SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Convergence Analysis of ILC Input Sequence for Underdetermined Linear Systems

Dong SHEN*, Jian HAN & Youqing WANG

College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, *P.R. China*

Appendix A Stochastic Approximation Algorithm

The results presented here are cited from [1]. The stochastic approximation algorithm is

$$x_{k+1} = x_k + b_k(h(x_k) + \omega_k) \tag{A1}$$

where $h : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz, $\{b_k\}$ is a positive sequence satisfying $\sum_{k=1}^{\infty} b_k = \infty$ and $\sum_{k=1}^{\infty} b_k^2 < \infty$, and ω_k represents the noise. Assume the following assumptions hold.

Assumption 1. $\{\omega_k\}$ is a square-integrable martingale difference sequence with respect to the σ-fields $\{\mathcal{F}_k\}$ with $\mathcal{F}_k = \sigma(x_0, \omega_1, \cdots, \omega_k)$, satisfying $\mathbb{E}(\|\omega_{k+1}\|^2 | \mathcal{F}_k) \leqslant L(1 + \|x_k\|^2)$ a.s. for some $L > 0$.

Assumption 2. $\forall u$, $h_\infty(u) = \lim_{c\uparrow\infty} h(cu)/c$ exists and the o.d.e. $\dot{x} = h_\infty(x)$ has origin as its globally asymptotically stable equilibrium.

Assumption 3. $S \triangleq \{x \in \mathbb{R}^n : h(x) = 0\} \neq \emptyset$. Besides, $\exists$ a continuously differentiable Lyapunov function: $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}$ such that $\nabla\mathcal{L}(x)^T h(x) < 0$ for $x \notin S$.

Then $x_k \to S$ almost surely (a.s.) as k approaches to infinity.

Appendix B Proofs for the Stochastic Version of ILC Algorithm

Consider the lifted linear system

$$Y_k = HU_k + Y_k^0 + \xi_k. \tag{B1}$$

The stochastic version of the update law is defined as follows,

$$U_{k+1} = U_k + a_k e_{\alpha_k,k} H_{\alpha_k}. \tag{B2}$$

From (B2) one has

$$U_{k+1} = U_k + a_k\theta_{\alpha_k}^T(Y_d - HU_k)H_{\alpha_k} - a_k\theta_{\alpha_k}^T\xi_k H_{\alpha_k}.$$

Note that U_k belongs to the space $\mathcal{R}_H$, $\forall k$. Thus, there exists a unique vector ϕ_k such that $U_k = H^T\phi_k$. Consequently, the convergence of U_k is equivalent to the convergence of ϕ_k. The recursion of ϕ_k is given as

$$\phi_{k+1} = \phi_k + a_k\Gamma(Y_d - HH^T\phi_k) - a_k M_{\alpha_k}\xi_k + a_k(M_{\alpha_k} - \Gamma)(Y_d - HH^T\phi_k). \tag{B3}$$

where $M_{\alpha_k} \in \mathbb{R}^{m\times m}$ is a matrix with 1 in its α_k-th diagonal position and zero elsewhere, and $\Gamma \triangleq \mathbb{E}M_{\alpha_k} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_m\}$.

Proof of Lemma 1. To show the convergence, it is sufficient to verify the validity of Assumptions 1-3 in Appendix A.

Denote $\zeta_k = (M_{\alpha_k} - \Gamma)(Y_d - HH^T\phi_k)$, and then define a nondecreasing σ-algebra $\mathcal{F}_k = \sigma(\xi_1, \cdots, \xi_k, \zeta_1, \cdots, \zeta_k)$. Noticing that ξ_k is independent of M_{α_k}, it is easy to find that Assumption 1 holds.

The regression function in (B3) is $h(\phi) = \Gamma(Y_d - HH^T\phi)$, thus $\lim_{c\uparrow\infty} h(c\phi)/c = -\Gamma HH^T\phi$. Then it is not difficult to obtain that the origin is the globally asymptotically stable equilibrium of the ordinary differential equation (o.d.e.) $\dot{\phi} = -\Gamma HH^T\phi$. This verifies the condition Assumption 2.

