The Annals of Applied Probability
2016, Vol. 26, No. 6, 3630–3658
DOI: 10.1214/16-AAP1187

CENTRAL LIMIT THEOREMS OF A RECURSIVE STOCHASTIC ALGORITHM WITH APPLICATIONS TO ADAPTIVE DESIGNS

BY LI-XIN ZHANG[1]

Zhejiang University

Stochastic approximation algorithms have been the subject of an enormous body of literature, both theoretical and applied. Recently, Laruelle and Pagès [*Ann. Appl. Probab.* **23** (2013) 1409–1436] presented a link between the stochastic approximation and response-adaptive designs in clinical trials based on randomized urn models investigated in Bai and Hu [*Stochastic Process. Appl.* **80** (1999) 87–101; *Ann. Appl. Probab.* **15** (2005) 914–940], and derived the asymptotic normality or central limit theorem for the normalized procedure using a central limit theorem for the stochastic approximation algorithm. However, the classical central limit theorem for the stochastic approximation algorithm does not include all cases of its regression function, creating a gap between the results of Laruelle and Pagès [*Ann. Appl. Probab.* **23** (2013) 1409–1436] and those of Bai and Hu [*Ann. Appl. Probab.* **15** (2005) 914–940] for randomized urn models. In this paper, we establish new central limit theorems of the stochastic approximation algorithm under the popular Lindeberg condition to fill this gap. Moreover, we prove that the process of the algorithms can be approximated by a Gaussian process that is a solution of a stochastic differential equation. In our application, we investigate a more involved family of urn models and related adaptive designs in which it is possible to remove the balls from the urn, and the expectation of the total number of balls updated at each stage is not necessary a constant. The asymptotic properties are derived under much less stringent assumptions than those in Bai and Hu [*Stochastic Process. Appl.* **80** (1999) 87–101; *Ann. Appl. Probab.* **15** (2005) 914–940] and Laruelle and Pagès [*Ann. Appl. Probab.* **23** (2013) 1409–1436].

1. Introduction. Stochastic approximation (SA) algorithms, which have progressively gained sway thanks to the development of computer science and automatic control theory, have been the subject of many studies. An SA algorithm is also used in clinical trials to solve the dose-finding problem [see, e.g., Cheung (2010) and the citations therein]. The basic frameworks of SA algorithms and their theoretical results can be found in classical textbooks such as those by Benveniste, Métivier and Priouret (1990), Duflo (1996, 1997), Kushner and Clark (1978) and

Received March 2014; revised February 2016.
[1]Supported by Grants from the NSF of China (No. 11225104), the 973 Program (No. 2015CB352302) and the Fundamental Research Funds for the Central Universities.
MSC2010 subject classifications. Primary 60F05, 62L20; secondary 60F15, 60F17.
Key words and phrases. Stochastic approximation algorithms, central limit theorem, urn model, adaptive design, Gaussian approximation, the ODE method.

Kushner and Yin (2003). In this paper, we consider the following recursive SA algorithm defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geq 0}, \mathsf{P})$

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \frac{\mathbf{h}(\boldsymbol{\theta}_n)}{n+1} + \frac{\Delta\mathbf{M}_{n+1} + \mathbf{r}_{n+1}}{n+1}, \tag{1.1}$$

where $\boldsymbol{\theta}_n$ is a row vector in $\mathbb{R}^d$, the regression function $\mathbf{h}: \mathbb{R}^d \to \mathbb{R}^d$ is a real vector-valued function, $\boldsymbol{\theta}_0$ is a finite random vector, $\mathbf{M}_0 = \mathbf{0}$, $\{\Delta\mathbf{M}_n, \mathscr{F}_n; n \geq 1\}$ is a sequence of martingale differences and $\mathbf{r}_n$ is a remainder term.

Very recently, Laruelle and Pagès (2013) presented a link between this SA algorithm and the response-adaptive randomization process in clinical trials based on the randomized Generalized Friedman Urn [GFU, also known as a generalized Pólya urn (GPU) in the literature] models investigated in Bai and Hu (1999, 2005). They derived the almost sure (a.s.) convergence and the joint asymptotic normality or central limit theorem (CLT) of the normalized procedure for both the urn compositions and the assignments by applying SA theory. Higueras et al. (2003, 2006) also showed that the urn compositions can be written as an SA algorithm under some extra assumptions, including that the total number of balls added to the urn at each stage is the same. However, they did not consider the procedure of assignments.

The main tool used by Laruelle and Pagès (2013) to derive the asymptotic normality of GPU models is the CLT for an SA algorithm. Various types of results on the CLT of $\boldsymbol{\theta}_n$ have been established in the literature under certain conditions, especially when $\mathbf{r}_n \equiv \mathbf{0}$, and they can thus be found in classical textbooks such as that by Kushner and Yin (2003), page 330. For results in a more general framework, one can refer to Pelletier (1998). Let $\boldsymbol{\theta}^*$ be an equilibrium point of $\{\mathbf{h} = \mathbf{0}\}$. Assume that the function $\mathbf{h}$ is differentiable at $\boldsymbol{\theta}^*$ and that all of the eigenvalues of $D\mathbf{h}(\boldsymbol{\theta}^*) =: (\partial h_i(\boldsymbol{\theta}^*)/\partial\theta_j; i, j = 1, \dots, d)$ have positive real parts. Denote $\rho = \mathrm{Re}(\lambda_{\min})$, where $\lambda_{\min}$ is the eigenvalue of $D\mathbf{h}(\boldsymbol{\theta}^*)$ with the lowest real part. In considering the CLT, $\rho > 1/2$ is usually assumed as a basic condition. The following CLT can be found in Duflo (1997), Benveniste, Métivier and Priouret (1990) and Kushner and Yin (2003) [cf. Theorem A.2 of Laruelle and Pagès (2013)] with different groups of conditions.

THEOREM 1.1. *Let $\boldsymbol{\theta}^*$ be an equilibrium point of $\{\mathbf{h} = \mathbf{0}\}$. Suppose that $\boldsymbol{\theta}_n \to \boldsymbol{\theta}^*$ a.s. and assume that for some $\delta > 0$,*

$$\begin{aligned} &\sup_{n\geq 0} \mathsf{E}\big[\|\boldsymbol{\Delta}\mathbf{M}_{n+1}\|^{2+\delta} | \mathscr{F}_n\big] < +\infty \qquad a.s. \\ &\mathsf{E}\big[(\boldsymbol{\Delta}\mathbf{M}_{n+1})^{\mathrm{t}} \boldsymbol{\Delta}\mathbf{M}_{n+1} | \mathscr{F}_n\big] \to \boldsymbol{\Gamma} \qquad a.s., \end{aligned} \tag{1.2}$$

where $\boldsymbol{\Gamma}$ is a deterministic symmetric positive semidefinite matrix and for an $\varepsilon > 0$,

$$(n+1)\mathsf{E}\big[\|\mathbf{r}_{n+1}\|^2 \mathbb{I}_{\{\|\boldsymbol{\theta}_n - \boldsymbol{\theta}^*\| \leq \varepsilon\}}\big] \to 0. \tag{1.3}$$

Suppose $\rho := \mathrm{Re}(\lambda_{\min}) > 1/2$. *Then*

$$\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*) \xrightarrow{\mathscr{D}} N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{1.4}$$

where

$$\boldsymbol{\Sigma} := \int_0^\infty \left(e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}_d/2)u}\right)^{\mathrm{t}} \boldsymbol{\Gamma} e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}_d/2)u}\, du, \tag{1.5}$$

and $\mathbf{I}_d$ *is a* $d \times d$*-identity matrix.*

In the cases of $\rho = 1/2$ and $0 < \rho < 1/2$, partial results have been established when $D\mathbf{h}(\boldsymbol{\theta}^*)$ is diagonal. For example, Duflo (1997), cf. Theorem 2.2.12 showed that if $D\mathbf{h}(\boldsymbol{\theta}^*) = \rho\mathbf{I}_d$, the CLT holds with rate $\sqrt{\frac{n}{\log n}}$ when $\rho = 1/2$, and $n^\rho(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)$ almost surely converges to a random vector when $0 < \rho < 1/2$. Laruelle and Pagès (2013) summarized this kind of results to their Theorem A.2 and applied them to GPU, but they missed the condition that $D\mathbf{h}(\boldsymbol{\theta}^*)$ is diagonal, thus the results in their Theorem 2.2(b) and (c) are not consistent with those in Theorems 2.2 and 3.2 of Bai and Hu (2005). The main purpose of this paper is to establish the CLT for a general matrix $D\mathbf{h}(\boldsymbol{\theta}^*)$. We find that in the cases of $\rho = 1/2$ and $0 < \rho < 1/2$, the results for a general matrix $D\mathbf{h}(\boldsymbol{\theta}^*)$ are much more complex than those for a diagonal matrix.

In the next section, we establish general asymptotic results on the SA algorithm (1.1) under the popular Lindeberg condition, which is less restrictive than (1.2). From these results, we find that the limiting behavior of the SA algorithm depends on not only the value of the eigenvalue $\lambda_{\min}$ but also the multiplicity of this eigenvalue. Moreover, $n^\rho(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)$ does not converge in general when $\rho < 1/2$. Further, in Section 3, we prove that the process of the algorithms can be approximated almost surely by a Gaussian process when $\rho \le 1/2$ under a condition a little more stringent than the Lindeberg condition, and the Gaussian process is a solution of a stochastic differential equation.

As an application of SA theory, in Section 4, we derive the asymptotic properties of an important class of response-adaptive designs in clinical trials based on the randomized GFU. Laruelle and Pagès (2013) provided a clever way to study the asymptotic normality of randomized urn models. Motivated by their idea, as an application of the new SA theory, in Section 4, we retrieve the a.s. convergence and the asymptotic normality of the randomized GFU models under assumptions much less stringent than those in Bai and Hu (1999, 2005). We investigate a more involved family of urn models in which it is possible for the balls of each type to be removed from the urn, and the expectation of the total number of balls updated at each stage is not necessarily a constant. The asymptotic property of such urns is stated as an open problem in Hu and Rosenberger (2006), page 158, and examples of models featuring the removal of balls can be found in Hu and Rosenberger (2006), Janson (2004), Zhang et al. (2011), etc. For this general framework, the first problem is to show the a.s. convergence. The methods of Bai and Hu (1999, 2005) and Higueras et al. (2003, 2006) do not work because they depend heavily on the assumption that the total number of balls or the expectation of the total number of balls updated at each stage is a constant. We show that the ordinary differential equation (ODE) method proposed by Laruelle and Pagès (2013) is valid to prove the a.s. convergence, although in their original proof, such an assumption is also needed. However, the ODE is no longer a linear equation, as it was in Laruelle and Pagès (2013). The convergence rate of the urn model depends on the second-largest eigenvalue λ_{sec} and the largest eigenvalue $\lambda_{\max}$ of the urn's limiting generating matrix. When the ratio $\lambda_{\mathrm{sec}}/\lambda_{\max}$ of these two eigenvalues is large ($>1/2$), the asymptotic property is also an unsolved problem [cf. Hu and Rosenberger (2006), page 158]. In Section 4, a clear answer to this open problem is provided.

Finally, some basic results on the convergence of the recursive algorithm and multi-dimensional martingales are given in the Appendix.

In the sequel to this paper, the Euclidean norm of a vector $\mathbf{x} = (x_1, \dots, x_d)$ is defined to be $\|\mathbf{x}\| = \sqrt{\sum_j x_j^2}$, and the norm of a matrix $\mathbf{M}$ is defined to be $\|\mathbf{M}\| = \sup\{\|\mathbf{xM}\| : \|\mathbf{x}\| = 1\}$. $\mathbf{1} = (1, \dots, 1)$ denotes the unit row vector in R^d. $\mathbf{x}^{\mathrm{t}}$ denotes the transpose of $\mathbf{x}$. For a function $\mathbf{f}(t) : \mathbb{R}^d \to \mathbb{R}$, $\dot{\mathbf{f}}(t)$ denotes its derivative, and for a function $\mathbf{f}(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^d$, $D\mathbf{f}(\mathbf{x})$ denotes the matrix of its partial derivatives with the (i, j)th element being $\partial f_i(\mathbf{x})/\partial x_j$. Further, for two positive sequences $\{a_n\}$ and $\{b_n\}$ and a sequence of vectors $\{\mathbf{v}_n\}$, we write $a_n = O(b_n)$ if there is a constant C such that $a_n \le Cb_n$, $a_n \sim b_n$ if $a_n/b_n \to 1$, $a_n \approx b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$, $\mathbf{v}_n = O(b_n)$ if there is a constant C such that $\|\mathbf{v}_n\| \le Cb_n$, and $\mathbf{v}_n = o(b_n)$ if $\|\mathbf{v}_n\|/b_n \to 0$.

2. Central Limit Theorems.

In this section, we consider the central limit theorem of the SA logarithm (1.1). We first need some assumptions. The first two are on the differentiability of the function $\mathbf{h}(\cdot)$.

ASSUMPTION 2.1. Let $\boldsymbol{\theta}^*$ be an equilibrium point of $\{\mathbf{h} = \mathbf{0}\}$. Assume that function $\mathbf{h}$ is differentiable at $\boldsymbol{\theta}^*$ and that all of the eigenvalues of $D\mathbf{h}(\boldsymbol{\theta}^*)$ have positive real parts.

Under Assumption 2.1, we have that $\mathbf{h}(\boldsymbol{\theta}^*) = \mathbf{0}$,

$$\mathbf{h}(\boldsymbol{\theta}) = \mathbf{h}(\boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)D\mathbf{h}(\boldsymbol{\theta}^*) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|) \qquad \text{as } \boldsymbol{\theta} \to \boldsymbol{\theta}^*, \tag{2.1}$$

and $D\mathbf{h}(\boldsymbol{\theta}^*)$ has the following Jordan canonical form:

$$\mathbf{T}^{-1} D\mathbf{h}(\boldsymbol{\theta}^*)\mathbf{T} = \operatorname{diag}(\mathbf{J}_1, \mathbf{J}_2, \dots, \mathbf{J}_s),$$

where

$$\mathbf{J}_t = \begin{pmatrix} \lambda_t & 1 & 0 & \dots & 0 \\ 0 & \lambda_t & 1 & \dots & 0 \\ \vdots & \dots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_t & 1 \\ 0 & 0 & 0 & \dots & \lambda_t \end{pmatrix}_{\nu_t \times \nu_t} = \lambda_t \mathbf{I}_{\nu_t} + \overline{\mathbf{J}}_{\nu_t},$$

where $\mathbf{I}_{\nu_t}$ is a $\nu_t \times \nu_t$-identity matrix and $\mathrm{Sp}(D\mathbf{h}(\boldsymbol{\theta}^*)) = \{\lambda_1, \dots, \lambda_s\}$ is the set of eigenvalues of $D\mathbf{h}(\boldsymbol{\theta}^*)$. Let $\rho = \min\{\mathrm{Re}(\lambda), \lambda \in \mathrm{Sp}(D\mathbf{h}(\boldsymbol{\theta}^*))\}$ and $\nu = \max\{\nu_t : \mathrm{Re}(\lambda_t) = \rho\}$.

