Assume the Taylor decomposition for function u, $\mathrm{u}(x + \Delta x) = \mathrm{u}(x) + \mathrm{u}'(\theta)\Delta x$ for θ between x and $x + \Delta x$.

Theorem 2 *Let x_t be defined by (1) and (2) for which almost sure convergence assumptions can be verified. Besides, one can also verify Assumptions B3.3, B3.4 e B4.3. If $\gamma(s) = 1/s$ then*

$$\sqrt{t}(x_t - x^*) \xrightarrow{d} \mathrm{N}(0, V) \tag{5}$$

where $\xrightarrow{d}$ denotes convergence in distribution, and V is a positive definite matrix and unique solution of the Lyapunov equation (see Theorem 3 in Section 4)

$$\left(\frac{I}{2} - (1/\mathrm{E}_0)\varphi'(x^*)\right)(-V) + (-V)\left(\frac{I}{2} - (1/\mathrm{E}_0)\varphi'(x^*)\right)^T = (1/\mathrm{E}_0)^2 S_\xi\,. \tag{6}$$

Comment 5 *The explicit solution of equation (6) is*

$$(-V) = -\int_0^\infty e^{W\cdot t} S e^{W^T\cdot t} dt$$

where $W = \frac{I}{2} - (1/\mathrm{E}_0)\varphi'(x^*)$, V is positive definite. Demonstration of this result can be find, for example, in Theorem 12.3.3 in Lancaster e Tismenetsky [3].

2 Proof of *almost sure* convergence

Demonstration of the almost sure convergence follows the work for the unidimensional case by Plakhov e Cruz (2004) [6]

Without loss of generality we suppose $x^* = 0$ so $\varphi(x^*) = 0$.

Lemma 1 *For each $\epsilon > 0$ exists $m = m(\epsilon)$ such that, almost surely, it occurs (i) exists t such that $|x_t| < \epsilon$, or (ii) exists t such that $|x_t| < R$ and $s_t \leq m$. (Remember that R is defined in B3.2)*

Proof. Choose $\epsilon > 0$ and define the stopping time

$$\tau = \tau(\epsilon, m) = \inf\{t : |x_t| < \epsilon \text{ or } (|x_t| < R \text{ and } s_t \leq m)\}.$$

Our aim is to prove that for some m we have $\mathrm{P}(\tau = \infty) = 0$.

Consider the sequence $\mathrm{E}_t = \mathrm{E}[V(x_t)\, \mathbb{I}(t < \tau)]$.

We introduce the simplified notation $V(x_t) = V_t$, $\mathbb{I}(t < \tau) = \mathbb{I}_t$, $\nabla V(x_t) = \nabla_t$, $\gamma(s_t) = \gamma_t$, and using that $\mathbb{I}_t \leq \mathbb{I}_{t-1}$, we obtain

$$\mathrm{E}_t - \mathrm{E}_{t-1} = \mathrm{E}[V_t I_t - V_{t-1}\, \mathbb{I}_{t-1}] \leq \mathrm{E}[(V_t - V_{t-1})\, \mathbb{I}_{t-1}]. \tag{7}$$

Using Taylor expansion

$$V_t = V(x_{t-1} - \gamma_{t-1} y_t) = V_{t-1} - \gamma_{t-1} y_t^T \nabla_{t-1} + \frac{1}{2}\gamma_{t-1}^2 y_t^T \nabla^2 V_{t-1}(x') y_t,$$

where x' is a point between x_t and x_{t-1}. Replacing y_t for $\varphi_{t-1} + \xi_t$ and, in agreement with B3.1, one obtains

$$V_t - V_{t-1} \leq -\gamma_{t-1}\varphi_{t-1}^T \nabla_{t-1} - \gamma_{t-1}\xi_t^T \nabla_{t-1} + \frac{1}{2}\gamma_{t-1}^2 (\varphi_{t-1}^T M \varphi_{t-1} + \xi_t^T M \xi_t). \tag{8}$$

Using (7) and (8) and observing that each values γ_{t-1}, φ_{t-1}, $\mathbb{I}_{t-1}$ is determined by x_{t-1} and s_{t-1} and so, mutually independent of ξ_t (Condition B1.1),

$$\begin{aligned}
\mathrm{E}_t - \mathrm{E}_{t-1} \leq \\
\leq\ & \mathrm{E}[-\gamma_{t-1}\varphi_{t-1}^T \nabla_{t-1} - \gamma_{t-1}\xi_t^T \nabla_{t-1} + \frac{1}{2}\gamma_{t-1}^2(\varphi_{t-1}^T M \varphi_{t-1} + \xi_t^T M \xi_t)\, \mathbb{I}_{t-1}] = \\
=\ & \mathrm{E}[-\gamma_{t-1}\varphi_{t-1}^T \nabla_{t-1}] + \mathrm{E}[-\gamma_{t-1}\xi_t^T \nabla_{t-1}] + \\
& \mathrm{E}[\frac{1}{2}\gamma_{t-1}^2(\varphi_{t-1}^T M \varphi_{t-1})\, \mathbb{I}_{t-1}] +
\end{aligned}$$

$$\mathrm{E}[\frac{1}{2}\gamma_{t-1}^2\, \mathbb{I}_{t-1}] \cdot \mathrm{E}[\xi_t^T M \xi_t]$$

then using

- $\mathrm{E}[-\gamma_{t-1}\xi_t^T \nabla_{t-1}] = 0$;
- $\mathrm{E}[\xi_t^T M \xi_t] \leq \mathrm{tr}(S_\xi M)$;

we have

$$\mathrm{E}_t - \mathrm{E}_{t-1} \leq \mathrm{E}[-\varphi_{t-1}^T \nabla_{t-1} + \frac{1}{2}\gamma_{t-1}(\varphi_{t-1}^T M \varphi_{t-1} + \mathrm{tr}(S_\xi M)))\gamma_{t-1}\, \mathbb{I}_{t-1}]\,. \tag{9}$$

If $\mathbb{I}_{t-1} = 1$, then *(i)* $|x_t| \geq R$, or *(ii)* $|x_t| \geq \epsilon$ and $s_t \geq m$. In case *(i)*, using B3.2, one obtains

$$-\varphi_{t-1}^T \nabla_{t-1} + \frac{1}{2}\gamma_{t-1}(\varphi_{t-1}^T M \varphi_{t-1} + \mathrm{tr}(S_\xi M)) \leq -\beta_0\,. \tag{10}$$

In case *(ii)* is valid that $\gamma_t < \gamma(m)$ and define $\delta_\epsilon := \inf\{\varphi(x)^T \nabla V(x),\ \text{for all } |x| \geq \epsilon\}$. In this context

$$\begin{aligned}
-\varphi_{t-1}^T \nabla_{t-1} + \frac{1}{2}\gamma_{t-1}(\varphi_{t-1}^T M \varphi_{t-1} + \mathrm{tr}(S_\xi M)) \leq \\
\leq\ & -\delta_\epsilon + \frac{1}{2}\gamma(m)(\varphi_{t-1}^T M \varphi_{t-1} + \mathrm{tr}(S_\xi M)) := -\beta(\epsilon, m)
\end{aligned} \tag{11}$$

We choose m such that $\beta(\epsilon, m) > 0$ and denote $\beta = \inf\{\beta_0, \beta(\epsilon, m)\}$. So, in both cases, the expression between parentesis in right side of (9) is less than $-\beta \cdot \gamma_{t-1}\, \mathbb{I}_{t-1}$ and so

$$\mathrm{E}_t - \mathrm{E}_{t-1} \leq -\beta \cdot \mathrm{E}[\gamma_{t-1}\, \mathbb{I}_{t-1}].$$

Using that $s_t \leq s_0 + t\mathrm{u}_+$ and $\mathrm{E}\, \mathbb{I}_t = \mathrm{P}(t < \tau)$ one have

$$\mathrm{E}_t - E_{t-1} \leq -\beta\, \gamma(s_0 + t\mathrm{u}_+)\, \mathrm{P}(t < \tau);$$

by $\mathrm{P}(j < \tau) \geq \mathrm{P}(t < \tau)$ when $j < t$ and, using induction argument,

$$E_t \leq E_1 - \beta \mathrm{P}(t < \tau) \sum_{j=0}^{t-1} \gamma(s_0 + j\mathrm{u}_+)\,.$$

where $\tilde{E}_0 := E(V(x_0)\, \mathbb{I}(0 < \nu)) < \infty$ by Assumption B1.4.

Function V is positive for $x \neq x^*$, so $E_t \geq 0$, and from here it follows

$$\mathrm{P}(t < \tau) < \frac{\tilde{E}_0}{\beta \sum_{j=0}^{t-1} \gamma(s_0 + j\mathrm{u}_+)}.$$

When $t \to \infty$ and using $\sum_{j=0}^{\infty} \gamma(s_0 + j\mathrm{u}_+) = \infty$ (inferred from Assumption B2.2), one can conclude that $P(\tau = \infty) = 0$.