* Corresponding author (email: shendong@mail.buct.edu.cn)

According to Assumption 3, the zero root of $h(\cdot)$ is $S = \{(HH^T)^{-1}Y_d\} = \{\phi_d\}$, which is obviously non-null. Define the continuously differentiable Lyapunov function be $\mathcal{L}(\phi) = \|H^T(\phi - \phi_d)\|^2$. Then $\mathcal{L}(\phi) = 0$ if and only if $\phi = \phi_d$, because H^T is of full-column rank. Besides, one has $\nabla\mathcal{L}(\phi) = 2HH^T(\phi - \phi_d)$, which results in that

$$\begin{aligned}\nabla\mathcal{L}^T(\phi)h(\phi) =&2(\phi - \phi_d)^T(HH^T)^T\Gamma(Y_d - HH^T\phi)\\ =&-2(\phi - \phi_d)^T(HH^T)^T\Gamma HH^T(\phi - \phi_d).\end{aligned}$$

This shows the validity of Assumption 3.

Thus the proof is completed with the help of stochastic approximation convergence result in Appendix A.

Then, by Lemma 1 and the relationship between U_k and ϕ_k, Theorem 1 can be established. In the following, we give the proof for Corollary 1.

Proof of Corollary 1. Noticing (B1), one is with no difficulty in deriving that

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|Y_d - Y_k\|^2 =&\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|Y_d - HU_k - \xi_k\|^2\\ =&\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|Y_d - HU_k\|^2(1 + o(1)) + \limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|\xi\|^2\end{aligned}$$

where $o(1) \to 0$ as k goes to infinity. It is obvious that the last term $\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|\xi\|^2$ is finite by A3 and could not be further improved by any input sequence. Thus the index $\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|Y_d - Y_k\|^2$ achieves it minimum when $\limsup_{n\to\infty}\sum_{k=1}^{n}\frac{1}{n}\|Y_d - HU_k\|^2(1 + o(1)) = 0$. Consequently, it is sufficient to show that $\widehat{E}_k = Y_d - HU_k \to 0$ as k goes to infinity, while the latter holds obviously according to Theorem 1.

Appendix C Proofs for the Deterministic Version of ILC Algorithm

The corresponding deterministic ILC update law is defined as

$$U_{k+1} = U_k + a_k\frac{1}{m}\left(e_{1,k}H_1 + e_{2,k}H_2 + \cdots + e_{m,k}H_m\right). \tag{C1}$$

By basic calculations, one has

$$U_{k+1} = U_k + \frac{a_k}{m}H^TE_k. \tag{C2}$$

The associated recursion of ϕ_k is as follows,

$$\phi_{k+1} = \phi_k + \frac{a_k}{m}(Y_d - HH^T\phi_k) - \frac{a_k}{m}\xi_k. \tag{C3}$$

The proofs for Lemma 2 and Theorem 2 can be conducted similar to those for Lemma 1 and Theorem 1. Here we provide a direct analysis for Corollary 2.

Proof of Corollary 2. Multiplying both sides of (C2) by H and then subtracting from Y_d, one has

$$\begin{aligned}Y_d - HU_{k+1} =&Y_d - HU_k - \frac{a_k}{m}HH^TE_k\\ =&Y_d - HU_k - \frac{a_k}{m}HH^T(Y_d - HU_k - \xi_k).\end{aligned}$$

Define $\chi_k \triangleq Y_d - HU_k$, then last equation could be rewritten as

$$\chi_{k+1} = \chi_k + a_k(-\frac{1}{m}HH^T\chi_k + \frac{1}{m}HH^T\xi_k).$$

The regression function is defined as $h(\chi) = -\frac{1}{m}HH^T\chi$, where HH^T is a positive definite matrix. Using similar steps of Lemma 1, one is easy to show that $\chi_k \to 0$ a.s. as k goes to infinity based on the convergence conclusion in Appendix A. Consequently, the averaged index $\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\|Y_d - Y_k\|^2$ achieves its minimum.

