1-1-2011

Robustness, Weak Stability, and Stability in Distribution of Adaptive Filteringalgorithms Under Model Mismatch

Ben G. Fitzpatrick
Loyola Marymount University, ben.fitzpatrick@lmu.edu

G. Yin
Wayne State University

Le Yi Wang
Wayne State University

MULTISCALE MODEL. SIMUL.
Vol. 9, No. 1, pp. 183–207

ROBUSTNESS, WEAK STABILITY, AND STABILITY IN DISTRIBUTION OF ADAPTIVE FILTERING ALGORITHMS UNDER MODEL MISMATCH*

BEN G. FITZPATRICK†, G. YIN‡, AND LE YI WANG§

Abstract. This work is concerned with robustness, convergence, and stability of adaptive filtering (AF) type algorithms in the presence of model mismatch. The algorithms under consideration are recursive and have inherent multiscale structure. They can be considered as dynamic systems, in which the "state" changes much more slowly than the perturbing noise. Beyond the existing results on adaptive algorithms, model mismatch significantly affects convergence properties of AF algorithms, raising issues of algorithm robustness. Weak convergence and weak stability (i.e., recurrence) under model mismatch are derived. Based on the limiting stochastic differential equations of suitably scaled iterates, stability in distribution is established. Then algorithms with decreasing step sizes and their convergence properties are examined. When input signals are large, identification bias due to model mismatch will become large and unacceptable. Methods for reducing such bias are introduced when the identified models are used in regulation problems.

Key words. adaptive filtering, model mismatch, recurrence, stability in distribution

AMS subject classifications. 93E11, 93E12, 93E15, 93E24, 60F17, 62L20

DOI. 10.1137/100797059

1. Introduction.

In control applications ranging from flight systems to noise cancelation to laser beam pointing and tracking, adaptive estimation, filtering, and control techniques play a crucial role. In particular, high-energy laser systems, free-space optical communications, laser welding and cutting, optical data storage, and scanning optical lithography rely on well-controlled beams operating with uncertain, difficult-to-characterize disturbances. Moreover, the dynamic response of the control system to laser steering inputs (such as fast steering mirrors) can be challenging to determine a priori, and physical modeling (as in flight control systems) is often insufficient. In such cases, adaptive estimation techniques, typically implemented through recursive stochastic algorithms, are crucial ingredients of the control design and implementation. Motivated by the laser beam pointing and tracking efforts of [7, 8, 28, 29], this work examines adaptive filtering (AF) algorithms in the presence of model mismatch. Applying AF algorithms, we often encounter model mismatch issues that inevitably lead to the problem of robustness under uncertainties of model function

*Received by the editors June 1, 2010; accepted for publication (in revised form) November 23, 2010; published electronically January 20, 2011.
http://www.siam.org/journals/mms/9-1/79705.html

†Tempest Technologies, 8939 South Sepulveda Blvd., Suite 506, Los Angeles, CA 90045, and Department of Mathematics, Loyola Marymount University, Los Angeles, CA 90045 (fitzpatrick@tempest-tech.com). Research of this author was supported in part by the Joint Technology Office and the Air Force Office of Scientific Research through the Multidisciplinary Research Initiative grant F49620-02-1-0319 and the Air Force Office of Scientific Research under grant FA9550-09-1-0524.

‡Department of Mathematics, Wayne State University, Detroit, MI 48202 (gyin@math.wayne.edu). The research of this author was supported in part by the National Science Foundation under DMS-0907753 and in part by the Air Force Office of Scientific Research under FA9550-10-1-0210.

§Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202 (lywang@wayne.edu). Research of this author was supported in part by the National Science Foundation under DMS-0624849 and in part by the Air Force Office of Scientific Research under FA9550-10-1-0210.

forms. Our goal is to derive conditions under which stability and robustness can be guaranteed. Note that the algorithms under consideration are recursive and have inherent multiscale structure. They can be considered as dynamic systems, in which the "state" changes much more slowly than the perturbing noise.

Recursive least squares (RLS), least mean squares, or general AF algorithms have been examined extensively in the literature. Recursive AF algorithms employ parameterized models, whose parameters are constructed using stochastic parametric optimization. In the literature, it has been established that under appropriate conditions on noise, by using the technique of centering and interpolation, the estimation errors can be expressed as a scaled or normalized error sequence whose weak limit is a diffusion process. There are many important results available concerning convergence, rates of convergence, and asymptotic efficiency; see, for instance, [1, 3, 6, 10, 11, 22, 23, 25, 26, 27, 36, 38, 41]. Emerging applications have also been reported in wireless communications and multiuser detection; see [30, 31, 40], among others.

In practical applications, the parametric model is often an approximation of a nonlinear system, leading to model mismatch in the AF algorithms. Introduction of model mismatch affects significantly the convergence properties of AF algorithms. Unlike observation noises or unmodeled dynamics (as referred to in the system identification literature [32, 33, 34, 35]), model mismatch carries distinct features. Uncertainties due to stochastic observation noises can be reduced by averaging. Unmodeled dynamics mentioned above are linear multiplicative uncertainty, represent terms of remote past which are independent of the model parameters, and can be reduced by time separation [32, 33, 39]. Model mismatch, however, is a nonlinear and nonmultiplicative uncertainty and is a function of the true model parameter and inputs. Consequently, analysis of its impact on identification accuracy and convergence and methods of its reduction require different approaches.

This paper establishes robustness of AF algorithms under model mismatch uncertainty. We offer at least partial answers to some intriguing questions: Given an AF algorithm and input signal, how many model mismatch errors can the system tolerate before losing convergence? How much bias will model mismatch cause on parameter estimation? What types of convergence are robust under model mismatch uncertainty? How can identification bias from model mismatch be reduced?

This paper will establish asymptotic results that resolve these issues. Our results show that robustness of an AF algorithm can be established quantitatively from the covariance of its regressors. Explicit bounds on estimation bias due to model mismatch are derived. Robustness of AF algorithms in the sense of weak convergence, weak stability, and convergence in distribution is presented.

Traditional convergence analysis of AF algorithms often concentrates on strong or mean square convergence of estimates, which is related to local asymptotic stability of certain related limit systems. As noted in [20], in many applications a relaxed notion of stability, called weak stability, is more useful. It is a recurrence property of the estimates and similar to the notion of Lagrange stability (bounded inputs producing bounded outputs) in deterministic dynamical systems. Weak stability establishes the property of an estimate to return to a bounded region infinitely often (the property known as recurrence). The work here follows the notion of recurrence in diffusion theory in [14, 15, 37] and the very recent work on switching diffusion processes [43, Chapter 3]; see also related references in [5, 16].

Stability in distribution under the framework of Lyapunov stability represents the following scenario: if the initial probability distribution of the underlying process is close to the limit distribution, then the distribution of the process asymptotically approaches that of the limit distribution. Consequently, the limit distribution can be reliably used to study statistical properties of the estimates. In this paper, we investigate robustness of AF algorithms in terms of stability in distribution when model mismatch is present. This is accomplished by using scaled and centered sequences and by studying their convergence robustness.

The rest of the paper is organized as follows. We begin in section 2 with the setup of AF problems. Also included are discussions on the associated limiting dynamical systems, such as ordinary differential equations (ODEs) and stochastic differential equations (SDEs) as well as a brief summary of related existing results on AF algorithms. Robustness of AF algorithms in weak convergence under model mismatch uncertainty is studied in section 3. Our results may be viewed as an expression of a "stability margin" against model mismatch uncertainty. Section 4 concentrates on the recurrence or weak stability of our recursive algorithms. First, criteria for recurrence of the limit dynamics are obtained by a Lyapunov function approach. Then recurrence of the recursive algorithms is obtained via a perturbed Lyapunov function method. Section 5 continues our study on robust stability in distribution of AF algorithms. The method of perturbed Lyapunov functions is used to develop robustness regions. Section 6 concludes the paper with some discussions on extensions of the main results of this paper. Implications of unbounded signals and decreasing stepsizes are discussed. When input signals are large, identification bias due to model mismatch will become large and unacceptable. Methods for reducing such bias are introduced when the identified models are used in regulation problems.

2. Preliminary considerations.

2.1. The basic problem structure. AF algorithms have been used quite frequently in various applications such as estimation, adaptive control, signal processing, and related fields. The most commonly used (and best understood) AF algorithms start with the linear regression observation structure

$$y_n = \varphi_n' \theta_* + \zeta_n, \tag{2.1}$$

where y_n is a one-dimensional output, $\varphi_n \in \mathbb{R}^r$ is known as the regressor, θ_* is an r-dimensional parameter, and $\{\zeta_n\}$ is a scalar sequence of random noise. This formulation will be used throughout the paper.

In using this structure, one aims to estimate the unknown parameter vector θ_* from observation sequences $\{(y_n, \varphi_n)\}$. In principle, an estimate can be derived by minimizing a cost function $L(\theta)$ such as $L(\theta) = E|y_n - \varphi_n' \theta|^2$.

For computational efficiency, we prefer estimation algorithms that recursively update the parameter estimates as new data are collected over "batch" optimization approaches. For example, a gradient-based recursive algorithm takes the form

$$\theta_{n+1} = \theta_n + a_n \varphi_n (y_n - \varphi_n' \theta_n), \tag{2.2}$$

in which $\{a_n\}$ is a sequence of positive scalars, known as stepsizes, satisfying $\sum_n a_n = \infty$, and $a_n \to 0$ as $n \to \infty$. Alternatively, we may use a constant stepsize algorithm

$$\theta_{n+1} = \theta_n + \varepsilon \varphi_n (y_n - \varphi_n' \theta_n), \tag{2.3}$$

where $\varepsilon > 0$ is a small constant stepsize. There is a wide variety of applications that are formulated or can be recast using the observation structure (2.1) and the form of AF problems. As observed in [22, Chapter 1], the traditional least squares algorithm can be approximated by the AF algorithms. Furthermore, in Chapter 3 of the aforementioned reference, a number of problems are presented, including pattern classification, autoregressive and moving average systems (see also [25, 26] and references therein), adaptive noise cancelation and disturbance rejection, antenna array processing, adaptive equalizers, and adaptive echo cancelation, among others. One of the noise cancelation methods is the application in [12, 36] to the problem of getting a good measurement of the heartbeat of a fetus, which is quite faint in comparison with the power of the beat of the mother's heart. References on wireless communications, multiuser detection, and spread code optimizations can also be found in [17, 41], among others. All of these use mean squares type criteria and utilize AF type algorithms.