□

Lemma 2 *For each $\epsilon > 0$ and $m > 0$ exists δ positive such that if $|x_0| < R$ and $s_0 \leq m$ then*

$$\mathrm{P}(\textit{exists } t, |x_t| < \epsilon) \geq \delta\,.$$

Proof. We consider function V defined in Assumptions B4. Let

$$\bar{\epsilon} = \inf\{V(x), |x| \geq \epsilon\}, \text{ and}$$
$$\bar{R} = \sup\{V(x), |x| \leq R\}$$

then $|x_0| \leq R \Rightarrow V(x_0) \leq \bar{R}$ and $V(x) < \bar{\epsilon} \Rightarrow |x| < \epsilon$.

We will show that $V(x_t) < \bar{\epsilon}$ for some t. Denote $V_t := V(x_t)$ and considering the decomposition

$$V_t = V_0 \frac{V_1}{V_0} \frac{V_2}{V_1} \cdots \frac{V_t}{V_{t-1}}$$

First define the deterministic process with constant step $\rho \leq \gamma(0)$

$$z_t = z_{t-1} - \rho\varphi(z_{t-1}), \quad t = 1, 2, \ldots$$

and by Assumption B3.1, exists $V(\cdot)$ such that $\{V(z_t)\}$ converges monotonically to zero. Using Taylor expansion

$$\begin{aligned}
V(z_t) &= V(z_{t-1} - \rho\varphi(z_{t-1})) = \\
&= V(z_{t-1}) - \rho\varphi(z_{t-1})^T \nabla V(z_{t-1}) + \\
&\quad + \frac{\rho^2}{2}\varphi(z_{t-1})^T \nabla^2 V(z')\varphi(z_{t-1}) \\
&= V(z_{t-1}) - \rho \times \\
&\quad (\varphi(z_{t-1})^T \nabla V(z_{t-1}) - \frac{\rho}{2}\varphi(z_{t-1})^T \nabla^2 V(z')\varphi(z_{t-1}))
\end{aligned}$$

for a certain vector z' between z_t and z_{t-1}. Define

$$U(z,\rho) := \frac{1}{V(z)} \times \Big(\varphi(z)^T \nabla V(z) - \frac{\rho}{2}\varphi(z)^T \nabla^2 V(z')\varphi(z)\Big)$$

where z' is a point between z and $z - \rho\varphi(z)$ and, since $V(z_t)$ decreases monotonically, then it is necessary that $U(\cdot,\cdot) > 0$. Define

$$\bar{U} := \inf_{\substack{\epsilon \le |z| \le R \\ \rho \le \gamma(0)}} U(z,\rho)$$

where $\bar{U}$ is a positive constant because $U(\cdot,\cdot) > 0$ in $\epsilon \le |z| \le R$ and $\rho \le \gamma(0)$.

Now, we consider Taylor expansion using the original process

$$\begin{aligned} V(x_t) &= V(x_{t-1} - \gamma(s_{t-1})\varphi(x_{t-1}) - \gamma(s_{t-1})\xi_t)) \\ &= V(x_{t-1} - \gamma(s_{t-1})\varphi(x_{t-1})) - \\ &\quad -\gamma(s_{t-1})\xi_t^T \nabla V(x_{t-1} - \gamma(s_{t-1})\varphi(x_{t-1})) + \frac{\gamma(s_{t-1})}{2}\xi_t^T \nabla V^2(x'')\xi_t \end{aligned}$$

and defining $\zeta_t := |\xi_t|$ we have for the last term

$$\begin{aligned} -\gamma(s_{t-1})\xi_t^T \nabla V(x_{t-1} - \gamma(s_{t-1})\varphi(x_{t-1})) + \frac{\gamma^2(s_{t-1})}{2}\xi_t^T \nabla^2 V(x'')\xi_t &\le \\ \gamma(0)\zeta_t |\nabla V(x_{t-1} - \gamma(s_{t-1})\varphi(x_{t-1}))| + \frac{\gamma^2(0)}{2}\zeta_t^2 M &\le \\ \zeta_t C_\xi & \end{aligned}$$

with the following justification

1. imposing $\zeta_t < 1$;

2. given $\epsilon \le |x| \le R$ then x_{t-1} and $\varphi(x_{t-1})$ are vectors from a closed and limited set and $\gamma(s_{t-1}) \le \gamma(0)$, so $\nabla V(x_{t-1} - \gamma(s_{t-1})\varphi(x_{t-1}))$ could be bounded.

From definition of function $U(\cdot,\cdot)$,

$$V(x_t) \le V(x_{t-1})(1 - \gamma(s_{t-1}) \cdot U(x_{t-1}, \gamma(s_{t-1}))) + \zeta_t \cdot C_\xi$$

and using $1/V(x) \le 1/\bar{\epsilon}$, for $\epsilon \le |x| \le R$, and that $\gamma(s_{t-1}) > \gamma(m + (t-1)\cdot u_+)$,

$$\begin{aligned} \frac{V_t}{V_{t-1}} &= 1 - \gamma(s_{t-1}) \cdot \bar{U} + \zeta_t \cdot C_\xi/\bar{\epsilon} \le \\ &\le 1 - \gamma(m + (t-1)u_+) \cdot \bar{U} + \zeta_t \cdot C_\xi/\bar{\epsilon}\,. \end{aligned}$$

Denoting $G_t := 1 - \gamma(m + (t-1)u_+) \cdot \bar{U}$ we have $G_t < 1$. Divergence of the series $\sum_t \gamma(m + t \cdot u_+)$ implies that the productory $\prod_{i=1}^{t-1} G_i$ goes to zero. Using that $G_t \le \sqrt{G_t} < 1$ one can choose ζ_t such that

$$G_t + \zeta_t \cdot C_\xi/\bar{\epsilon} \le \sqrt{G_t} < 1 \tag{12}$$

and

$$\frac{V_t}{V_{t-1}} \le \sqrt{G_t}$$

whenever that $\epsilon \le |x_{t-1}| \le R$ and $|\xi_t| < \zeta_t < 1$. We choose n such that $\bar{R}\prod_{i=1}^{n-1}\sqrt{G_t} < \bar{\epsilon}$ and suppose we have $|x_0| < R$, $s_0 \le m$ and $|\xi_t| < \zeta_t$ when $1 \le t \le n-1$. Then, for some $t \in \{1, \dots, n\}$, $|x_t| < \epsilon$ with probability superior to

$$\delta := \mathrm{P}(|\xi_1| < \zeta_1, |\xi_2| < \zeta_2, \dots, |\xi_n| < \zeta_n),$$

since from Assumption B1.2 $P(\xi_t \in I) > 0$, for any I.

□

From Lemmas 1 and 2 we have for each $\epsilon > 0$ that exists $\delta > 0$ such that for arbitrary initial conditions x_0, s_0, s_1

$$\mathrm{P}(\text{for some } t, |x_t| < \epsilon) > \delta.$$

Then, we can choose a positive integer number $n = n(x_0, s_0, s_1)$ such that

$$\mathrm{P}(\text{for some } t \le n, |x_t| < \epsilon) > \delta/2\,.$$

Denote $\bar{p} = \sup \mathrm{P}(\text{for each } t, |x_t| \ge \epsilon)$, being the supremum over all initial conditions x_0, s_0, s_1. Fix x_0, s_0, s_1; then

$$\begin{aligned} &\mathrm{P}(\text{for each } t, |x_t| \ge \epsilon) = \\ &= \mathrm{P}(\text{for each } t > n, |x_t| \ge \epsilon \,\Big|\, \text{for each } t \le n, |x_t| \ge \epsilon) \cdot \mathrm{P}(\text{for each } t \le n, |x_t| \ge \epsilon) \le \\ &\le \bar{p}\,(1 - \delta/2). \end{aligned} \tag{13}$$

Taking supremum of the L.S. of (13) over all triple (x_0, s_0, s_1) and denote it by $\bar{p}$. Then, we obtain the inequality $\bar{p} \le \bar{p}\,(1-\delta/2)$ from which $\bar{p} = 0$. So, we obtain the following Lemma

Lemma 3 *For each $\epsilon > 0$, almost surely exists t such that $|x_t| < \epsilon$.*

Lemma 4 *Choose $\epsilon > 0$ and $\eta > 0$. Then, exists $\epsilon_1 > 0$ and $\delta > 0$ such that if $|x_0| < \epsilon_1$ then*

$$\mathrm{P}(\textit{for some } t,\ |x_t| < \epsilon \textit{ and } s_t \ge \eta) > \delta\,.$$

Proof. Starting by $x_t = x_0 - \sum_{i=1}^{t} \gamma_{i-1} y_i$ and using Taylor expansion,

$$\begin{aligned} V(x_t) &= V(x_0 - \sum_{i=1}^{t} \gamma_{i-1} y_i) \le \\ &\le V(x_0) + |\nabla V(x_0)| \sum_{i=1}^{t} \gamma_{i-1} |y_i| \cos(y_i, \nabla V(x_0)) + C_1 |\sum_{i=1}^{t} \gamma_{i-1} y_i|^2\,. \end{aligned}$$