Appendix D Illustrative Simulations

In order to show the convergence of the input sequence, we consider the following stochastic point-to-point ILC problem. The system is described by

$$\begin{aligned}x_k(t+1) =&Ax_k(t) + Bu_k(t) + w_k(t+1),\\ y_k(t) =&Cx_k(t) + v_k(t),\end{aligned}$$

with matrices A, B, and C being

$$A = \begin{bmatrix} 0.5 & -0.07 & -0.26\\ -0.03 & 0.38 & -0.3\\ -0.1 & -0.13 & 0.4\end{bmatrix},\quad B = \begin{bmatrix} -1 & 0.1 & 0\\ 0 & -1 & 0.4\\ 0.2 & 0 & -1\end{bmatrix},\quad C = \begin{bmatrix} 1.8 & 0.5 & 0.3\\ 0.4 & 1.6 & 1.2\end{bmatrix}.$$

Here it is obvious that CB is of full-row rank, then the input and output sequences can be rewritten as super-vectors $U_k = [u_k^T(0), u_k^T(1), \cdots, u_k^T(N-1)]^T$, $Y_k = [y_k^T(1), y_k^T(2), \cdots, y_k^T(N)]^T$, and the transfer matrix

$$G = \begin{bmatrix} CB & 0 & \cdots & 0 \\ CAB & CB & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{N-1}B & CA^{N-2}B & \cdots & CB \end{bmatrix}.$$

Then one has the following relationship between input and output $Y_k = GU_k + \eta_k^0 + \epsilon_k$, where η_k^0 is the response to initial conditions, defined as $\eta_k^0 = [(CA)^T \quad (CA^2)^T \quad \cdots \quad (CA^N)^T]^T x_k(0)$ while $x_k(0)$ denotes the initial state. The stochastic noise term ϵ_k denotes a combination of $w_k(t)$ and $v_k(t)$.

For simple illustration, let $N = 6$, then $G \in \mathbb{R}^{12\times 18}$, $Y_k \in \mathbb{R}^{12}$ and $U_k \in \mathbb{R}^{18}$. The initial state is simply assumed to be zero, i.e., $x_k(0) = 0$. Suppose the reference points are placed to the 2nd, 4th, 5th, and 12th dimensions of Y_k. Denote the vector stacked by these four points as Y_k^{ob}. It is obvious that $Y_k^{ob} \in \mathbb{R}^4$. Then a matrix Φ is constructed as a sparse matrix with $\Phi_{1,2}$, $\Phi_{2,4}$, $\Phi_{3,5}$, $\Phi_{4,12}$ being 1 and other entries being 0, where $\Phi_{i,j}$ denotes the element at the ith row and jth column. Consequently, one has

$$Y_k^{ob} = \Phi G U_k + \Phi \eta_k^0 + \Phi \epsilon_k.$$

Note that $\Phi G \in \mathbb{R}^{4\times 18}$, then the underdetermined linear system (B1) has been established. Moreover, the stochastic noise $\Phi\epsilon_k$ is assumed to be normal distributed with covariance $0.1^2 I$ in this simulation.

The arbitrary given reference trajectory is $Y_d = [3\ 2\ 1.5\ 1]^T$. Then, in this point-to-point control problem, we have 18 inputs but only 4 outputs, which is underdetermined. It is easy to find out that the optimal U^* solving the optimization problem is $U^* = [-0.3939\ -1.7883\ -0.2628\ -0.4248\ -0.7637\ -0.0702\ -0.6241\ -0.1365\ 0.0100\ -0.0138\ -0.0480\ 0.0690\ -0.0148\ -0.1180\ 0.0767\ -0.0449\ -0.4375\ -0.1570]^T$.