When we consider the case of $\rho \le 1/2$, we need a condition a little more stringent than (2.1).

ASSUMPTION 2.2. Suppose that Assumption 2.1 is satisfied, $\mathbf{h}(\boldsymbol{\theta}^*) = \mathbf{0}$ and

$$\mathbf{h}(\boldsymbol{\theta}) = \mathbf{h}(\boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*) D\mathbf{h}(\boldsymbol{\theta}^*) + o(\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|^{1+\varepsilon}) \qquad \text{as } \boldsymbol{\theta} \to \boldsymbol{\theta}^* \tag{2.2}$$

for some $\varepsilon > 0$.

We show the CLT under the following conditional Lindeberg's condition, which is popular in the study of the CLT for martingales.

ASSUMPTION 2.3. Suppose that the following Lindeberg's condition is satisfied:

$$\frac{1}{n} \sum_{m=1}^{n} \mathsf{E}\left[\|\Delta \mathbf{M}_m\|^2 \mathbb{I}\{\|\Delta \mathbf{M}_m\| \ge \varepsilon \sqrt{n}\} | \mathscr{F}_{m-1}\right] \to 0 \qquad \text{a.s.} \quad \text{or} \quad \text{in } L_1, \forall \varepsilon > 0. \tag{2.3}$$

Further, assume that

$$\frac{1}{n} \sum_{m=1}^{n} \mathsf{E}\left[(\Delta \mathbf{M}_m)^{\mathrm{t}} \Delta \mathbf{M}_m | \mathscr{F}_{m-1}\right] \to \boldsymbol{\Gamma} \qquad \text{a.s.} \quad \text{or} \quad \text{in } L_1, \tag{2.4}$$

where $\boldsymbol{\Gamma}$ is a symmetric positive semidefinite random matrix.

In Assumption 2.3, $\boldsymbol{\Gamma}$ is a $\mathscr{F}_\infty(= \bigvee_n \mathscr{F}_n)$ measurable random matrix, which was assumed to be deterministic in Bai and Hu (1999, 2005), Pelletier (1998) and Laruelle and Pagès (2013). Although $\boldsymbol{\Gamma}$ is usually deterministic in practice, we consider the general martingales, as in Hall and Heyde (1980). Our main results are the following two theorems on the limiting properties of the sequence $\{\boldsymbol{\theta}_n\}$ in the cases of $0 < \rho < 1/2$ and $\rho = 1/2$.

THEOREM 2.1. *Suppose that* $\boldsymbol{\theta}_n \to \boldsymbol{\theta}^*$ *a.s., Assumptions* 2.2 *and* 2.3 *are satisfied, and* $\rho = 1/2$. *Further, for the remainder term* $\mathbf{r}_n$ *we assume that*

$$\sum_{k=1}^{n} \mathbf{r}_k = o(\sqrt{n/\log n}) \qquad a.s. \tag{2.5}$$

or

$$\sum_{m=1}^{n} \frac{\|\mathbf{r}_m\|}{\sqrt{m}} = o(\sqrt{\log n}) \qquad a.s. \quad or \quad in\ L_1. \tag{2.6}$$

Then

$$\frac{\sqrt{n}}{(\log n)^{\nu-1/2}}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*) \xrightarrow{D} N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}) \qquad (stably), \tag{2.7}$$

where

$$\widetilde{\boldsymbol{\Sigma}} = \lim_{n\to\infty} \frac{1}{(\log n)^{2\nu-1}} \int_0^{\log n} \big(e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}_d/2)u}\big)^{\mathrm{t}} \boldsymbol{\Gamma} e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}_d/2)u}\, du, \tag{2.8}$$

and $N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})$ *denotes a mixing normal distribution with the conditional characteristic function* $f(\mathbf{t}) = \exp\{-\frac{1}{2}\mathbf{t}\widetilde{\boldsymbol{\Sigma}}\mathbf{t}^t\}$ *for given* $\widetilde{\boldsymbol{\Sigma}}$. *Moreover,* $\widetilde{\boldsymbol{\Sigma}}$ *satisfies*

$$\big(\mathbf{T}^{\star \mathrm{t}}\widetilde{\boldsymbol{\Sigma}}\mathbf{T}\big)_{ij} = \frac{1}{((\nu-1)!)^2}\frac{1}{2\nu-1}\mathbf{t}_{a1}^{\star}\boldsymbol{\Gamma}\mathbf{t}_{b1}^{\mathrm{t}}, \tag{2.9}$$

whenever $i = \nu_1 + \cdots + \nu_a$, $j = \nu_1 + \cdots + \nu_b$ *and* $\lambda_a = \lambda_b$, $\mathrm{Re}(\lambda_a) = 1/2$, $\nu_a = \nu_b = \nu$, *and* $(\mathbf{T}^{\star \mathrm{t}}\widetilde{\boldsymbol{\Sigma}}\mathbf{T})_{ij} = 0$ *otherwise. Here,* $\mathbf{x}^\star$ *is the conjugate vector of a complex vector* $\mathbf{x}$ *and* $\mathbf{t}_{a1}^{\mathrm{t}}$ *is the first column vector of the ath block in* $\mathbf{T} = [\ldots, \mathbf{t}_{a1}^{\mathrm{t}}, \ldots, \mathbf{t}_{a\nu_a}^{\mathrm{t}}, \ldots]$. *Further, let* $\mathbf{r}_{a\nu_a}$ *be the last row vector of the ath block in* $\mathbf{T}^{-1} = [\ldots, \mathbf{r}_{a1}^{\mathrm{t}}, \ldots, \mathbf{r}_{a\nu_a}^{\mathrm{t}}, \ldots]^{\mathrm{t}}$. *Then* $\mathbf{r}_{a\nu_a}$ *and* $\mathbf{t}_{a1}^{\mathrm{t}}$ *are respectively the left and right eigenvectors of* $\mathbf{H}$ *with respect to the eigenvalue* λ_a, *and*

$$\widetilde{\boldsymbol{\Sigma}} = \frac{1}{((\nu-1)!)^2}\frac{1}{2\nu-1} \sum_{a,b:\lambda_a=\lambda_b, \mathrm{Re}(\lambda_a)=1/2, \nu_a=\nu_b=\nu} \big(\mathbf{r}_{a\nu_a}^{\mathrm{t}}\mathbf{t}_{a1}\big)^{\star}\boldsymbol{\Gamma}\big(\mathbf{t}_{b1}^{\mathrm{t}}\mathbf{r}_{b\nu_b}\big).$$

THEOREM 2.2. *Suppose that* $\boldsymbol{\theta}_n \to \boldsymbol{\theta}^*$ *a.s., Assumption* 2.2 *is satisfied with* $0 < \rho < 1/2$. *Further, assume that*

$$\sum_{m=1}^{n} \mathsf{E}\big[(\Delta\mathbf{M}_m)^{\mathrm{t}}\Delta\mathbf{M}_m | \mathscr{F}_{m-1}\big] = O(n) \qquad a.s. \quad or \quad in\ L_1, \quad and \tag{2.10}$$

$$\sum_{k=1}^{n} \mathbf{r}_k = o\big(n^{1-\rho-\delta_0}\big) \qquad a.s.\ for\ some\ \delta_0 > 0. \tag{2.11}$$

Then there are complex random variables $\xi_1, \ldots, \xi_s$ *such that*

$$\frac{n^\rho}{(\log n)^{\nu-1}}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*) - \sum_{a:\mathrm{Re}(\lambda_a)=\rho, \nu_a=\nu} e^{-\mathrm{i}\,\mathrm{Im}(\lambda_a)\log n}\xi_a\mathbf{e}_a\mathbf{T}^{-1} \to \mathbf{0} \qquad a.s.,$$

where $\mathrm{i}=\sqrt{-1}$, $\mathbf{e}_a=(\mathbf{0},\dots,\mathbf{0},0,\dots,0,1,\mathbf{0},\dots,\mathbf{0})$ *is the vector such that the* ν_a*th element of its block* a *is* 1 *and other elements are zero*, *and* $\mathbf{e}_a\mathbf{T}^{-1}=\mathbf{r}_{a\nu_a}$ *is a right eigenvector of* $\mathbf{H}$ *with respect to the eigenvalue* λ_a.

When $\rho>1/2$, the CLT is classical and can be found in the literature under various groups of settings. Moreover, the stepsize $\frac{1}{n+1}$ can be more general. One can refer to Benveniste, Métivier and Priouret (1990), Duflo (1997), Kushner and Yin (2003), cf. Theorem 2.1, Chapter 10, Pelletier (1998), etc. Here, in considering applications to a general framework of GPU models, we present the following example under the Lindeberg condition.

THEOREM 2.3. *Suppose that* $\boldsymbol{\theta}_n\to\boldsymbol{\theta}^*$ *a.s.*, *Assumptions* 2.1 *and* 2.3 *are satisfied*, *and* $\rho>1/2$. *Further, for the remainder term* $\mathbf{r}_n$ *we assume that*

$$\sum_{k=1}^n \mathbf{r}_k=o(\sqrt{n})\qquad a.s.\quad or\quad in\ L_1. \tag{2.12}$$

Then

$$\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta}^*)\xrightarrow{D}N(\mathbf{0},\widetilde{\boldsymbol{\Sigma}})\qquad (stably), \tag{2.13}$$

where

$$\widetilde{\boldsymbol{\Sigma}}=\int_0^\infty \bigl(e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}_d/2)u}\bigr)^{\mathrm{t}}\boldsymbol{\Gamma}e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}_d/2)u}\,du. \tag{2.14}$$

REMARK 2.1. The condition (2.2) cannot be weakened to (2.1) in Theorems 2.1 and 2.2. The convergence rates in conditions (2.5) or (2.6) cannot be weakened in Theorem 2.1.

From the proof of Theorems [cf. (2.22) and (2.25)], we have the following corollary on the rate of the a.s. convergence.

COROLLARY 2.1. *Suppose that* $\boldsymbol{\theta}_n\to\boldsymbol{\theta}^*$ *a.s.*, *Assumption* 2.1 *is satisfied with* $\rho>0$. *Further assume that condition* (2.10) *in Theorem* 2.2 *is satisfied*, *and* $\frac{1}{n}\sum_{k=1}^n\mathbf{r}_k=o(n^{-\frac12\wedge\rho+\delta})$ *a.s. for all* $\delta>0$. *Then*

$$\boldsymbol{\theta}_n-\boldsymbol{\theta}^*=o\bigl(n^{-\frac12\wedge\rho+\delta}\bigr)\qquad a.s.\ for\ all\ \delta>0.$$

Now, we give the proof of Theorems 2.1–2.3. Write $\mathbf{H}=D\mathbf{h}(\boldsymbol{\theta}^*)$,

$$\mathbf{H}(\boldsymbol{\theta})=\begin{cases}\mathbf{H}+\dfrac{(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^{\mathrm{t}}}{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|}\dfrac{\mathbf{h}(\boldsymbol{\theta})-\mathbf{h}(\boldsymbol{\theta}^*)-(\boldsymbol{\theta}-\boldsymbol{\theta}^*)\mathbf{H}}{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|}, & \text{if } \boldsymbol{\theta}\neq\boldsymbol{\theta}^*,\\ \mathbf{H}, & \text{if } \boldsymbol{\theta}=\boldsymbol{\theta}^*.\end{cases}$$

Then $\mathbf{H}(\boldsymbol{\theta}) \to \mathbf{H}$ as $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$ and

$$\mathbf{h}(\boldsymbol{\theta}) = \mathbf{h}(\boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)\mathbf{H}(\boldsymbol{\theta}).$$

Let $\mathbf{H}_{n+1} = \mathbf{H}(\boldsymbol{\theta}_n)$, $\boldsymbol{\Pi}_n^n = \mathbf{I}_d$ and for all $0 \le m \le n-1$

$$\boldsymbol{\Pi}_m^n = \left(\mathbf{I}_d - \frac{\mathbf{H}_{m+1}}{m+1}\right) \cdots \left(\mathbf{I}_d - \frac{\mathbf{H}_n}{n}\right), \qquad \widetilde{\boldsymbol{\Pi}}_m^n = \prod_{j=m+1}^{n} \left(\mathbf{I}_d - \frac{\mathbf{H}}{j}\right). \tag{2.15}$$

Then $\mathbf{H}_n \to \mathbf{H}$ a.s. as $n \to \infty$. It follows that for all $1 \le m \le n-1$, $\|\boldsymbol{\Pi}_m^n\| \le C_\delta (n/m)^{-\rho+\delta}$ a.s. and $\|\boldsymbol{\Pi}_0^n\| \le C_\delta n^{-\rho+\delta}$ a.s. by Proposition B.1(i) in the Appendix. By (1.1),

$$\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}^* = (\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\left(\mathbf{I}_d - \frac{\mathbf{H}_{n+1}}{n+1}\right) + \frac{\Delta \mathbf{M}_{n+1} + \mathbf{r}_{n+1}}{n+1}. \tag{2.16}$$

It follows that

$$\boldsymbol{\theta}_n - \boldsymbol{\theta}^* = (\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*)\boldsymbol{\Pi}_0^n + \sum_{m=1}^{n} \frac{\Delta \mathbf{M}_m}{m} \boldsymbol{\Pi}_m^n + \sum_{m=1}^{n} \frac{\mathbf{r}_m}{m} \boldsymbol{\Pi}_m^n. \tag{2.17}$$

If we write $\mathbf{s}_n = \sum_{m=1}^n \mathbf{r}_m$, then the last term is

$$\begin{aligned} \sum_{m=1}^{n} \frac{\mathbf{r}_m}{m} \boldsymbol{\Pi}_m^n &= \frac{\mathbf{s}_n}{n} \boldsymbol{\Pi}_n^n + \sum_{m=1}^{n-1} \mathbf{s}_m \left(\frac{1}{m} \boldsymbol{\Pi}_m^n - \frac{1}{m+1} \boldsymbol{\Pi}_{m+1}^n\right) \\ &= \frac{\mathbf{s}_n}{n} \boldsymbol{\Pi}_n^n + \sum_{m=1}^{n-1} \mathbf{s}_m \frac{\mathbf{I}_d - \mathbf{H}_{m+1}}{m(m+1)} \boldsymbol{\Pi}_{m+1}^n. \end{aligned} \tag{2.18}$$

PROOF OF THEOREMS 2.1 AND 2.3. First, we consider the case of $\rho = 1/2$. Suppose the conditions in Theorem 2.1 are satisfied. At first, (2.5) or (2.6) will imply that

$$\sum_{m=1}^{n} \mathbf{r}_m = o(n^{1/2+\delta}) \qquad \text{a.s. } \forall \delta > 0. \tag{2.19}$$

In fact, it is sufficient to show that (2.6) implies (2.19). Note that

$$\sum_{m=1}^{n} \|\mathbf{r}_m\| \le \sqrt{n} \sum_{m=1}^{n} \frac{\|\mathbf{r}_m\|}{\sqrt{m}} = o(\sqrt{n \log n}) \qquad \text{a.s.} \quad \text{or} \quad \text{in } L_1. \tag{2.20}$$

Assume that the above inequality holds in the sense of L_1. Then

$$\sum_k \mathsf{P}\left(\sum_{m=1}^{2^{k+1}} \|\mathbf{r}_m\| \ge \varepsilon (2^k)^{1/2+2\delta}\right) \le C \sum_k \frac{(2^{k+1} \log 2^{k+1})^{1/2}}{(2^k)^{1/2+2\delta}} < \infty.$$

It follows that for $2^k \le n \le 2^{k+1}$,

$$\frac{\sum_{m=1}^{n}\|\mathbf{r}_m\|}{n^{1/2+2\delta}} \le \frac{\sum_{m=1}^{2^{k+1}}\|\mathbf{r}_m\|}{(2^k)^{1/2+2\delta}} \to 0 \qquad \text{a.s.}$$

(2.19) is true.

