On small perturbations of stable Markov operators: unbounded case

Bernard Delyon[*] Anatoli Juditsky[†]

March 30, 1999

Abstract

We consider the problem of estimating the bounds for generic expressions of the type $E[\phi(\gamma, X_1)...\phi(\gamma, X_n)]$, where (X_i) is a *non necessarily bounded* Markov chain, ϕ is a smooth function, and γ is a small parameter. We show that when the process (X_i) is exponentially ergodic, some tight bounds can be obtained by small perturbation of the transition operator of the chain. The result is then applied to prove exponential convergence of matrix products and exponential inequalities for Markov chains.

1 Introduction

The objective of this work is to provide simple estimates for generic expressions of the type

$$E_x[\phi(\gamma, X_0)...\phi(\gamma, X_{n-1})f(X_n)] \tag{1}$$

for small values of the perturbation parameter γ. Here (X_i) is a Markov chain with initial condition x and transition operator P, f is some function, and ϕ is a function with some regularity w.r.t. γ such that

$$\phi(0, x) = 1.$$

One rather classical way to do this consists in rewriting the quantity in (1) as

$$\int P_\gamma^n(f)(x)\mu(dx)$$

where $\mu(dx)$ is the unit mass at x an P_γ is the operator

$$P_\gamma f(x) = \phi(\gamma, x)Pf(x) = \phi(\gamma, x)E[f(X_1)|X_0 = x]$$

[*]IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France. e-mail:name@irisa.fr

[†]This is the author to whom the correspondence should be sent.

(this change of operator has been used in particular in [4] and [12]). Under standard conditions on the Markov chain, P can be rewritten as

$$P = \pi + R$$

where π is the invariant measure and R has a spectral radius strictly smaller than one. It has been proved (cf. [4]) that the use of standard perturbation results allows to extend continuously this decomposition to P_γ when the function $\phi(\gamma, x)$ is bounded (or "almost" bounded as in [12]).

The main result of this note is an extension of these results for the case of unbounded ϕ (Proposition 1).

A direct applications of this result are exponential inequalities (cf. section 3.1) and large or moderate deviations bounds for Markov chains.

The estimate of the product (1) can also be used to show the L_p stability of matrices products. For instance, the following problem (cf. section 3.3) often arises in the in the analysis of parameter tracking algorithms (cf. [9], [8], [1] etc.): consider the process $(\Delta_n)_{n\geq 0}$ defined with the difference equation

$$\Delta_n = (I - \gamma L_n)\Delta_{n-1} + e_n, \quad \Delta_0 \in \mathbf{R}^d$$

where e_n is a random vector, $L_n \in \mathbf{R}^{d\times d}$ is a random matrix and γ is a small parameter. If we denote

$$\Phi_j^n = \prod_{i=j}^n (I - \gamma L_i) = (I - \gamma L_n)...(I - \gamma L_j), \quad n \geq 0, j \geq 1$$

$(\prod_{i=n+1}^n (I - \gamma L_i) = I)$, we get the following explicit formula for Δ_n:

$$\Delta_n = \Phi_1^n \Delta_0 + \sum_{j=1}^n \Phi_{j+1}^n e_j, \quad n \geq 0.$$

Then it is clear that the behavior of Δ_n heavily depends on the control of the matrix product Φ_j^n. One way to control Δ_n is to provide some almost sure estimates for the product. This way we obtain the convergence in distribution for Δ_n as $\gamma \to 0$ and $n \to \infty$ under quite general assumptions on (L_i) and (e_i) [1]. In order to obtain the estimates of $E|\Delta_n|^p$, $0 < p < \infty$, a bound for $E|\Phi_j^n|^q$ for some $q > p$ reveals crucial. Important progress has been achieved recently in the estimation of this quantity for a variety of algorithms (cf. [5], [8], [6] among others). However, there are no adequate estimates available for the classical LMS algorithm when $L_i = Y_i Y_i^T$, where Y_i is a component of an unbounded Markov chain. Our aim in this note is to provide the estimates of the type

$$E\left|\prod_{i=j}^n (I - \gamma Y_i Y_i^T)\right|^q \leq K(1 - \kappa\gamma)^n$$

when $Y_i = f(X_i)$ is a component of the geometrically ergodic Markov chain (X_i).