The initial input U_0 is simply set to be zero. The learning step is $a_k = 6.5/k$.

In order to make a comparison, we run both the stochastic version and deterministic version of the ILC algorithms and then put the results together. For the stochastic version, the random variable α_k is assumed to be uniformly distributed on $\{1, 2, 3, 4\}$ for simplicity. The tracking errors of the selected four outputs, i.e., the 2nd, 4th, 5th, and 12th outputs, are shown in Figure D1, where the blue dashed lines and the red solid lines correspond to the stochastic and deterministic versions, respectively. The followings could be found from Figure D1. The stochastic version needs more iterations to obtain a good tracking performance as the information of every iteration is not fully used. In comparison, the deterministic version uses full tracking information generated in every iteration, which leads to a fast convergence speed. Moreover, the tracking deviation of the stochastic version is larger than that of the deterministic version for the first 40 iterations.

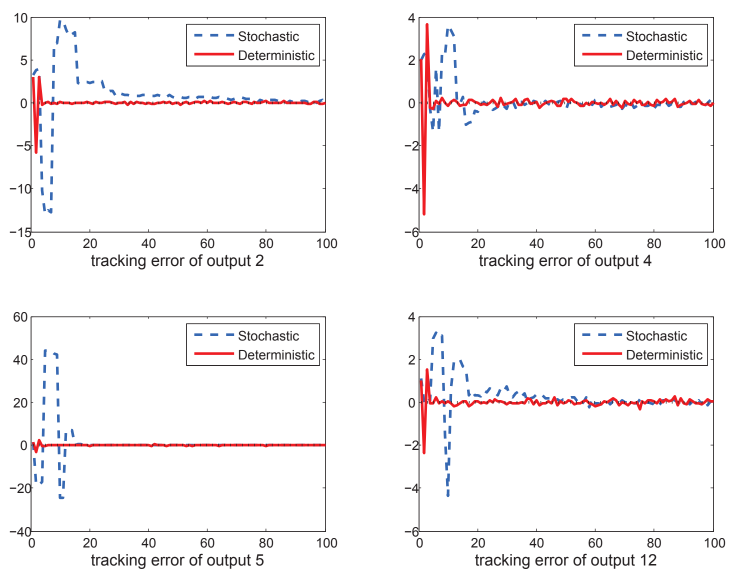

Figure D1 Convergence of selected outputs.

The convergence of the former nine inputs is shown in Figure D2 and Figure D3 for stochastic and deterministic versions, respectively. Comparing these two figures, it is observed that the deterministic case converges faster and closer to the desired values. The convergence of latter nine inputs is similar and thus is omitted for saving space.

References

1 G. Thoppe, V.S. Borkar, D. Manjunath. A stochastic Kaczmarz algorithm for network tomography. Automatica, 2014, 50(3): 910-914