To guarantee the increase in step counter s_t required by this Lemma we consider two conical symmetrical sections where vectors y_t will stay and where we impose a maximum and a minimum length for $|y_t|$, $y_I \le |y_t| \le y_{II}$, with y_I, y_{II} to be defined. We take x_0 as a reference point with gradient $\nabla_0 := \nabla V(x_0)$. As we will see, we are interested in limiting the internal product

$$y^T \nabla V(x_0) = |y_t| \cdot |\nabla_0| \cdot \cos(y_t, \nabla_0)$$

We choose y_{odd} belongs to the conical section on the opposite side of vector ∇_0 and y_{even} to the conical section. We choose a value θ for the internal angle of the cone centrered in vector ∇_0 with θ belonging to $(0, \pi/2)$. In this case $\cos(y_t, \nabla_0)$ is limited by

$$-1 \quad \le \cos(y_t, \nabla_0) \le \quad -\cos(\theta), \quad t \text{ odd}, \tag{14}$$

$$\cos(\theta) \quad \le \cos(y_t, \nabla_0) \le \quad 1, \quad t \text{ even}. \tag{15}$$

Using (14) and (15) we have

$$-y_{II} \le |y_t| \cos(y_1, \nabla_0) \le -y_I \cos(\theta), \quad \text{odd case}, \tag{16}$$

$$y_I \cos(\theta) \le |y_t| \cos(y_2, \nabla_0) \le y_{II}, \quad \text{even case}. \tag{17}$$

It is possible to show $V(x_t) < \bar{\epsilon}$ if we prove

$$V(x_0) \quad < \quad \bar{\epsilon}/3; \tag{18}$$

$$\left| \sum_{i=1}^{t} \gamma_{i-1} |y_i| |\nabla_0| \cos(y_i, \nabla_0) \right| \quad < \quad \bar{\epsilon}/3; \tag{19}$$

$$C_1 \Big| \sum_{i=1}^{t} \gamma_{i-1} y_i \Big|^2 \quad < \quad \bar{\epsilon}/3. \tag{20}$$

From (18) we can estimate ϵ_1 by Assumption B3.3.

From (20) we conclude

$$C_1 \Big| \sum_{i=1}^{t} \gamma_{i-1} y_i \Big|^2 \le C_1 y_{II}^2 \sum_{i=1}^{\infty} \gamma_{i-1}^2 < \bar{\epsilon}/3 \tag{21}$$

and from where we can choose y_{II} (by Assumption B2.2 the series is convergent).

Because y_t belongs to symmmetrical conical sections,

$$\text{u}(-y_t^T y_{t-1}) \le \text{u}(y_I^2 \cos(\pi - \theta)) = \text{u}(-y_I^2 \cos\theta), \quad t = 1, 2, \ldots, n-1$$

therefore

$$s_t \ge (t-2)\text{u}(-y_I^2 \cos\theta), \quad t = 3, 4, \ldots, n. \tag{22}$$

To satisfy $s_t \ge \eta$ required by this Lemma's statement, we assume $y_I \ge y_{II}/2$, and

$$n - 2 \ge \frac{\eta}{\text{u}(-(y_{II}^2/4)\cos\theta)} \tag{23}$$

obtained from (22).

Developing the L.S. of (19) we have by (16) and (17),

$$\begin{aligned} -y_{II} \sum_{\substack{i=1 \\ \text{(odd)}}}^{t} \gamma_{i-1} + y_I \cos(\theta) \sum_{\substack{i=1 \\ \text{(even)}}}^{t} \gamma_{i-1} \leq & \\ \leq \quad \sum_{i=1}^{t} \gamma_i |y_i| \, |\nabla_0| \, \cos(y_i, \nabla_0) \leq & \qquad (24) \\ \leq \quad -y_I \cos(\theta) \sum_{\substack{i=1 \\ \text{(odd)}}}^{t} \gamma_{i-1} + y_{II} \sum_{\substack{i=1 \\ \text{(even)}}}^{t} \gamma_{i-1} \, . & \end{aligned}$$

Odd sum is bigger than even sum if we start at $i = 1$. So

$$\left| \sum_{i=1}^{t} \gamma_{i-1} |y_i| \, |\nabla_0| \, \cos(y_i, \nabla_0) \right| \leq y_{II} \sum_{\substack{i=1 \\ \text{(odd)}}}^{t} \gamma_{i-1} - y_I \cos(\theta) \sum_{\substack{i=1 \\ \text{(even)}}}^{t} \gamma_{i-1} \qquad (25)$$

Using (25), Condition (19) is satisfied if

$$y_{II} \sum_{\substack{i=1 \\ \text{(odd)}}}^{t} \gamma_{i-1} - y_I \cos(\theta) \sum_{\substack{i=1 \\ \text{(even)}}}^{t} \gamma_{i-1} \leq \bar{\epsilon}/3 \qquad (26)$$

where we can choose $y_I \geq y_{II}/2$.

For each iteration t the values of $\varphi(x_t) := \varphi_t$, y_I, y_{II}, θ are known. Let

$$v_t := \frac{(\varphi_{t-1} + \xi_t)^T \nabla_0}{|y_t| \cdot |\nabla_0|}$$

and the conditions that define the admissible region for each random vector ξ_t are

$$\begin{aligned} & y_I \leq |\varphi_{t-1} + \xi_t| \leq y_{II} \\ & \pi \leq \cos^{-1}(v_t) \leq \pi - \theta, \quad t \text{ odd} \\ & 0 \leq \cos^{-1}(v_t) \leq \theta, \quad t \text{ even.} \end{aligned} \qquad (27)$$

We define δ_1 as the smallest probability of the regions defined in each iteration $t = 1, \ldots, n$ and define $\delta := \delta_1^n$. Probability δ_1 is positive by Assumption B1.3.

□

From Lemmas 3 and 4 it follows that for each $\epsilon > 0$ and $\eta > 0$ the probability that for some t, $|x_t| < \epsilon$ and $s_t \geq \eta$ be greater than a positive δ, will depend only on ϵ and η. Repeating the argument of Lemma 3 we have

Lemma 5 *For each $\epsilon > 0$ and $\eta > 0$, almost surely exists t such that $|x_t| < \epsilon$ and $s_t \geq \eta$.*

We define the stopping time $\tau(\epsilon) = \inf\{t : |x_t| \geq \epsilon\}$.

Lemma 6 *For each $0 < \theta < \mathrm{E}_0$ exists a constant $\epsilon_0 > 0$ and a sequence π_n such that $\lim_{n\to\infty} \pi_n = 0$ and*

$$\mathrm{P}(s_t > s_0 + t\theta - n \textit{ for each } t < \tau(\epsilon_0)) > 1 - \pi_n.$$

Proof. We will show that

$$\mathrm{P}(\text{exists } t < \tau(\epsilon_0) \text{ such that } s_t \leq s_0 + t\theta - n) \leq \pi_n \to 0\,.$$

From B4.2 it follows that for some ω_0 positive exists $\mathrm{E}_{\omega_0} > \theta$ where $\mathrm{E}_{\omega_0} = \mathrm{E}[u(X^{(\omega_0)})]$ and

$$X^{(\omega_0)} = \inf_{\substack{|\varphi_1| \leq \omega_0 \\ |\varphi_2| \leq \omega_0}} [-(\xi_1 + \varphi_1)^T (\xi_2 + \varphi_2)]. \tag{28}$$

We choose ϵ_0 such that

$$\sup_{|x| < \epsilon_0} |\varphi(x)| \leq \omega_0$$

and define the sequence $\{\tilde{s}_t\}$ by

$$\tilde{s}_0 = s_0; \quad \tilde{s}_t = \tilde{s}_{t-1} + \mathrm{u}(X_t^{(\omega_0)}) \tag{29}$$

where

$$X_t^{(\omega_0)} = \inf_{\substack{|\varphi_{t-1}| \leq \omega_0 \\ |\varphi_{t-2}| \leq \omega_0}} [-(\xi_t + \varphi_{t-1})^T (\xi_{t-1} + \varphi_{t-2})]. \tag{30}$$

Comparing (29) and (30) with (2), for $t < \tau(\epsilon_0)$, we obtain

$$\tilde{s}_t \leq s_t. \tag{31}$$

From (29) it follows that

$$\tilde{s}_t - s_0 = t\mathrm{E}_{\omega_0} + \mathbb{I}_t^{\mathrm{even}} + \mathbb{I}_t^{\mathrm{odd}} \tag{32}$$

where

$$\mathbb{I}_t^{\mathrm{even}} = \sum_{\substack{i=1 \\ (i \text{ even})}}^{t} [\mathrm{u}(X_t^{(\omega_0)}) - \mathrm{E}_{\omega_0}], \quad \mathbb{I}_t^{\mathrm{odd}} = \sum_{\substack{i=1 \\ (i \text{ odd})}}^{t} [\mathrm{u}(X_t^{(\omega_0)}) - \mathrm{E}_{\omega_0}]$$

where $\mathbb{I}_t^{\mathrm{even}}$ and $\mathbb{I}_t^{\mathrm{odd}}$ are sums of independent and identically distributed random variables with mean zero and variance linear with t.