2 Main result

Notations We consider a topoligical space $\mathcal{X}$ (typically a part of $\mathbf{R}^d$), endowed with its Borel σ-field. For any positive function V on $\mathcal{X}$, we will denote

$$\begin{aligned} \mathcal{C}_V &= \{f \text{ continuous}, \ \sup_x \frac{|f(x)|}{V(x)} < \infty\} \\ \|f\|_V &= \sup_x \frac{|f(x)|}{V(x)} \quad f \in C_V \end{aligned}$$

and the operator norm

$$\|P\|_V = \sup_{f \neq 0} \frac{\|Pf\|_V}{\|f\|_V}.$$

We consider the assumption :

(A0) P is a Markov transition probability with a unique invariant probability π such that, for some function $V \geq 1$, $\|P\|_V < \infty$ and the spectral radius of $P - \pi$ in $\mathcal{C}_V$ is smaller than one :

$$\|P^n - \pi\|_V \leq C\rho^n \qquad \rho < 1$$

Proposition 1 *We suppose that P satisfies assumption A0.*

Let $\phi(\gamma, x)$ be a parametrized function of x and set

$$\phi_i(x) = \frac{d^i\phi}{d\gamma^i}(0, x).$$

We assume that

$$\begin{aligned} \phi(0, x) &= 1 \\ |\frac{d^i\phi}{d\gamma^i}(\gamma, x)|PV(x) &\leq K_0V(x), \qquad 0 \leq \gamma \leq \gamma_0, \qquad i = 1, 2, 3 \end{aligned} \tag{2}$$

for some $K_0, \gamma_0 > 0$. Then the operator P_γ defined by

$$P_\gamma f(x) = \phi(\gamma, x)Pf(x)$$

satisfies for γ small enough

$$P_\gamma = \Pi_\gamma + R_\gamma$$

where R_γ has a spectral radius uniformly bounded away from 1 on $\mathcal{C}_V$, $\Pi_\gamma R_\gamma = R_\gamma\Pi_\gamma = 0$ and Π_γ is a rank 1 operator

$$\Pi_\gamma f(x) = \pi_\gamma(f)f_\gamma(x).$$

π_γ is a measure such that $|\pi_\gamma|(V) < \infty$ and $f_\gamma(x) = \Pi_\gamma(1)$ is the eigenvector. We have the expansion of the eigenvalue $\lambda_\gamma = \pi_\gamma(f_\gamma(x))$ of Π_γ

$$\lambda_\gamma = 1 + \gamma\pi(\phi_1) + \frac{\gamma^2\sigma^2}{2} + O(\gamma^3)$$

where
$$\sigma^2 = \pi(\phi_2) + 2\sum_{n=1}^{\infty}\left(\pi(\phi_1 P^n \phi_1) - \pi(\phi_1)^2\right).$$
If the condition in (2) is only satisfied for $i = 1, 2$, we still have the expansion

$$\begin{aligned} \lambda_\gamma &= 1 + \gamma\pi(\phi_1) + O(\gamma^2) \\ f_\gamma(x) &= 1 + O(\gamma V(x)). \end{aligned}$$

Comments: Assumption A0 is satisfied for a large class of Markov chains. Consider the assumption :

(A1) P is a Feller transition kernel on $\mathcal{X}$. There exists a measure ψ, whose support has a non-empty interior such that for any ψ-positive set A

$$P_x(X_n \in A \text{ for some } n \geq 1) > 0 \tag{3}$$

for all x, and the chain X_n is aperiodic. Furthermore, there exist positive real numbers ϵ and C and a measurable function V, $V \geq 1$, bounded on compact sets, which tends to infinity, such that

$$\int V(y)P(x, dy) \leq (1-\epsilon)V(x) + C \tag{4}$$

for all x.

Thanks to theorems 6.0.1(iii), 6.2.5(ii) and 15.0.1(iii) of [11], assumption A1 implies A0, with the same function V. A simple consequence of the proposition is the

Corollary 1 *We suppose that P satisfies assumption A0.*

Let $\phi(\gamma, x)$ be a parametrized function of x; we assume that

$$\begin{aligned} \phi(0, x) &= 1 \\ \left|\frac{d^i\phi}{d\gamma^i}(\gamma, x)\right| PV(x) &\leq K_0 V(x), \qquad 0 \leq \gamma \leq \gamma_0, \qquad i = 1, 2 \end{aligned}$$

for some $K_0, \gamma_0 > 0$. Then for any function f, $\|f\|_V < \infty$

$$|E_x[\phi(\gamma, X_0)...\phi(\gamma, X_{n-1})f(X_n)]| \leq C(\pi(\phi(\gamma, .)) + C\gamma^2)^n \|f\|_V V(x) \tag{5}$$

for some constant C depending on ϕ any n and γ small enough.