In contrast, one can verify that condition (2.10) or (2.4) implies that

$$\mathbf{M}_n = o\big(n^{1/2+\delta}\big) \qquad \text{a.s. for all } \delta > 0. \tag{2.21}$$

Recall (2.17), (2.18) and $\mathbf{H}_n \to \mathbf{H}$ a.s. as $n \to \infty$. We have

$$\begin{aligned}
\boldsymbol{\theta}_n - \boldsymbol{\theta}^* &= (\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*)\boldsymbol{\Pi}_0^n + \frac{\mathbf{M}_n + \mathbf{s}_n}{n}\boldsymbol{\Pi}_n^n + \sum_{m=1}^{n-1}(\mathbf{M}_m + \mathbf{s}_m)\frac{\mathbf{I}_d - \mathbf{H}_{m+1}}{m(m+1)}\boldsymbol{\Pi}_{m+1}^n \\
&= o\big(n^{-\rho+\delta}\big) + \frac{o(n^{1/2+\delta})}{n} + \sum_{m=1}^{n-1}\frac{o(m^{1/2+\delta})}{m(m+1)}\left(\frac{n}{m}\right)^{-\rho+\delta} \qquad (2.22)\\
&= o\big(n^{-1/2+2\delta}\big) \qquad \text{a.s. } \forall \delta > 0.
\end{aligned}$$

According to condition (2.2), we have

$$\begin{aligned}
\mathbf{r}_{n+1}^* &=: (\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\mathbf{H} - \big(\mathbf{h}(\boldsymbol{\theta}_n) - \mathbf{h}(\boldsymbol{\theta}^*)\big) \\
&= o\big(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}^*\|^{1+\varepsilon}\big) = o\big(n^{-1/2-\varepsilon/4}\big) \qquad \text{a.s.}
\end{aligned} \tag{2.23}$$

and

$$\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}^* = (\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\left(\mathbf{I}_d - \frac{\mathbf{H}}{n+1}\right) + \frac{\Delta\mathbf{M}_{n+1} + \mathbf{r}_{n+1} + \mathbf{r}_{n+1}^*}{n+1}.$$

It follows that

$$\begin{aligned}
\boldsymbol{\theta}_n - \boldsymbol{\theta}^* &= (\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*)\widetilde{\boldsymbol{\Pi}}_0^n + \sum_{m=1}^{n}\frac{\Delta\mathbf{M}_m}{m}\widetilde{\boldsymbol{\Pi}}_m^n + \sum_{m=1}^{n}\frac{\mathbf{r}_m}{m}\widetilde{\boldsymbol{\Pi}}_m^n + \sum_{m=1}^{n}\frac{\mathbf{r}_m^*}{m}\widetilde{\boldsymbol{\Pi}}_m^n \\
&=: (\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*)\widetilde{\boldsymbol{\Pi}}_0^n + \boldsymbol{\zeta}_n + \boldsymbol{\eta}_n + \boldsymbol{\eta}_n^*,
\end{aligned}$$

where $\widetilde{\boldsymbol{\Pi}}_m^n$ is defined as in (2.15), and $\|\widetilde{\boldsymbol{\Pi}}_m^n\| \le C(n/m)^{-1/2}\log^{\nu-1}(n/m)$ by Proposition B.1(i) in the Appendix. Thus,

$$\boldsymbol{\eta}_n^* = O(1)\sum_{m=1}^{n}\frac{m^{-1/2-\varepsilon/4}}{m}\left(\frac{n}{m}\right)^{-1/2}\log^{\nu-1}\frac{n}{m} = o\big(n^{-1/2}(\log n)^{\nu-1/2}\big) \qquad \text{a.s.}$$

If (2.6) is satisfied, then

$$\begin{aligned}
\boldsymbol{\eta}_n &= O(1)\sum_{m=1}^{n}\frac{\|\mathbf{r}_m\|}{m}\left(\frac{n}{m}\right)^{-1/2}\log^{\nu-1}\frac{n}{m} \\
&= o\big(n^{-1/2}(\log n)^{\nu-1/2}\big) \qquad \text{a.s.} \quad \text{or} \quad \text{in } L_1.
\end{aligned}$$

If (2.5) is satisfied, then we also have

$$\begin{aligned}\boldsymbol{\eta}_n &= \frac{\mathbf{s}_n}{n}\widetilde{\boldsymbol{\Pi}}_n^n + \sum_{m=1}^{n-1} \mathbf{s}_m \frac{\mathbf{I}_d - \mathbf{H}}{m(m+1)} \widetilde{\boldsymbol{\Pi}}_{m+1}^n \\ &= O(1) \sum_{m=1}^{n} \frac{o(\sqrt{m/\log m})}{m^2} \left(\frac{n}{m}\right)^{-1/2} \log^{\nu-1} \frac{n}{m} \\ &= O(1) n^{-1/2} \log^{\nu-1} n \sum_{m=1}^{n} \frac{o(1)}{m\sqrt{\log m}} \\ &= o\big(n^{-1/2} (\log n)^{\nu-1/2}\big) \qquad \text{a.s.}\end{aligned}$$

At last, $\boldsymbol{\zeta}_n$ is a sum of martingale differences. By verifying the Lindeberg condition and checking the variance, we can show that

$$\frac{\sqrt{n}}{(\log n)^{\nu-1/2}} \boldsymbol{\zeta}_n \xrightarrow{D} N(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}}) \qquad \text{(stably)}$$

via the CLT for martingales [cf. Corollary 3.1 of Hall and Heyde (1980)]. The above convergence is stated in Proposition B.2 in the Appendix. The proof of Theorem 2.1 is complete.

Now, we consider the case of $\rho > 1/2$. Suppose Assumptions 2.1, 2.3 and (2.12) are satisfied. It is obvious that the first term of (2.17) is $O(1)n^{-\rho+\delta} = o(n^{-1/2})$ a.s., and the last term is

$$\sum_{m=1}^{n} \frac{\mathbf{r}_m}{m} \boldsymbol{\Pi}_m^n = \frac{o(\sqrt{n})}{n} + \sum_{m=1}^{n-1} \frac{o(\sqrt{m})}{m(m+1)} \left(\frac{n}{m}\right)^{-\rho+\delta} = o\big(n^{-1/2}\big)$$

in probability by (2.12) and (2.18). The middle term of (2.17) is a sum of weighted martingale differences. Unfortunately, we cannot apply the CLT for martingales directly because $\{\frac{\Delta \mathbf{M}_m}{m} \boldsymbol{\Pi}_m^n; m = 1, \dots, n\}$ is not an array of martingale differences. We can show that the random weight $\boldsymbol{\Pi}_m^n$ can be replaced by the nonrandom weight $\widetilde{\boldsymbol{\Pi}}_m^n$, that is,

$$\sqrt{n} \sum_{m=1}^{n} \frac{\Delta \mathbf{M}_m}{m} (\boldsymbol{\Pi}_m^n - \widetilde{\boldsymbol{\Pi}}_m^n) \to \mathbf{0} \qquad \text{in probability.} \tag{2.24}$$

Now, $\{\frac{\Delta \mathbf{M}_m}{m} \widetilde{\boldsymbol{\Pi}}_m^n; m = 1, \dots, n\}$ is an array of martingale differences. By verifying the Lindeberg condition and checking the variance, we can show that

$$\sqrt{n} \sum_{m=1}^{n} \frac{\Delta \mathbf{M}_m}{m} \widetilde{\boldsymbol{\Pi}}_m^n \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}) \qquad \text{(stably)}$$

via the CLT for martingales. The above convergence and (2.24) are stated in Proposition B.2 in the Appendix. Thus, (2.13) is proved. The proof is now complete. $\square$

PROOF OF THEOREM 2.2. Recall (2.17) and $\mathbf{H}_n \to \mathbf{H}$ a.s. as $n \to \infty$. By (2.11) and (2.21), we have

$$\begin{aligned}
&\boldsymbol{\theta}_n - \boldsymbol{\theta}^* \\
&\quad = (\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*)\boldsymbol{\Pi}_0^n + \frac{\mathbf{M}_n + \mathbf{s}_n}{n}\boldsymbol{\Pi}_n^n + \sum_{m=1}^{n-1}(\mathbf{M}_m + \mathbf{s}_m)\frac{\mathbf{I}_d - \mathbf{H}_{m+1}}{m(m+1)}\boldsymbol{\Pi}_{m+1}^n \\
&\quad = o\big(n^{-\rho+\delta}\big) \\
&\qquad + \frac{o(n^{1/2+\delta/2}) + o(n^{1-\rho})}{n} + \sum_{m=1}^{n-1}\frac{o(m^{1/2+\delta/2}) + o(m^{1-\rho})}{m(m+1)}\left(\frac{n}{m}\right)^{-\rho+\delta} \\
&\quad = o\big(n^{-\rho+\delta}\big) \qquad \text{a.s.}
\end{aligned}$$

It follows that

$$n^{\rho-\delta}\big(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*\big) \to \mathbf{0} \qquad \text{a.s. for all } \delta > 0. \tag{2.25}$$

According to (2.2), we can rewrite (1.1) as

$$\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}^* = (\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\left(\mathbf{I}_d - \frac{\mathbf{H}}{n+1}\right) + \frac{\mathbf{r}_{n+1}^*}{n+1},$$

where $\mathbf{r}_{n+1}^* = o(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}^*\|^{1+\varepsilon}) + \Delta\mathbf{M}_{n+1} + \mathbf{r}_{n+1}$. From (2.11), (2.21) and (2.25), it follows that $\mathbf{s}_n^* =: \sum_{k=1}^n \mathbf{r}_k^* = o(n^{1-\rho-\delta})$ a.s. for some $\delta > 0$. Recall that $\mathbf{H}$ has the Jordan canonical form $\mathbf{T}^{-1}\mathbf{H}\mathbf{T} = \mathrm{diag}(\mathbf{J}_1, \dots, \mathbf{J}_s)$, with $\mathbf{J}_a = \lambda_a \mathbf{I}_{\nu_a} + \overline{\mathbf{J}}_{\nu_a}$. Denote $(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\mathbf{T} := \mathbf{y}_n = (\mathbf{y}_{n,1}, \dots, \mathbf{y}_{n,s})$, $\mathbf{r}_n^*\mathbf{T} := \widetilde{\mathbf{r}}_n = (\widetilde{\mathbf{r}}_{n,1}, \dots, \widetilde{\mathbf{r}}_{n,s})$, $\mathbf{s}_n^*\mathbf{T} := \widetilde{\mathbf{s}}_n = (\widetilde{\mathbf{s}}_{n,1}, \dots, \widetilde{\mathbf{s}}_{n,s})$. Then

$$\mathbf{y}_{n+1,a} = \mathbf{y}_{n,a}\left(\mathbf{I}_{\nu_a} - \frac{\mathbf{J}_a}{n+1}\right) + \frac{\widetilde{\mathbf{r}}_{n+1,a}}{n+1}.$$

Write

$$\widetilde{\boldsymbol{\Pi}}_k^{n,a} = \prod_{j=k+1}^n \left(\mathbf{I}_{\nu_a} - \frac{\mathbf{J}_a}{j}\right).$$

Then

$$\|\widetilde{\boldsymbol{\Pi}}_0^{n,a}\| \le Cn^{-\mathrm{Re}(\lambda_a)}(\log n)^{\nu_a-1}, \qquad \|\widetilde{\boldsymbol{\Pi}}_k^{n,a}\| \le C\left(\frac{n}{k}\right)^{-\mathrm{Re}(\lambda_a)}\left(\log\frac{n}{k}\right)^{\nu_a-1},$$

$1 \le k \le n$. If $\mathrm{Re}(\lambda_a) < 1$, then $\widetilde{\boldsymbol{\Pi}}_0^{n,a} n^{\mathbf{J}_a} \to \mathbf{A}_a$, $n^{-\mathbf{J}_a}(\widetilde{\boldsymbol{\Pi}}_0^{n,a})^{-1} \to \mathbf{A}_a^{-1}$ for an invertible matrix $\mathbf{A}_a$, and $\|\widetilde{\boldsymbol{\Pi}}_0^{n,a}\| \approx n^{-\mathrm{Re}(\lambda_a)}(\log n)^{\nu_a-1}$. We have

$$\begin{aligned}
\mathbf{y}_{n,a} &= \mathbf{y}_{0,0}\widetilde{\boldsymbol{\Pi}}_0^{n,a} + \sum_{k=1}^n \frac{\widetilde{\mathbf{r}}_{k,a}}{k}\widetilde{\boldsymbol{\Pi}}_k^{n,a} \\
&= \mathbf{y}_{0,0}\widetilde{\boldsymbol{\Pi}}_0^{n,a} + \frac{\widetilde{\mathbf{s}}_{n,a}}{n}\widetilde{\boldsymbol{\Pi}}_n^{n,a} + \sum_{k=1}^{n-1}\frac{\widetilde{\mathbf{s}}_{k,a}}{k}\frac{\mathbf{I}_{\nu_a} - \mathbf{J}_a}{k+1}\widetilde{\boldsymbol{\Pi}}_{k+1}^{n,a}.
\end{aligned}$$

If $\mathrm{Re}(\lambda_a) > \rho$, then

$$\begin{aligned}\|\mathbf{y}_{n,a}\| &= O(1)n^{-\mathrm{Re}(\lambda_a)}(\log n)^{\nu_a-1} + o(n^{-\rho-\delta}) \\ &\quad + \sum_{k=1}^{n-1} \frac{o(k^{-\rho-\delta})}{k+1}\left(\frac{n}{k}\right)^{-\mathrm{Re}(\lambda_a)}\left(\log\frac{n}{k}\right)^{\nu_a-1} \\ &= o(n^{-\rho-\kappa}) \qquad \text{a.s.}\end{aligned}$$

for some $0 < \kappa < \mathrm{Re}(\lambda_a) - \rho$. If $\mathrm{Re}(\lambda_a) = \rho$, then

$$\begin{aligned}\|\mathbf{y}_{n,a}\| &= O(1)n^{-\rho}(\log n)^{\nu_a-1} + o(n^{-\rho-\delta}) + \sum_{k=1}^{n-1} \frac{o(k^{-\rho-\delta})}{k+1}\left(\frac{n}{k}\right)^{-\rho}\left(\log\frac{n}{k}\right)^{\nu_a-1} \\ &= O(1)n^{-\rho}(\log n)^{\nu_a-1} = o(n^{-\rho}(\log n)^{\nu-1}) \qquad \text{a.s. when } \nu_a < \nu.\end{aligned}$$