Note that the above algorithms do not rely on knowledge of the statistics of y_n and φ_n. If the observation sequence $\{(y_n, \varphi_n)\}$ is stationary and independent of ζ_n and

$$E\varphi_n\varphi_n' = R > 0, \;\; E\varphi_n y_n = q,$$

where $R > 0$ is meant to be a symmetric positive definite matrix [13], then θ_*, the minimizer of $L(\theta) = E|y_n - \varphi_n'\theta|^2$, is, in fact, the unique solution to the Wiener–Hopf equation $R\theta_* = q$ or $\theta_* = R^{-1}q$. In the following subsection, we review some of the relevant results for these algorithms. The interested reader may consult any number of excellent expositions, such as [1, 3, 10, 21, 22, 23, 26, 27, 36, 38, 40, 41], for more information.

Departing from the traditional AF problem setting, we are interested in systems with model mismatch

$$y_n = \varphi_n'\theta_* + \widetilde{\Delta}_n(\theta_*, \varphi_n) + \zeta_n, \tag{2.4}$$

where $\widetilde{\Delta}_n(\theta_*, \varphi_n)$ represents the model mismatch term. Model mismatch is a distinct type of uncertainty from typical consideration of additive measurement noise and reduced-order model approximation. Many practical systems are nonlinear and infinite dimensional, and the finite dimensional linear structure (2.1) is an approximation. In the literature of system identification [32, 33, 34], a system is often parameterized as

$$y_n = \varphi_n'\theta_* + \widetilde{\varphi}_n'\widetilde{\theta}_* + \zeta_n,$$

where the regressor $\widetilde{\varphi}_n'$ represents the signals of remote past and $\widetilde{\theta}_*$ are different parameters from θ_*, which is unknown but known only to be bounded. That is, for infinite dimensional linear systems, their finite dimensional representation introduces a specific type of uncertainty. Such uncertainty in the literature of system identification is referred to as unmodeled dynamics. This form of unmodeled dynamics is linear and multiplicative (namely, its size is proportional to the input magnitude) uncertainty. In this paper, the model mismatch on which we concentrate can be a nonlinear function on the modeled parameter θ_* and recent input. While unmodeled linear dynamics have been extensively treated in the literature, studies of impact of model mismatch, especially nonlinear mismatch, on adaptive filters are relatively new and will be the focus of this paper.

Now suppose the true system is really given by (2.4), but we actually used $y_n = \varphi_n'\theta_* + \zeta_n$ in the formulation instead. Then an error is introduced. The question is this: how much will this mismodeling error impact the estimation? Such a question is an important issue in the robustness consideration. The desired results give us a sense of insurance on the tolerance level we can endure. Understanding this robustness is our main concern in this paper.

Example 2.1. Consider a nonlinear system $y_n = (u_n + \lambda u_n^2)\theta_* + \zeta_n = u_n\theta_* + \lambda u_n^2\theta_* + \zeta_n$, where λ is a small positive constant. Here u_n is the input and $\widetilde{\Delta}_n = \lambda u_n^2\theta_*$ is the model mismatch term. It is noted that this term depends on the input and the unknown parameter. In this case, $\varphi_n = u_n$.

It is tempting to include the model mismatch $\widetilde{\Delta}_n$ in the disturbance term ζ_n and to treat them as a new combined disturbance term. However, unlike ζ_n, $\widetilde{\Delta}_n$ is usually a function of θ_* and φ_n and cannot be reduced by averaging. As a result, it can potentially cause estimation bias. Thus, separating these two sources of uncertainties is crucial to understanding the impact of model mismatch on the adaptive filter, which is the focus of this paper. We first summarize some basic results and approaches that we will extend to analyze the model mismatch problem.

2.2. A brief review of existing results. Recursive algorithms of AF type are dynamical systems, whose asymptotic behavior is of key interest. To study the asymptotic properties of (2.3), we use weak convergence methods (see, e.g., [22]). We assume that the observation y_n is given by (2.3), where $\{\zeta_n\}$ is a stationary sequence with 0 mean. Furthermore, we assume that the following conditions hold.

(A1) The sequence $\{\varphi_n'\varphi_n\}$ and $\{\varphi_n\zeta_n\}$ is stationary and uniformly ϕ-mixing (see [2, p. 166] as well as [4]) such that
- (a) $E\varphi_n\varphi_n' = R$ a symmetric positive definite matrix;
- (b) $E\varphi_n\zeta_n = 0$;
- (c) the mixing measure of the above processes, $\widetilde{\psi}_n$, satisfies $\sum_n \widetilde{\psi}_n^{1/2} < \infty$.

Note that the term "uniform mixing" is taken from [4]. Such a process is simply called ϕ-mixing in [2]. The essence of a mixing process is that the remote past and distant future of the process are asymptotically independent. This property also requires the signals be bounded. We address the issue of unbounded signals in the last section of the paper.

By (2.3),

$$\theta_{n+1} = \theta_n + \varepsilon\varphi_n\zeta_n - \varepsilon\varphi_n\varphi_n'(\theta_n - \theta_*).$$

By virtue of (A1), the sequences $\{\varphi_n\varphi_n'\}$ and $\{\varphi_n\zeta_n\}$ are strongly ergodic. It follows that, for each $m \geq 0$,

$$(2.5)\quad \frac{1}{n}\sum_{j=m}^{n+m-1}\varphi_j\varphi_j' \to R, \quad \frac{1}{n}\sum_{j=m}^{n+m-1}\varphi_j\zeta_j \to 0 \text{ with probability one (w.p.1).}$$

In fact, for the weak convergence analysis, we need only the limits in (2.5) to hold in the sense of convergence in probability. In addition, the mixing condition implies that

$$(2.6)\qquad \sqrt{\varepsilon}\sum_{j=0}^{\lfloor t/\varepsilon\rfloor-1}\varphi_j\zeta_j \text{ converges weakly to } \sigma w(t),$$

where

$$(2.7)\qquad \sigma\sigma' = E\varsigma_0\varsigma_0' + \sum_{j=1}^{\infty}E\varsigma_j\varsigma_0' + \sum_{j=1}^{\infty}E\varsigma_0\varsigma_j', \quad \varsigma_j = \phi_j\zeta_j.$$

Here $\lfloor z \rfloor$ denotes the integer part of z, and $w(\cdot)$ is a standard Brownian motion. To apply the weak convergence theory of [22], we extend the estimator sequence to continuous time by taking a piecewise constant interpolation, defined by

$$\theta^\varepsilon(t) = \theta_n \quad \text{for} \quad t \in [\varepsilon n, \varepsilon n + \varepsilon). \tag{2.8}$$

We suppose, for simplicity, the initial iterate θ_0 is independent of ε. Below we give some results for this AF algorithm. The proof can be constructed using the idea in [21]; see also the general results on stochastic approximation [22].

First, under (A1), $\theta^\varepsilon(\cdot)$ converges weakly to $\theta(\cdot)$, which is a solution of the differential equation

$$\dot{\theta} = -R\theta + q, \quad \theta(0) = \theta_0, \tag{2.9}$$

where $q = R\theta_*$.

Note that the significance of (2.9) is that its stationary point θ_* is precisely the quantity we seek: the limiting process is deterministic, and it converges in time to the desired "true" parameter. It is also interesting to note that the long time nature of this "two-layer" asymptotic result can be reduced. That is, let t_ε be a sequence such that $t_\varepsilon \to \infty$ as $\varepsilon \to 0$. Then it can be shown that $\theta^\varepsilon(\cdot + t_\varepsilon)$ converges weakly to θ_* (or in probability) as $\varepsilon \to 0$.

Next we define $X_n = (\theta_n - \theta_*)/\sqrt{\varepsilon}$. We can further show that there is an N_ε such that $\{X_n : n \geq N_\varepsilon\}$ is tight. Similar to $\theta^\varepsilon(\cdot)$, define

$$X^\varepsilon(t) = X_n \quad \text{for} \quad t \in [\varepsilon(n - N_\varepsilon), \varepsilon(n + 1 - N_\varepsilon)). \tag{2.10}$$

Then we can show that $X^\varepsilon(\cdot)$ converges weakly to $X(\cdot)$, which is a solution of the SDE

$$dX = -RXdt + \sigma dw, \tag{2.11}$$

where $\sigma\sigma'$ is symmetric and nonnegative definite and $w(\cdot)$ is a standard Brownian motion. The first result (2.9) is consistency, while this second result (2.11) is a form of asymptotic normality.

This paper differs from the existing references in several aspects. This work focuses on weak stability (recurrence, positive recurrence, and ergodicity; see [37]) and stability in distribution. In accordance with [24], a deterministic system $\dot{x} = g(t, x)$, which satisfies appropriate conditions, is Lagrange stable if the solutions are ultimately uniformly bounded. When stochastic systems (such as diffusions) are considered, almost sure boundedness excludes many systems. Thus, in lieu of such boundedness, one seeks stability in a certain weak sense [37]. In this paper we will establish general criteria for positive recurrence and weak stability by employing Lyapunov function and perturbed Lyapunov function methods.

3. Consistency and asymptotic normality under model mismatch. The AF algorithm of (2.3) is designed for the specific dynamics of (2.1). In this section, we seek extensions of the standard results of the previous section to the model mismatch case. To begin, we discuss the nature of mismatches we will consider.

We assume that the model mismatch is represented by an error on the output observations

$$y_n = \varphi_n' \theta_* + \widetilde{\Delta}_n + \zeta_n, \tag{3.1}$$

where $\widetilde{\Delta}_n$ represents the model mismatch and φ_n and ζ_n are as in the previous section satisfying (A1). The recursive AF algorithm remains the same as (2.3). For convergence analysis, we rewrite the algorithm as

$$\theta_{n+1} = \theta_n + \varepsilon\varphi_n(\varphi_n'\theta_* - \varphi_n'\theta_n + \zeta_n) + \varepsilon\Delta_n. \tag{3.2}$$

The last term in (3.2) is the model mismatch with $\Delta_n = \varphi_n\widetilde{\Delta}_n$.

Remark 3.1. Note that the model mismatch Δ_n (respectively, $\widetilde{\Delta}_n$) depends on θ_* as well as φ_n. Since Δ_n is unknown, it cannot be used for parameter updating. That is, (2.3) is still the algorithm we use: parameter estimates are computed directly from the observed data. However, for convergence analysis, the effect of Δ_n must be included, and hence (3.2) is a suitable form to have an explicit term on model mismatch.

