THE LIMIT OF STOCHASTIC GRADIENT ALGORITHM FOR IDENTIFYING SYSTEMS NOT PERSISTENTLY EXCITED

CHEN HANFU (陈翰馥) AND GUO LEI (郭 雷)
(Institute of Systems Science, Academia Sinica, Beijing)
Received May 28, 1985.

Consider the multi-input multi-output stochastic system

$$y_n + A_1 y_{n-1} + \cdots + A_p y_{n-p} = B_1 u_{n-1} + \cdots + B_q u_{n-q} + w_n + C_1 w_{n-1} + \cdots + C_r w_{n-r}, \tag{1}$$

where the orders p, q, r are assumed known but the parameter

$$\theta^\tau = [-A_1 \cdots - A_p B_1 \cdots B_q C_1 \cdots C_r] \tag{2}$$

is unknown. Arbitrarily given initial values θ_0 and φ_0, the estimate θ_n for θ based on $\{u_i, i \leqslant n-1\}$ and $\{y_i, i \leqslant n\}$ is recursively given by the stochastic gradient algorithm

$$\theta_{n+1} = \theta_n + (\varphi_n / r_n)(y^\tau_{n+1} - \phi^\tau_n \theta_n), \tag{3}$$

$$\varphi^\tau_n = [y^\tau_n \cdots y^\tau_{n-p+1} u^\tau_n \cdots u^\tau_{n-q+1} y^\tau_n - \varphi^\tau_{n-1}\theta_{n-1} \cdots y^\tau_{n-r+1} - \varphi^\tau_{n-r}\theta_{n-r}], \tag{4}$$

$$r_n = 1 + \sum_{i=1}^{n} \|\varphi_i\|^2, \quad r_0 = 1. \tag{5}$$

Let z be the shift-back operator and let

$$C(z) = I + C_1 z + \cdots + C_r z^r. \tag{6}$$

Then $C(z) w_n$ is the system noise. Assume that $w_i = 0$, $i < 0$, w_i is $\mathscr{F}_i$-measurable and

$$E(W_n / \mathscr{F}_{n-1}) = 0, \; E(\|W_n\|^2 / \mathscr{F}_{n-1}) \leqslant C_0 r^\varepsilon_{n-1}, \; C_0 > 0, \; \varepsilon \in [0, 1), \; \forall n \geqslant 1, \tag{7}$$

where $\{\mathscr{F}_i\}$ is a family of nondecreasing σ-algebras.

For the case $r = 0$, the system noise is uncorrelated, then φ_n does not contain the last elements $y^\tau_i - \varphi^\tau_{i-1}\theta_{i-1}$, $n \leqslant i \leqslant n - r + 1$.

In the sequel $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$ always denote the maximum and minimum eigenvalues of a matrix X respectively. In the consistency consideration of θ_n the persistent excitation condition is usually invoked[1-4], i.e.

$$\lambda_{\max}\left(\sum_{i=1}^{n} \varphi_i \varphi^\tau_i\right) \Big/ \lambda_{\min}\left(\sum_{i=1}^{n} \varphi_i \varphi^\tau_i\right) \leqslant k < \infty, \tag{8}$$

with k possibly depending on ω. For the simple case $r = 0$ even if the above mentioned ratio of maximum to minimum eigenvalues diverges with a certain rate the least square estimates can still be consistent for single-input and single-output[5] and for multi-input and multi-output systems respectively[6]. For the $r>0$ case the authors have proved[7,8] that the estimates given by the stochastic gradient algorithm are strongly consistent for a class of systems not persistently excited. Hence, we cite some results which will be needed in our discussion. Set

$$\varphi_n^0 = [y_n^\tau \cdots y_{n-p+1}^\tau u_n^\tau \cdots u_{n-q+1}^\tau w_n^\tau \cdots w_{n-r+1}^\tau]^\tau, \tag{9}$$

$$\Phi^0(n+1, i) = \left(I - \frac{\varphi_n^0 \varphi_n^{0\tau}}{r_n^0}\right)\Phi^0(n, i), \Phi^0(i, i) = I \tag{10}$$

$$r_n^0 = 1 + \sum_{i=1}^n \|\varphi_i^0\|^2, \ r_0^0 = 1, \tag{11}$$

$$\Phi(n+1, i) = \left(I - \frac{\varphi_n \varphi_n^\tau}{r_n}\right)\Phi(n, i), \ \Phi(i, i) = I. \tag{12}$$