Finally, consider the $\mathbf{y}_{n,a}$ with $\mathrm{Re}(\lambda_a) = \rho$ and $\nu_a = \nu$. Note that

$$\mathbf{y}_{n,a}(\widetilde{\mathbf{\Pi}}_0^{n,a})^{-1} = \mathbf{y}_{0,a} + \frac{\widetilde{\mathbf{s}}_{n,a}}{n}(\widetilde{\mathbf{\Pi}}_0^{n,a})^{-1} + \sum_{k=1}^{n-1} \frac{\widetilde{\mathbf{s}}_{k,a}}{k}\frac{\mathbf{I}_{\nu_a} - \mathbf{J}_a}{k+1}(\widetilde{\mathbf{\Pi}}_0^{k+1,a})^{-1}.$$

Observe that

$$\frac{\widetilde{\mathbf{s}}_{n,a}}{n}(\widetilde{\mathbf{\Pi}}_0^{n,a})^{-1} = o(n^{-\rho-\delta})n^{\rho}(\log n)^{\nu_a-1} \to 0 \qquad \text{a.s.,}$$

$$\sum_{k=1}^{\infty}\left\|\frac{\widetilde{\mathbf{s}}_{k,a}}{k}\frac{\mathbf{I}_{\nu_a} - \mathbf{J}_a}{k+1}(\widetilde{\mathbf{\Pi}}_0^{k+1,a})^{-1}\right\| \le c\sum_{k=1}^{\infty}\frac{O(k^{1-\rho-\delta})}{k^2}k^{\rho}(\log k)^{\nu_a-1} < \infty \qquad \text{a.s.}$$

It follows that

$$\mathbf{y}_{n,a}(\widetilde{\mathbf{\Pi}}_0^{n,a})^{-1} \to \mathbf{y}_{0,a} + \sum_{k=1}^{\infty} \frac{\widetilde{\mathbf{s}}_{k,a}}{k}\frac{\mathbf{I}_{\nu_a} - \mathbf{J}_a}{k+1}(\widetilde{\mathbf{\Pi}}_0^{k+1,a})^{-1} \qquad \text{a.s.}$$

Thus,

$$\mathbf{y}_{n,a}n^{\mathbf{J}_a} \to \boldsymbol{\xi}_a =: \left[\mathbf{y}_{0,a} + \sum_{k=1}^{\infty} \frac{\widetilde{\mathbf{s}}_{k,a}}{k}\frac{\mathbf{I}_{\nu_a} - \mathbf{J}_a}{k+1}(\widetilde{\mathbf{\Pi}}_0^{k+1,a})^{-1}\right]\mathbf{A}_a \qquad \text{a.s.}$$

It follows that

$$\begin{aligned}\mathbf{y}_{n,a} &= (\boldsymbol{\xi}_a + o(1))n^{-\mathbf{J}_a} = (\boldsymbol{\xi}_a + o(1))n^{-\lambda_a}\exp\{-\overline{\mathbf{J}}_{\nu_a}\log n\} \\ &= \sum_{j=0}^{\nu_a-1}(\boldsymbol{\xi}_a + o(1))n^{-\lambda_a}\frac{(-\overline{\mathbf{J}}_{\nu_a})^j(\log n)^j}{j!} \\ &= \xi_{a,\nu_a-1}n^{-\lambda_a}(-1)^{\nu_a-1}\frac{(\log n)^{\nu_a-1}}{(\nu_a-1)!}(0,\dots,0,1) + o(n^{-\rho}(\log n)^{\nu-1}) \\ &= \xi_{a,\nu_a-1}n^{-\mathrm{i}\,\mathrm{Im}(\lambda_a)}(-1)^{\nu_a-1}\frac{n^{-\rho}(\log n)^{\nu_a-1}}{(\nu_a-1)!}(0,\dots,0,1) + o(n^{-\rho}(\log n)^{\nu-1}),\end{aligned}$$

because $(\bar{\mathbf{J}}_{\nu_a})^{\nu_a} = \mathbf{0}$. Denote $\xi_a = \xi_{a,\nu_a-1}(-1)^{\nu_a-1}\frac{1}{(\nu_a-1)!}$. Then

$$\frac{n^\rho}{(\log n)^{\nu-1}}\mathbf{y}_n - \sum_{a:\mathrm{Re}(\lambda_a)=\rho,\nu_a=\nu} e^{-\mathrm{i}\,\mathrm{Im}(\lambda_a)\log n}\xi_a\mathbf{e}_a \to \mathbf{0} \qquad \text{a.s.},$$

$$\frac{n^\rho}{(\log n)^{\nu-1}}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*) - \sum_{a:\mathrm{Re}(\lambda_a)=\rho,\nu_a=\nu} e^{-\mathrm{i}\,\mathrm{Im}(\lambda_a)\log n}\xi_a\mathbf{e}_a\mathbf{T}^{-1} \to \mathbf{0} \qquad \text{a.s.}$$

The proof is complete. □

3. Gaussian process approximation. Write $\mathbf{H} = D\mathbf{h}(\boldsymbol{\theta}^*)$. Suppose that $\mathbf{B}(t)$ is a d-dimensional standard Brownian motion that is independent of $\boldsymbol{\Gamma}$. Let $\mathbf{G}_t$ be a solution of the following differential equation:

$$d\mathbf{G}(t) = -\frac{\mathbf{G}(t)}{t}\mathbf{H}\,dt + \frac{d\mathbf{B}(t)}{t}\boldsymbol{\Gamma}^{1/2}, \qquad \mathbf{G}(1) = \mathbf{G}_1. \tag{3.1}$$

It can be verified that

$$\mathbf{G}(t) = \int_1^t \frac{d\mathbf{B}(x)\boldsymbol{\Gamma}^{1/2}}{x}\left(\frac{x}{t}\right)^{\mathbf{H}} + \mathbf{G}_1 t^{-\mathbf{H}}, \qquad t > 0.$$

When $\mathbf{G}_1 = \mathbf{0}$, for a given $\boldsymbol{\Gamma}$, $\mathbf{G}(t)$ is a Gaussian process with the variance-covariance matrix

$$\mathrm{Var}\{\mathbf{G}(t)\} = \frac{1}{t}\int_0^{\log t} \left(e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}/2)u}\right)^{\mathrm{t}}\boldsymbol{\Gamma}e^{-(D\mathbf{h}(\boldsymbol{\theta}^*)-\mathbf{I}/2)u}\,du. \tag{3.2}$$

It is obvious that for a given $\boldsymbol{\Gamma}$, the limit variabilities in (2.14) and (2.8) are, respectively,

$$\lim_{t\to\infty} t\,\mathrm{Var}\{\mathbf{G}(t)\} \quad \text{and} \quad \lim_{t\to\infty}\frac{t}{\log^{2\nu-1} t}\mathrm{Var}\{\mathbf{G}(t)\}.$$

The next theorem shows that $\boldsymbol{\theta}_n - \boldsymbol{\theta}^*$ can be approximated by the Gaussian process $\mathbf{G}(t)$ under certain conditions. From the Gaussian approximation, we can obtain the law of the iterated logarithm for $\boldsymbol{\theta}_n - \boldsymbol{\theta}^*$ and the functional central limit theorem for the process $\boldsymbol{\theta}_{[nt]} - \boldsymbol{\theta}^*$.

THEOREM 3.1. *Suppose that Assumption 2.2 is satisfied, $\boldsymbol{\theta}_n \to \boldsymbol{\theta}^*$ a.s. and*

$$\sum_{m=1}^n \mathbf{r}_m = o\left(n^{1/2-\varepsilon_0}\right) \qquad \textit{a.s.}, \tag{3.3}$$

$$\sum_{m=1}^\infty \mathsf{E}\left[\|\Delta\mathbf{M}_m\|^2 I\{\|\Delta\mathbf{M}_m\|^2 \ge m^{1-\varepsilon_0}\}|\mathscr{F}_{m-1}\right]/m^{1-\varepsilon_0} < \infty \qquad \textit{a.s.,} \quad \textit{and} \tag{3.4}$$

$$\sum_{m=1}^n \mathsf{E}\left[(\Delta\mathbf{M}_m)^{\mathrm{t}}\Delta\mathbf{M}_m|\mathscr{F}_{m-1}\right] = n\boldsymbol{\Gamma} + o\left(n^{1-\varepsilon_0}\right) \qquad \textit{a.s.} \quad \textit{or} \quad \textit{in } L_1 \tag{3.5}$$

for some $0 < \varepsilon_0 < 1$, *where* $\mathbf{\Gamma}$ *is a symmetric positive semidefinite matrix that is* $\mathscr{F}_m$*-measurable for some* m. *Then* [possibly in an enlarged probability space with the process $\{(\boldsymbol{\theta}_n, \mathbf{M}_n, \mathbf{r}_n); n \geq 1\}$ being redefined without changing its distribution] *there is a* d*-dimensional standard Brownian motions* $\mathbf{B}(t)$ *that is independent of* $\mathbf{\Gamma}$, *such that*

$$\boldsymbol{\theta}_n - \boldsymbol{\theta}^* = \mathbf{G}(n) + o\big(n^{-1/2-\tau}\big) \qquad \textit{a.s. for some } \tau > 0, \tag{3.6}$$

when $\rho > 1/2$, *and*

$$\boldsymbol{\theta}_n - \boldsymbol{\theta}^* = \mathbf{G}(n) + O\big(n^{-1/2}\log^{\nu-1} n\big) \qquad \textit{a.s.} \tag{3.7}$$

when $\rho = 1/2$, *where* $\mathbf{G}(t)$ *is the solution of equation* (3.1).

REMARK 3.1. The proof of Gaussian approximation is based on the strong approximation theorems for multivariate martingales. The condition that $\mathbf{\Gamma}$ is $\mathscr{F}_m$-measurable for some m is given by Eberlein (1986), Monrad and Philipp (1991) and Zhang (2004) to establish the strong approximation theorems for multivariate martingales. For general random $\mathbf{\Gamma}$, the strong approximation is unknown. In practice, $\mathbf{\Gamma}$ is usually assumed to be deterministic.

PROOF OF THEOREM 3.1. Note conditions (3.4) and (3.5). By Theorem 1.3 of Zhang (2004), possibly in an enlarged probability space with the process $\{(\boldsymbol{\theta}_n, \mathbf{M}_n, \mathbf{r}_n); n \geq 1\}$ redefined without changing its distribution, there is d-dimensional standard Brownian motions $\mathbf{B}(t)$ independent of $\mathbf{\Gamma}$, such that

$$\mathbf{M}_n = \mathbf{B}(n)\mathbf{\Gamma}^{1/2} + o\big(n^{1/2-\tau}\big) \qquad \text{a.s. for some } \tau > 0. \tag{3.8}$$

Let $\mathbf{G}(t)$ be the solution of equation (3.1). By some elementary calculation we can write

$$\begin{aligned}
\mathbf{G}(n+1) - \mathbf{G}(n) &= -\int_n^{n+1} \frac{\mathbf{G}(x)}{x}\,dx\mathbf{H} + \int_n^{n+1} \frac{d\mathbf{B}(x)}{x}\mathbf{\Gamma}^{1/2} \\
&= -\frac{\mathbf{G}(n)}{n+1}\mathbf{H} + \frac{[\mathbf{B}(n+1) - \mathbf{B}(n)]\mathbf{\Gamma}^{1/2} + \boldsymbol{\delta}_{n+1}}{n+1}
\end{aligned}$$

with

$$\sum_{m=1}^{\infty} \boldsymbol{\delta}_m \qquad \text{being convergent a.s.} \tag{3.9}$$

According to (1.1), we have

$$\begin{aligned}
&\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}^* \\
&\quad = (\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\Big(\mathbf{I}_d - \frac{\mathbf{H}}{n+1}\Big) + \frac{\Delta\mathbf{M}_{n+1} + [(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\mathbf{H} - \mathbf{h}(\boldsymbol{\theta}_n)] + \mathbf{r}_{n+1}}{n+1}.
\end{aligned}$$

It follows that the sequence $\{\boldsymbol{\theta}_n - \boldsymbol{\theta}^* - \mathbf{G}(n)\}$ satisfies

$$\boldsymbol{\theta}_{n+1} - \boldsymbol{\theta}^* - \mathbf{G}(n+1) = (\boldsymbol{\theta}_n - \boldsymbol{\theta}^* - \mathbf{G}(n))\left(\mathbf{I}_d - \frac{\mathbf{H}}{n+1}\right) + \frac{\mathbf{r}^*_{n+1}}{n+1}$$

with

$$\mathbf{r}^*_{n+1} = \mathbf{r}_{n+1} - \boldsymbol{\delta}_{n+1} + \left[(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\mathbf{H} - \mathbf{h}(\boldsymbol{\theta}_n)\right] + \Delta\left(\mathbf{M}_{n+1} - \mathbf{B}(n+1)\boldsymbol{\Gamma}^{1/2}\right).$$

According to (2.23),

$$(\boldsymbol{\theta}_n - \boldsymbol{\theta}^*)\mathbf{H} - \mathbf{h}(\boldsymbol{\theta}_n) = o\left(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}^*\|^{1+\varepsilon}\right) = o\left(n^{-1/2-\varepsilon/4}\right) \qquad \text{a.s.} \tag{3.10}$$

From (3.3), (3.8), (3.9) and (3.10), it follows that

$$\mathbf{s}^*_n =: \sum_{m=1}^{n} \mathbf{r}^*_m = o\left(n^{1/2-\tau}\right) \qquad \text{a.s. for some } \tau > 0.$$

Recall that $\widetilde{\boldsymbol{\Pi}}^n_m = \sum_{j=m+1}^{n}(\mathbf{I}_d - \frac{\mathbf{H}}{j})$ and $\|\widetilde{\boldsymbol{\Pi}}^n_m\| \le C_0(\frac{n}{m})^{-\rho}\log^{\nu-1}\frac{n}{m}$ by Proposition B.1(i). Following the lines in (2.17) and (2.18), we conclude that

$$\begin{aligned}
&\boldsymbol{\theta}_n - \boldsymbol{\theta}^* - \mathbf{G}(n) \\
&\quad = (\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*)\widetilde{\boldsymbol{\Pi}}^n_0 + \frac{\mathbf{s}^*_n}{n}\widetilde{\boldsymbol{\Pi}}^n_n + \sum_{m=1}^{n-1} \mathbf{s}^*_m \frac{\mathbf{I}_d - \mathbf{H}}{m(m+1)}\widetilde{\boldsymbol{\Pi}}^n_{m+1} \\
&\quad = O(1)n^{-\rho}\log^{\nu-1} n + \frac{o(n^{1/2-\tau})}{n} + \sum_{m=1}^{n-1} \frac{o(m^{1/2-\tau})}{m(m+1)}\left(\frac{n}{m}\right)^{-\rho}\log^{\nu-1} n \\
&\quad = O\left(n^{-(1/2+\tau)\wedge\rho}\log^{\nu-1} n\right) \qquad \text{a.s.}
\end{aligned}$$