(A2) The model mismatch Δ_n is parameterized by a smooth function of the parameter and regressor $\Delta_n = \widetilde{f}(\theta_*, \varphi_n)$ so that there is a smooth function $f(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^r$ satisfying for each positive integer m, and we have

$$\frac{1}{n}\sum_{j=m}^{n+m-1} E_m\widetilde{f}(\theta_*, \varphi_j) \to f(\theta_*) \text{ in probability as } n \to \infty,$$

where E_m denotes the conditional expectation with respect to the σ-algebra generated by $\{\varphi_j, \zeta_j : j \le m\}$.

Under conditions (A1) and (A2), define $\theta^\varepsilon(t) = \theta_n$ for $t \in [n\varepsilon, n\varepsilon + \varepsilon)$ with θ_n given by (3.2). We can show by the weak convergence analysis (see [22, Chapter 8]) that $\theta^\varepsilon(\cdot)$ converges weakly to $\theta(\cdot)$, which can be characterized by the solution of the ODE

$$\dot{\theta}(t) = -R\theta(t) + f(\theta_*) + q, \tag{3.3}$$

where $q = R\theta_*$. The convergence to the ODE limit (3.3) is valid when $\varepsilon \to 0$, $n \to \infty$, but εn remains to be bounded. Because the dynamic system (3.3) is linear in θ, it has a unique equilibrium point θ_b,

$$-R\theta_b + f(\theta_*) + q = 0 \quad \text{or} \quad \theta_b = R^{-1}(f(\theta_*) + q). \tag{3.4}$$

As a result, the identification bias due to the model mismatch is $R^{-1}f(\theta_*)$.

Example 3.2. Consider the system in Example 2.1. Here $\widetilde{\Delta}_n = \lambda u_n^2\theta_*$ and $\varphi_n = u_n$. Then $\widetilde{f}(\theta_n, \varphi_n) = u_n\widetilde{\Delta}_n = \lambda u_n^3\theta_*$. Suppose that u_n and ζ_n are stationary and that $Eu_n^2 = R$, $Eu_n^3 = \delta$, $Ey_nu_n = q_1$, $Eu_n\zeta_n = 0$. Then $f(\theta_*) = \lambda\delta\theta_*$. Using (3.2) with the above specifications, we obtain the limit ODE given by

$$\dot{\theta}(t) = -R\theta(t) + R\theta_* + \lambda\delta\theta_*.$$

Thus the θ_b is given by $\theta_b = \theta_* + R^{-1}\lambda\delta\theta_*$. Compared to the result without model mismatch, the model mismatch causes a bias in estimation, and the bias is given by $R^{-1}\lambda\delta\theta_*$.

The behavior of (3.2) when $\varepsilon \to 0$ and $\varepsilon n \to \infty$ is of considerable additional interest. The next claim establishes such a result.

THEOREM 3.3. *Suppose that* (A1) *and* (A2) *hold. Then, for any sequence* $t_\varepsilon \to \infty$ *as* $\varepsilon \to 0$, $\theta^\varepsilon(t_\varepsilon + \cdot)$ *converges weakly to* θ_b.

Proof. For any $T < \infty$, define $\Theta^\varepsilon(t) = \theta^\varepsilon(t_\varepsilon + \cdot)$ and $\Theta_T^\varepsilon(t) = \theta^\varepsilon(t_\varepsilon + \cdot - T)$. Consider the process $\{\Theta^\varepsilon(\cdot), \Theta_T^\varepsilon(\cdot)\}$, and denote the limit by $(\Theta(\cdot), \Theta_T(\cdot))$. It is easily seen that $\Theta(0) = \Theta_T(T)$. We then have

$$\begin{aligned}\Theta_T(T) &= \exp(-RT)\Theta_T(0) + \int_0^T \exp\{-R(T-t)\}(f(\theta_*)+q)dt \\ &\to R^{-1}(f(\theta_*)+q) = \theta_b \text{ as } T\to\infty,\end{aligned}$$

where θ_b is defined in (3.4). □

Example 3.4. Consider the system in Example 3.2 with $\lambda = 0.05$. The true parameter $\theta_* = 10$. The input u_n is an independently and identically distributed (i.i.d.) uniformly distributed process on $[0,6]$. Then $R = Eu_1^2 = 12$, $q = R\theta_* = 120$, $\delta = Eu_1^3 = 54$, and $f(\theta_*) = \lambda\delta\theta_* = 27$. The model mismatch causes a bias in estimation, and the bias is given theoretically by $R^{-1}\lambda\delta\theta_* = 27/12 = 2.25$.

We now run the recursive algorithm (2.3). The noise $\{\zeta_n\}$ is an i.i.d. sequence of normal random variables with mean 0 and variance $\sigma^2 = 25$. The initial estimate $\theta_1 = 40$.

Case 1. *εn is bounded.* Suppose that the stepsize ε is selected as $\varepsilon = 1/n$, where n is the total data window for simulation. For the cases of $n_1 = 50, \varepsilon_1 = 1/50$; $n_2 = 100, \varepsilon_2 = 1/100$; $n_3 = 200, \varepsilon_3 = 1/200$; and $n_4 = 500, \varepsilon_4 = 1/500$, the simulation results are shown in Figure 1. The final values of the estimates are 12.087 (bias 2.087), 12.357 (bias 2.357), 12.143 (bias 2.143), and 12.446 (bias 2.446) for the four data windows, all with an estimation bias close to the theoretical value 2.25.

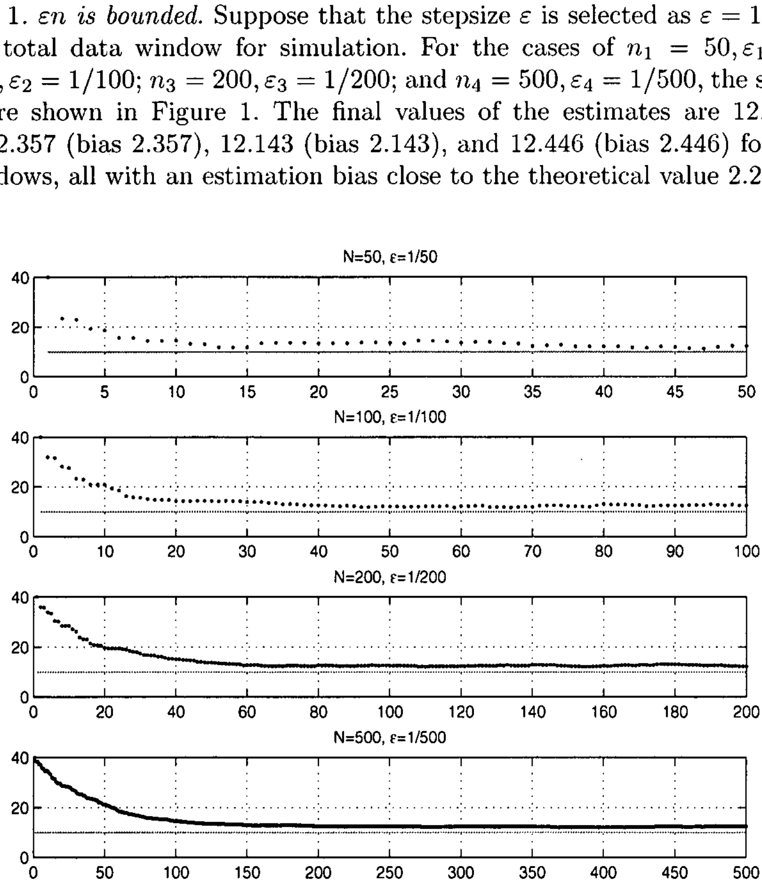

FIG. 1. *Estimation results under recursive algorithms with bounded εn.*

Case 2. *εn is unbounded.* Suppose that the stepsize ε is selected as $\varepsilon = 1/n^{0.8}$, where N is the total data window for simulation. For the cases of $n_1 = 50, \varepsilon_1 = 1/50^{0.8}$; $n_2 = 100, \varepsilon_2 = 1/100^{0.8}$; $n_3 = 200, \varepsilon_3 = 1/200^{0.8}$; and $n_4 = 500, \varepsilon_4 = 1/500^{0.8}$, the simulation results are shown in Figure 2. The final values of the estimates are 10.941

(bias 0.941), 11.371 (bias 1.371), 12.050 (bias 2.05), and 12.293 (bias 2.293) for the four data windows. Since the stepsize is larger than in Case 1, the convergence speeds are lower here. As a result, $n_4 = 500$ represents an asymptotical bias of 2.293 which is close to the theoretical value 2.25.

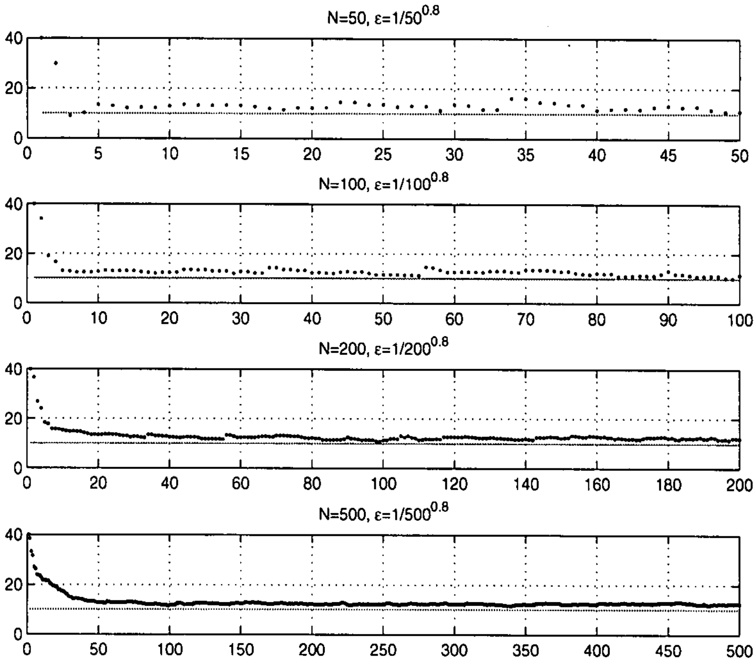

FIG. 2. *Estimation results under recursive algorithms with unbounded εn.*

To proceed, define

$$\pi_n = \varphi_n \zeta_n + (\varphi_n \varphi_n' - R)(\theta_* - \theta_b) + [f(\theta_*, \varphi_n) - f(\theta_*)]. \tag{3.5}$$

The mixing assumption implies that the $\{\pi_n\}$ is a mixing process. Define

$$\widetilde{\pi}_n = \begin{pmatrix} \pi_n^1 \\ \pi_n^2 \\ \pi_n^3 \end{pmatrix} \in \mathbb{R}^{3r \times 1},$$

where

$$\begin{aligned} \pi_n^1 &= \varphi_n \zeta_n, \\ \pi_n^2 &= (\varphi_n \varphi_n' - R)(\theta_* - \theta_b), \\ \pi_n^3 &= f(\theta_*, \varphi_n) - f(\theta_*). \end{aligned}$$

That is, we pile up the three r-dimensional random vectors π_n^1, π_n^2, and π_n^3, respectively. The three vectors represent the random mixing sequence due to noise, bias, and mismatch, respectively. Then the well-known functional central limit theorem enables us to obtain that as $\varepsilon \to 0$, $\sqrt{\varepsilon} \sum_{j=0}^{\lfloor t/\varepsilon \rfloor - 1} \widetilde{\pi}_j$ converges weakly to a $3r$-dimensional Brownian motion with covariance matrix

$$\widetilde{\Sigma} = E\widetilde{\pi}_0 \widetilde{\pi}_0' + \sum_{j=1}^{\infty} E\widetilde{\pi}_j \widetilde{\pi}_0' + \sum_{j=1}^{\infty} E\widetilde{\pi}_0 \widetilde{\pi}_j'.$$

Using the notation of partitioned vectors, we can write π_n as

$$\pi_n = (I_r, I_r, I_r)\begin{pmatrix} \pi_n^1 \\ \pi_n^2 \\ \pi_n^3 \end{pmatrix},$$

where I_r is an r-dimensional identity matrix. Using the weak convergence to the Brownian motion mentioned above, together with Slutsky's theorem, we arrive at the following lemma.