Comment 6 *Both variables $\mathbb{I}_t^{even}$ e $\mathbb{I}_t^{odd}$ are asymptotical normal however they are dependent from each others. We use the following argument to estimate the probability of their sum: $X+Y < a$ implies $X < a/2$ or $Y < a/2$ where X and Y are random variables and a a real constant. Then,*

$$\mathrm{P}(X + Y < a) \leq \mathrm{P}(X < a/2) + \mathrm{P}(Y < a/2) \simeq 2\mathrm{P}(X < a/2).$$

So, using that $\mathrm{Var}\, \mathbb{I}_t^{\mathrm{even}} = t \cdot V_{\mathbb{I}_1}$, we have

$$\mathrm{P}(\mathbb{I}_t^{\mathrm{even}} + \mathbb{I}_t^{\mathrm{odd}} < 2a) \lesssim 2\mathrm{P}(\mathbb{I}_t^{\mathrm{even}} < a) \leq 2\Phi(\frac{a}{\sqrt{t}\sqrt{V}_{\mathbb{I}_1}}). \tag{33}$$

From the event $s_t \leq s_0 + t\theta - n$, we know that $\tilde{s}_t \leq s_t$ for $t < \tau(\epsilon_0)$. It follows

$$\begin{aligned} \tilde{s}_t &\leq s_0 + t\theta - n \Leftrightarrow \\ s_0 + t\mathrm{E}_{\omega_0} + \mathbb{I}_t^{\mathrm{even}} + \mathbb{I}_t^{\mathrm{odd}} &\leq s_0 + t\theta - n \Leftrightarrow \\ \mathbb{I}_t^{\mathrm{even}} + \mathbb{I}_t^{\mathrm{odd}} &\leq -t(\mathrm{E}_{\omega_0} - \theta) - n\,. \end{aligned} \tag{34}$$

Comment 7 *We will use the following argument, where* $\{X_i, i = 1, \ldots\}$ *is a sequence of random variables,*

$$\mathrm{P}(\textit{exists } t < \tau \textit{ such that } X_t < a) \leq \sum_{i=1}^{\tau} \mathrm{P}(X_i < a) \leq \sum_{i=1}^{\infty} \mathrm{P}(X_i < a)\,. \tag{35}$$

By (33), (34) and (35) it follows

$$\begin{aligned}
\mathrm{P}(\text{exists } t < \tau(\epsilon_0) \text{ such that } s_t \leq s_0 + t\theta - n) &\leq \\
\mathrm{P}(\text{exists } t < \tau(\epsilon_0) \text{ such that } \mathbb{I}_t^{\text{even}} + \mathbb{I}_t^{\text{odd}} \leq -t(\mathrm{E}_{\omega_0} - \theta) - n) &\leq \\
\sum_{i=1}^{\infty} \mathrm{P}(\mathbb{I}_i^{\text{even}} + \mathbb{I}_i^{\text{odd}} \leq -i(\mathrm{E}_{\omega_0} - \theta) - n) &\lesssim \\
2\sum_{i=1}^{\infty} \mathrm{P}(\frac{\mathbb{I}_i^{\text{even}}}{\sqrt{iV_I}} \leq -\sqrt{i}\frac{\mathrm{E}_{\omega_0} - \theta}{\sqrt{V}_I} - \frac{n}{\sqrt{iV_I}}) &\leq \\
2\sum_{i=1}^{\infty} \Phi(-\sqrt{i}K_1 - \frac{n}{\sqrt{i}}K_2) &:= \pi_n
\end{aligned}$$

for certain constants $K_1 > 0$ and $K_2 > 0$. Last series is convergent and so $\pi_n \to 0$, then

$$\begin{aligned}
\pi_n := \mathrm{P}(&\text{exists } t \text{ such that } \mathbb{I}_t^{\text{even}} + \mathbb{I}_t^{\text{odd}} \leq \\
&\leq -n - t(\mathrm{E}_{\omega_0} - \theta)) \to 0 \text{ when } n \to \infty.
\end{aligned}$$

□

Now, choose θ and ϵ_0 as in Lemma 6, and arbitrarily positive values $\epsilon < \epsilon_0$ and n, and define the stopping time

$$\nu = \nu(n, \epsilon) = \inf\{t : |x_t| \geq \epsilon \text{ or } s_t \leq s_0 - n + t\theta\}$$

and choose $\epsilon_1 > 0$ such that

$$\sup_{|x|<\epsilon_1} V(x) < \frac{1}{2} \inf_{|x|>\epsilon} V(x).$$

Lemma 7 *Let* $|x_0| < \epsilon_1$*, so*

$$\mathrm{P}(\nu < \infty) \leq K \int_{s_0-n-1}^{\infty} \gamma^2(s)ds + \pi_n,$$

where K *is a constant depending on* ϵ*.*

Proof. Using (8) on Lemma 1,

$$V_t - V_{t-1} \leq -\gamma_{t-1}\varphi_{t-1}^T \nabla V_{t-1} - \gamma_{t-1}\xi_t^T \nabla V_{t-1} + 1/2\gamma_{t-1}^2(\varphi_{t-1}^T M\varphi_{t-1} + \xi_t^T M\xi_t)$$

and let $V_t - V_0 \leq I'_t + I''_t$ where

$$\begin{aligned}
I'_t &= \left|\sum_{i=1}^{t} \gamma_{i-1}\varphi_{i-1}^T \nabla V_{i-1} + \gamma_{i-1}\xi_i^T \nabla V_{i-1}\right| \\
I''_t &= 1/2\sum_{i=1}^{t} \gamma_{i-1}^2(\varphi_{i-1}^T M\varphi_{i-1} + \xi_i^T M\xi_i).
\end{aligned}$$

Let $\delta := (1/2)\inf_{|x|>\epsilon} V(x)$. For $|x_t| > \epsilon$ then $V_t - V_0 > \delta$, therefore,

$$I'_t + I''_t \geq V_t - V_0 > \delta,$$

implying $I'_t > \delta/2$ or $I''_t > \delta/2$. We wish to estimate $\mathrm{P}(\nu < \infty)$. Denote

$$P' = \mathrm{P}(I'_\nu\, \mathbb{I}(\nu < \infty) > \delta/2)$$
$$P'' = \mathrm{P}(I''_\nu\, \mathbb{I}(\nu < \infty) > \delta/2)$$

and using Lemma 6,

$$\mathrm{P}(\nu < \epsilon) \leq \pi_n + P' + P''. \tag{36}$$

Using Markov's inequality (for example, [9, p. 59]), $\mathbb{I}^2(\cdot) = \mathbb{I}(\cdot)$, and $\mathbb{I}(i-1 < \nu < \infty) < \mathbb{I}(i-1 < \nu)$,

$$\begin{aligned} P' &\leq \frac{4}{\delta^2}\mathrm{E}[{I'_\nu}^2\, \mathbb{I}^2(\nu < \infty)] = \\ &= \frac{4}{\delta^2}\mathrm{E}\left[\left(\sum_{i=1}^{\nu-1}\gamma_{i-1}(\varphi_{i-1}^T + \xi_i^T)\nabla V_{i-1}\right)^2 \cdot\, \mathbb{I}(\nu < \infty)\right] \\ &= \frac{4}{\delta^2}\sum_{i,j=1}^{\infty}\mathrm{E}[\gamma_{i-1}(\varphi_{i-1}^T + \xi_i^T)\nabla V_{i-1}\, \mathbb{I}(i-1 < \nu) \times \\ &\qquad \times\gamma_{j-1}(\varphi_{j-1}^T + \xi_j^T)\nabla V_{j-1}\, \mathbb{I}(j-1 < \nu)]. \end{aligned}$$

Recall that variables γ_{i-1}, V_{i-1}, $\mathbb{I}(i-1 < \nu)$ and ξ_i are mutually independent. We conclude that terms with $i \neq j$ are zero. So,

$$P' \leq \frac{4}{\delta^2}\sum_{i=1}^{\infty}\mathrm{E}[\gamma_{i-1}^2(\varphi_{i-1}^T\nabla V_{i-1})^2(\xi_i^T\nabla V_{i-1})^2\, \mathbb{I}(i-1 < \nu)] \leq K'\mathrm{E}\sum_{i=1}^{\nu-1}\gamma_{i-1}^2 \tag{37}$$

where K' is a constant that verifies

$$(4/\delta^2)\cdot \sup_{|x|<\epsilon}(\varphi_{i-1}^T\nabla V_{i-1})^2 \cdot \sup_{|x|<\epsilon}\mathrm{E}[\xi_i^T\nabla V_{i-1}]^2 < K'.$$

Using $\mathrm{P}(X > \delta/2) \leq \frac{\mathrm{E}|X|}{2/\delta}$,

$$P'' \leq \frac{2}{\delta}(1/2)\mathrm{E}[\sum_{i=1}^{\nu-1}\gamma_{i-1}^2(\varphi_{i-1}^T M\varphi_{i-1} + \xi_i^T M\xi_i)] \leq K''\sum_{i=1}^{\nu-1}\gamma_{i-1}^2 \tag{38}$$

where K'' verifies

$$(2/\delta)\sup_{|x|<\epsilon}\varphi_{t-1}^T M\varphi_i + \mathrm{E}\xi_i^T M\xi_i < K''$$

using $\mathrm{E}\xi\xi^T := S_\xi$.