The proof is complete. □

4. Urn models. Urn models have long been considered powerful mathematical instruments in many areas, including the physical sciences, biological sciences, social sciences and engineering [Johnson and Kotz (1977), Kotz and Balakrishnan (1997)]. The Pólya urn (also known as the Pólya–Eggenberger urn) model was originally proposed to model the problem of contagious diseases [Eggenberger and Pólya (1923)]. Since then, there have been numerous generalizations and extensions. Among them, the GFU (also known as the generalized Pólya urn or GPU in the literature) is the most popular [see Athreya and Karlin (1968), Athreya and Ney (1972), Janson (2004); etc.]. In clinical trial studies, response-adaptive designs for randomizing treatments to patients aim at detecting "on-line" which treatment should be assigned to more patients while retaining enough randomness to preserve the basis of treatments. A large family of adaptive designs is based on the GFU [Bai and Hu (1999, 2005), Hu and Rosenberger (2006), Hu and Zhang

(2004), Smythe (1996), Wei (1979), Wei and Durham (1978), Zhang, Hu and Cheung (2006), Zhang et al. (2011); etc.]. In this model, the adaptive approach relies on the cumulative information provided by the responses to previous patients' treatments to adjust treatment allocation to the new patients. The idea of this modeling is that the urn contains balls of d different types representative of the treatments. At the beginning, the urn contains $\mathbf{Y}_0 = (Y_{0,1}, \dots, Y_{0,d}) \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ balls, where Y_{0k} denotes the number of balls of type k, $k = 1, \dots, d$. At stage m ($m = 1, 2, \dots$), a ball is drawn from the urn with instant replacement. If the ball is of type k, then the mth patient is allocated to treatment k, and additional $D_{k,q}(m)$ balls of type q, $q = 1, \dots, d$, are added to the urn, where $D_{k,q}(m)$ may be a function of another random variable $\boldsymbol{\xi}(m)$ and also may be a function of urn compositions and the results of draws from previous stages. The random vector $\boldsymbol{\xi}(m)$ is usually the response of the mth patient. This procedure is repeated throughout n stages. After n draws and generations, the urn composition is denoted by the row vector $\mathbf{Y}_n = (Y_{n,1}, \dots, Y_{n,d})$, where $Y_{n,k}$ is the number of balls of type k in the urn after the nth draw. This relation can be written as the following recursive formula:

$$\mathbf{Y}_n = \mathbf{Y}_{n-1} + \mathbf{X}_n \mathbf{D}_n, \tag{4.1}$$

where $\mathbf{D}_n = (D_{k,q}(n))_{k,q=1}^d$ and $\mathbf{X}_n$ is the result of the nth draw, distributed according to the urn composition at the previous stage, that is, if the nth draw is a type k ball, then the kth component of $\mathbf{X}_n$ is 1 and other components are 0. The matrices' $\mathbf{D}_n$s' are named as the adding rules. The conditional expectations $\mathbf{H}_n = (\mathsf{E}[D_{k,q}(n)|\mathscr{F}_{n-1}])_{k,q=1}^d$, for given the history sigma field $\mathscr{F}_{n-1}$ generated by the urn compositions $\mathbf{Y}_1, \dots, \mathbf{Y}_{n-1}$, the results of draws $\mathbf{X}_1, \dots, \mathbf{X}_{n-1}$ and $\boldsymbol{\xi}(1), \dots, \boldsymbol{\xi}(n-1)$ of all previous stages, $n = 1, 2, \dots,$ are named as the generating matrices. When $\mathbf{D}_n$, $n = 1, 2, \dots,$ are independent and identically distributed, the GFU model is usually said to be homogeneous. In such a case, $\mathbf{H}_n = \mathbf{H}$ are identical and nonrandom and the adding rule $\mathbf{D}_n$ is merely a function of the $\boldsymbol{\xi}(n)$. In the general heterogeneous cases, both $\mathbf{D}_n$ and $\mathbf{H}_n$ depend on the entire history of all of the stages.

Write $\mathbf{N}_n = (N_{n,1}, \dots, N_{n,d})$, where $N_{n,k}$ is the number of times that a type k ball is drawn in the first n stages. Also, in an adaptive design based on this urn model, $N_{n,k}$ is the number of patients being assigned to treatment k after n assignments. Obviously,

$$\mathbf{N}_n = \sum_{k=1}^{n} \mathbf{X}_k. \tag{4.2}$$

Athreya and Karlin (1967, 1968) first considered the asymptotic properties of the homogeneous GFU model and conjectured that $\mathbf{N}_n$ is asymptotically normal. Janson (2004) established the functional CLTs of $\mathbf{Y}_n$ and $\mathbf{N}_n$ for a homogenous case in which the numbers of each type of balls were assumed to be integers. Bai and Hu (2005) established the asymptotic normality for the nonhomogeneous GFU model under the following conditions:

$$\mathbf{H}_n \to \mathbf{H} \quad \text{a.s.}, \qquad \mathbf{H} = (H_{k,j})_{d\times d}, \qquad H_{k,j} \ge 0, \tag{4.3}$$

$$\sup_{n\ge 1} \mathsf{E}\big[\|\mathbf{D}_n\|^{2+\varepsilon}|\mathscr{F}_{n-1}\big] < +\infty \qquad \text{a.s. for some } \varepsilon > 0, \tag{4.4}$$

$$\mathrm{Cov}\big[\{D_{q,k}(n), D_{q,l}(n)\}|\mathscr{F}_{n-1}\big] \to V_{qkl} \qquad \text{a.s.}, q, k, l = 1, \dots d, \tag{4.5}$$

$$\sum_{m=1}^{\infty} \frac{\|\mathbf{H}_m - \mathbf{H}\|_\infty}{\sqrt{m}} < \infty \qquad \text{a.s.}, \tag{4.6}$$

$$n\mathsf{E}\|\mathbf{H}_n - \mathsf{E}\mathbf{H}_n\|^2 \to 0 \qquad \text{a.s.}, \tag{4.7}$$

$$\mathbf{H}_n \mathbf{1}^{\mathrm{t}} = \alpha \mathbf{1}^{\mathrm{t}} \qquad \text{with } \mathbf{1} = (1, \dots, 1) \qquad \text{for some } \alpha > 0, \tag{4.8}$$

and $\lambda_{\mathrm{sec}} \le \alpha/2$, where λ_{sec} is the second largest real part of the eigenvalues of $\mathbf{H}$. Higueras et al. (2006) also considered the asymptotic normality of the urn compositions $\mathbf{Y}_n$ under the condition that

$$n\mathsf{E}\|\mathbf{H}_n - \mathbf{H}\|^2 \to 0, \tag{4.9}$$

which is weaker than Bai and Hu's conditions (4.6) and (4.7). However, Higueras et al. (2006) only considered the case $\lambda_{\mathrm{sec}} < \alpha/2$ and assumed an extra assumption; namely, that $\mathbf{D}_n \mathbf{1}' = \alpha \mathbf{1}'$, which is stricter than (4.8).

Laruelle and Pagès (2013) derived the joint asymptotic distribution of the vector $(\mathbf{Y}_n, \mathbf{N}_n)$ and weakened conditions (4.6) and (4.7) to (4.9). Moreover, the results only held when $\lambda_{\mathrm{sec}} < \alpha/2$. In the study of adaptive designs driven by urn models, $\lambda_{\mathrm{sec}} \le \alpha/2$ is a very stringent condition even when $d = 3$ [cf. Chapter 4 of Hu and Rosenberger (2006)]. The limit properties for $\lambda_{\mathrm{sec}} > \alpha/2$ and for the case that (4.8) is not satisfied are stated as open problems in Hu and Rosenberger (2006), page 158. In this section, we derive the joint asymptotic distribution of $(\mathbf{Y}_n, \mathbf{N}_n)$ by applying our new results on the SA algorithm (1.1). We consider both the cases of $\lambda_{\mathrm{sec}} \le \alpha/2$ and $\lambda_{\mathrm{sec}} > \alpha/2$. We also remove condition (4.8) and weaken condition (4.4) to the conditional Lindeberg condition.

Before we state the results, we first need some more notation and assumptions. To include various cases, we allow the numbers of balls to be nonintegers and negative. For example, $D_{k,l}(n) < 0$ means that $|D_{k,l}(n)|$ balls of type l are removed from the urn when a ball of type of k is drawn. We assume that a type of ball with a negative number will never be selected and so

$$\mathsf{P}(X_{n,k} = 1|\mathscr{F}_{n-1}) = \frac{Y_{n-1,k}^+}{|\mathbf{Y}_{n-1}^+|}.$$

Here, $Y_{n,k}^+ = \max\{Y_{n,k}, 0\}$ is the positive part of $Y_{n,k}$, $\mathbf{Y}_n^+ = (Y_{n,1}^+, \dots, Y_{n,d}^+)$, $|\mathbf{Y}_n^+| = \sum_{k=1}^d Y_{n,k}^+$ and $\frac{0}{0}$ is defined as $\frac{1}{d}$, which means that each type of ball is selected with equal probability when the urn has no balls with a positive number. In this general framework, the urn allows negative and/or non-integer numbers of balls, removal and nonhomogeneous updating. In considering the asymptotic properties, we need two assumptions on the adding rules.

ASSUMPTION 4.1. Suppose that there is a deterministic matrix $\mathbf{H} = (H_{q,k})_{q,k}^d$ with $H_{q,k} \geq 0$ for $q \neq k$ such that

$$\sum_{m=1}^{n} \|\mathbf{H}_m - \mathbf{H}\| = o(n) \qquad \text{a.s.} \tag{4.10}$$

Further, assume that $\mathbf{H}$ has a single largest eigenvalue $\alpha > 0$ and the corresponding left eigenvector $\mathbf{v} = (v_1, \dots, v_d)$ and right eigenvector $\mathbf{u}^{\mathrm{t}} = (u_1, \dots, u_d)^{\mathrm{t}}$ such that $\sum_k v_k = \sum_k v_k u_k = 1$ and $v_k > 0, u_k > 0, k = 1, \dots, d$.

Without loss of generality, we assume that $\alpha = 1$ throughout this paper. Otherwise, we may consider $\mathbf{Y}_m/\alpha, \mathbf{D}_m/\alpha$ instead.

Assumption 4.1 means that the updating is asymptotically stable and that on average, a draw will not generate the removal of the undrawn balls to avoid urn extinction, although balls of any type can be dropped from the urn at each specific stage.

When $\mathbf{H}$ satisfies the conditions in Assumption 4.1, we let $\lambda_2, \dots, \lambda_s$ be the other eigenvalues of $\mathbf{H}$ and suppose that $\mathbf{H}$ has the following Jordan canonical form decomposition:

$$\operatorname{diag}(1, \mathbf{J}_2, \dots, \mathbf{J}_s) \tag{4.11}$$

with $\mathbf{J}_t = \lambda_t \mathbf{I}_{\nu_t} + \overline{\mathbf{J}}_{\nu_t}$, where ν_t is the order of the Jordan block $\mathbf{J}_t$. Denote by $\lambda_{\mathrm{sec}} = \max\{\mathrm{Re}(\lambda_2), \dots, \mathrm{Re}(\lambda_s)\}$ and $\nu = \max\{\nu_t : \mathrm{Re}(\lambda_t) = \lambda_{\mathrm{sec}}\}$.

ASSUMPTION 4.2. Let $V_{qkl}(n) =: \mathrm{Cov}[\{D_{qk}(n), D_{ql}(n)\} | \mathscr{F}_{n-1}]$, $q, k, l = 1, \dots, d$, and denote by $\mathbf{V}_{nq} = (V_{qkl}(n))_{k,l=1}^d$. Suppose that

$$\frac{1}{n} \sum_{m=1}^{n} \mathsf{E}\big[\|\mathbf{D}_m\|^2 I\{\|\mathbf{D}_m\| \geq \varepsilon\sqrt{n}\} | \mathscr{F}_{m-1}\big] \to 0 \qquad \text{a.s.} \quad \text{or} \quad \text{in } L_1, \forall \varepsilon > 0, \tag{4.12}$$

$$\frac{1}{n} \sum_{m=1}^{n} \mathbf{V}_{mq} \to \mathbf{V}_q \qquad \text{a.s.} \quad \text{or} \quad \text{in } L_1 \text{ for all } q = 1, \dots, d, \tag{4.13}$$

where $\mathbf{V}_q = (V_{qkl})_{k,l=1}^d$, $q = 1, \dots, d$, are $d \times d$ symmetric positive semidefinite random matrices.

4.1. *Convergence results.* An important step in showing the a.s. convergence of $\mathbf{Y}_n/n$ and $\mathbf{N}_n/n$ in Bai and Hu (1999, 2005) and Laruelle and Pagès (2013) is the convergence of $\frac{\sum_{k=1}^d Y_{n,k}}{n}$, which is derived from the observation that

$$\left\{\sum_{k=1}^{d}\left(\Delta Y_{n,k} - \mathsf{E}[\Delta Y_{n,k}|\mathscr{F}_{n-1}]\right); n \geq 1\right\}$$

is a martingale difference sequence where $\Delta Y_{n,k} = Y_{n,k} - Y_{n-1,k}$ and

$$\sum_{k=1}^{d} \mathsf{E}[\Delta Y_{n,k}|\mathscr{F}_{n-1}] = \mathsf{E}\big[\mathbf{X}_n\mathbf{D}_n\mathbf{1}^t|\mathscr{F}_{n-1}\big] = \frac{\mathbf{Y}_{n-1}^{+}}{|\mathbf{Y}_{n-1}^{+}|}\mathbf{H}_n\mathbf{1}^{\mathrm{t}} = \alpha.$$

The last equality above is due to condition (4.8). When (4.8) is not satisfied, there is not an easy way to directly show the convergence of $\frac{\sum_{k=1}^d Y_{n,k}}{n}$. Next, we modify the ODE method proposed by Laruelle and Pagès (2013) to prove the convergence of $\mathbf{Y}_n/n$ and $\mathbf{N}_n/n$. The following theorem is the main result followed by its proof. Some of the basic tools in the ODE method that we used in the proof are presented in the Appendix.