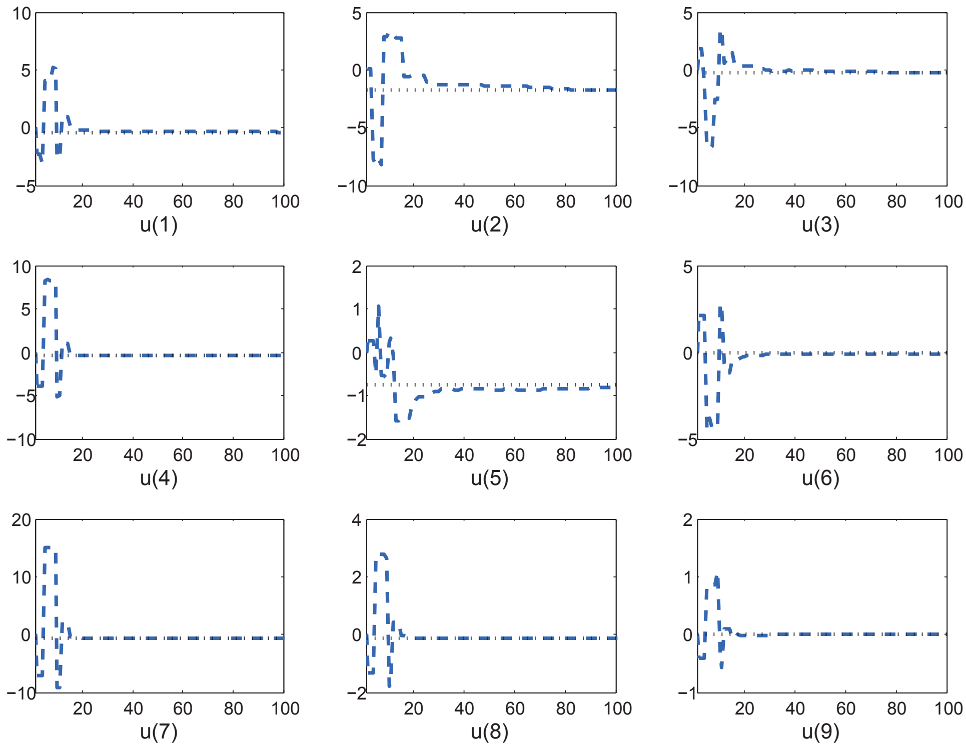

Figure D2 Convergence of the first 9 inputs for the stochastic version.

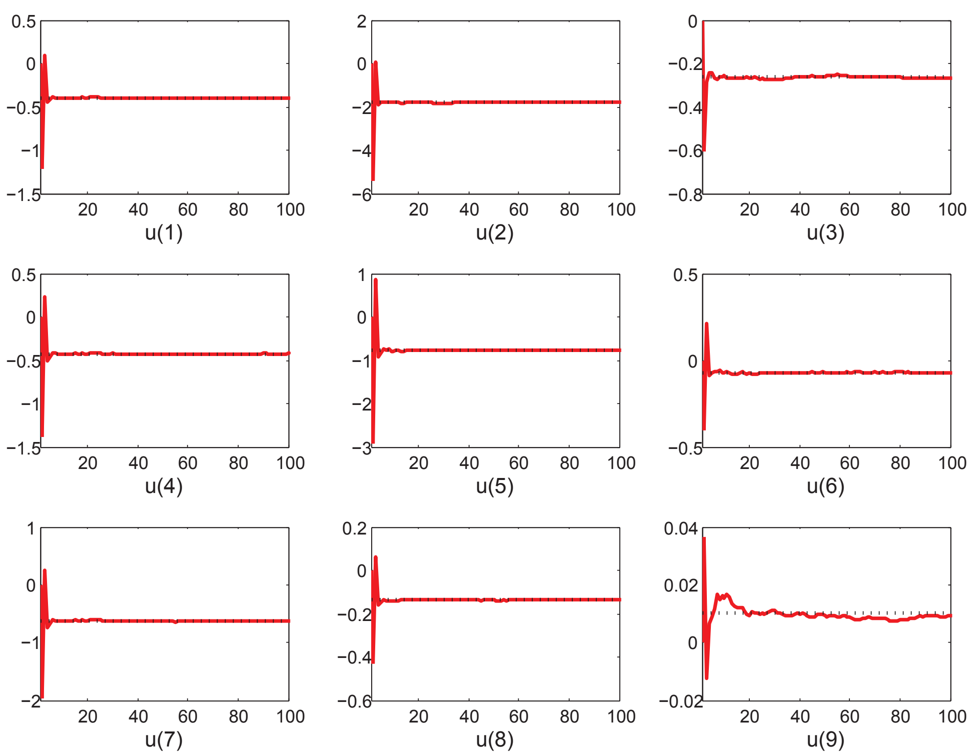

Figure D3 Convergence of the first 9 inputs for the deterministic version.