For $t < \nu$, $s_t > s_0 + t\theta - n$, then $\gamma_t < \gamma(s_0 - n + t\theta)$, and

$$\mathrm{E}\left[\sum_{i=1}^{\nu-1}\gamma_i^2\right] < \sum_{i=1}^{\infty}\gamma^2(s_0 - n + i\theta) \leq \frac{1}{\theta}\int_{s_0-n-1}^{\infty}\gamma^2(s)ds. \tag{39}$$

Taking $K = \theta^{-1}(K' + K'')$, from (36), (37), (38) and (39) we obtain Lemma 7.

$\square$

Now, choose positive $\epsilon < \epsilon_0$ and choose n and η such that $1 - \pi_n - K\int_{\eta-n-1}^{\infty}\gamma^2(s)ds =: \delta$ be positive. Choose also $\epsilon_1 = \epsilon_1(\epsilon)$ as defined above. In agreement with Lemmas 5 and 7, *almost surely* exists t_0 such that $|x_{t_0}| < \epsilon_1$, $s_{t_0} \geq \eta$, and the probability for all $t \geq t_0$, $|x_t| < \epsilon$ exceeds δ.

We define the sequence of stopping times $\tau_1 = 1$,

$$\tau_{i+1} = \inf\{\tau > \tau_i : |x_\tau| \geq \epsilon, \text{ and for some } \tau_i \leq t < \tau,\, |x_t| < \epsilon_1 \text{ and } s_t > \eta\}, \quad i = 1, 2, \ldots .$$

We have

$$\mathrm{P}(\tau_{i+1} = \infty \,|\, \tau_i < \infty) \geq \delta,$$

from

$$\mathrm{P}(\tau_{i+1} < \infty) = \mathrm{P}(\tau_{i+1} < \infty \,|\, \tau_i < \infty)\,\mathrm{P}(\tau_i < \infty) \leq (1 - \delta)\,\mathrm{P}(\tau_i < \infty).$$

So, $\mathrm{P}(\tau_i < \infty) \to 0$ quando $i \to \infty$; implying that *almost surely* $i_0 = \sup\{i : \tau_i < \infty\}$ is finite.

In accordance to Lemma 5, *almost surely* exists $t_0 \geq \tau_{i_0}$ such that $|x_{t_0}| < \epsilon_1$ and $s_{t_0} > \eta$; from here we conclude that $|x_t| < \epsilon$ when $t > t_0$. Theorem 1 is proved. $\square$

3 Proof of the asymptotical normality

The central idea of the proof follows the work of Delyon and Juditsky (1993) [1].

Lemma 8 (Delyon e Juditsky [1]) *Let (ν_t) be a random sequence of real numbers such that $\nu_t \to 0$ almost surely when $t \to \infty$. Then exists a deterministic sequence (a_t) such that*

$$a_t \to 0 \quad \textit{and} \quad \nu_t/a_t \to 0 \quad \textit{almost surely.} \tag{40}$$

In what follows o and O have the standard deterministic meaning however many times they represent stochastic random variables belonging to $\mathcal{F}_t$ $\sigma-$algebra of events.

Lemma 9 *Let $\{z_i, i = 1, \ldots\}$ be a sequence of non-negative random variables verifying $z_i \to 0$ almost surely, and let $\{|\xi_i|\}$, be a sequence of iid random variables with finite variances. Possibly, variables z_i and ξ_i are dependent. Then*

$$\sum_{i=1}^{t} z_i\,|\xi_i| = o(t)$$

almost surely.

Proof. From Lemma 8 there exists a deterministic sequence $\{a_i\}$ such that $z_i/a_i \to 0$ *almost surely*. Then $0 \le z_i(\omega)/a_i < M(\omega)$ for each elementary event ω. Denote $\zeta_i := |\xi_i| - \mu$ where $\mu := \mathrm{E}(|\xi|)$, so $\mathrm{E}\zeta_i = 0$ and $\mathrm{Var}\zeta_i < \infty$.

Let $S_t = \sum_{i=1}^{t} a_i\zeta_i$. Then $S_t/t \to 0$ in probability by Chebychev inequality. Then, by Levy's Theorem (for example, [7] p. 211) $S_t/t \to 0$ *almost surely* because $\{a_i\zeta_i\}$ is a sequence of independent random variables. (The same result using Kronecker Lemma [7] because $\sum \mathrm{Var}(a_i\zeta_i/i) < \infty$.)

Then $S_t = o(t)$ *almost surely* and

$$\begin{aligned} \left|\sum_{i=1}^{t} \frac{z_i}{a_i} \cdot a_i \cdot |\xi_i|\right| &\le M(\omega) \cdot \sum_{i=1}^{t} a_i \cdot |\xi_i| \\ = M(\omega) \cdot \sum_{i=1}^{t} (a_i \cdot \zeta_i + a_i \cdot \mu_{|\xi|}) &= M(\omega) \cdot o(t) = o(t) \textit{ almost surely.} \end{aligned}$$

□

Recall definition of E_0 in Assumption B4.2.

Lemma 10 *Let s_0 and s_1 be random variables which are initial conditions of the process $\{s_t\}$, defined in (2). Then*

$$\gamma(s_t) = 1/s_t = \frac{1}{\mathrm{E}_0 t}(1 + o_t), \; \textit{almost surely} \tag{41}$$

where o_t is a random variable defined in $\mathcal{F}_t$ and for which $\lim_{t\to\infty} o_t = 0$ almost surely.

Proof. Assumption B4.3 permits the decomposition

$$\begin{aligned} \mathrm{u}(-y_{i-1}y_i) = \mathrm{u}(-(\varphi_{i-2}+\xi_{i-1})^T(\varphi_{i-1}+\xi_i)) = \\ = \mathrm{u}(-(\varphi_{i-2}+\xi_{i-1})^T(\varphi_{i-1}+\xi_i)) = \\ = \mathrm{u}(-\varphi_{i-2}^T\varphi_{i-1} - \varphi_{i-2}^T\xi_i - \varphi_{i-1}^T\xi_{i-1} - \xi_{i-1}^T\xi_i) = \\ = \mathrm{u}(-\xi_{i-1}^T\xi_i) + \mathrm{u}'(\theta_i) \times \left(-\varphi_{i-2}^T\varphi_{i-1} - \varphi_{i-2}^T\xi_i - \varphi_{i-1}^T\xi_{i-1}\right) \end{aligned} \tag{42}$$

where θ_i is a point between $-y_{i-1}^T y_i$ and $-\xi_{i-1}^T\xi_i$. We also have that function u' is limited and $\varphi(x_i) \to 0$ from where, by Lemma 9,

$$\sum_{i=1}^{t} \mathrm{u}'(\theta_i)\varphi_{i-2}^T\varphi_{i-1} = o(t) \tag{43}$$

$$\sum_{i=1}^{t} \mathrm{u}'(\theta_i)\varphi_{i-2}^T\xi_i = o(t) \tag{44}$$

$$\sum_{i=1}^{t} \mathrm{u}'(\theta_i)\varphi_{i-1}^T\xi_{i-1} = o(t)\,. \tag{45}$$

So, we have

$$\begin{aligned} s_t &= s_0 + s_1 + \sum_{i=1}^{t}(\mathrm{u}(-y_{i-1}^T y_i) - \mathrm{u}(-\xi_{i-1}^T\xi_i)) + \\ &\quad + \sum_{\text{even}}^{t}\mathrm{u}(-\xi_{i-1}^T\xi_i) + \sum_{\text{odd}}^{t}\mathrm{u}(-\xi_{i-1}^T\xi_i) \\ &= s_0 + s_1 + \Delta U_t + P_t + I_t. \end{aligned}$$

By (43), (44) and (45)

$$\Delta U_t = \sum_{i=1}^{t}(\mathrm{u}(-y_{i-1}y_i) - \mathrm{u}(-\xi_{i-1}\xi_i)) = o(t) \textit{ almost surely}.$$

Each of the sums P_t and I_t is composed of independent terms of mean E_0 and finite variance. By the law of iterated logarithm

$$P_t + I_t = \mathrm{E}_0 t + \mathrm{o}(\sqrt{t\log\log t})\,.$$

Using $\lim_{t\to\infty} s_0/t = 0$ *almost surely*, also for s_1, we have

$$s_t = s_0 + s_1 + \mathrm{E}_0 t + t o_t + \mathrm{o}(\sqrt{t\log\log t}) = (\mathrm{E}_0 + o_t)t,$$

almost surely. Then

$$\begin{aligned} s_t &= (\mathrm{E}_0 + o_t)t = \mathrm{E}_0 t\left(\frac{1}{1-\frac{o_t}{\mathrm{E}_0+o_t}}\right) = \\ &= \mathrm{E}_0 t\left(\frac{1}{1+o_t}\right). \end{aligned}$$

□

Demonstration of Theorem 2 We choose $x^* = 0$. From last Section, we have shown the *almost surely* convergence of $x_t \to 0$ and in Lemma 10 we shown the mean beahaviour of $s_t = \mathrm{E}_0 t(\frac{1}{1+o_t})$ where $o_t \to 0$ *almost surely*.