Theorem 1. (i) *If $C(z) - \frac{1}{2}I$ is strictly positive real, then $\Phi(n, 0)\underset{n\to\infty}{\longrightarrow}0$ is equivalent to $\Phi^0(n, 0)\underset{n\to\infty}{\longrightarrow}0$.* (ii) *If $C(z) - \frac{1}{2}I$ is strictly positive real and $\Phi(n, 0)\underset{n\to\infty}{\longrightarrow}0$ (or $\Phi^0(n, 0)\underset{n\to\infty}{\longrightarrow}0$), then $\theta_n \underset{n\to\infty}{\longrightarrow}\theta$ a. s. for any θ_0, where θ_n is given by* (3). (iii) *For the case $r=0$,*

$$\theta_n \underset{n\to\infty}{\longrightarrow}\theta \text{ a. s. for any } \theta_0 \text{ iff } \Phi(n, 0)\underset{n\to\infty}{\longrightarrow}0.$$

Theorem 2. *Suppose that $r_n \underset{n\to\infty}{\longrightarrow}\infty$, $\lim_{n\to\infty} r_n/r_{n-1} < \infty$, and that*

$$\lambda_{\max}\left(\sum_{i=1}^n \varphi_i\varphi_i^\tau\right)\Big/\lambda_{\min}\left(\sum_{i=1}^n \varphi_i\varphi_i^\tau\right) \leqslant M(\log r_n)^{\frac{1}{4}}, \ \forall n \geqslant N_0, \tag{13}$$

for some N_0 and M possibly depending on ω. Then $\Phi(n, 0)\underset{n\to\infty}{\longrightarrow}0$.

Theorem 2 is an algebraic result which is true for any vector sequence $\{\varphi_n\}$ only if $\Phi(n, i)$ and r_n are defined by (5) and (12) respectively, for example, φ_n, $\Phi(n, 0)$ and r_n can be replaced by φ_n^0, $\Phi^0(n, 0)$ and r_n^0 respectively. This theorem shows that the estimate given by the stochastic gradient algorithm converges to θ if the condition number of $\sum_{i=1}^n \varphi_i\varphi_i^\tau$ increases as $n\to\infty$ not faster than $(\log r_n)^{\frac{1}{4}}$. The question is in order to guarantee $\Phi(n, 0)\underset{n\to\infty}{\longrightarrow}0$ whether the condition number of $\sum_{i=1}^n \varphi_i\varphi_i^\tau$ can be allowed to grow faster than the rate mentioned above and what the limit is. The following theorem asserts that in order for Theorem 2 to hold the order $\frac{1}{4}$ of $\log r_n$ on the right-hand side of (13) cannot be enlarged to greater than 1.

Theorem 3. *For any $\delta > 0$ there exists a vector sequence $\{\varphi_n\}$ such that*

$$\lambda_{\max}\left(\sum_{i=1}^n \varphi_i\varphi_i^\tau\right)\Big/\lambda_{\min}\left(\sum_{i=1}^n \varphi_i\varphi_i^\tau\right) \leqslant M(\log r_n)^{1+\delta}, \tag{14}$$

and $\Phi(n, 0)\underset{n\to\infty}{\not\longrightarrow}0$, where r_n and $\Phi(n, 0)$ are defined by (5) *and* (12) *respectively.*

Proof. Suppose the contrary is true, i. e. (14) implies $\Phi(n, 0)\underset{n\to\infty}{\longrightarrow}0$.

Let $p=1, q=r=0$ and A_1 be stable in (1). For $\{w_n\}$ in addition to (7) we suppose that

$$\sup_i E(\|w_i\|^2/\mathscr{F}_{i-1}) < \infty, \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} w_i w_i^\tau = R > 0. \tag{15}$$

Let $\{\zeta_n\}$ be a sequence of iid random vectors independent of $\{w_n\}$ and with the same dimension as w_n, such that $E\zeta_n = 0$, $E\zeta_n\zeta_n^\tau = I$, $\|\zeta_n\| \leqslant c$ with c being a constant, $\forall n \geqslant 1$. Set

$$v_n = \zeta_n/(\log n)^{(1+\delta/2)}, \delta > 0, \tag{16}$$

$$\varphi_n^\tau = [y_n^\tau, v_n^\tau], \ \varphi_n^{v\tau} = [0, v_n^\tau]. \tag{17}$$