Proof of Proposition 1 and Corollary 1 The assumptions on $\phi(\gamma, x)$ imply that $\gamma \longrightarrow P_\gamma$ is a continuous function (with the operator-norm $\|\cdot\|_V$) and

$$P_\gamma = (1 + \gamma\phi_1 + \gamma^2\phi_2)P + O(\gamma^3), \tag{6}$$

where $O()$ is considered w.r.t. the norm $\|\cdot\|_V$. Since P_γ is a continuous function of γ, Theorem 3.16 p.212 of [7] (cf also IV.3.5 p.213) implies that the decomposition

$$P = \pi + R$$

extends continuously in the neighborhood of $\gamma = 0$ into

$$P_\gamma = \Pi_\gamma + R_\gamma$$

where

$$\Pi_\gamma = -\frac{1}{2i\pi}\int_\Gamma (P_\gamma - zI)^{-1} z\, dz \tag{7}$$

(Γ is a small contour in the complex plane around 1) is a rank 1 operator. Note that the convergence of the integral in (7) and its continuity with respect to γ follows from Assumption A0: 1 is an isolated point of the spectrum of P (we refer to [7] for details). Π_γ is a continuous function of γ and so is R_γ; the spectral radius of R_γ is thus an upper semi-continuous function of γ (as mentioned in [7]). Since $\Pi_\gamma(1) \neq 0$ for γ small enough, we can take

$$f_\gamma = \Pi_\gamma(1).$$

Equations (7) and (6) imply that f_γ and $\lambda_\gamma = \pi_\gamma(f_\gamma)$ (the largest eigenvalue of P_γ) can be expanded as

$$\begin{aligned} f_\gamma(x) &= 1 + \gamma f_1(x) + \gamma^2 f_2(x) + O(\gamma^3 V(x)) \\ \lambda_\gamma &= 1 + \gamma\lambda_1 + \gamma^2\lambda_2 + O(\gamma^3). \end{aligned}$$

The terms of the expansion are easily identified by expanding both sides of

$$P_\gamma(f_\gamma) = \lambda_\gamma f_\gamma.$$

We obtain

$$\begin{aligned} \phi_1 + Pf_1 &= \lambda_1 + f_1 \\ \phi_2 + \phi_1 Pf_1 + Pf_2 &= \lambda_2 + \lambda_1 f_1 + f_2. \end{aligned}$$

The first equation implies

$$\lambda_1 = \pi(\phi_1)$$

and f_1 is a solution of the Poisson equation and expresses as

$$f_1 = C + \sum_{n=0}^{\infty} P^n(\phi_1 - \lambda_1)$$

where C is some constant. Second equation leads to

$$\lambda_2 = \pi(\phi_2 + \phi_1 P f_1 - \lambda_1 f_1).$$

Finally

$$\begin{aligned}
1 + \gamma\lambda_1 + \gamma^2\lambda_2 &= 1 + \gamma\pi(\phi_1) + \gamma^2\pi(\phi_2) + \gamma^2\pi(\phi_1 P f_1 - \lambda_1 f_1) \\
&= 1 + \gamma\pi(\phi_1) + \gamma^2\pi(\phi_2) + \gamma^2\pi(\phi_1 P f_1 - \phi_1 C) \\
&= 1 + \gamma\pi(\phi_1) + \gamma^2\pi(\phi_2) + \gamma^2 \sum_{n=1}^{\infty} \pi(\phi_1(P^n\phi_1 - \lambda_1)) \\
&= 1 + \gamma\pi(\phi_1) + \gamma^2\pi(\phi_2) + \gamma^2 \sum_{n=1}^{\infty} \pi(\phi_1 P^n \phi_1) - \pi(\phi_1)^2.
\end{aligned}$$

For Corollary 1, note that

$$\begin{aligned}
|E_x[\phi(\gamma, X_0)...\phi(\gamma, X_{n-1}) f(X_n)]| &= |P_\gamma^n f(x)| \\
&= |\Pi_\gamma^n f(x) + R_\gamma^n f(x)| \\
&\leq (\|\Pi_\gamma^n f\|_V + \|R_\gamma^n f\|_V) V(x) \\
&\leq K(\lambda_\gamma^{n-1} + \rho(R_\gamma)^n)\|f\|_V V(x)
\end{aligned}$$

∎

3 Applications

3.1 Exponential inequalities

An application of Proposition1 is the estimation of the probability of large deviations for sums

$$S_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i),$$

where (X_i) is a Markov chain. Such estimates can be easily obtained using so-called exponential inequalities ([2]). We show here how such an inequality can be obtained using the result of Proposition 1.