By Lemma 8 we can conclude that there exists a sequence (a_t) of positive non random numbers such that

$$a_t \to 0 \quad \text{and} \quad |o_t|/a_t \to 0, \quad |x_t|/a_t \to 0 \quad \textit{almost surely}. \tag{46}$$

Comment 8 *We provide an explanantion for the above fact. We can make $\theta_t := |o_t| + |x_t|$ and then $\theta_t \to 0$ almost surely. Then exists $a_t \to 0$, deterministicaly, such that $\theta_t/a_t \to 0$ almost surely. From here it follows $|o_t|/a_t \to 0$ and $|x_t|/b_t \to 0$ almost surely.*

We define the stopping times

$$\tau_R = \inf\{t : |o_t| \geq R|a_t|\}, \quad \sigma_R = \inf\{t : |x_t| \geq R|a_t|\} \tag{47}$$

for $R > 0$ and

$$\nu = \min(\tau_R, \sigma_R)\,. \tag{48}$$

From Lemma 8 and from (46) we conclude that for each $\epsilon > 0$ we can choose $R < \infty$ such that

$$\mathrm{P}(\nu = \infty) \geq 1 - \epsilon. \tag{49}$$

In this way, with a probability so large as we want we have a deterministic bound common to $|o_t|$ and $|x_t|$.

Now, consider the similar process to the algorithm in (1) but with deterministic step $\gamma_t = 1/(\mathrm{E}_0 t)$ applied to the function $\varphi(x) = \alpha x$ (α is the derivative of φ in x^*),

$$z_t = z_{t-1} - \frac{1}{\mathrm{E}_0 t}(\alpha z_{t-1} + \xi_t), \quad z_0 = x_0. \tag{50}$$

Asymptotical properties of this process are known (for example, Nevel'son e Has'minskii [4]). So

$$\begin{aligned} z_t t^{1/2-\epsilon} \to 0, \textit{almost surely},\ \text{for each } \epsilon > 0, \\ \mathrm{E}|z_t|^2 \leq K/t, \quad K > 0 \\ \sqrt{t} z_t \xrightarrow{\mathrm{d}} N(0, V). \end{aligned} \tag{51}$$

where V is the matrix defined in (6).

Based on Lemma 15 in the reference Section, Lemma 13 will show that, assimptotically, $\sqrt{t}x_t$ and $\sqrt{t}z_t$ will have the same limiting distribuition, described in (51). □

Lemma 11 *Consider the following recursive formula, where $b > 0$, a_0 are real numbers,*

$$0 \leq a_{t+1} \leq (1 - \frac{b}{t})a_t + \mathrm{O}((t^{-1}), \quad t = 1, 2, \ldots\,. \tag{52}$$

Then $a_t \to 0$.

Proof. Consider the recursive sequence, where ϵ is a positive real number,

$$0 \leq A_{t+1} \leq (1 - \frac{b}{t})A_t + \epsilon/t, \quad t = t_0, t_0 + 1, \ldots .$$

Then

$$0 \leq A_{t+1} \leq A_t - \frac{bA_t - \epsilon}{t}, \quad t = t_0, t_0 + 1, \ldots .$$

or

$$0 \leq bA_{t+1} - \epsilon \leq bA_t - \epsilon - b\frac{bA_t - \epsilon}{t}, \quad t = t_0, t_0 + 1, \ldots .$$

We write $B_t = bA_t - \epsilon$ and

$$B_{t+1} = B_t(1 - b/t)$$

so $B_t \to 0$, therefore $A_t \to \epsilon/b$.

Lemma's sequence is

$$0 \leq a_{t+1} \leq (1 - \frac{b}{t})a_t + \mathrm{O}((1)/t, \quad t = 1, 2, \ldots .$$

for which we choose $\epsilon > 0$ such that $o(1) < \epsilon$ if $t \geq t_0$ for some t_0. We define

$$A_{t+1} = (1 - \frac{b}{t})A_t + \epsilon/t, \quad t = t_0, t_0 + 1, \ldots$$

and $A_{t_0} = a_{t_0}$. Now, we show $0 \leq a_t \leq A_t$ using an induction argument. Suppose $A_t - a_t \geq 0$ for $t \geq t_0$. For $t + 1$

$$A_{t+1} - a_{t+1} = (1 - \frac{b}{t})(A_t - a_t) + (\epsilon - o(1))/t$$

verifying that $A_{t+1} - a_{t+1} \geq 0$ using hypothesis. Then $0 \leq a_t \leq A_t$.

With $A_t \to \epsilon/b$ and since we can choose a small enough ϵ, we conclude that $A_t \to 0$ and therefore $a_t \to 0$.

□

Lemma 12 *Let A be a positive definite matrix and symmetrical, a, b, c and d real vectors. Then*

$$\begin{aligned}(a + b + c + d)^T A(a + b + c + d) \quad \leq \quad & a^T Aa + \\ & +3(b^T Ab + c^T Ac + d^T Ad) + \\ & +a^T Ab + b^T Aa + \\ & +2a^T A(c + d) .\end{aligned}$$

Proof. From

$$\begin{aligned}(a - b)^T A(a - b) \quad &= \quad a^T Aa + b^T Ab - a^T Ab - b^T Aa \geq 0 \Leftrightarrow \\ &\Leftrightarrow \quad a^T Ab + b^T Aa \leq a^T Aa + b^T Ab\end{aligned}$$

we have

$$\begin{aligned}(a+b)^T A(a+b) &= a^T Aa + b^T Ab + a^T Ab + b^T Aa \\ &\le a^T Aa + b^T Ab + a^T Aa + b^T Ab \\ &= 2(a^T Aa + b^T Ab)\,.\end{aligned}$$

In a similar way

$$\begin{aligned}(a+b+c)^T A(a+b+c) &= a^T Aa + b^T Ab + c^T Ac + \\ &\quad (a^T Ab + b^T Aa) + (a^T Ac + c^T Aa) + \\ &\quad (b^T Ac + c^T Ab) \\ &\le a^T Aa + b^T Ab + c^T Ac + \\ &\quad (a^T Aa + b^T Ab) + (a^T Aa + c^T Ac) + \\ &\quad (b^T Ab + c^T Ac) \\ &= 3(a^T Aa + b^T Ab + c^T Ac)\,.\end{aligned}$$

So,

$$\begin{aligned}(a+b+c+d)^T A(a+b+c+d) &= (a+(b+c+d))^T A(a+(b+c+d)) \\ &= a^T Aa + a^T A(b+c+d) + \\ &\quad (b+c+d)^T Aa + (b+c+d)^T A(b+c+d) \\ &\le a^T Aa + 3(b^T Ab + c^T Ac + d^T Ad) + \\ &\quad a^T Ab + b^T Aa + 2a^T A(c+d)\,.\end{aligned}$$

□

Lemma 13 *Let* $\Delta_t := x_t - z_t$*. Then* $\sqrt{t}\Delta_t \overset{pr}{\to} 0$*.*

Proof. From Lemma 10, $\gamma_t = \frac{1}{s_t} = \frac{1}{E_0 t}(1+o_t)$ where o_t is a random variable of $\mathcal{F}_t$ which converges to 0 *almost surely*. Then, from (1), (2) with $\gamma_t = 1/s_t$,

$$x_{t+1} = x_t - \frac{1}{E_0 t}(1+o_t)(\varphi(x_t) + \xi_{t+1}) \tag{53}$$

and

$$x_{t+1} = x_t - \frac{1}{E_0 t}\varphi(x_t) - \frac{1}{E_0 t}\xi_{t+1} - \frac{o_t}{E_0 t}\varphi(x_t) - \frac{o_t}{E_0 t}\xi_{t+1}\,.$$

From Assumption B3.4,

$$\varphi(x) = (\varphi(x) - \varphi'(0)x) + \varphi'(0)x\,,$$

so

$$\begin{aligned} x_{t+1} = x_t &- \frac{1}{E_0 t}\varphi'(0)x_t - \frac{1}{E_0 t}\xi_{t+1} - \frac{o_t}{E_0 t}\xi_{t+1} - \\ &- \frac{1}{E_0 t}\left(o_t\varphi(x_t) + \varphi(x_t) - \varphi'(0)x_t\right) . \end{aligned}$$