THEOREM 4.1. *Suppose that*

$$\sum_{m=1}^{n} \mathbf{V}_{mq} = O(n) \qquad a.s. \quad or \quad in\ L_1\ for\ all\ q = 1, \dots, d, \tag{4.14}$$

and Assumption 4.1 is satisfied with $\lambda_{\mathrm{sec}} < 1$. Then

$$\frac{\mathbf{Y}_n}{n} \to \mathbf{v} \qquad a.s. \quad and \quad \frac{\mathbf{N}_n}{n} \to \mathbf{v}. \tag{4.15}$$

REMARK 4.1. It is easily seen that (4.14) is implied by either $\sup_m \mathsf{E}[\|\mathbf{D}_m\|^2|\mathscr{F}_{m-1}] < \infty$ a.s. or $\sup_m \mathsf{E}\|\mathbf{D}_m\|^2 < \infty$.

PROOF. To prove this theorem, we note that $\mathbf{Y}_{n+1} = \mathbf{Y}_n + \mathbf{X}_{n+1}\mathbf{D}_{n+1}$ and $\mathsf{E}[\mathbf{X}_{n+1}|\mathscr{F}_n] = \frac{\mathbf{Y}_n^+}{|\mathbf{Y}_n^+|}$. Let $\Delta\mathbf{M}_{n,1} = \mathbf{X}_n - \mathsf{E}[\mathbf{X}_n|\mathscr{F}_{n-1}]$ and $\Delta\mathbf{M}_{n,2} = \mathbf{X}_n(\mathbf{D}_n - \mathsf{E}[\mathbf{D}_n|\mathscr{F}_{n-1}])$. We have

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{\mathbf{Y}_n^+}{|\mathbf{Y}_n^+|}\mathbf{H} + \Delta\mathbf{M}_{n+1,1}\mathbf{H} + \Delta\mathbf{M}_{n+1,2} + \mathbf{X}_{n+1}(\mathbf{H}_{n+1} - \mathbf{H}). \tag{4.16}$$

Under assumption (4.10), we have $\sum_{m=1}^n \mathbf{X}_m(\mathbf{H}_m - \mathbf{H}) = o(n)$ a.s. It can be verified that $\mathbf{M}_{n,1}\mathbf{H} + \mathbf{M}_{n,2} = o(n)$ a.s. by assumption (4.14). Thus,

$$\begin{aligned}\mathbf{Y}_n &= \mathbf{Y}_0 + \sum_{m=0}^{n-1}\frac{\mathbf{Y}_m^+}{|\mathbf{Y}_m^+|}\mathbf{H} + (\mathbf{M}_{n,1}\mathbf{H} + \mathbf{M}_{n,2}) + \sum_{m=1}^{n}\mathbf{X}_m(\mathbf{H}_m - \mathbf{H})\\ &= \sum_{m=0}^{n-1}\frac{\mathbf{Y}_m^+}{|\mathbf{Y}_m^+|}\mathbf{H} + o(n) \qquad \text{a.s.}\end{aligned}$$

It follows that

$$\limsup_{n\to\infty}\frac{|Y_{n,k}|}{n}\le\limsup_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\frac{|\sum_{q=1}^{d}Y_{m,q}^{+}H_{q,k}|}{|\mathbf{Y}_m^{+}|}\le\max_{q,k}|H_{q,k}|\qquad\text{a.s.}$$

and

$$\liminf_{n\to\infty}\frac{\mathbf{Y}_n\mathbf{u}^{\mathrm{t}}}{n}=\liminf_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}\frac{\mathbf{Y}_m^{+}\mathbf{u}^{\mathrm{t}}}{|\mathbf{Y}_m^{+}|}\ge\min_k u_k>0\qquad\text{a.s.}$$

Let Θ^{∞} be the set of limiting values of $\frac{\mathbf{Y}_n^{+}}{n}$ as $n\to\infty$. Then

$$\Theta^{\infty}\subset\Big\{\boldsymbol{\theta}=(\theta_1,\dots,\theta_d):\boldsymbol{\theta}\mathbf{u}^{\mathrm{t}}>0,0\le\boldsymbol{\theta}_k\le\max_{q,l}|H_{q,l}|,k=1,\dots,d\Big\}.\tag{4.17}$$

Next, we show that

$$\mathbf{Y}_n^{+}-\mathbf{Y}_n=o(n)\qquad\text{a.s.}\tag{4.18}$$

Note that $|\mathbf{Y}_n^{+}|\ge c\mathbf{Y}_n^{+}\mathbf{u}^{\mathrm{t}}\ge c\mathbf{Y}_n\mathbf{u}^{\mathrm{t}}\to\infty$ a.s. as $n\to\infty$. Without loss of generality, we assume that $|\mathbf{Y}_n^{+}|>0$ for all n. Then $X_{m+1,k}=0$ if $Y_{m,k}<0$. For n and k, let $l_n=\max\{l\le n:Y_{l,k}\ge 0\}$ be the largest integer for which $Y_{l,k}\ge 0$. Then

$$\begin{aligned}
Y_{n,k}&=Y_{l_n,k}+\sum_{m=l_n+1}^{n}\sum_{q=1}^{d}X_{m,q}D_{q,k}(m)\\
&=Y_{l_n,k}+\sum_{m=l_n+1}^{n}\sum_{q=1}^{d}X_{m,q}\big[D_{q,k}(m)-H_{q,k}(m)\big]\\
&\quad+\sum_{m=l_n+1}^{n}\sum_{q=1}^{d}X_{m,q}\big[H_{q,k}(m)-H_{q,k}\big]+\sum_{m=l_n+1}^{n}\sum_{q=1}^{d}X_{m,q}H_{q,k}\\
&=Y_{l_n,k}+\sum_{m=l_n+1}^{n}\sum_{q=1}^{d}X_{m,q}H_{q,k}+o(n)\ge X_{l_n+1,k}H_{k,k}+o(n)\qquad\text{a.s.}
\end{aligned}$$

because $H_{q,k}\ge 0$ if $q\ne k$ and $X_{m,k}=0$ for $m=l_n+2,\dots,n$. It follows that $\liminf_{n\to\infty}\frac{Y_{n,k}}{n}\ge 0$ a.s., and then (4.18) follows.

Now, write $\boldsymbol{\theta}_n=\frac{\mathbf{Y}_n^{+}}{n}$ and

$$\begin{aligned}
\mathbf{r}_n&=(\Delta\mathbf{M}_{n,1}\mathbf{H}+\Delta\mathbf{M}_{n,2})+\mathbf{X}_n(\mathbf{H}_n-\mathbf{H})+\big(\Delta\mathbf{Y}_n^{+}-\Delta\mathbf{Y}_n\big),\\
\mathbf{s}_n&=\sum_{m=1}^{n}\mathbf{r}_m=(\mathbf{M}_{n,1}\mathbf{H}+\mathbf{M}_{n,2})+\sum_{m=1}^{n}\mathbf{X}_m(\mathbf{H}_m-\mathbf{H})+\big(\mathbf{Y}_n^{+}-\mathbf{Y}_n\big)\\
&\quad-\mathbf{M}_{0,1}\mathbf{H}-\mathbf{M}_{0,2}-\big(\mathbf{Y}_0^{+}-\mathbf{Y}_0\big).
\end{aligned}$$

Then, from (4.16) and (4.18) we conclude that $\boldsymbol{\theta}_n$ is bounded with a probability of one and satisfies the SA algorithm (1.1) with

$$\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{\theta}\left(\mathbf{I}_d - \frac{\mathbf{H}}{|\boldsymbol{\theta}|}\right) \quad \text{and} \quad \frac{\mathbf{s}_n}{n} \to \mathbf{0} \qquad \text{a.s.},$$

where $|\boldsymbol{\theta}| = \sum_{k=1}^{d} |\theta_k|$. It is obvious that $\mathbf{h}(\boldsymbol{\theta})$ is a continuous function on $\{\boldsymbol{\theta} : \boldsymbol{\theta}\mathbf{u}^{\mathrm{t}} > 0\}$.

By Theorem A.1 (a) and Remark A.1, the set Θ^{∞} of the limiting values of $\boldsymbol{\theta}_n$ is a.s. a compact connected set, stable by the flow of the ordinary differential equation (ODE):

$$\dot{\boldsymbol{\theta}} = -\mathbf{h}(\boldsymbol{\theta}).$$

It is obvious that $\mathbf{h}(\mathbf{v}) = 0$. By Theorem A.2, $\Theta =: \{\boldsymbol{\theta} : \boldsymbol{\theta}\mathbf{u}^{\mathrm{t}} > 0\}$ is a region of attraction of the above ODE for $\mathbf{v}$. Moreover, Θ is a neighborhood of $\mathbf{v}$. Further, $\Theta^{\infty} \subset \Theta$ by (4.17). By Theorem A.1 (b), we conclude that

$$\frac{\mathbf{Y}_n^+}{n} = \boldsymbol{\theta}_n \to \mathbf{v} \qquad \text{a.s.}$$

Accordingly, $\mathbf{Y}_n/n \to \mathbf{v}$ a.s., $|\mathbf{Y}_n^+|/n \to |\mathbf{v}| = 1$ a.s.

Finally,

$$\begin{aligned} \mathbf{N}_n &= \mathbf{N}_{n-1} + \big(\mathbf{X}_n - \mathsf{E}[\mathbf{X}_n|\mathscr{F}_{n-1}]\big) + \frac{\mathbf{Y}_{n-1}^+}{|\mathbf{Y}_{n-1}^+|} \\ &= \cdots = \mathbf{M}_{n,1} - \mathbf{M}_{0,1} + \sum_{m=0}^{n-1} \frac{\mathbf{Y}_m^+}{|\mathbf{Y}_m^+|}. \end{aligned} \tag{4.19}$$

It follows that

$$\lim_{n\to\infty} \frac{\mathbf{N}_n}{n} = \lim_{n\to\infty} \frac{\mathbf{M}_{n,1}}{n} + \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \frac{\mathbf{Y}_m^+/m}{|\mathbf{Y}_m^+|/m} = \mathbf{v} \qquad \text{a.s.}$$

The proof is complete. $\square$

4.2. *Limiting distribution.* We apply Theorems 2.1–2.3 to show the rates of convergence. First, we show that the random vector $(\frac{\mathbf{Y}_n}{n}, \frac{\mathbf{N}_n}{n})$ satisfies equation (1.1). By (4.16) and (4.19), we have

$$\mathbf{Y}_{n+1} = \mathbf{Y}_n + \frac{\mathbf{Y}_n^+}{n} \frac{\mathbf{H}}{|\mathbf{Y}_n^+/n|} + \Delta\mathbf{M}_{n+1,1}\mathbf{H} + \Delta\mathbf{M}_{n+1,2} + \mathbf{X}_{n+1}(\mathbf{H}_{n+1} - \mathbf{H})$$

and

$$\begin{aligned} \mathbf{N}_{n+1} &= \mathbf{N}_n + \Delta\mathbf{M}_{n+1,1} + \frac{\mathbf{Y}_n^+}{n} \frac{\mathbf{I}_d}{|\mathbf{Y}_n^+/n|} \\ &= \mathbf{N}_n + \Delta\mathbf{M}_{n+1,1} + \left(\frac{\mathbf{Y}_n^+}{n} - \mathbf{v}\right) \frac{n}{|\mathbf{Y}_n^+|} (\mathbf{I}_d - \mathbf{1}^{\mathrm{t}}\mathbf{v}) + \mathbf{v}. \end{aligned} \tag{4.20}$$

Write $\boldsymbol{\theta}_n = (\boldsymbol{\theta}_n^{(1)}, \boldsymbol{\theta}_n^{(2)}) = (\frac{\mathbf{Y}_n^+}{n}, \frac{\mathbf{N}_n}{n})$, $\boldsymbol{\theta}^* = (\mathbf{v}, \mathbf{v})$,

$$\Delta\mathbf{M}_n = (\Delta\mathbf{M}_{n+1,1}\mathbf{H} + \Delta\mathbf{M}_{n+1,2}, \Delta\mathbf{M}_{n+1,1})$$

and

$$\mathbf{r}_{n+1} = \big(\mathbf{X}_{n+1}(\mathbf{H}_{n+1} - \mathbf{H}) + \Delta\big(\mathbf{Y}_{n+1}^+ - \mathbf{Y}_{n+1}\big), \mathbf{0}\big).$$

Then $\boldsymbol{\theta}_n$ satisfies SA algorithm (1.1):

$$\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n - \frac{\mathbf{h}(\boldsymbol{\theta}_n)}{n+1} + \frac{\Delta\mathbf{M}_{n+1} + \mathbf{r}_{n+1}}{n+1}, \tag{4.21}$$

with

$$\mathbf{h}(\boldsymbol{\theta}) = \boldsymbol{\theta}\begin{pmatrix} \mathbf{I}_d - \dfrac{\mathbf{H}}{|\boldsymbol{\theta}^{(1)}|} & -\dfrac{\mathbf{I}_d}{|\boldsymbol{\theta}^{(1)}|} \\ \mathbf{0} & \mathbf{I}_d \end{pmatrix}.$$

For $\mathbf{r}_{n+1}$, by noting that $Y_{n,q}$ is positive eventually, and thus $Y_{n,q} = Y_{n,q}^+$ eventually due to Theorem 4.1 and the fact that $v_q > 0$, we have

$$\mathbf{Y}_n^+ - \mathbf{Y}_n = O(1) \qquad \text{a.s.} \tag{4.22}$$

It follows that

$$\sum_{m=1}^n \|\mathbf{r}_m\| = O(1)\sum_{m=1}^n \|\mathbf{H}_m - \mathbf{H}\| + O(1) \qquad \text{a.s.} \tag{4.23}$$

For $\Delta\mathbf{M}_{n+1}$, write $\boldsymbol{\Sigma}_{n,1} = \operatorname{diag}(\frac{\mathbf{Y}_n}{n}) - \frac{\mathbf{Y}_n^{\mathrm{t}}}{n}\frac{\mathbf{Y}_n}{n}$, $\boldsymbol{\Sigma}_{n,2} = \sum_{q=1}^d \frac{Y_{n,q}}{n}\mathbf{V}_{n+1,q}$. We have

$$\mathsf{E}\big[(\Delta\mathbf{M}_{n+1})^{\mathrm{t}}\Delta\mathbf{M}_{n+1}|\mathscr{F}_n\big] = \begin{pmatrix} \mathbf{H}^{\mathrm{t}}\boldsymbol{\Sigma}_{n,1}\mathbf{H} + \boldsymbol{\Sigma}_{n,2} & \mathbf{H}^{\mathrm{t}}\boldsymbol{\Sigma}_{n,1} \\ \boldsymbol{\Sigma}_{n,1}\mathbf{H} & \boldsymbol{\Sigma}_{n,1} \end{pmatrix}.$$