Theorem 1 *Let $(X_n)_{n\geq 0}$ be a Markov chain with transition probability kernel P satisfying assumption A0, and $g(\cdot)$ be a function such that $\|g\|_V < \infty$, $\pi(g) = 0$ and there is $K < \infty$ and $\gamma_0 > 0$ such that*

$$(1 + |g(x)|^3) e^{\gamma_0 g(x)} PV(x) \leq KV(x).$$

Then for any $0 < \gamma \le \gamma_0$

$$E_x[e^{\gamma(g(X_0)+g(X_1)+\ldots+g(X_n))}] = (1+\gamma V(x)O(1))\lambda_\gamma^n + \rho^n V(x)O(1) \tag{8}$$

where

$$\begin{aligned} \lambda_\gamma &= e^{\frac{\gamma^2\sigma^2}{2}+O(\gamma^3)} \\ \sigma^2 &= \pi(g^2) + 2\sum_{n=1}^{\infty}\pi(gP^n g), \end{aligned}$$

$\rho < 1$ *and* $|O(1)| \le C$ *where* C *does not depend on* n, γ *or* x; λ_γ *is the eigenvalue of Proposition 1.*

Proof: We put

$$\phi(\gamma, x) = e^{\gamma g(x)}.$$

Then

$$\pi(\phi(\gamma, x)) = 1 + \frac{\gamma^2}{2}\pi(g^2) + O(\gamma^3 V(x)).$$

Note that the assumptions of Proposition 1 hold, so $P_\gamma(x) = e^{\gamma g(x)}P(x)$ can be decomposed as follows:

$$P_\gamma = \Pi_\gamma + R_\gamma.$$

Note that there exists n_0 such that $R^{n_0} < 1$ and the continuity of the operator norm implies that there exists $\rho < 1$ and γ_0 such that $R_\gamma^{n_0} < \rho^{n_0}$ for any $\gamma \le \gamma_0$. Setting $f(x) \equiv 1$, we have

$$\begin{aligned} E_x[e^{\gamma(g(X_0)+g(X_1)+\ldots+g(X_n))}f(X_{n+1})] &= \Pi_\gamma^{n+1}f(x) + R_\gamma^{n+1}f(x) \\ &= \lambda_\gamma^n f_\gamma(x) + R_\gamma^{n+1}f(x) \\ &= \lambda_\gamma^n(1+\gamma O(V(x))) + \rho^n O(V(x)). \end{aligned}$$

3.2 Large deviations

Using the result above we obtain the following large deviations bound:

Theorem 2 *Suppose that the assumptions of theorem 1 hold. Then for any fixed* x *and some* $\epsilon_0 > 0$ *we have*

$$\log P_x(g(X_1) + \ldots. + g(X_n) > \lambda) = -\frac{\lambda^2}{2n\sigma^2}(1+\epsilon^{1/4}O(1)), \quad \epsilon = \max(\frac{\lambda}{n}, \frac{n}{\lambda^2})$$

for any $\lambda \in [\sqrt{n/\epsilon_0}, \epsilon_0 n]$.

Note that bound on $O(1)$ *does not depend on* n, λ *or* ϵ.

Proof: We set $Z_i = g(X_i)$. Note first that inequality (8) implies that

$$E_x[e^{\gamma(Z_0+Z_1+\ldots+Z_n)}] = e^{\frac{\gamma^2\sigma^2 n}{2}+n\gamma^3 O(1)}$$

if x is fixed and $C/\sqrt{n} < \gamma < \gamma_0$ for some C (the latter condition implies that $\gamma < C^{-1}n\gamma^3$). Now we can rewrite it as follows:

$$E_x[e^{\gamma(\frac{Z_0+Z_1+...+Z_n}{\sigma\sqrt{n}}-\alpha)}] \quad = \quad e^{-\frac{\alpha^2}{2}}e^{\frac{(\gamma-\alpha)^2}{2}+\frac{\gamma^3}{\sqrt{n}}O(1)} \quad \text{if} \quad C < \gamma < \gamma_0\sqrt{n}\sigma. \tag{9}$$

The derivation the large deviations principle from the exponential bound (9) is based on approximations of the indicator $1_{Y>0}$ by a sum of exponentials. The following inequality holds for $r > 1, \alpha, \delta > 0$:

$$4r^{-2-\alpha/\delta}(e^{\delta Y}-1)(r-e^{\delta Y})e^{\alpha Y} \leq 1_{Y>0} \leq e^{\alpha Y}. \tag{10}$$

The first inequality is obtained by the following simple observation: the left hand side of (10) is negative if $e^{\delta Y}$ is outside $[1, r]$. Else

$$4r^{-2-\alpha/\delta}(e^{\delta Y}-1)(r-e^{\delta Y})e^{\alpha Y} \quad \leq \quad 4r^{-2-\alpha/\delta}\left(\frac{r-1}{2}\right)^2 e^{\alpha Y} \leq 1.$$

We shall apply inequality (10) with

$$\begin{aligned} Y &= \frac{Z_1+...+Z_n}{\sigma\sqrt{n}} - \alpha \\ \alpha &= \frac{\lambda}{\sigma\sqrt{n}}, \quad \delta^2 = \left(\frac{\lambda^3}{n^2}+1\right)\frac{1}{\sqrt{\epsilon}\sigma^2}, \quad r = e^{2\delta^2}. \end{aligned}$$