LEMMA 3.5. *$\sqrt{\varepsilon}\sum_{j=0}^{t/\varepsilon-1}\pi_j$ converges weakly to a Brownian motion $\widetilde{w}(\cdot)$ with covariance $\widehat{\Sigma}t$, where*

$$\widehat{\Sigma} = (I_r, I_r, I_r)\widetilde{\Sigma}\begin{pmatrix} I_r \\ I_r \\ I_r \end{pmatrix}.$$

Equivalently, we can write $\widetilde{w}(t)$ as $\widehat{\Sigma}^{1/2}w(t)$, where $w(\cdot)$ is a standard Brownian motion. For simplicity of exposition, we assume the tightness of the sequence $\{(\theta_n - \theta_b)/\sqrt{\varepsilon} : n \geq N_\varepsilon\}$ in the next result. This tightness will be established in Theorem 4.2.

THEOREM 3.6. *Assume conditions* (A1)–(A2) *hold. Define $X_n = (\theta_n - \theta_b)/\sqrt{\varepsilon}$, where the iteration* (3.2) *defines θ_n. Assume that there is an $N_\varepsilon > 0$ such that $\{X_n : n \geq N_\varepsilon\}$ is tight. Define $X^\varepsilon(\cdot)$ by $X^\varepsilon(t) = X_n$ for $t \in [\varepsilon(n - N_\varepsilon), \varepsilon(n - N_\varepsilon + 1))$. Then $X^\varepsilon(\cdot)$ converges weakly to $X(\cdot)$, which is the solution of*

$$dX = -RX\,dt + \widehat{\Sigma}^{1/2}dw. \tag{3.6}$$

Proof. Then

$$X_{n+1} = X_n - \varepsilon\varphi_n\varphi_n' X_n + \sqrt{\varepsilon}\pi_n. \tag{3.7}$$

It follows that

$$X^\varepsilon(t+s) - X^\varepsilon(t) = -\varepsilon\sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1}\varphi_j\varphi_j' X_j + \sqrt{\varepsilon}\sum_{j=t/\varepsilon}^{(t+s)/\varepsilon-1}\pi_j. \tag{3.8}$$

In the above and what follows, for example, t/ε is meant to be $\lfloor t/\varepsilon \rfloor$, the integer part of t/ε, similarly for $(t+s)/\varepsilon$. However, for notational simplicity, we will not use the floor function notation henceforth. The rest of the argument can be done as in [22, Chapter 10]. The main idea is that by the martingale averaging approach, we can show that the first term on the right of the equality sign of (3.8) tends to $-\int_t^{t+s} RX(\widetilde{s})d\widetilde{s}$ and the second term goes to $\int_t^{t+s}\widehat{\Sigma}^{1/2}dw(\widetilde{s})$. A few details are omitted for brevity. $\square$

Define $\Sigma^* = R^{-1}\widehat{\Sigma}R^{-1}$, which can be considered as an optimal one in the sense discussed in [22, Chapter 11]. To illustrate the optimality, we may consider a sequence of decreasing stepsizes $\{\varepsilon_n\}$ and the associated algorithm

$$\theta_{n+1} = \theta_n + \varepsilon_n\varphi_n(\varphi_n'\theta_* - \varphi_n'\theta_n + \zeta_n) + \varepsilon_n\Delta_n.$$

Taking $\varepsilon_n = O(1/n^\gamma)$, it can be shown [42] that $n^{\gamma/2}(\theta_n - \theta_b)$ is asymptotically normal. The scaling factor $n^{\gamma/2}$, together with the associated covariance, give us the rate of convergence. Among the different $\varepsilon_n = O(1/n^\gamma)$, $\gamma = 1$ gives us the best scaling. Next we consider

$$\theta_{n+1} = \theta_n + \frac{\Gamma}{n}\varphi_n(\varphi_n'\theta_* - \varphi_n'\theta_n + \zeta_n) + \frac{\Gamma}{n}\Delta_n.$$

Treating Γ as a matrix-valued parameter, we may carry out an optimization task to find the optimal covariance Σ^* and show that $n^{1/2}(\theta_n - \theta_b)$ is asymptotically normal with asymptotic covariance $R^{-1}\widehat{\Sigma}R^{-1}$.

As was observed, in addition to the noise effect because of $\varphi_n\zeta_n$, the diffusion term $\widehat{\Sigma}$ is caused by the bias $(\varphi_n\varphi_n' - R)(\theta_* - \theta_b)$ and the model mismatch $f(\theta_*, \varphi_n) - f(\theta_*)$. A natural question is: can we reduce the variation due to these factors? An effective way of resolving the problem is to take an iterate averaging as in [22, Chapter 11] with a minimal window of averaging. To proceed, let m_ε be a sequence of integers satisfying $m_\varepsilon \to \infty$ as $\varepsilon \to 0$. Define

$$\begin{aligned} \overline{\theta}_n &= \sum_{j=m_\varepsilon}^{m_\varepsilon+n-1} \theta_j, \\ \overline{X}^\varepsilon(t) &= \frac{\sqrt{\varepsilon}}{t} \sum_{j=m_\varepsilon}^{m_\varepsilon+(t/\varepsilon)-1} (\theta_j - \theta_b). \end{aligned} \tag{3.9}$$

LEMMA 3.7. *For each fixed t, as $\varepsilon \to 0$, $\overline{X}^\varepsilon(t)$ converges in distribution to $N(0, \Sigma^*/t + O(1/t^2))$, a normal random variable with mean 0 and covariance $\Sigma^*/t + O(1/t^2)$, where $\Sigma^* = R^{-1}\widehat{\Sigma}R^{-1}$.*

Proof. The proof is essentially in [22, section 11.1.2]. We omit the details for brevity. □

As in the case of a typical estimation problem, the quality of approximation is judged by precision (bias) and variability (variance). By means of (3.3), if we fix the parametrization, there will be a term contributing to the bias and a term contributing to the model mismatch. From the definition of π_n, in evaluating the quality of approximation, we are able to reduce the variation or variance. An effective way is to use an iterate averaging; see [22, Chapter 11]. Lemma 3.7 says essentially that. As a consequence, we have $\overline{\theta}_n \sim N(\theta_b, \Sigma^*/(t/\varepsilon))$. That is, $\overline{\theta}_n$ is distributed asymptotically normally with a mean θ_b and a covariance Σ^*/number of iterates with the window of averaging.

4. Weak stability. We now turn to the results of interest: the stability of the estimation process with respect to perturbation. In this section, we show that the scaled sequence of errors is recurrent or weakly stable. To proceed, let us first recall the definition of recurrence for diffusion process. For simplicity, we deal only with time-homogeneous processes. Consider a diffusion process $Y(\cdot)$ given by

$$dY(t) = g_1(Y(t))dt + g_2(Y(t))dw, \tag{4.1}$$

where $w(\cdot)$ is an $\mathbb{R}^r$-valued standard Brownian motion and $g_1(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^r$ and $g_2(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^{r\times r}$ are functions satisfying suitable conditions. Denote by $Y^y(\cdot)$ the solution satisfying initial condition $Y(0) = y$. The process $Y^y(t)$ is said to be *regular* if, for any $0 < T < \infty$,

$$P\{\sup_{0\le t\le T} |Y^y(t)| = \infty\} = 0. \tag{4.2}$$

That is, a process is regular if it has no finite explosion time. Let β_n be the first exit time of the process $Y^y(t)$ from the bounded set $\{\widetilde{y} : |\widetilde{y}| < n\} \times \mathcal{M}$, that is,

$$\beta_n = \inf\{t : |Y^y(t)| = n\}. \tag{4.3}$$

Then the sequence $\{\beta_n\}$ is monotonically increasing and hence has a (finite or infinite) limit. It is not difficult to see that the process $Y^y(t)$ is regular if and only if

$$\beta_n \to \infty \text{ w.p.1 as } n \to \infty. \tag{4.4}$$

Let U be an open subset of $\mathbb{R}^r$ with compact closure, and let τ_U^y be the first hitting time of the set U. That is, $\tau_U^y = \inf\{t : Y^y(t) \in U\}$. A regular process $Y^y(\cdot)$ is *recurrent* with respect to U if $P(\tau_U^y < \infty) = 1$ for any $y \in U^c$, where U^c denotes the complement of U. A recurrent process with finite mean recurrence time for some set U ($U \subset \mathbb{R}^r$ being a bounded open set with compact closure) is said to be *positive recurrent* with respect to U; otherwise, the process is *null recurrent* with respect to U. Thus, recurrence means that if the process starts somewhere outside an open set, it will return to this set. If not only the process returns to the compact set but also the expected return time is finite, the process is positive recurrent. As demonstrated in [14], the property of recurrence, in fact, is independent of the open set U chosen. That is, if $Y^y(\cdot)$ is recurrent (respectively, positive recurrent) with respect to U, then, for any $\widetilde{U}$ (open subset of $\mathbb{R}^r$ with compact closure), $Y^y(\cdot)$ is recurrent (respectively, positive recurrent) with respect to $\widetilde{U}$. Thus, from now on, we will only say a process $Y^y(\cdot)$ is recurrent or positive recurrent without specifying the set chosen. In what follows, we first establish recurrence of the limit SDE and then use such recurrence to study the recurrence of the algorithm.