Then, under Assumptions 4.1 and 4.2,

$$\frac{1}{n}\sum_{m=1}^n \mathsf{E}\big[(\Delta\mathbf{M}_m)^{\mathrm{t}}\Delta\mathbf{M}_m|\mathscr{F}_{m-1}\big] \to \boldsymbol{\Gamma} \qquad \text{a.s.} \quad \text{or} \quad \text{in } L_1,$$

where

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{H}^{\mathrm{t}}\boldsymbol{\Sigma}_1\mathbf{H} + \boldsymbol{\Sigma}_2 & \mathbf{H}^{\mathrm{t}}\boldsymbol{\Sigma}_1 \\ \boldsymbol{\Sigma}_1\mathbf{H} & \boldsymbol{\Sigma}_1 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_1 = \operatorname{diag}(\mathbf{v}) - \mathbf{v}^{\mathrm{t}}\mathbf{v}, \qquad \boldsymbol{\Sigma}_2 = \sum_{q=1}^d v_q\mathbf{V}_q.$$

Finally, for $\mathbf{h}(\boldsymbol{\theta})$, it is easily seen that $\mathbf{h}(\boldsymbol{\theta})$ is twice differentiable at $\boldsymbol{\theta}^*$ with

$$D\mathbf{h}(\boldsymbol{\theta}^*) = \begin{pmatrix} \mathbf{I}_d - (\mathbf{H} - \mathbf{1}^{\mathrm{t}}\mathbf{v}) & -(\mathbf{I}_d - \mathbf{1}^{\mathrm{t}}\mathbf{v}) \\ \mathbf{0} & \mathbf{I}_d \end{pmatrix}.$$

Obviously, the system of the eigenvalues of both $D\mathbf{h}(\boldsymbol{\theta}^*)$ and $\mathbf{I}_d - (\mathbf{H} - \mathbf{1}^{\mathrm{t}}\mathbf{v})$ is $\{1, 1-\lambda_2, \dots, 1-\lambda_t\}$. Thus, $\rho = \min\{\operatorname{Re}(1), \operatorname{Re}(1-\lambda_2), \dots, \operatorname{Re}(1-\lambda_t)\} = 1 -$

$\lambda_{\rm sec}$. Further, it can be shown that if $\lambda_a \neq 0$, then the largest order of Jordan blocks of both $D\mathbf{h}(\boldsymbol{\theta}^*)$ and $\mathbf{I}_d - (\mathbf{H} - \mathbf{1}^{\rm t}\mathbf{v})$ with respect to their eigenvalue $1 - \lambda_a$ is the same as the largest order of Jordan blocks of $\mathbf{H}$ with respect to its eigenvalue λ_a. Hence, by applying Theorems 2.3 and 2.1 we have the following central limit theorems for $(\mathbf{Y}_n, \mathbf{N}_n)$.

THEOREM 4.2. *Suppose that Assumptions 4.1 and 4.2 are satisfied.*

(i) *Further, assume that $\lambda_{\rm sec} < 1/2$ and*

$$\sum_{m=1}^{n} \|\mathbf{H}_m - \mathbf{H}\| = o(\sqrt{n}) \qquad a.s. \quad or \quad in\ L_1. \tag{4.24}$$

Then

$$\sqrt{n}\left(\frac{\mathbf{Y}_n}{n} - \mathbf{v}, \frac{\mathbf{N}_n}{n} - \mathbf{v}\right) \xrightarrow{D} N(\mathbf{0}, \widetilde{\mathbf{\Sigma}}) \qquad (stably),$$

where

$$\widetilde{\mathbf{\Sigma}} = \int_0^{\infty} (e^{-\mathbf{Q}u})^{\rm t} \mathbf{\Gamma} e^{-\mathbf{Q}u}\, du$$

and

$$\mathbf{Q} = \begin{pmatrix} \mathbf{H} - \mathbf{1}^{\rm t}\mathbf{v} - \mathbf{I}_d/2 & \mathbf{I}_d - \mathbf{1}^{\rm t}\mathbf{v} \\ \mathbf{0} & -\mathbf{I}_d/2 \end{pmatrix}.$$

(ii) *Assume that $\lambda_{\rm sec} = 1/2$ and*

$$\sum_{m=1}^{n} \frac{\|\mathbf{H}_m - \mathbf{H}\|}{\sqrt{m}} = o(\sqrt{\log n}) \qquad a.s. \quad or \quad in\ L_1. \tag{4.25}$$

Then

$$\frac{\sqrt{n}}{(\log n)^{\nu - 1/2}}\left(\frac{\mathbf{Y}_n}{n} - \mathbf{v}, \frac{\mathbf{N}_n}{n} - \mathbf{v}\right) \xrightarrow{D} N(\mathbf{0}, \widetilde{\mathbf{\Sigma}}) \qquad (stably),$$

where

$$\widetilde{\mathbf{\Sigma}} = \lim_{n\to\infty} \frac{1}{(\log n)^{2\nu-1}} \int_0^{\log n} (e^{-\mathbf{Q}u})^{\rm t} \mathbf{\Gamma} e^{-\mathbf{Q}u}\, du.$$

REMARK 4.2. It can be verified that (4.6), which is the condition of Bai and Hu (2005), implies (4.24) and (4.25). (4.24) is also weaker than (4.9), which is condition (2.11) in Laruelle and Pagès (2013). Further, it can be verified that either (4.24) or (4.25) implies

$$\sum_{m=1}^{n} \|\mathbf{H}_m - \mathbf{H}\| = o(n^{1-\varepsilon_0}) \qquad \text{a.s. for some } \varepsilon_0 > 0$$

[cf. the proof of (2.19)]. Thus, condition (4.10) can be removed from the theorems.

THEOREM 4.3. *Suppose that Assumptions 4.1 and (4.14) are satisfied. Further, assume that $\lambda_{\text{sec}} > 1/2$ and that*

$$\sum_{m=1}^{n} \|\mathbf{H}_m - \mathbf{H}\| = o(n^{\lambda_{\text{sec}}-\delta_0}) \qquad \text{a.s. for some } \delta_0 > 0. \tag{4.26}$$

Then there are random complex variables $\xi_2, \dots, \xi_s$ and nonzero linearly independent left eigenvectors $\mathbf{l}_2, \dots, \mathbf{l}_s$ of $\mathbf{H}$ with $\mathbf{l}_a\mathbf{H} = \lambda_a \mathbf{l}_a$ such that

$$\begin{aligned} &\frac{n^{1-\lambda_{\text{sec}}}}{(\log n)^{\nu-1}}\left(\frac{\mathbf{N}_n}{n} - \mathbf{v}\right) \\ &\quad - \sum_{a:\mathrm{Re}(\lambda_a)=\lambda_{\text{sec}}, \nu_a=\nu} e^{\mathrm{i}\,\mathrm{Im}(\lambda_t)\log n}\xi_a \mathbf{l}_a(\mathbf{I} - \mathbf{1}'\mathbf{v}) \to \mathbf{0} \qquad \text{a.s.} \end{aligned} \tag{4.27}$$

and

$$\frac{n^{1-\lambda_{\text{sec}}}}{(\log n)^{\nu-1}}\left(\frac{\mathbf{Y}_n}{n} - \mathbf{v}\right) - \frac{n^{1-\lambda_{\text{sec}}}}{(\log n)^{\nu-1}}\left(\frac{\mathbf{N}_n}{n} - \mathbf{v}\right)\mathbf{H} \to \mathbf{0} \qquad \text{a.s.} \tag{4.28}$$

PROOF. We apply Theorem 2.2 to prove this theorem. Assume that $\mathbf{T}$ is a matrix such that

$$\mathbf{T}^{-1}[\mathbf{I}_d - (\mathbf{H} - \mathbf{1}^{\mathrm{t}}\mathbf{v})]\mathbf{T} = \mathrm{diag}(1, (1-\lambda_2)\mathbf{I}_{\nu_2} + \overline{\mathbf{J}}_{\nu_2}, \dots, (1-\lambda_s)\mathbf{I}_{\nu_s} + \overline{\mathbf{J}}_{\nu_s}).$$

By (4.21), we have

$$\frac{\mathbf{Y}_{n+1}^+}{n+1} = \frac{\mathbf{Y}_n^+}{n} - \frac{\mathbf{h}_1(\frac{\mathbf{Y}_n^+}{n})}{n+1} + \frac{\Delta\mathbf{M}_{n+1,1}\mathbf{H} + \Delta\mathbf{M}_{n+1,2} + \mathbf{r}_{n+1}^{(1)}}{n+1},$$

where $\mathbf{h}_1(\boldsymbol{\theta}^{(1)}) = \mathbf{I}_d - \frac{\mathbf{H}}{|\boldsymbol{\theta}^{(1)}|}$ is twice-differentiable with $D\mathbf{h}_1(\mathbf{v}) = \mathbf{I}_d - (\mathbf{H} - \mathbf{1}^{\mathrm{t}}\mathbf{v})$. Condition (2.10) is satisfied by assumption (4.14), and (2.11) is satisfied by (4.23) and assumption (4.26). Thus, by Theorem 2.2, there are complex random variables $\xi_2, \dots, \xi_s$ such that

$$\frac{n^{1-\lambda_{\text{sec}}}}{(\log n)^{\nu-1}}\left(\frac{\mathbf{Y}_n^+}{n} - \mathbf{v}\right) - \sum_{a:\mathrm{Re}(\lambda_a)=\lambda_{\text{sec}}, \nu_a=\nu} e^{\mathrm{i}\,\mathrm{Im}(\lambda_a)\log n}\xi_a \mathbf{e}_a \mathbf{T}^{-1} \to \mathbf{0} \qquad \text{a.s.}$$

From (4.20) and the above convergence, we have

$$\begin{aligned} &\mathbf{N}_n - n\mathbf{v} \\ &\quad = \mathbf{M}_{n,1} + \sum_{m=1}^{n-1}\left(\frac{\mathbf{Y}_m^+}{m} - \mathbf{v}\right)\left(\frac{m}{\mathbf{Y}_m^+} - 1\right)(\mathbf{I}_d - \mathbf{1}^{\mathrm{t}}\mathbf{v}) + \sum_{m=1}^{n-1}\left(\frac{\mathbf{Y}_m^+}{m} - \mathbf{v}\right)(\mathbf{I}_d - \mathbf{1}^{\mathrm{t}}\mathbf{v}) \\ &\quad = O(\sqrt{n\log\log n}) + \sum_{m=1}^{n-1}\left(O(1)\frac{(\log m)^{\nu-1}}{m^{1-\lambda_{\text{sec}}}}\right)^2 \end{aligned}$$

$$+\sum_{m=1}^{n-1}\frac{(\log m)^{\nu-1}}{m^{1-\lambda_{\rm sec}}}\sum_{a:\mathrm{Re}(\lambda_a)=\lambda_{\rm sec},\nu_a=\nu}\big[e^{\mathrm{i}\,\mathrm{Im}(\lambda_a)\log m}\xi_a\mathbf{e}_a\mathbf{T}^{-1}+o(1)\big](\mathbf{I}_d-\mathbf{1}^{\rm t}\mathbf{v})$$

$$=n^{\lambda_{\rm sec}}(\log n)^{\nu-1}\bigg[o(1)+\sum_{a:\mathrm{Re}(\lambda_a)=\lambda_{\rm sec},\nu_a=\nu}e^{\mathrm{i}\,\mathrm{Im}(\lambda_a)\log n}\lambda_a^{-1}\xi_a\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{I}_d-\mathbf{1}^{\rm t}\mathbf{v})\bigg]\quad\text{a.s.}$$

Write $\mathbf{l}_a=\lambda_a^{-1}\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{I}-\mathbf{u}'\mathbf{v})$ if $\lambda_a\neq 0$ and $\mathbf{l}_a=\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{I}-\mathbf{u}'\mathbf{v})$ if $\lambda_a=0$. Note that $\mathbf{e}_a\mathbf{T}^{-1}$ is a left eigenvector of $\mathbf{H}-\mathbf{1}'\mathbf{v}$ with respect to the eigenvalue λ_a. We conclude that $\mathbf{l}_a\mathbf{H}=\lambda_a^{-1}\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{H}-\mathbf{u}'\mathbf{v})=\lambda_a^{-1}\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{H}-\mathbf{1}'\mathbf{v})(\mathbf{I}-\mathbf{u}'\mathbf{v})=\lambda_a\mathbf{l}_a$ if $\lambda_a\neq 0$ and $\mathbf{l}_a\mathbf{H}=\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{H}-\mathbf{1}'\mathbf{v})(\mathbf{I}-\mathbf{u}'\mathbf{v})=\mathbf{0}$ if $\lambda_a=0$. Further, $\mathbf{l}_a(\mathbf{I}-\mathbf{1}'\mathbf{v})=\lambda_a^{-1}\xi_a\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{I}_d-\mathbf{1}^{\rm t}\mathbf{v})$ and $\mathbf{l}_a(\mathbf{I}-\mathbf{1}'\mathbf{v})\mathbf{H}=\lambda_a^{-1}\mathbf{e}_a\mathbf{T}^{-1}(\mathbf{H}-\mathbf{1}^{\rm t}\mathbf{v})=\mathbf{e}_a\mathbf{T}^{-1}$ when $\lambda_a\neq 0$. (4.27) is proved, and (4.28) is also proved by noting (4.22). Finally, the linear independence of $\mathbf{l}_2,\dots,\mathbf{l}_s$ is due to the linear independence of the system $\{\mathbf{v}(=\mathbf{e}_1\mathbf{T}^{-1}),\mathbf{e}_2\mathbf{T}^{-1},\dots,\mathbf{e}_s\mathbf{T}^{-1}\}$. □

REMARK 4.3. When $\lambda_{\rm sec}>1/2$, Bai and Hu (2005) showed that $\mathbf{Y}_n-n\mathbf{v}=O(n^{\lambda_{\rm sec}}\log^{\nu-1}n)$ in probability. Now, by Theorem 4.3, $\mathbf{Y}_n-n\mathbf{v}=O(n^{\lambda_{\rm sec}}\times\log^{\nu-1}n)$ a.s. and $\mathbf{N}_n-n\mathbf{v}=O(n^{\lambda_{\rm sec}}\log^{\nu-1}n)$ a.s. Further, if all eigenvalues with $\mathrm{Re}(\lambda_t)=\lambda_{\rm sec}$ and $\nu_t=\nu$ are real, then both $(\mathbf{Y}_n-n\mathbf{v})/(n^{\lambda_{\rm sec}}\log^{\nu-1}n)$ and $(\mathbf{N}_n-n\mathbf{v})/(n^{\lambda_{\rm sec}}\log^{\nu-1}n)$ a.s. converge toward a finite random vector.