Take B_1 of compatible dimension and set

$$\theta^\tau = [-A_1, B_1]. \tag{18}$$

Then

$$y_{n+1} = \theta^\tau(\varphi_n - \varphi_n^v) + w_{n+1}. \tag{19}$$

We now estimate θ by the algorithm given by (3) and (5) with φ_n defined by (17). It is clear that θ_n remains invariant whatever B_1 values, since the right-hand side of (19) is independent of B_1. Hence θ_n cannot converge to θ given by (18). Then the theorem will be proved if we can show $\tilde{\theta}_n \triangleq \theta - \theta_n \xrightarrow[n\to\infty]{} 0$, under the converse assumption. It is obvious that

$$\tilde{\theta}_{n+1} = \tilde{\theta}_n - (\varphi_n/r_n)(\varphi_n^\tau\theta - \varphi_n^{v\tau}\theta - \varphi_n^\tau\theta_n + w_{n+1}^\tau) = \Phi(n, 0)\tilde{\theta}_0 + \sum_{j=0}^{n} \Phi(n+1, j+1) \frac{\varphi_j\varphi_j^{v\tau}}{r_j}\theta + \sum_{j=0}^{n} \Phi(n+1, j+1) \frac{\varphi_j w_{j+1}^\tau}{r_j}. \tag{20}$$

We first prove

$$\lim_{n\to\infty} \frac{1}{n} (\log n)^{1+\delta} \lambda_{\min}\left(\sum_{i=1}^{n} \varphi_i\varphi_i^\tau\right) > 0. \tag{21}$$

If (21) were not true, there would exist $\{\alpha_{n_k}\}, \{\beta_{n_k}\}$ such that (22)

$$\|\alpha_{n_k}\|^2 + \|\beta_{n_k}\|^2 = 1,$$

$$\frac{1}{n_k} (\log n_k)^{1+\delta} \sum_{i=1}^{n_k} (\alpha_{n_k}^\tau y_i + \beta_{n_k}^\tau v_i)^2 \xrightarrow[k\to\infty]{} 0. \tag{23}$$

Let $\{\varepsilon_n\}$ be a Martingale difference sequence, ε_n and x_{n+1} be $\mathscr{F}_n$-measurable and $\sup_n E(\|\varepsilon_n\|^2/\mathscr{F}_{n-1}) < \infty$. Then the following estimate takes place[5]

$$\sum_{i=1}^{n} x_i\varepsilon_i^\tau = O\left(\left[\sum_{i=1}^{n} \|x_i\|^2\right]^\beta\right), \quad \forall \beta > \frac{1}{2}. \tag{24}$$

Since A_1 is stable from (15) we know

$$\sum_{i=1}^{n} \|y_i\|^2 = O(n). \tag{25}$$

By (24) there is a constant $c_1 > 0$ such that

$$\left|\sum_{i=1}^{n_k} \alpha_{n_k}^\tau y_i \beta_{n_k}^\tau v_i\right| = O\left(\sum_{i=1}^{n_k} \|y_i\|^2\right)^\beta \leqslant c_1 n_k^\beta, \ \frac{1}{2} < \beta < 1.$$

Hence by (23) we have

$$\frac{1}{n_k}(\log n_k)^{1+\delta}\left[\alpha_{nk}^{\tau}\sum_{i=1}^{n_k} y_i y_i^{\tau}\alpha_{n_k} + \beta_{nk}^{\tau}\sum_{i=1}^{n_k} v_i v_i^{\tau}\beta_{nk}\right]\underset{k\to\infty}{\longrightarrow} 0. \tag{26}$$

By the Burkholder inequality and the C_r-inequality we know that there exist constants $c_2>0$, $c_3>0$, such that

$$E\left\|\sum_{i=2}^{n}(v_i v_i^{\tau}-(1/\log^{1+\delta}i)I)\right\|^{2+\frac{\delta}{2}} \leqslant c_2 E\left(\sum_{i=2}^{n}\|v_i v_i^{\tau}-(1/\log^{1+\delta}i)I\|^2\right)^{1+\frac{\delta}{4}}$$

$$\leqslant c_2 n^{\frac{\delta}{4}} E\sum_{i=2}^{n}\|v_i v_i^{\tau}-(1/\log^{1+\delta}i)I\|^{2+\frac{\delta}{2}} \leqslant c_3 n^{1+\frac{\delta}{4}}. \tag{27}$$

By (27) and the Borel-Cantelli lemma it follows that

$$\frac{1}{n^{\gamma}}\sum_{i=2}^{n}(v_i v_i^{\tau}-(1/\log^{1+\delta}i)I)\underset{n\to\infty}{\longrightarrow}0, \tag{28}$$

for any

$$\gamma\in\left(\left(2+\frac{\delta}{4}\right)\Big/\left(2+\frac{\delta}{2}\right),\ 1\right).$$