4.1. Regularity and recurrence of associated SDE. Associated with the SDE (3.6), there is a differential operator, known as the generator of the diffusion process, given by

$$\mathcal{L}g(x) = \frac{1}{2}\text{tr}\left(\widehat{\Sigma}\nabla^2 g(x)\right) - \nabla g'(x)Rx \tag{4.5}$$

for any real-valued function $g(\cdot)$ that is twice continuously differentiable with respect to x, where $\widehat{\Sigma}_{ij}$ and $-(Rx)_i$ denote the ijth entry and ith component, respectively, of the matrix $\sigma\sigma'$ and the vector $-Rx$. The following lemma is a simple application of the general results for diffusion processes in [14].

THEOREM 4.1. *Assume that* (A1) *and* (A2) *hold. Consider the diffusion given in* (3.6). *The following assertions hold:*

(a) *The process $X^x(\cdot)$ is regular.*

(b) *The process $X^x(\cdot)$ is positive recurrent.*

Proof. To prove (a), we first note that since the SDE is linear, the Lipschitz condition and the linear growth condition are automatically satisfied. In accordance with [14], we need to consider only a function $V : \mathbb{R}^r \mapsto \mathbb{R}$ $V(x) = |x|$. Note that $V(\cdot)$ is twice continuously differentiable in $\mathbb{R}^r - N_{\delta_0}(0)$, where $N_{\delta_0}(0)$ is a deleted neighborhood of the origin. On this set,

$$\frac{\partial V(x)}{\partial x_i} = \frac{x_i}{|x|}, \; \frac{\partial^2 V(x)}{\partial x_i x_j} = \frac{\delta_{ij}|x| - \frac{x_i x_j}{|x|}}{|x|^2}.$$

Then it is fairly easily to show that, for some $c > 0$,

$$\begin{aligned} &\mathcal{L}V(y) \le cV(y) \\ &\inf_{|y|>K_0} V(y) \to \infty \quad \text{as} \quad K_0 \to \infty. \end{aligned} \tag{4.6}$$

The general criterion for regularity of diffusion processes [14] then yields the desired regularity.

Define the Lyapunov function

$$V(x) = \frac{1}{2}x'x. \tag{4.7}$$

This quadratic function is radially unbounded, and condition (4.6) is satisfied. We omit the details here since a finer estimate similar to this will be used in the proof of part (b).

To prove (b), we again use the Lyapunov function given by (4.7). It is readily checked that

$$V_x(x) = x, \quad V_{xx}(x) = I. \tag{4.8}$$

Thus,

$$\mathcal{L}V(x) = -x'Rx + \frac{1}{2}\text{tr}\left(\widehat{\Sigma}\right). \tag{4.9}$$

By noting the positive definiteness of R,

$$-x'Rx \le -\lambda_{\min}(R)|x|^2. \tag{4.10}$$

Then for all $x \in \{z : |z|^2 > (\lambda_{\max}(\widehat{\Sigma})/\lambda_{\min}(R))\}$,

$$\mathcal{L}V(x) \le \left(-\lambda_{\min}(R)|x|^2 + \frac{1}{2}\text{tr}\left(\widehat{\Sigma}\right)\right) < -\alpha \tag{4.11}$$

for some $\alpha > 0$. The desired positive recurrence follows. □

4.2. Regularity and recurrence of the adaptive algorithm. In this section, we study the recurrence of the iterates given by (3.2). In a way, this presents an effort to assess properties of the approximation errors as a centered and scaled sequence. We define the δ neighborhood of θ_0 by $N_\delta(\theta_0) = \{\theta : |\theta - \theta_0| < \delta\}$. Denoting $\widetilde{\theta}_n = \theta_n - \theta_b$, we have the following theorem.

THEOREM 4.2. *Suppose that* (A1)–(A2) *hold. Then*

$$EV(\widetilde{\theta}_{n+1}) = O(\varepsilon) \quad \textit{for} \quad n \quad \textit{sufficiently large.}$$

Proof. We first note that

$$\begin{aligned} \widetilde{\theta}_{n+1} &= \widetilde{\theta}_n - \varepsilon R\widetilde{\theta}_n + \varepsilon\pi_n \\ &= \widetilde{\theta}_n - \varepsilon R\widetilde{\theta}_n + \varepsilon\{\varphi_n\zeta_n + (\varphi_n\varphi_n' - R)(\theta_* - \theta_b) + [f(\theta_*, \varphi_n) - f(\theta_*)]\}. \end{aligned} \tag{4.12}$$

To proceed, use $\mathcal{F}_n$ to denote the σ-algebra generated by $\{\varphi_j, \zeta_j : j < n\}$, and denote the conditional expectation with respect to $\mathcal{F}_n$ by E_n. Using $V(\widetilde{\theta}) = \widetilde{\theta}'\widetilde{\theta}/2$, straightforward calculation yields

$$\begin{aligned} &E_nV(\widetilde{\theta}_{n+1}) - V(\widetilde{\theta}_n) \\ &\quad = -\varepsilon\widetilde{\theta}_n'R\widetilde{\theta}_n + \varepsilon\widetilde{\theta}_n'\pi_n + \varepsilon^2|-R\widetilde{\theta}_n + \pi_n|^2 \\ &\quad = -\varepsilon\widetilde{\theta}_n'R\widetilde{\theta}_n + \varepsilon\widetilde{\theta}_n'\pi_n + O(\varepsilon^2)(V(\widetilde{\theta}_n) + 1)). \end{aligned} \tag{4.13}$$

To complete the proof, we need to get order of magnitude estimates of the terms above. Toward that end, we introduce some perturbations, a process which will lead to the result of interest. Define

$$\begin{aligned} V_1^\varepsilon(\widetilde{\theta},n) &= \varepsilon \sum_{j=n}^{\infty} E_n \widetilde{\theta}' \varphi_j \zeta_j, \\ V_2^\varepsilon(\widetilde{\theta},n) &= \varepsilon \sum_{j=n}^{\infty} E_n \widetilde{\theta}' (\varphi_j \varphi_j' - R)(\theta_* - \theta_b), \\ V_3^\varepsilon(\widetilde{\theta},n) &= \varepsilon \sum_{j=n}^{\infty} E_n \widetilde{\theta}' [\widetilde{f}(\theta_*, \varphi_j) - f(\theta_*)]. \end{aligned} \tag{4.14}$$

By virtue of the well-known mixing inequalities (see [2, p. 166]) and the familiar inequality $2ab \le (a^2 + b^2)$ for any two real numbers a and b, we see that there is a constant $K > 0$ such that

$$\begin{aligned} |V_1^\varepsilon(\widetilde{\theta},n)| &\le \varepsilon l \sum_{j=n}^{\infty} E_n \widetilde{\theta}' \varphi_j \zeta_j \Big| \le \varepsilon \sum_{j=n}^{\infty} |\widetilde{\theta}| |E_n \varphi_j \zeta_j| \\ &\le K\varepsilon(|\widetilde{\theta}|^2 + 1) \le K\varepsilon(V(\widetilde{\theta}) + 1). \end{aligned} \tag{4.15}$$

Likewise, it is easily verified that

$$\begin{aligned} |V_2^\varepsilon(\widetilde{\theta},n)| &\le K\varepsilon(1 + V(\widetilde{\theta})), \\ |V_3^\varepsilon(\widetilde{\theta},n)| &\le K\varepsilon(1 + V(\widetilde{\theta})). \end{aligned} \tag{4.16}$$

Next, note that

$$\begin{aligned} & E_n V_1^\varepsilon(\widetilde{\theta}_{n+1}, n+1) - V_1^\varepsilon(\widetilde{\theta}_n, n) \\ &= E_n V^\varepsilon(\widetilde{\theta}_{n+1}, n+1) - E_n V_1^\varepsilon(\widetilde{\theta}_n, n+1) + E_n V_1^\varepsilon(\widetilde{\theta}_n, n+1) - V_1^\varepsilon(\widetilde{\theta}_n, n) \\ &= \varepsilon \sum_{j=n+1}^{\infty} (\widetilde{\theta}_{n+1} - \widetilde{\theta}_n)' \varphi_j \zeta_j - \varepsilon \widetilde{\theta}_n' \varphi_n \zeta_n \\ &= O(\varepsilon^2)(V(\widetilde{\theta}_n) + 1) - \varepsilon \widetilde{\theta}_n' \varphi_n \zeta_n. \end{aligned} \tag{4.17}$$

Similarly, we obtain

$$\begin{aligned} & E_n V_2^\varepsilon(\widetilde{\theta}_{n+1}, n+1) - V_2^\varepsilon(\widetilde{\theta}_n, n) \\ &= E_n V_2^\varepsilon(\widetilde{\theta}_{n+1}, n+1) - E_n V_2^\varepsilon(\widetilde{\theta}_n, n+1) + E_n V_2^\varepsilon(\widetilde{\theta}_n, n+1) - V_2^\varepsilon(\widetilde{\theta}_n, n) \\ &= O(\varepsilon^2)(V(\widetilde{\theta}_n) + 1) - \varepsilon \widetilde{\theta}_n' (\varphi_n \varphi_n' - R)(\theta_* - \theta_b) \end{aligned} \tag{4.18}$$

and

$$\begin{aligned} & E_n V_3^\varepsilon(\widetilde{\theta}_{n+1}, n+1) - V_3^\varepsilon(\widetilde{\theta}_n, n) \\ &\quad = O(\varepsilon^2)(V(\widetilde{\theta}_n) + 1) - \varepsilon \widetilde{\theta}_n' (\widetilde{f}(\theta_*, \varphi_n) - f(\theta_*)). \end{aligned} \tag{4.19}$$

Define

$$V^\varepsilon(\widetilde{\theta}_n, n) = V(\widetilde{\theta}_n) + V_1^\varepsilon(\widetilde{\theta}_n, n) + V_2^\varepsilon(\widetilde{\theta}_n, n) + V_3^\varepsilon(\widetilde{\theta}_n, n). \tag{4.20}$$

It is easily seen that

$$V^\varepsilon(\widetilde{\theta}_n, n) = V(\widetilde{\theta}_n) + O(\varepsilon)(1 + V(\widetilde{\theta}_n)).$$

Then upon cancelation, we obtain

$$\begin{aligned} E_n V^\varepsilon(\widetilde{\theta}_{n+1}, n+1) - V^\varepsilon(\widetilde{\theta}_n, n) \\ \leq -\varepsilon\widehat{\lambda} V(\widetilde{\theta}_n) + O(\varepsilon^2)(V(\widetilde{\theta}_n) + 1). \end{aligned} \tag{4.21}$$