APPENDIX A: ODE METHODS FOR THE RECURSIVE ALGORITHM

THEOREM A.1 (Kushner–Clark). *Consider the following recursive procedure*:

$$\boldsymbol{\theta}_{n+1}=\boldsymbol{\theta}_n-\gamma_{n+1}\mathbf{h}(\boldsymbol{\theta}_n)+\gamma_{n+1}\mathbf{r}_{n+1},\tag{RP}$$

where $\mathbf{h}$ *is a continuous function and* $\{\gamma_n\}$ *is a positive sequence that tends toward zero, such that* $\sum_{n=1}^{\infty}\gamma_n$ *diverges.*

(a) *We suppose that sequence* $\{\boldsymbol{\theta}_n\}$ *is bounded, and for all* $T>0$,

$$\lim_{n\to\infty}\sum_{j\le m(n,T)}\bigg\|\sum_{k=n}^{j}\gamma_{k+1}\mathbf{r}_{k+1}\bigg\|=0,\tag{A.1}$$

where $m(n,T)=\inf\{k:k\ge n,\gamma_{n+1}+\cdots+\gamma_{k+1}\ge T\}$. *Then the set* Θ^∞ *of the limiting values of* $\boldsymbol{\theta}_n$ *is a compact connected set, stable by the flow of the ODE*:

$$\mathrm{ODE}_h\equiv\dot{\boldsymbol{\theta}}=-\mathbf{h}(\boldsymbol{\theta}).\tag{ODE1}$$

(b) *Further, let* Θ *be a region of attraction for* $\boldsymbol{\theta}^*$, *where* $\boldsymbol{\theta}^*$ *is a zero of* $\mathbf{h}$, *that is, the following properties are satisfied*:

(i) *For any solution of* (ODE1), *if* $\boldsymbol{\theta}(0) \in \Theta$, *then* $\boldsymbol{\theta}(s) \in \Theta$ *for all* $s \geq 0$;
(ii) *if* $\boldsymbol{\theta}$ *is a solution of* (ODE1) *for which* $\boldsymbol{\theta}(0) \in \Theta$, *then*

$$\boldsymbol{\theta}(s) \to \boldsymbol{\theta}^* \qquad \text{as } s \to +\infty; \quad \text{and}$$

(iii) *given* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that* $\boldsymbol{\theta}(0) \in \Theta$ *and* $\|\boldsymbol{\theta}(0) - \boldsymbol{\theta}^*\| \leq \delta$ *imply* $\|\boldsymbol{\theta}(s) - \boldsymbol{\theta}^*\| \leq \varepsilon$ *for all* $s \geq 0$.

Suppose that Θ *is a neighborhood of* $\boldsymbol{\theta}^*$. *We assume the framework of part* (a). *If sequence* $\{\boldsymbol{\theta}_n\}$ *returns infinitely often to a compact subset of* Θ, *then it tends toward* $\boldsymbol{\theta}^*$.

This is known as the Kushner–Clark theorem and can be found in the book by Duflo (1997), page 318. A similar theorem is obtained by Ljung (1977). Variants and improvements have been proposed in classical textbooks by Duflo (1996, 1997), Kushner and Clark (1978) and Kushner and Yin (2003), in addition to some papers [see, e.g., Fort and Pagès (1996)].

REMARK A.1. If $\gamma_n \equiv \frac{1}{n}$ and $\frac{1}{n}\sum_{k=1}^n \mathbf{r}_k \to 0$, then (A.1) is satisfied.
In fact, let $\mathbf{s}_n = \sum_{k=1}^n \mathbf{r}_k$. Then for $j \leq m(n,T)$, we have $\sum_{k=n+1}^j \frac{1}{k} \leq T$ and

$$\sum_{k=n}^{j} \frac{\mathbf{r}_{k+1}}{k+1} = \sum_{k=n+1}^{j} \frac{\mathbf{s}_k}{k+1}\frac{1}{k} + \frac{\mathbf{s}_{j+1}}{j+1} - \frac{\mathbf{s}_n}{n+1}.$$

It follows that

$$\begin{aligned}\max_{j \leq m(n,T)} \left\| \sum_{k=n}^{j} \frac{\mathbf{r}_{k+1}}{k+1} \right\| &\leq \sup_{m \geq n} \frac{\|\mathbf{s}_m\|}{m} \sum_{k=n+1}^{m(n,T)} \frac{1}{k} + 2 \sup_{m\geq n} \frac{\|\mathbf{s}_m\|}{m} \\ &\leq (T+2) \sup_{m \geq n} \frac{\|\mathbf{s}_m\|}{m} \to 0 \qquad \text{a.s. as } n \to \infty.\end{aligned}$$

THEOREM A.2. *Let* $\mathbf{H}$ *be a matrix satisfying Assumption* 4.1. *Suppose that* $\mathbf{u}^{\mathrm{t}} > 0$ *and* $\mathbf{v} > 0$ *are, respectively, the right and left eigenvectors of* $\mathbf{H}$ *with respect to the largest eigenvalue* 1 *with* $\mathbf{v}\mathbf{1}^{\mathrm{t}} = 1$ *and* $\mathbf{v}\mathbf{u}^{\mathrm{t}} = 1$. *Consider the ordinary differential equation*

$$\dot{\boldsymbol{\theta}} = -\boldsymbol{\theta}\left(\mathbf{I}_d - \frac{\mathbf{H}}{|\boldsymbol{\theta}|}\right), \qquad \boldsymbol{\theta}(0) = \boldsymbol{\theta}_0, \tag{ODE2}$$

where $|\boldsymbol{\theta}| = \sum_{k=1}^d |\theta_k|$. *Then*, $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta}\mathbf{u}^{\mathrm{t}} > 0\}$ *is a region of attraction for* $\mathbf{v}$.

APPENDIX B: BASIC RESULTS FOR MATRICES AND MARTINGALES

PROPOSITION B.1. *Let $\{\mathbf{H}_n\}$ be a sequence of real matrices and $\mathbf{H} = D\mathbf{h}(\boldsymbol{\theta}^*)$. Write $\mathbf{\Pi}_m^n = \prod_{j=m+1}^n (\mathbf{I}_d - \frac{\mathbf{H}_j}{j})$ and $\widetilde{\mathbf{\Pi}}_m^n = \prod_{j=m+1}^n (\mathbf{I}_d - \frac{\mathbf{H}}{j})$ for all $1 \le m \le n-1$. Then*:

(i) $\|\widetilde{\mathbf{\Pi}}_m^n\| \le C_0(\frac{n}{m})^{-\rho}\log^{\nu-1}\frac{n}{m} \le C_\delta(\frac{n}{m})^{-\rho+\delta}$ *for all* $\delta > 0$;

(ii) *If* $\mathbf{H}_n \to \mathbf{H}$ *as* $n \to \infty$, *then for all* $\delta > 0$, $\|\mathbf{\Pi}_m^n\| \le C_\delta(\frac{n}{m})^{-\rho+\delta}$ *and* $\mathbf{\Pi}_m^n - \widetilde{\mathbf{\Pi}}_m^n = o(1)(\frac{n}{m})^{-\rho+\delta}$ *as* $n \ge m \to \infty$;

(iii) *If* $\sum_{j=1}^\infty \frac{\|\mathbf{H}_j - \mathbf{H}\|}{j}(\log j)^{\nu-1} < \infty$, *then* $\|\mathbf{\Pi}_m^n\| \le C(\frac{n}{m})^{-\rho}\log^{\nu-1}\frac{n}{m}$ *and* $\mathbf{\Pi}_m^n - \widetilde{\mathbf{\Pi}}_m^n = o(1)(\frac{n}{m})^{-\rho}\log^{\nu-1}\frac{n}{m}$ *as* $n \ge m \to \infty$; *and*

(iv) $\max_{x\in[m-c,m+c]}\|\widetilde{\mathbf{\Pi}}_m^n - (\frac{n}{x})^{-\mathbf{H}}\| = o(1)(\frac{n}{m})^{-\rho}\log^{\nu-1}\frac{n}{m}$ *as* $n \ge m \to \infty$.

Here, for a positive number a, $a^{\mathbf{H}}$ *is defined as* $a^{\mathbf{H}} = e^{\mathbf{H}\log a} = \sum_{j=0}^\infty \frac{1}{j!}(\log a)^j \mathbf{H}^j$.

PROPOSITION B.2. *Suppose that Assumption 2.3 is satisfied. Write $\mathbf{H} = D\mathbf{h}(\boldsymbol{\theta}^*)$, $\mathbf{\Pi}_m^n = \prod_{j=m+1}^n (\mathbf{I}_d - \frac{\mathbf{H}_j}{j})$ and $\widetilde{\mathbf{\Pi}}_m^n = \prod_{j=m+1}^n (\mathbf{I}_d - \frac{\mathbf{H}}{j})$ and*

$$\boldsymbol{\zeta}_n = \sum_{m=1}^n \frac{\Delta\mathbf{M}_m}{m}\widetilde{\mathbf{\Pi}}_m^n.$$

(i) *If* $\rho > 1/2$ *and* $\mathbf{H}_n \to \mathbf{H}$ *a.s., then*

$$\sqrt{n}\sum_{m=1}^n \frac{\Delta\mathbf{M}_m}{m}\mathbf{\Pi}_m^n - \boldsymbol{\zeta}_n \to \mathbf{0} \qquad \textit{in probability},$$
$$\sqrt{n}\boldsymbol{\zeta}_n \xrightarrow{D} N(\mathbf{0}, \mathbf{\Sigma}) \qquad (\textit{stably}), \tag{B.1}$$

where

$$\mathbf{\Sigma} = \int_0^\infty \big(e^{-(\mathbf{H}-\mathbf{I}_d/2)u}\big)^{\mathrm{t}}\mathbf{\Gamma}e^{-(\mathbf{H}-\mathbf{I}_d/2)u}\,du.$$

(ii) *If* $\rho = 1/2$, *then*

$$\frac{\sqrt{n}}{(\log n)^{\nu-1/2}}\boldsymbol{\zeta}_n \xrightarrow{D} N(\mathbf{0}, \widetilde{\mathbf{\Sigma}}) \qquad (\textit{stably}),$$

where

$$\widetilde{\mathbf{\Sigma}} = \lim_{n\to\infty}\frac{1}{(\log n)^{2\nu-1}}\int_0^{\log n}\big(e^{-(\mathbf{H}-\mathbf{I}_d/2)u}\big)^{\mathrm{t}}\mathbf{\Gamma}e^{-(\mathbf{H}-\mathbf{I}_d/2)u}\,du$$

satisfies (2.9).

The proofs of Propositions A.2, B.1 and B.2 appear in the supplementary material [Zhang (2016)].

Acknowledgments. Special thanks go to the anonymous referees, the Associate Editor and the Editors for their constructive comments, which led to a much improved version of this paper.

SUPPLEMENTARY MATERIAL

Supplement to "Central Limit Theorems of a recursive stochastic algorithm with applications to adaptive designs"(DOI: 10.1214/16-AAP1187SUPP; .pdf). The proofs of basic results stated in the Appendices are given.

REFERENCES

ATHREYA, K. B. and KARLIN, S. (1967). Limit theorems for the split times of branching processes. *J. Math. Mech.* **17** 257–277. MR0216592

ATHREYA, K. B. and KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Statist.* **39** 1801–1817. MR0232455

ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York. MR0373040

BAI, Z. D. and HU, F. (1999). Asymptotic theorems for urn models with nonhomogeneous generating matrices. *Stochastic Process. Appl.* **80** 87–101. MR1670107

BAI, Z.-D. and HU, F. (2005). Asymptotics in randomized urn models. *Ann. Appl. Probab.* **15** 914–940. MR2114994

BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. *Applications of Mathematics* (*New York*) **22**. Springer, Berlin. MR1082341

CHEUNG, Y. K. (2010). Stochastic approximation and modern model-based designs for dose-finding clinical trials. *Statist. Sci.* **25** 191–201. MR2789989

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

DUFLO, M. (1997). *Random Iterative Models*. *Applications of Mathematics* (*New York*) **34**. Springer, Berlin. MR1485774

EBERLEIN, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.* **14** 260–270. MR0815969

EGGENBERGER, F. and PÓLYA, G. (1923). Über die statistik verketetter vorgänge. *Z. Angew. Math. Mech.* **1** 279–289.

FORT, J.-C. and PAGÈS, G. (1996). Convergence of stochastic algorithms: From the Kushner–Clark theorem to the Lyapounov functional method. *Adv. in Appl. Probab.* **28** 1072–1094. MR1418247

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

HIGUERAS, I., MOLER, J., PLO, F. and SAN MIGUEL, M. (2003). Urn models and differential algebraic equations. *J. Appl. Probab.* **40** 401–412. MR1978099

HIGUERAS, I., MOLER, J., PLO, F. and SAN MIGUEL, M. (2006). Central limit theorems for generalized Pólya urn models. *J. Appl. Probab.* **43** 938–951. MR2274628

HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, Hoboken, NJ. MR2245329

HU, F. and ZHANG, L.-X. (2004). Asymptotic normality of urn models for clinical trials with delayed response. *Bernoulli* **10** 447–463. MR2061440

JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.* **110** 177–245. MR2040966

JOHNSON, N. L. and KOTZ, S. (1977). *Urn Models and Their Application*: *An Approach to Modern Discrete Probability Theory*. Wiley, New York. MR0488211

KOTZ, S. and BALAKRISHNAN, N. (1997). Advances in urn models during the past two decades. In *Advances in Combinatorial Methods and Applications to Probability and Statistics*. *Stat*. *Ind*. *Technol*. 203–257. Birkhäuser, Boston, MA. MR1456736

KUSHNER, H. J. and CLARK, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. *Applied Mathematical Sciences* **26**. Springer, New York. MR0499560

KUSHNER, H. J. and YIN, G. G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **35**. Springer, New York. MR1993642

LARUELLE, S. and PAGÈS, G. (2013). Randomized urn models revisited using stochastic approximation. *Ann*. *Appl*. *Probab*. **23** 1409–1436. MR3098437

LJUNG, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Trans*. *Automat*. *Control* **AC-22** 551–575. MR0465458

MONRAD, D. and PHILIPP, W. (1991). Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales. *Probab*. *Theory Related Fields* **88** 381–404. MR1100898

PELLETIER, M. (1998). Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann*. *Appl*. *Probab*. **8** 10–44. MR1620405

SMYTHE, R. T. (1996). Central limit theorems for urn models. *Stochastic Process*. *Appl*. **65** 115–137. MR1422883

WEI, L. J. (1979). The generalized Pólya's urn design for sequential medical trials. *Ann*. *Statist*. **7** 291–296.

WEI, L. J. and DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J*. *Amer*. *Statist*. *Assoc*. **73** 840–843.

ZHANG, L. (2004). Strong approximations of martingale vectors and their applications in Markov-chain adaptive designs. *Acta Math*. *Appl*. *Sin*. *Engl*. *Ser*. **20** 337–352. MR2064011

ZHANG, L.-X. (2016). Supplement to "Central Limit Theorems of a recursive stochastic algorithm with applications to adaptive designs." DOI:10.1214/16-AAP1187SUPP.

ZHANG, L.-X., HU, F. and CHEUNG, S. H. (2006). Asymptotic theorems of sequential estimation-adjusted urn models. *Ann*. *Appl*. *Probab*. **16** 340–369. MR2209345

ZHANG, L.-X., HU, F., CHEUNG, S. H. and CHAN, W. S. (2011). Immigrated urn models—theoretical properties and applications. *Ann*. *Statist*. **39** 643–671. MR2797859

SCHOOL OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
ZHEDA ROAD, NO. 38
HANG ZHOU, 310027
P.R. CHINA
E-MAIL: stazlx@zju.edu.cn