Note that δ is small compared to α:

$$\frac{\delta^2}{\alpha^2} \quad = \quad \frac{1}{\sqrt{\epsilon}}\left(\frac{\lambda}{n}+\frac{n}{\lambda^2}\right) \leq 2\sqrt{\epsilon} \tag{11}$$

and

$$\frac{\alpha^3}{\sqrt{n}} \leq \frac{\sqrt{\epsilon}}{\sigma}\delta^2.$$

Using (10) and the second inequality of (9), we obtain the upper bound

$$P(Y \geq 0) \quad \leq \quad E[e^{\alpha Y}] \leq e^{-\frac{\alpha^2}{2}+O(\alpha^3/\sqrt{n})} \leq e^{-\frac{\alpha^2}{2}(1+O(\epsilon))}.$$

The lower bound requires more effort:

$$\begin{aligned} P(Y \geq 0) \quad &\geq \quad 4r^{-2-\alpha/\delta}E[-e^{(2\delta+\alpha)Y} + (r+1)e^{(\delta+\alpha)Y} - re^{\alpha Y}] \\ &\geq \quad 4r^{-2-\alpha/\delta}e^{-\frac{\alpha^2}{2}}\left\{-e^{2\delta^2+O(\alpha^3/\sqrt{n})} + (r+1)e^{\delta^2/2+O(\alpha^3/\sqrt{n})} - re^{O(\alpha^3/\sqrt{n})}\right\} \\ &\geq \quad 4r^{-2-\alpha/\delta}e^{-\frac{\alpha^2}{2}+O(\alpha^3/\sqrt{n})}\left\{(e^{2\delta^2}+1)e^{\delta^2/2} - 2e^{2\delta^2+O(\alpha^3/\sqrt{n})}\right\} \\ &\geq \quad 4r^{-2-\alpha/\delta}e^{-\frac{\alpha^2}{2}(1+O(\epsilon))}\left\{(e^{2\delta^2}+1)e^{\delta^2/2} - 2e^{2\delta^2+O(\sqrt{\epsilon}\delta^2)}\right\} \end{aligned}$$

Since δ is larger than $(\sqrt{\epsilon}\sigma^2)^{-1}$,

$$(e^{2\delta^2}+1)e^{\delta^2/2} - 2e^{2\delta^2+O(\sqrt{\epsilon}\delta^2)} \geq 1/4$$

for an ϵ small enough. Thus

$$P(Y \geq 0) \quad \geq \quad e^{-4\delta^2-2\alpha\delta}e^{-\frac{\alpha^2}{2}(1+O(\epsilon))} \geq e^{-\frac{\alpha^2}{2}(1+O(\epsilon^{1/4}))}$$

(here we used (11)). ∎

3.3 Matrix products

We come back to the problem exposed in the introduction: suppose that $(X_n)_{n\geq 1}$, $X_n \in \mathbf{R}^d$ is a Markov chain with transition probability kernel P. Let Y_n be a component of X_n. Consider for some $n \geq 1$ the (right) matrix products

$$\Phi_1^n = (I - \gamma Y_n Y_n^T)...(I - \gamma Y_1 Y_1^T) = \prod_{i=1}^{n} (I - \gamma Y_i Y_i^T) \tag{12}$$

and

$$\Psi_1^n = \prod_{i=1}^{n} (I - \gamma \frac{Y_i Y_i^T}{1 + |Y_i|^2}).$$

We have the following consequence of Proposition 1

Theorem 3 *Suppose that assumption A1 holds true and that*

$$\lim_{n\to\infty} E\left(\frac{Y_n Y_n^T}{1 + |Y_n|^2}\right) > 0. \tag{13}$$

Then for any $p < \infty$ there exist positive constants γ_0, κ and $C < \infty$ such that for any $\gamma \leq \gamma_0$

$$E_x |\Psi_1^n|^p \leq C \exp(-\kappa\gamma n) V(x).$$

Proof: The technical problem here is that we have to reduce the problem to the estimation of a product of random numbers (instead of matrices); this is done by grouping the matrix product into small subproducts whose norm has expectation smaller than one (each of these norm is now our $\phi(\gamma, X_k)$, with a sligth modification due to the regularity condition require for ϕ), and the problem is reduced to the estimation of the product of these norms. More precisely, for some p large enough, set

$$\begin{aligned} X'_k &= (X_k, X_{k-1}, ..X_{k-p+1}) \\ \phi(\gamma, X'_k) &= \lambda_{max}(I - \gamma \frac{Y_i Y_i^T}{1 + |Y_i|^2}) + \gamma^2 K_p \\ f &= |\Pi_{i=qp+1}^{n} (I - \gamma \frac{Y_i Y_i^T}{1 + |Y_i|^2})| \end{aligned}$$