Note that there is a $\kappa_1 > 1$ such that we can make

$$-\varepsilon\widehat{\lambda} V(\widetilde{\theta}_n) + O(\varepsilon^2) V(\widetilde{\theta}_n) \leq -\kappa_1 \widehat{\lambda}\varepsilon V(\widetilde{\theta}_n).$$

Furthermore, using the bounds in (4.15) and (4.16), we obtain

$$\begin{aligned} E_n V^\varepsilon(\widetilde{\theta}_{n+1}, n+1) - V^\varepsilon(\widetilde{\theta}_n, n) \\ \leq -\kappa_1 \widehat{\lambda}\varepsilon V^\varepsilon(\widetilde{\theta}_n, n) + O(\varepsilon^2)(1 + V(\widetilde{\theta}_n)). \end{aligned} \tag{4.22}$$

Iterating on (4.22), we finally arrive at

$$EV^\varepsilon(\widetilde{\theta}_{n+1}, n+1) \leq (1 - \kappa_1\widehat{\lambda}\varepsilon)^n EV^\varepsilon(\widetilde{\theta}_0, 0) + \sum_{j=0}^{n} (1 - \kappa_1 \widehat{\lambda}\varepsilon)^j O(\varepsilon^2).$$

Since

$$\sum_{j=0}^{n} (1 - \kappa_1 \widetilde{\lambda}\varepsilon)^j = O(1/\varepsilon),$$

we arrive at that, for sufficiently large n,

$$EV^\varepsilon(\widetilde{\theta}_{n+1}, n+1) \leq O(\varepsilon).$$

Using (4.15) and (4.16) again, we also obtain

$$EV(\widetilde{\theta}_{n+1}) \leq O(\varepsilon) \quad \text{for sufficiently large } n.$$

The desired result thus follows. □

Similar to the definition of regularity for the continuous-time processes, we say that the sequence of recursively defined iterates $\{Z_n\}$ is regular if, for any $0 < N < \infty$,

$$P(\sup_{0 \leq n \leq N} |Z_n| = \infty) = 0.$$

It is again a notion of no finite explosion time. As a direct consequence of Theorem 4.2, by using Chebyshev's inequality, we obtain the following corollary.

COROLLARY 4.3. *Under the conditions of Theorem* 4.2, *the sequences* $\{\theta_n\}$, $\{\widetilde{\theta}_n\}$, *and* $\{X_n : n \geq N_\varepsilon\} = \{\widetilde{\theta}_n/\sqrt{\varepsilon} : n \geq N_\varepsilon\}$ *are regular.*

Note that Theorem 4.2 is more or less a tightness result for the centered and scaled sequence $\{\theta_n - \theta_*\}$. To proceed, we obtain another lemma, which is in the direction of the recurrence of $\{(\theta_n - \theta_*)/\varepsilon\}$. The proof is based on a perturbed Lyapunov function method.

To proceed, we use the idea in [14]; see also the treatment in [19]. Let

$$U = \{x : V(x) \leq \varpi\},$$

where $\varpi > 0$. Let N_ε be such that, for all $n \geq N_\varepsilon$, we have $EV^\varepsilon(n) = O(\varepsilon)$ and $EV(\widetilde{\theta}_n) = O(\varepsilon)$. Denote by

$$\tau^o = \inf\{n \geq N_\varepsilon : \widetilde{\theta}_n \in U^c\}, \tag{4.23}$$

i.e., τ^o is the exit time after N_ε, i.e., the first time the iterate is not in U. Let

$$\tau = \inf\{n \geq \tau^o : \widetilde{\theta}_n \in U\}. \tag{4.24}$$

That is, τ is the first return time after τ^o.

THEOREM 4.4. *Under the conditions of Theorem* 4.2,

$$E(\tau - \tau^o) < \infty. \tag{4.25}$$

Proof. Define the Lyapunov function $V(\cdot)$ and the perturbations $V_1^\varepsilon(\cdot)$, $V_2^\varepsilon(\cdot)$, $V_3^\varepsilon(\cdot)$, and $V^\varepsilon(\cdot)$ as in Theorem 4.2. To simplify the notation, we adopt the conventions $V^\varepsilon(n)$ and $V_i^\varepsilon(n)$ to denote $V^\varepsilon(\widetilde{\theta}, n)$, $V_i^\varepsilon(\widetilde{\theta}, n)$, and $V^\varepsilon(n) = V^\varepsilon(\widetilde{\theta}_n, n)$, respectively. Note that, for all n with $\tau^o \leq n < \tau$, $V(\widetilde{\theta}_n) \geq \varpi$. Then by virtue of (4.22), for some $\alpha_0 > 0$,

$$\begin{aligned} &E[V^\varepsilon(\tau) - V^\varepsilon(\tau^o)] \\ &\quad \leq E \sum_{j=\tau^o}^{\tau-1} E_{\tau^o}[-\varepsilon\alpha_0 V(\widetilde{\theta}_j) + O(\varepsilon^2)] \\ &\quad \leq [-\varepsilon\alpha_0\varpi + O(\varepsilon^2)]E(\tau - \tau^o) \\ &\quad \leq -\frac{\varepsilon\alpha_0\varpi}{2}E(\tau - \tau^o). \end{aligned} \tag{4.26}$$

In the last step above, we have used that, for $\varepsilon > 0$ small enough, we can make

$$-\alpha_0\varepsilon\varpi + O(\varepsilon^2) \leq -\alpha_0\varepsilon\varpi/2.$$

Clearly the nonnegativity of $V(\cdot)$ and the order estimates on $V_i^\varepsilon(n)$ imply that $EV^\varepsilon(\tau) \geq 0$. Thus, we obtain

$$E(\tau - \tau^o) \leq \frac{2}{\varepsilon\alpha_0\varpi}EV^\varepsilon(\tau^o) \leq \frac{2}{\alpha_0\varpi}O(1).$$

In the last step above, we have used the fact that $EV^\varepsilon(\tau^o) = O(\varepsilon)$. This implies that $E(\tau - \tau^o) < \infty$ as desired. □

Remark 4.5. Consider now the iterates returning to a bounded set repeatedly, a situation which is inspired by [18]; see also the similar ideas in treating diffusion processes in [14]. Denote $K_i = \{\widetilde{\theta} : V(\widetilde{\theta}) \leq \varpi_i\}$ for $i = 1, 2$ with $\varpi_2 > \varpi_1$, and let τ_1^o be the first exit time of the iterates from K_1, i.e., $\tau_1^o = \inf\{n : \widetilde{\theta}_n \notin K_1\}$. $\tau_1 = \min\{n : n > \tau_1^o, \widetilde{\theta}_{\tau_2} \in K_1\}$. Note that by virtue of the estimates (4.15) and (4.16) and owing to the fact $V(\widetilde{\theta}_n) \geq 0$, we have that, for each $\widetilde{\theta}$,

$$V_i^\varepsilon(\widetilde{\theta}, n) \geq -O(\varepsilon), \ i = 1, 2, \tag{4.27}$$

and hence,

$$|V^\varepsilon(n)| \geq |V(\widetilde{\theta}_n)| - O(\varepsilon) \geq -O(\varepsilon). \tag{4.28}$$

In addition, $\{V^\varepsilon(n \wedge \tau_1)\}$ is a supermartingale, and (4.22) implies that, for some $\kappa > 0$,

$$E_n V^\varepsilon(n+1) - V^\varepsilon(n) \leq -\kappa\varepsilon V^\varepsilon(n) + O(\varepsilon^2) \tag{4.29}$$

for sufficiently large n. This, together with (4.28) and (4.29), leads to that, for any n satisfying $\tau_1^o \le n+1 < \tau_2$,

$$E_{\tau_1^o} V^\varepsilon(n+1) - V^\varepsilon(\tau_1^o) \le -\alpha_1 \varepsilon \sum_{k=\tau_1^o}^{n} V(\widetilde{\theta}_k) + O(\varepsilon^2) \le O(\varepsilon^2) \tag{4.30}$$

for some $\alpha_1 > 0$. By using the lower bound (4.27), we finally obtain

$$E_{\tau_1^o} V^\varepsilon(n+1) \le O(\varepsilon) + O(\varepsilon^2) = O(\varepsilon).$$

Furthermore, we also obtain $E_{\tau_1^o} V(\widetilde{\theta}_{n+1}) = O(\varepsilon)$. That is, $\widetilde{\theta}_n \in K_2$ with probability one.

To summarize the results obtained, we have the following:

(i) The iterates $\widetilde{\theta}_n$ are positive recurrent in the sense of Theorem 4.4.

(ii) Let K_i, τ_i^o, and τ_i be defined as above with $i = 1, 2$. Then eventually $\widetilde{\theta}_n$ remains in K_2 w.p.1 for any $\varpi_2 > \varpi_1 > 0$.

4.3. Remarks on path excursion. The results obtained in Theorem 4.4 (in particular the first return time to the set U) can be readily extended to treat returning to the given set in subsequent instances. To be more specific, we can define the set U as in the last section. Define

$$\tau_1^o = \inf\{n \ge N_\varepsilon : \widetilde{\theta}_n \in U^c\}.$$

That is, τ_1^o is the exit time after N_ε, i.e., the first time the iterate is not in U. Let

$$\tau_1 = \inf\{n \ge \tau_1^o : \widetilde{\theta}_n \in U\}.$$

That is, τ_1 is the first return time after τ_1^o. Next define inductively

$$\begin{aligned} \tau_k^o &= \inf\{n \ge \tau_{k-1} : \widetilde{\theta}_n \in U^c\}, \\ \tau_k &= \inf\{n \ge \tau_k^o : \widetilde{\theta}_n \in U\}. \end{aligned}$$

Following exactly the same approach as in the last section, we obtain

$$E(\tau_k - \tau_k^o) < \infty. \tag{4.31}$$

Thus, the iterates will return to the set U infinitely often.

5. Stability in distribution. In the previous section, we obtained results on weak stability or recurrence of the sequences associated with (3.2). In this section, we consider stability in distribution. As can be seen that in the limit SDE (3.6), $\widehat{\Sigma}$ is nonnegative definite and the diffusion matrix does not depend on x. Thus, it does not admit trivial solution. The usual notion of stability frequently used is roughly the convergence of the nontrivial solution to that of the trivial one for large time t. However, at this point, due to the nondegeneracy, such a notion is not appropriate. We will need to consider the probability distributions. We aim to compare the sequence of probability distributions of the recursive algorithm and invariant distribution of limit of the suitably scaled sequence. First we recall the definition of stability; then we find the sufficient conditions that guarantee the stability in distribution for the recursive algorithm. This section is motivated by the work on asymptotic distribution of [19, pp. 153–156].