Define

$$v_t := o_t \frac{\varphi(x_t)}{|x_t|} + \frac{\varphi(x_t) - \varphi'(0)x_t}{|x_t|}$$

and for $t \le \nu$ we have $|x_t| \le Ra_t$ and $|o_t| \le Ra_t$

$$\begin{aligned} |v_t| &\le Ra_t \sup_x \frac{|\varphi(x)|}{|x|} + \sup_{|x|\le Ra_t} \frac{|\varphi(x_t) - \varphi'(0)x_t|}{|x_t|} \le \\ &\le Ra_t M + o(1) := c_t . \end{aligned} \tag{54}$$

We note that $c_t \to 0$ where c_t is a positive decreasing sequence and

$$x_{t+1} = x_t - \frac{1}{E_0 t}\varphi'(0)x_t - \frac{1}{E_0 t}\xi_{t+1} - \frac{o_t}{E_0 t}\xi_{t+1} - \frac{1}{E_0 t}v_t|x_t| .$$

Considering the algorithm for z_t

$$\begin{aligned} z_{t+1} &= z_t &- \frac{1}{E_0 t}(\varphi'(0)z_t + \xi_{t+1}) = \\ &= z_t &- \frac{1}{E_0 t}\varphi'(0)z_t - \frac{1}{E_0 t}\xi_{t+1} \end{aligned}$$

and

$$\begin{aligned} x_{t+1} &= x_t - \frac{1}{E_0 t}\varphi'(0)x_t - \frac{1}{E_0 t}\xi_{t+1} - \frac{o_t}{E_0 t}\xi_{t+1} - \frac{1}{E_0 t}v_t|x_t|, \\ z_{t+1} &= z_t - \frac{1}{E_0 t}\varphi'(0)z_t - \frac{1}{E_0 t}\xi_{t+1} \end{aligned}$$

from where

$$\Delta_{t+1} = \Delta_t - \frac{1}{E_0 t}\varphi'(0)\Delta_t - \frac{1}{E_0 t}v_t|x_t| - \frac{o_t}{E_0 t}\xi_{t+1} .$$

We wish to show that $\sqrt{t}\Delta_t = \sqrt{t}(x_t - z_t) \overset{\text{pr}}{\to} 0$ and for that porpouse we define $V_t := \Delta_t^T A \Delta_t$ where A is a definite positive matrix to be specified.

First we show that $\mathrm{E}[tV_t\, \mathbb{I}(t < \nu)] \to 0$ and by Theorem 5, p. 24, follows $\sqrt{t}(x_t - z_t) \overset{\text{pr}}{\to} 0$. So,

$$\begin{aligned} V_{t+1} &= \Delta_{t+1}^T A \Delta_{t+1} = \\ &= (\Delta_t - \frac{1}{E_0 t}\varphi'(0)\Delta_t - \frac{1}{E_0 t}v_t|x_t| - \frac{o_t}{E_0 t}\xi_{t+1})^T \cdot \\ &\quad \cdot A \cdot \\ &\quad (\Delta_t - \frac{1}{E_0 t}\varphi'(0)\Delta_t - \frac{1}{E_0 t}v_t|x_t| - \frac{o_t}{E_0 t}\xi_{t+1}) \end{aligned}$$

or, after transposition,

$$\begin{aligned} V_{t+1} &= \Delta_{t+1}^T A \Delta_{t+1} = \\ &= (\Delta_t^T - \frac{1}{E_0 t}\Delta_t^T \varphi'(0)^T - \frac{1}{E_0 t}v_t^T|x_t| - \frac{o_t}{E_0 t}\xi_{t+1}^T) \cdot \\ &\quad \cdot A \cdot \\ &\quad (\Delta_t - \frac{1}{E_0 t}\varphi'(0)\Delta_t - \frac{1}{E_0 t}v_t|x_t| - \frac{o_t}{E_0 t}\xi_{t+1}) . \end{aligned}$$

To estimate V_{t+1} we use Lemma 12 to obtain

$$V_{t+1} \leq V_t + B_t + C_t + D_t$$

with B_t, C_t and D_t to be specified and Using $\mathbb{I}(t+1 < \nu) \leq \mathbb{I}(t < \nu)$ we estimate $\mathrm{E}[(t+1)V_{t+1}\, \mathbb{I}(t+1 < \nu)]$ by

$$\begin{aligned}
\mathrm{E}[(t+1)V_{t+1}\, \mathbb{I}(t+1 < \nu)] \quad &\leq \quad \mathrm{E}[(t+1)V_t\, \mathbb{I}(t < \nu)] \\
&\quad\ +\mathrm{E}[(t+1)B_t\, \mathbb{I}(t < \nu)] \\
&\quad\ +\mathrm{E}[(t+1)C_t\, \mathbb{I}(t < \nu)] \\
&\quad\ +\mathrm{E}[(t+1)D_t\, \mathbb{I}(t < \nu)]\,.
\end{aligned}$$

Considering times when $t \leq \nu$ we have $|x_t| \leq Ra_t$ and $|o_t| \leq Ra_t$. For B_t, considering $t < \nu$,

$$\begin{aligned}
B_t &= \frac{3}{E_0^2 t^2}\left(\Delta_t^T \varphi'(0)^T A \varphi'(0)\Delta_t + |x_t|^2 v_t^T A v_t + o_t^2 \xi_{t+1}^T A \xi_{t+1}\right) \\
&\leq \frac{3}{E_0^2}\frac{1}{t^2}\left(K_1 \cdot V_t + |v_t|^2 \cdot |x_t|^2 \cdot |A| + o_t^2 |A||\xi_{t+1}|^2\right) \\
&\leq \frac{3}{E_0^2}\frac{1}{t^2}\left(K_1 \cdot V_t + c_t^2 \cdot R^2 a_t^2 \cdot |A| + R^2 a_t^2 \cdot |\xi_{t+1}|^2 \cdot |A|\right) \\
&\leq \frac{3}{E_0^2}\frac{1}{t^2}\left(K_1 \cdot V_t + o(1) + o(1) \cdot |\xi_{t+1}|^2\right)
\end{aligned}$$

where K_1 is a positive constant such that

$$\Delta_t^T \varphi'(0)^T A \varphi'(0)\Delta_t \leq K_1 \Delta_t^T A \Delta_t = K_1 V_t.$$

From

$$(t+1)B_t \leq \frac{3(t+1)}{E_0^2}\frac{1}{t^2}\left(K_1 \cdot V_t + o(1) + o(1) \cdot |\xi_{t+1}|^2\right)$$

and using

- $\frac{3(t+1)}{E_0^2}\frac{1}{t^2} \leq \frac{K_3}{t}$, for some positive constant K_3;
- $\frac{3(t+1)}{E_0^2}\frac{1}{t^2} o(1) = o(t^{-1})$;
- $\mathrm{E}[|\xi_{t+1}|^2] = tr(S_\xi)$;

we have

$$\mathrm{E}[(t+1)B_t\, \mathbb{I}(t \leq \nu)] = \frac{K_3}{t} V_t + o(t^{-1})\,.$$

Now we expand C_t,

$$\begin{aligned}
C_t &= \Delta_t^T A \frac{-1}{E_0 t}\varphi'(0)\Delta_t + \frac{-1}{E_0 t}\Delta_t^T \varphi'(0) A \Delta_t = \\
&= \frac{-1}{t}\Delta_t^T (A\varphi'(0)/E_0 + \varphi'(0)^T/E_0 A)\Delta_t\,.
\end{aligned}$$

Aiming and estimate of C_t in a useful way we find a matrix A which verifies $A\varphi'(0)/E_0 + \varphi'(0)^T/E_0 A = I + A$ and we use also $I + A \geq (1+\beta)A$ for a real positive constant β. We write, for $A = A^T$,

$$\begin{aligned}
A\varphi'(0)/E_0 + \varphi'(0)^T/E_0 A &= I + A \Leftrightarrow \\
\varphi'(0)^T/E_0 A + A\varphi'(0)/E_0 &= I + A \Leftrightarrow \\
\varphi'(0)^T/E_0 A - \frac{A}{2} + A\varphi'(0)/E_0 - \frac{A}{2} &= I \Leftrightarrow \\
(\varphi'(0)^T/E_0 - \frac{I}{2})A + A(\varphi'(0)/E_0 - \frac{I}{2}) &= I
\end{aligned}$$

and for use Lyapunov's result (Theorem 3) we write the last equality as

$$(\frac{I}{2} - \varphi'(0)^T/E_0)A + A(\frac{I}{2} - \varphi'(0)/E_0) = -I$$

where, from Assumption B3.3, $\frac{I}{2} - \varphi'(0)/E_0$ is negative definite, therefore solution A exists and is positive definite. Finalizing,

$$\begin{aligned}
C_t &= \frac{-1}{t}\Delta_t^T(A\varphi'(0)/E_0 + \varphi'(0)^T/E_0 A)\Delta_t \\
&= \frac{-1}{t}\Delta_t^T(A + I)\Delta_t \\
&\leq -(1+\beta)\frac{1}{t}V_t
\end{aligned}$$

We estimate the last term D_t

$$D_t = \frac{-1}{E_0 t}(2\Delta_t^T A v_t \cdot |x_t| + 2\Delta_t^T A o_t \xi_{t+1})\,.$$

Recall that we are considering $t < \nu$ and because we can't use $|\Delta_t| \leq V_t$ we follow this

- $x_t = \Delta_t + z_t$ from where $|x_t|^2 \leq |\Delta_t|^2 + |z_t|^2$;
- $2|\Delta_t|^2 \leq K_2 V_t$ (2 by convenience) for a certain positive constant K_2.