(where q is such that $0 \leq n - qp < p$, and K_p is suitably chosen so that the next equation holds) and

$$|\Psi_1^n| \leq f\ \Pi_{k=1}^{q} |\Pi_{i=kp-p+1}^{kp} (I - \gamma \frac{Y_i Y_i^T}{1 + |Y_i|^2})| \leq f\ \Pi_{k=1}^{q} \phi(\gamma, X'_{kp})$$

Thanks to lemma 1 the process $(X'_k)_{k=p,2p,..}$, satisfies assumption A0. Equation2 is obviously satisfied for $i = 1, 2$ and we can apply Proposition 1 to the Markov chain $(X'_n)_{n=p,2p,..}$ and equation (5) rewrites

$$E_x |\Psi_1^n|^p \leq C(\pi(\phi(\gamma, X'_k) + C\gamma^2)^q V(x). \tag{14}$$

As an elementary consequence of (13) one has the existence of p, γ_0 and $0 < \epsilon < 1$ such that for $\gamma < \gamma_0$

$$\pi(\phi(\gamma, X'_k)) \leq 1 - \epsilon\gamma. \tag{15}$$

Equations (14, 15) lead immediately to the result. ∎

The study of the product of matrices $(I - \gamma Y_i Y_i^T)$ is somewhat more involved. We have, however, the following

Theorem 4 *Let $(Y_n)_{n\geq 1}$, $Y_n \in \mathbf{R}^d$ be defined with the following stochastic model:*

$$X_n = AX_{n-1} + Bw_n, \quad X_0 = x \in \mathbf{R}^r, \quad w_n \in \mathbf{R}^s \tag{16}$$

$$Y_n = CX_n. \tag{17}$$

We assume that the matrix $A \in \mathbf{R}^{r\times r}$ is stable (the spectral radius $\rho(A) < 1$), and $B \in \mathbf{R}^{r\times s}$ and $C \in \mathbf{R}^{d\times r}$ are such that the rank of the matrix

$$M = C[I, A, A^2, ..., A^{r-1}]B \tag{18}$$

is equal to d.

(w_n), $n \geq 1$, is a sequence of independent and identically distributed random vectors, $E[w_1 w_1^T] > 0$. We assume that $E[\exp(\alpha|w_1|)] < \infty$ for some $\alpha > 0$ and the distribution of w_1 possesses an absolutely continuous component in some neighborhood of zero.

Then for any $p < \infty$ and $\epsilon > 0$ there exist positive constants γ_0, κ and $C < \infty$ such that for any $\gamma \leq \gamma_0$

$$E_x[|\Phi_1^n|^p] \leq C\exp(-\kappa\gamma n)e^{\epsilon|x|}.$$

Note that in particular equations (16),(17) can represent any gaussian process (Y_n) with rational spectrum.

Proof: We need first the following lemmae:

Lemma 1 *Let $(X_n)_{n\geq 1}$ be a Markov chain satisfying assumption A0, then for any $p > 1$ the process $X'_n = (X_{n-p+1}, ...X_{n-1}, X_n)$, $n \geq p$, satisfies also assumption A0 where now $\rho' = \rho$, π' is the invariant measure of the chain (X'_n) :*

$$\pi'(f) = \pi(E[f(X_1, ..., X_p)|X_1 = x])$$

and the weight function is

$$V'(x_1, ...x_p) = V(x_1) + V(x_2) + ... + V(x_p).$$

The choice

$$V'(x_1, ...x_p) = V(x_p)$$

is also convenient.

Note : The first choice of V' seems more complicated but has the superiority of leading to a larger space C_V.

Proof: Denote by P' the transition kernel of X'; we have for any function f and any $x' = (x'_1, ...x'_p)$

$$\begin{aligned} |P'^n f(x') - \pi' f| &= |E[\, f(X_{n+1}, ..., X_{n+p}) \mid (X_1, ...X_p) = x'] - \pi' f| \\ &= |E[\, E[f(X_{n+1}, ..., X_{n+p})|X_{n+1}] \mid (X_1, ...X_p) = x'] - \pi' f| \end{aligned}$$

and setting

$$\phi(x) = E[f(X_{n+1}, ..., X_{n+p})|X_{n+1} = x]$$

we get

$$\begin{aligned} |P'^n f(x') - \pi' f| &= |E[\phi(X_{n+1}) \mid (X_1, ...X_p) = x'] - \pi\phi| \\ &= |P^{n+1-p}\phi(x'_p) - \pi\phi| \\ &\le C\rho^n \|\phi\|_V V(x'_p) \\ &\le C\rho^n \|\phi\|_V V'(x'). \end{aligned}$$