5.1. Stability in distribution of associated SDE. First let us recall the definition of stability in distribution for a diffusion process.

DEFINITION 5.1. *The process* $Y(t)$ *given in* (4.1) *is said to be stable in distribution if there exists a probability measure* $\overline{\mu}(\cdot)$ *such that the distribution* $\mu(t,dx)$ *of* (4.1) *converges weakly to* $\overline{\mu}(dx)$ *as* $t \to \infty$ *for each initial data* $y \in \mathbb{R}^r$.

Remark 5.2. Note that for a diffusion process, the distribution $\mu(t,dx)$ turns out to the transition probability $P(t,y,dx)$ with the initial data y. The definition of stability in distribution essentially says that if the initial "distribution" is not too far away from the stationary distribution, then the difference of time t-dependent distribution will not be far away from the stationary one. To be more precise, this property can be stated as if, for each $\delta > 0$ and arbitrary integer n_0, there exist an $\eta = \eta(\delta) > 0$ and n_0^η such that, for any $\rho_j \in C_0(\mathbb{R}^r)$ (continuous functions with compact support) with $j \le \max(n_0^\eta, n_0)$,

$$\left| \prod_{j=1}^{n_0^\eta} \int \rho_j(y)\mu(0,dy) - \prod_{j=1}^{n_0^\eta} \int \rho_j(y)\overline{\mu}(dy) \right| < \eta$$

implies that

$$\left| \prod_{j=1}^{n_0} \int \rho_j(y)\mu(t,dy) - \prod_{j=1}^{n_0} \int \rho_j(y)\overline{\mu}(dy) \right| < \delta.$$

LEMMA 5.3. *Assume that the conditions of Theorem* 4.1 *hold. The limit diffusion process* (3.6) *obtained from* (3.2) *is stable in distribution.*

Proof. We note that by virtue of Theorem 4.1, the limit diffusion process is positive recurrent. It then follows from [14] that the process is ergodic. That is, there is a unique invariant distribution. In fact, this limit distribution $\overline{\mu}(x)$ can be written as

$$\overline{\mu}(x) = \int \nu(y)dy,$$

where $\nu(\cdot)$ is the so-called invariant density of the limit distribution that can be obtained by solving the Kolmogorov forward equation

$$\mathcal{L}^*\nu = 0, \quad \int \nu(x)dx = 1, \tag{5.1}$$

where $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$. Thus the lemma is obtained. ☐

LEMMA 5.4. *Assume the conditions of Lemma* 5.3. *Then the process* $X^x(t)$, *with initial data* $X(0) = x$, *given by* (3.6) *is a Feller process. That is, for any bounded and continuous function* $h(\cdot)$, $u(x) = E_x h(X(t))$ *is continuous.*

Proof. The proof is standard: see, e.g., [9]. We thus omit the detailed argument. ☐

Remark 5.5. A set $\mathcal{C} \subset C_b(\mathbb{R}^r)$ is said to be convergence determining if $\int f d\mu_n \to \int f d\overline{\mu}$ as $n \to \infty$ for each $f \in \mathcal{C}$ implies that μ_n converges to $\overline{\mu}$ weakly. Denote by $C_b(\mathbb{R}^r)$ the space of real-valued bounded and continuous functions defined on $\mathbb{R}^r$. Then $C_b(\mathbb{R}^r)$ is convergence determining. For a proof of this result, see [4, p. 112].

In what follows, to highlight that the dynamics start at $X(0)$, by abusing notation slightly, we often write

$$\int \rho(x)\overline{\mu}(dx) \quad \text{as} \quad E_\mu \rho(X(0)).$$

LEMMA 5.6. *Let K_T be a set of $\mathbb{R}^r$-valued random variables, which is tight (the letter T signals the tightness). Under the conditions of Theorem 4.1, for any positive integer n_0, $0 = \Delta_1 < \Delta_2 < \cdots < \Delta_{n_0}$, and any $\rho_j(\cdot) \in C_b(\mathbb{R}^r)$, $j \le n_0$,*

$$E_{X(0)} \prod_{j=1}^{n_0} \rho_j(X(t+\Delta_j)) \to E_\mu \prod_{j=1}^{n_0} \rho_j(X(\Delta_j)) \tag{5.2}$$

uniformly in $X(0) \in K_T$ as $t \to \infty$.

Proof. We prove the assertion by induction on n_0. First, for $n_0 = 1$, $E_{X(0)}\rho_1(X(t))$ is bounded and continuous for each $t > 0$ since $\rho_1(\cdot) \in C_b(\mathbb{R}^r)$. Suppose that $X(0) \in K_T$ and the distribution with initial condition $X(0)$ is denoted by $\mu(0,\cdot)$. Then the tightness of $X(0)$ yields that

$$\begin{aligned} E_{X(0)}\rho_1(X(t)) &= \int \mu(0,dx) E_x \rho_1(X(t)) \\ &\to \int \mu(0,dx) E_\mu \rho_1(X(0)) \\ &= E_\mu \rho_1(X(0)) \end{aligned} \tag{5.3}$$

since $\int \mu(0,dx) = 1$. Moreover, the convergence is uniform in $X(0)$ since it belongs to a tight set.

Suppose that the assertion is true for $n_0 = \ell - 1$. We prove that it is also true for $n_0 = \ell$. In view of Lemmas 5.3 and 5.4, as $t \to \infty$, $\mu(t+\Delta_{\ell-1}, dx)$ converges weakly to $\mu(dx)$, and $E_x\rho_\ell(X(t+\Delta_\ell)) \to E_\mu \rho_\ell(X(\Delta_\ell))$. By noting the measurability of $\prod_{j=1}^{\ell-1} \rho_j(X(t+\Delta_j))$ with respect to the σ-algebra generated by $\{W(u) : u \le t+\Delta_{\ell-1}\}$, we have

$$\begin{aligned} &E_{X(t+\Delta_{\ell-1})}\rho_\ell(X(t+\Delta_\ell)) \\ &= \mu(t+\Delta_{\ell-1}, dx) E_x \rho_\ell(X(t+\Delta_\ell)) \\ &\to \int \mu(\Delta_{\ell-1}, dx) E_\mu \rho_\ell(X(\Delta_\ell)) \quad \text{as} \quad t \to \infty. \end{aligned} \tag{5.4}$$

Thus, using the induction hypothesis, we arrive at

$$\begin{aligned} &E_{X(0)} \prod_{j=1}^{\ell} \rho_j(X(t+\Delta_j)) \\ &= E_{X(0)} \prod_{j=1}^{\ell-1} \rho_j(X(t+\Delta_j)) E_{X(t+\Delta_{\ell-1})} \rho_\ell(X(t+\Delta_\ell)) \\ &\to \Big[E_\mu \prod_{j=1}^{\ell-1} \rho_j(X(\Delta_j))\Big] E_\mu \rho_\ell(X(\Delta_\ell)) \quad \text{as} \quad t \to \infty \\ &= E_\mu \prod_{j=1}^{\ell} \rho_j(X(\Delta_j)). \end{aligned} \tag{5.5}$$

Thus the proof of the lemma is concluded. □

5.2. Stability in distribution of the recursive algorithm.

THEOREM 5.7. *Consider algorithm (3.2). Assume the conditions of Theorem 4.1 are fulfilled. For arbitrary positive integer n_0, $\rho_j \in C(\mathbb{R}^r)$, $j \le n_0$, and for any $\Delta > 0$, there exist $t_0 < \infty$ and positive integer $\varepsilon_0 > 0$ such that, for all $t \ge t_0$ and $\varepsilon \le \varepsilon_0$,*

$$\left| E \prod_{j=1}^{n_0} \rho_j(X^\varepsilon(t+\Delta_j)) - \prod_{j=1}^{n_0} E_\mu \rho_j(X(\Delta_j)) \right| < \Delta. \tag{5.6}$$

Moreover, for any sequence $s_\varepsilon \to \infty$ as $\varepsilon \to 0$, the distribution of $((X^\varepsilon(s_\varepsilon + \Delta_1)), \ldots, (X^\varepsilon(s_\varepsilon + \Delta_{n_0}))$ *converges weakly to the stationary distribution of* $(X(\Delta_1), \ldots, X(\Delta_{n_0}))$.

Proof. Suppose that (5.6) were not true. There would exist a subsequence $\{\eta(\varepsilon)\}$ of $\{\varepsilon\}$ and a sequence $s_{\eta(\varepsilon)} \to \infty$ such that

$$\left| E \prod_{j=1}^{n_0} \rho_j(X^{\eta(\varepsilon)}(s_{\eta(\varepsilon)} + \Delta_j)) - E_\mu \prod_{j=1}^{n_0} \rho_j(X(\Delta_j)) \right| \geq \Delta > 0. \tag{5.7}$$

Fix $T > 0$ and choose a further subsequence $\{\delta(\eta)\} = \{\delta(\eta(\varepsilon))\}$ of $\{\eta(\varepsilon)\}$, and the corresponding sequence $(X^{\delta(\eta)}(\cdot))$ such that $(X^{\delta(\eta)}(s_{\delta(\eta)} - T))$ converges weakly to a random variable $X(0)$. Lemma 3.6 implies that $X^{\delta(\eta)}(s_{\delta(\eta)} - T + \cdot)$ converges weakly to $(X(\cdot))$ with initial condition $X(0)$. Moreover,

$$\begin{aligned} &E \prod_{j=1}^{n_0} \rho_j(X^{\delta(\eta)}(s_{\delta(\eta)} - T + T + \Delta_j)) \\ &\quad \to EE_{X(0)} \prod_{j=1}^{n_0} \rho_j(X(T + \Delta_j) \quad \text{as} \quad \delta(\eta) \to 0. \end{aligned}$$

Owing to the tightness of the limit diffusion, the collection of all possible $X(0)$ over all $T > 0$ and weakly convergent subsequence is tight. Thus, by Lemma 5.6, there exists $T_0 > 0$ such that, for all $T \geq T_0$,

$$\left| EE_{X(0)} \prod_{j=1}^{n_0} \rho_j(X(T + \Delta_j)) - E_\mu \prod_{j=1}^{n_0} \rho_j(X(\Delta_j)) \right| < \Delta/2,$$

which contradicts (5.7).

Using Lemma 5.6 again, part (i) of the theorem implies that $(X^\varepsilon(s_\varepsilon + \cdot))$ converges weakly to the random variable with the invariant distribution $\overline{\mu}(\cdot)$ as $s_\varepsilon \to \infty$. Thus part (ii) of the assertion also follows. □

6. Ramifications and further remarks. In this section, we discuss some generalizations, adjustments to the algorithm (2.3), and some observations concerning the results of the paper. This paper introduces the notion of model mismatch; one of the important parts of this section is devoted to mismatch and bias reduction.