Consequently,

$$\varliminf_{n\to\infty}\lambda_{\min}\left((\log^{1+\delta}n/n)\sum_{i=2}^{n}v_i v_i^{\tau}\right)=\varliminf_{n\to\infty}\left[\lambda_{\min}\left(\frac{\log^{1+\delta}n}{n}\sum_{i=2}^{n}\left(v_i v_i^{\tau}-\frac{1}{\log^{1+\delta}i}I\right)\right)\right.$$

$$\left.+\frac{\log^{1+\delta}n}{n}\sum_{i=2}^{n}\frac{1}{\log^{1+\delta}i}I\right]\geqslant 1$$

which together with (22) and (26) implies that $\|\beta_{n_k}\|\underset{k\to\infty}{\longrightarrow}0$, $\|\alpha_{n_k}\|\underset{k\to\infty}{\longrightarrow}1$.

By (24) and (25) it is easy to see that

$$\varliminf_{n\to\infty}\lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n}y_i y_i^{\tau}\right)\geqslant\lambda_{\min}(R)>0.$$

Hence for the sufficiently large k

$$\frac{1}{n_k}(\log n_k)^{1+\delta}\alpha_{nk}^{\tau}\sum_{i=1}^{n_k}y_i y_i^{\tau}\alpha_{nk}\geqslant\frac{1}{2}\lambda_{\min}(R)\|\alpha_{n_k}\|^2(\log n_k)^{1+\delta}\underset{k\to\infty}{\longrightarrow}\infty,$$

which contradicts (26). Thus (21) is verified.

From the boundedness of $\|\zeta_i\|$ and (25), it follows that

$$\sum_{i=1}^{n}\|\varphi_i\|^2=\sum_{i=1}^{n}(\|y_i\|^2+\|v_i\|^2)=O(n).$$

Then there is a constant $c_4>0$ such that

$$\lambda_{\max}\left(\sum_{i=1}^{n}\varphi_i\varphi_i^{\tau}\right)\Big/\lambda_{\min}\left(\sum_{i=1}^{n}\varphi_i\varphi_i^{\tau}\right)\leqslant c_4(\log n)^{1+\delta},$$

which would yield $\Phi(n,0)\underset{n\to\infty}{\longrightarrow}0$ by the converse assumption. From [7] we have

$$\sum_{j=0}^{n} \|\Phi(n+1, j+1)\varphi_j\|^2 / r_j \leqslant d,$$

where d denotes the dimension of φ_n. Then for the sufficiently large n, it follows that $r_n \geqslant \sum_{i=1}^{n} \|y_i\|^2 \geqslant \frac{n}{2} \lambda_{\min}(R)$, and that

$$\sum_{j=n}^{\infty} \|\varphi_j^v\|^2 / r_j \leqslant (2c^2 / \lambda_{\min}(R)) \sum_{j=n}^{\infty} 1/j \log^{1+\delta} j < \infty.$$

since $\|v_n\|^2 \leqslant c^2 / \log^{1+\delta} n$. Therefore, the second term on the right-hand side of (20) tends to 0 as $n \to \infty$ while the third term also goes to 0 as shown in [7]. All of this would imply the impossible convergence $\theta_n \xrightarrow[n\to\infty]{} \theta$. Thus the theorem is proved.

Remark 1. The result of the present paper does not exclude the possibility that an estimate given by an algorithm different from the stochastic gradient one can converge to the true value under the condition number growing faster than that mentioned in Theorem 3.

Remark 2. There are two open problems: (i) We conjecture that $\Phi^0(n, 0) \xrightarrow[n\to\infty]{} 0$ is necessary for consistency of θ_n for any θ_0 when $r > 0$ and $C(z) - \frac{1}{2} I$ is strictly positive real. (ii) Write the right-hand side of (13) as $M(\log r_n)^\alpha$. We guess that Theorem 2 remains valid for $\alpha \leqslant 1$.

References

[1] Ljung, L., *IEEE Trans. Autom. Control*, AC-**21** (1976), 5: 779—781.
[2] Moore. J. B., *Automatica*, **14** (1978), 5: 505—509.
[3] Chen, H. F., *Int. J. Control*, **34** (1981), 5: 921—936.
[4] ———, *Recursive Estimation and Control for Stochastic Systems*, John Wiley, New York, 1985.
[5] Lai, T. L. & Wei, C. Z., *The Annals of Statistics*, **10** (1982), 1: 154—166.
[6] Chen, H. F., *Scientia Sinica*, Ser. A, **25** (1982), 7: 771—784.
[7] Chen, H. F. & Guo, L., *J. Syst. Sci. and Math. Scis.*, **5** (1985), 2: 81—93.
[8] ———, *Acta Math. Appl. Sinica* (English Series), **2** (1985), 2.