Then,

$$\begin{aligned}
2\Delta_t^T A v_t \cdot |x_t| &\leq 2|\Delta_t| \cdot |x_t| \cdot |A| \cdot c_t \\
&\leq (|\Delta_t|^2 + |x_t|^2) \cdot |A| \cdot c_t \\
&\leq (2|\Delta_t|^2 + |z_t|^2) \cdot |A| \cdot c_t \\
&\leq (K_2 V_t + |z_t|^2) \cdot |A| \cdot c_t
\end{aligned}$$

We considering again the estimation of D_t

$$\begin{aligned}
D_t &\leq \frac{-1}{E_0 t}(2\Delta_t^T A v_t \cdot |x_t| + 2\Delta_t^T A o_t \xi_{t+1}) \leq \\
&\leq \frac{K_2}{E_0 t} \cdot |A| \cdot c_t \cdot V_t + \frac{1}{E_0 t} \cdot |A| \cdot c_t \cdot |z_t|^2 - \frac{2}{E_0 t}\Delta_t^T A o_t \xi_{t+1}\,.
\end{aligned}$$

Taking

- $\mathrm{E}[|z_t|^2] = K_4/t$, for some constant K_4;

Then

$$\begin{aligned}\mathrm{E}[(t+1)D_t] &= \frac{K_2(t+1)}{E_0 t}\cdot|A|\cdot c_t\cdot V_t\\ &\quad+\frac{t+1}{E_0 t}\cdot|A|\cdot c_t\cdot\frac{K_4}{t}\\ &\leq o(1)V_t+o(t^{-1})\end{aligned}$$

Now, putting all together, always considering $t<\nu$,

$$\begin{aligned}(t+1)V_{t+1} &\leq (t+1)V_t+\frac{K_3}{t}V_t+\\ &\quad o(t^{-1})-\frac{t+1}{t}(1+\beta)V_t+\\ &\quad o(1)V_t+o(t^{-1})\leq\\ &\leq V_t(t+1\frac{K_3}{t}-(1+\beta)\frac{t+1}{t}+o(1))+o(t^{-1})\leq\\ &\leq t\cdot V_t(1+\frac{1}{t}+\frac{K_3}{t^2}-(1+\beta)\frac{t+1}{t^2}+o(t^{-2}))+o(t^{-1})\leq\\ &\leq tV_t(1-(1+\beta)\frac{1}{t}+o(t^{-1}))+o(t^{-1})\leq\\ &\leq tV_t(1-(1+\beta+o(1))\frac{1}{t})+o(t^{-1})\leq\\ &\leq tV_t(1-(\beta/2)\frac{1}{t})+o(t^{-1})\,.\end{aligned}$$

It follows that,

$$\mathrm{E}[(t+1)V_{t+1}\,\mathbb{I}(t+1<\nu)]\leq\mathrm{E}[tV_t\,\mathbb{I}(t<\nu)]+o(t^{-1})$$

and by Lemma 12

$$\mathrm{E}[tV_t\,\mathbb{I}(t<\nu)]\to 0,$$

then, by Theorem 5,

$$tV_t\,\mathbb{I}(t<\nu)\overset{\mathrm{pr}}{\to}0,$$

or

$$\sqrt{t}(x_t-z_t)\,\mathbb{I}(t<\nu)\overset{\mathrm{pr}}{\to}0\,,$$

or even, by definition of convergence in probability,

$$\forall\eta>0\quad\mathrm{P}(|\sqrt{t}(x_t-z_t)\,\mathbb{I}(t<\nu)|<\eta)\to 1\,.$$

The following events are related by

$$\sqrt{t}(x_t-z_t)<\eta\Rightarrow\sqrt{t}(x_t-z_t)\,\mathbb{I}(t<\nu)<\eta$$

and by $P(\sqrt{t}(x_t-z_t)<\eta)\leq\mathrm{P}(\sqrt{t}(x_t-z_t)\,\mathbb{I}(t<\nu)<\eta)$ we have

$$\sqrt{t}(x_t-z_t)\overset{\mathrm{pr}}{\to}0\,.$$

□

4 Some standard results

Theorem 3 (A. M. Lyapunov, 1947 (cited in [3], Chap. 13.1)) *Let $U, W \in \mathbb{C}^{n\times n}$ and let W be positive definite.*

(a) If U is stable then the equation

$$UA + AU^* = W$$

as a unique solution A negavtive definite.

(b) If exists a negative definite matrix A satisfying the above equation then A is stable.

Comment 9 *Stable is when all eigenvalues are negative. When all eigenvalues are negative then the matrix is negative definite.*

Lemma 14 (Markov Inequality (for example, [9])) *Let Z a r.v. and $g : \mathbb{R} \to [0, \infty]$ a non decreasing function. Then*

$$\mathrm{E}g(Z) \geq \mathrm{E}(g(Z); Z \geq c) \geq g(c)\mathrm{P}(Z \geq c)$$

Theorem 4 (Martingale convergence, [9], Cap. 12) *Let M be a martingale for which $M_n \in \mathcal{L}^2, \forall n$. Then M is limitied in $\mathcal{L}^2$ iif*

$$\sum \mathrm{E}[(M_k - M_{k-1})^2] < \infty$$

and when this we have

$$M_n \to M_\infty \text{ almost surely and in } \mathcal{L}^2 .$$

Theorem 5 ([9], Chap. 13.7) *Let (X_n) be a sequence in $\mathcal{L}^1$ and $X \in \mathcal{L}^1$. Then $X_n \to X$ in $\mathcal{L}^1$, or similarly $\mathrm{E}(|X_n - X|) \to 0$, iif, the following conditions are verifyed,*

1. *$X_n \to X$ in probability;*
2. *the sequence (X_n) is uniformly integrable ($\forall \epsilon > 0 \exists K : \mathrm{E}[|X|; |X| > K] < \epsilon$).*

Lemma 15 (Slutsky's Theorem, [7] Sec.8.6) *If $|X_t - Z_t| \overset{pr}{\to} 0$ and X_t converges in distribution then Z_t converges in distribuition for the same limit.*

Theorem 6 (Kolmogorov Law of Iterated Logarithm [9]) *Let $X_1, X_2, \ldots$ be random variables independent and identically distributed with mean 0 and variance 1. Let $S_n := X_1 + \cdots + X_n$. Then, almost surely,*

$$\limsup \frac{S_n}{\sqrt{2n \log\log n}} \to +1, \qquad \liminf \frac{S_n}{\sqrt{2n \log\log n}} \to -1 .$$

References

[1] Bernard Delyon and Anatoli Juditsky. Accelerated stochastic approximation. *SIAM J. Optim.*, 3(4):868–881, 1993.

[2] Harry Kesten. Accelerated stochastic approximation. *Ann. Math. Stat.*, 29:41–59, 1958.

[3] Peter Lancaster and Miron Tismenetsky. *The theory of matrices. 2nd ed., with applications.* Computer Science and Applied Mathematics. Orlando etc.: Academic Press (Harcourt Brace Jovanovich, Publishers). XV, 570 p., 1985.

[4] M.B. Nevel'son and R.Z. Has'minskii. *Stochastic approximation and recursive estimation. Translated from the Russian by Israel Program for Scientific Translations. Translation edited by B. Silver.* Translations of Mathematical Monographs. Vol. 47. Providence, R.I.: American Mathematical Society. IV, 244 p., 1976.

[5] Alexander Plakhov and Luís Borges Almeida. Modified kesten algorithm. Work done in Instituto Superior Tcnico, Lisbon, Portugal., 2000.

[6] Alexander Plakhov and Pedro Cruz. A stochastic approximation algorithm with step size adaptation. *Journal of Mathematical Sciences – Special Volume "Aveiro Seminar on Control, Optimization and Graph Theory"*, 107:119–130, 2004.

[7] b029 Resnik, Sidney, 1999.

[8] Herbert Robbins and Sutton Monro. A stochastic approximation method. *Ann. Math. Stat.*, 22:400–407, 1951.

[9] David Williams. *Probability with martingales.* Cambridge University Press. XV, 251 p., 1991.