And

$$\begin{aligned} \|\phi\|_V &= \|E[f(X_{n+1}, ..., X_{n+p})|X_{n+1} = x]\|_V \\ &\le \|f\|_{V'} \|E[V'(X_{n+1}, ..., X_{n+p})|X_{n+1} = x]\|_V \\ &\le C'\|f\|_{V'} \end{aligned}$$

∎

Lemma 2 *Let $x \in \mathbf{R}^d$ and $y = Fx + w$, where the matrix F is stable and $w \in \mathbf{R}^d$ is a random variable independent of x with $E[\exp(\alpha|w|)] < \infty$ for some $\alpha > 0$. There exists a symmetric matrix $R > 0$ and constants $\beta > 0$ and $K < \infty$ such that*

$$Ee^{|Ry|} \le Ke^{(1-\beta)|Rx|}. \tag{19}$$

Proof: Since the matrix F is stable, we conclude by the Lyapunov theorem that there exists a symmetric matrix $P > 0$ such that, for instance,

$$F^T PF - P = -I.$$

This implies that $F^T PF \le (1 - \beta')P$ for some $\beta > 0$. Put $t = \alpha\rho^{-1/2}(P)$, where $\rho(P)$ is the spectral radius of P. Then if $R = tP^{1/2}$, where $P^{1/2}$ is a symmetric square root of P, we have

$$e^{|Ry|} = e^{t|P^{1/2}y|} \le e^{t|P^{1/2}Fx|}e^{t|P^{1/2}w|} \le e^{(1-\beta')^{1/2}|Rx|}e^{\alpha|w|}.$$

The statement of the lemma is obtained by taking the expectation with respect to w. ∎

Since the matrix M in (18) has full rank, using Theorem 1 from [3] we conclude that there exists $a > 0$ such that

$$\pi\left(\sum_{i=1}^{r} Y_i Y_i^T > aI\right) > 0,$$

and

$$E_\pi \lambda_{\min}\left(\sum_{i=1}^{r} Y_i Y_i^T\right) > 0. \tag{20}$$

Now we return to the matrix product (12). Denote $m = [n/r]$ and $l_n = mr$ (here $[\cdot]$ denotes the integer part). We decompose the product in the following way:

$$\Phi_1^n = \Phi_{l_n+1}^n \prod_{j=0}^{m-1} Q_j,$$

where

$$Q_j = \prod_{i=1}^{r} (I - \gamma Y_{jr+i} Y_{jr+i}^T).$$

Note that

$$\begin{aligned}
|Q_j|^p &= \left| I - \gamma \sum_{i=1}^{r} Y_{jr+i} Y_{jr+i}^T + \ldots + (-1)^r \gamma^r \prod_{i=1}^{r} Y_{jr+i} Y_{jr+i}^T \right|^p \\
&\leq \left| I - \gamma \sum_{i=1}^{r} Y_{jr+i} Y_{jr+i}^T + \gamma^2 K(r) \left(1 + \sum_{i=1}^{r} |Y_{jr+i}|^2\right)^r \right|^p \\
&\leq 1 - p\gamma \lambda_{\min}\left(\sum_{i=1}^{r} Y_{jr+i} Y_{jr+i}^T\right) + \gamma^2 K'(r) \left(1 + \sum_{i=1}^{r} |Y_{jr+i}|^2\right)^{pr}.
\end{aligned}$$

Hence

$$\left|\prod_{j=0}^{m-1} Q_j\right|^p \leq \prod_{j=0}^{m-1} |Q_j|^p \leq \prod_{j=1}^{m} \left(1 - \gamma p \lambda_{\min}\left(\sum_{i=1}^{r} Y_{jr+i} Y_{jr+i}^T\right) + \gamma^2 R_j\right),$$

where

$$R_j \leq K\left(1 + \sum_{i=1}^{r} |Y_{jr+i}|^2\right)^{2pr}.$$

In order to apply Proposition 1, we introduce the Markov chain $(\bar{X}_j)$, $j \geq 1$, $\bar{X}_j \in \mathbf{R}^{r^2}$, where

$$\bar{X}_j = (X_{jr}^T, X_{jr-1}^T, \ldots, X_{(j-1)r+1}^T)^T.$$

It satisfies the equation

$$\bar{X}_j = F\bar{X}_{j-1} + e_j,$$

where

$$F = \begin{pmatrix} A^r, & 0 & \ldots & 0 \\ A^{r-1}, & 0 & \ldots & 0 \\ & & \ldots & \\ A, & 0 & \ldots & 0 \end{pmatrix} \quad \text{and} \quad e_j = \begin{pmatrix} \sum_{i=0}^{r-1} A^i B w_{jr-i} \\ \sum_{i=1}^{r-1} A^i B w_{jr-i} \\ \ldots \\ B w_{(j-1)r+1} \end{pmatrix},$$

with the initial condition $\bar{X}_0 = \bar{x} = (x^T, 0, ..., 0)^T$. Note that the matrix F is stable and the random vectors e_m are mutually independent. From Lemma 2 we conclude that assumption A1 holds true for $(\bar{X}_j)$ with $V(\bar{X}) = e^{|R\bar{X}|}$ with some symmetric positive $R \in \mathbf{R}^{r^2 \times r^2}$. Furthermore, given $\epsilon > 0$ we can choose R such that, $\|R\| < \epsilon$ (by the Hölder inequality, we can substitute R with ϵR in (19) for $\epsilon \leq 1$), and for some $\beta > 0$

$$E(e^{|R\bar{X}_j|}|\mathcal{F}_{j-1}) = PV(\bar{X}_{j-1}) \leq Ke^{(1-\beta)|R\bar{X}_{j-1}|}. \tag{21}$$

We set

$$\phi(\gamma, \bar{X}_j) = 1 - \gamma p \lambda_{\min}\left(\sum_{i=1}^{r} X_{jr+i} X_{jr+i}^T\right) + \gamma^2 R_j.$$

Then due to (20)

$$\pi\phi'_\gamma(0, \bar{X}_j) \leq -\kappa \ \text{ for some } \ \kappa > 0,$$

and

$$\begin{aligned} |\phi'_\gamma(\gamma, \bar{X}_j)| &\leq K(1 + |\bar{X}_j|^{2pr}), \\ |\phi''_\gamma(\gamma, \bar{X}_j)| &\leq K(1 + |\bar{X}_j|^{2pr}). \end{aligned}$$

We have from (21) that (2) is satisfied with $i = 1, 2$. The last assertion of Proposition 1 with $f(\bar{X}_{m+1}) \equiv |\Phi^n_{l_n+1}|^p$ leads to

$$\begin{aligned} E_{\bar{x}}|\Phi^n_1|^p &\leq E_{\bar{x}} \prod_{j=1}^{m} |Q_j|^p \left|\Phi^n_{l_n+1}\right|^p \\ &\leq |E_{\bar{x}}[\phi(\gamma, \bar{X}_1)...\phi(\gamma, \bar{X}_m) f(\bar{X}_{m+1})]| \\ &\leq K(1 - \kappa\gamma)^m V(\bar{x}) \end{aligned}$$

for any initial condition $\bar{x}$ of the Markov chain $(\bar{X}_k)$. However, by construction of V we have

$$V(\bar{x}) = V(x, 0, ..., 0) \leq e^{\epsilon|x|}.$$

∎

References

[1] B. Delyon, A. Juditsky Asymptotical Study of Parameter Tracking Algorithms, SIAM J. Control and Opt., Vol. 33, No. 1, 323-345, 1995.

[2] M. Freidlin A. Ventzel, *Random Perturbations of Dynamical Systems.* New York, Springer–Verlag, 1984.

[3] M. Green and J.B. Moore, Persistence of excitation in linear systems. *Syst. Contr. Lett.*, V. 7, 351-360, 1986.

[4] Y.Guivarc'h, J. Hardy, Théorème limite pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov, *Ann. Inst. H. Poincaré*, Vol. 24, No.1, 73-88, 1988.

[5] L. Guo, L. Ljung And P. Priouret, Performance Analysis of Forgetting factor RLS, *Int. J. Adapt. Control Signal Process*, 7, No.6, 525-537, 1993

[6] A. Juditsky, P. Priouret A Robust Algorithm for Random Parameter Tracking, *IEEE-AC*, Vol. 39, No. 6, 1211-1221, 1994.

[7] T.Kato, *Perturbation theory for linear operators*, Springer-Verlag 1966.

[8] L. Ljung, P. Priouret, Remarks on the Mean Square Tracking Error, *Int.J. of Adapt. Control and Signal Process.*, 5, No.6, 395-403, 1991.

[9] L.Ljung, S.Gunnarson, Adaptation and Tracking in System Identification-A Survey, *Automatica*, Vol. 26, No. 1, pp. 7-21,1990.

[10] O.Macchi, E.Eweda, Bases théoriques pour l'égalisation adaptative en mode autodidacte, *Annales de télécommunication*, 41, No 5-6, mai-juin 1986.

[11] S.P.Meyn, R.L.Tweedie, *Markov chains and stochastic stability*, Springer-Verlag 1993.

[12] L. Wu, *Moderate deviations of dependent random variables related to CLT and LIL*, Prépublication N. 118, Lab. de Probabilité de l'Université Paris VI, 1992.