6.1. Unbounded signals. So far, we have mainly considered the case that the signals are correlated mixing type but are bounded up to this point. Next we elaborate on how unbounded signals may be treated. Instead of (A1), we consider the following assumption.

(A1′) The sequence $\{\phi_n, y_n\}$ is stationary such that $E\varphi_n\varphi_n' = R > 0$ and $E\varphi_n\zeta_n = 0$. In addition,

–

$$E|\phi_n|^{4+\Delta} < \infty, \; E|y_n|^{4+\Delta} < \infty \text{ for some } \Delta > 0. \tag{6.1}$$

– $\{\phi_n\phi_n' - R\}$ and $\{\phi_n y_n - q\}$ are moving average sequences of order m, i.e.,

$$\begin{aligned} \varphi_n\varphi_n' - R &= \sum_{i=0}^{m_0} C_i \xi_{n-i}^R, \\ \varphi_n\zeta_n &= \sum_{i=0}^{m_0} D_i \xi_{n-i}^q, \end{aligned} \tag{6.2}$$

where C_i, D_i, $i \le m_0$ are matrices with appropriate dimension and $\{\xi_n^R\}$ and $\{\xi_n^q\}$ are stationary martingale difference sequences.

Here the signals are allowed to be unbounded moving average type processes. Using the approach provided in the previous sections, the conclusions of the results developed thus far, including recurrence and stability in distribution, continue to hold for this unbounded signal case.

6.2. Decreasing stepsize algorithms. Although the analysis has been devoted to constant stepsize algorithms, the results obtained carry over to decreasing stepsize algorithms

$$\theta_{n+1} = \theta_n + a_n \varphi_n(\varphi_n' \theta_* - \varphi_n' \theta_n + \zeta_n) + a_n \Delta_n, \tag{6.3}$$

where $a_n \ge 0$, $a_n \to 0$, $\sum_n a_n = \infty$, and $\Delta_n = \varphi_n \widetilde{\Delta}_n$ with $\widetilde{\Delta}_n$ as in (3.2). Let us briefly comment on the needed modification below.

To link the discrete time with continuous time, we define

$$t_n = \sum_{j=0}^{n-1} a_k, \quad m(t) = \sup\{n, t_n \le t\}.$$

Here t_n is a "connector" that bridges the connection naturally, whereas $m(t)$ is an "inverse" that takes a continuous time back to the discrete moment. The interpolation sequences are defined as follows. Denoting $X_n = (\theta_n - \theta_b)/\sqrt{a_n}$, let $\theta^0(t)$ and $X^0(t)$ be the piecewise constant interpolations of θ_n and X_n on the interval $t \in [t_n, t_{n+1})$. Define the shift sequence by

$$\theta^n(t) = \theta^0(t + t_n) \quad \text{and} \quad X^n(t) = X^0(t + t_n).$$

Then we can proceed with the analysis as in the previous two sections. We obtain the same limit ODE and the same limit SDE. We can also establish an analog of Theorem 4.2. In this case, the statement will be changed to $EV(\widetilde{\theta}_{n+1}) = O(a_n)$. In addition, we can also show that Theorem 4.4 is still true, as is the recurrence time estimate (4.31). Furthermore, the results in Theorem 5.7 carry over to the decreasing stepsize case.

6.3. Bias reduction. Recall the observation equation (3.1), in which the bound on the model mismatch $\widetilde{\Delta}_n$ dictates the model mismatch leading bias, as stated in (3.4). Typically φ_n contains input u_k for $k \le n$, and the above observation equation represents a linearization effort in which the model mismatch term has less impact on the linear term when the input signal u_n, and hence φ_n, are "small." However, when u_k takes large values, the higher-order terms contained in $\widetilde{\Delta}_n$ will dominate and make the estimation bias unacceptably large for the identified model to be useful. In general, bias correction to accommodate nonlinearity in $\widetilde{\Delta}_n$ requires model structures that capture more higher-order terms, leading to complicated nonlinear identification problems. Here we show that if the identified model is intended for utility in a regulation problem and when the system nonlinearity has some structures, bias correction can be effectively achieved by a modified linear structure.

Assume in this subsection that the nonlinearity has the separation structure

$$\widetilde{\Delta}_n = h'(\varphi_n)\theta_* \tag{6.4}$$

and that $h(\cdot)$ is twice continuously differentiable. Note that $\theta_* \in \mathbb{R}^r$ and

$$\varphi_n' = [u_n, u_{n-1}, \dots, u_{n-r+1}].$$

In a regulation problem in which the set point is a constant, control actions will eventually lead to an asymptotically constant $u_n = c$. It is noted that if c is not small, identification bias may become very large. Apparently, when u_n is a constant, it is no longer persistently exciting, losing its capability in providing sufficient information for system identification. Assume that a small dither is added to the input, resulting in

$$u_n = c + \widetilde{\varepsilon} v_n, \tag{6.5}$$

where $|v_n| \leq 1$ and will be selected for identification experiments. For any small $\widetilde{\varepsilon} > 0$, this dither will have diminishing effects on control performance and hence is acceptable.

Under the input in (6.5), we have

$$\varphi_n = c\mathbb{1}_r + \widetilde{\varepsilon}\widetilde{\varphi}_n,$$

where $\mathbb{1}_r$ is the m dimensional vector of all 1's and $\widetilde{\varphi}_n = [v_n, v_{n-1}, \ldots, v_{n-r+1}]$ which is uniformly bounded for all n. Now by a truncated Taylor expansion of $h(\varphi_n)$ around $c\mathbb{1}_r$, we have

$$h(\varphi_n) = h(c\mathbb{1}_r + \widetilde{\varepsilon}\widetilde{\varphi}_n) = h(c\mathbb{1}_r) + \widetilde{\varepsilon}G(c\mathbb{1}_r)\widetilde{\varphi}_n + \widetilde{\varepsilon}^2\delta(\xi_n),$$

where $h(c\mathbb{1}_r)$ and $G(c\mathbb{1}_r)$ are unknown but constant vector and matrix, respectively, and $\widetilde{\varepsilon}^2\delta(\xi_n)$ represents the second-order remainder term in the Taylor expansion. The term $\delta(\xi_n)$ is uniformly bounded since ξ_n lies in a uniformly bounded neighborhood of $c\mathbb{1}_r$. Consequently, the observation equation may be expressed as

$$\begin{aligned} y_n &= (c\mathbb{1}_r' + \widetilde{\varepsilon}\widetilde{\varphi}_n')\theta_* + (h(c\mathbb{1}_r) + \widetilde{\varepsilon}G(c\mathbb{1}_r)\widetilde{\varphi}_n + \widetilde{\varepsilon}^2\delta(\xi_n))\theta_* + \zeta_n \\ &= (c\mathbb{1}_r' + h'(c\mathbb{1}_r))\theta_* + \widetilde{\varepsilon}\widetilde{\varphi}_n'(I + G'(c\mathbb{1}_r))\theta_* + \widetilde{\varepsilon}^2\delta(\xi_n)\theta_* + \zeta_n. \end{aligned}$$

Now define $b_* = (c\mathbb{1}_r' + h'(c\mathbb{1}_r))\theta_*$ and $\theta_*^0 = (I + G'(c\mathbb{1}_r))\theta_*$. Let $\widehat{\theta}_*' = [b_*, (\theta_*^0)']$ be the new unknown but true system parameter vector and $\psi_n' = [1, \widetilde{\varepsilon}\widetilde{\varphi}_n']$ be the new regressor. Then the observation equation becomes

$$y_n = \psi_n'\widetilde{\theta}_* + \widetilde{\varepsilon}^2\delta(\xi_n)\theta_* + \zeta_n. \tag{6.6}$$

Since this observation structure is the same as (3.1), all the algorithms and analysis in the previous sections are applicable to this modified linear structure with model mismatch term $\widetilde{\varepsilon}^2\delta(\xi_n)\theta_*$. The main idea for bias correction is to design v_k such that ψ_n satisfies persistent excitation conditions and then to choose $\widetilde{\varepsilon}$ sufficiently small for bias reduction.

Define

$$\widetilde{R} = E\psi_n\psi_n', \quad \widetilde{g} = E\psi_n\delta(\xi_n)\theta_*.$$

By the same arguments as in (3.4), we conclude that the identification bias after the bias correction scheme is $\widetilde{R}^{-1}\widetilde{\varepsilon}^2\widetilde{g}$.

Example 6.1. Consider the system in Example 3.4. Since $\widetilde{\Delta}_n = \lambda u_n^2\theta_*$, it satisfies the special structure condition imposed in this section with $h(\varphi_n) = \lambda u_n^2$. When $u_n = c + \widetilde{\varepsilon}v_n$,

$$h(\varphi_n) = \lambda(c + \widetilde{\varepsilon}v_n)^2 = \lambda c^2 + \widetilde{\varepsilon}\lambda 2cv_n + \widetilde{\varepsilon}^2\lambda v_n^2.$$

Suppose the dither is designed as a periodic signal with $v_1 = 0$ and $v_2 = 1$. Since $\lambda = 0.05$ and $\theta_* = 10$, from $\psi_n' = [1, \widetilde{\varepsilon} v_n]$ we have

$$\widetilde{R} = E\begin{bmatrix} 1 \\ \widetilde{\varepsilon} v_n \end{bmatrix} [1, \widetilde{\varepsilon} v_n] = \begin{bmatrix} 1 & \widetilde{\varepsilon}/2 \\ \widetilde{\varepsilon}/2 & \widetilde{\varepsilon}^2/2 \end{bmatrix},$$

$$\widetilde{\varepsilon}^2 \widetilde{g} = \widetilde{\varepsilon}^2 E\begin{bmatrix} 1 \\ \widetilde{\varepsilon} v_n \end{bmatrix} \lambda v_n^2 \theta_* = \begin{bmatrix} 0.25\widetilde{\varepsilon}^2 \\ 0.25\widetilde{\varepsilon}^3 \end{bmatrix},$$

which implies that the bias is

$$\begin{aligned} \widetilde{R}^{-1} \widetilde{\varepsilon}^2 \widetilde{g} &= \frac{4}{\widetilde{\varepsilon}^2} \begin{bmatrix} \widetilde{\varepsilon}^2/2 & -\widetilde{\varepsilon}/2 \\ -\widetilde{\varepsilon}/2 & 1 \end{bmatrix} \begin{bmatrix} 0.25\widetilde{\varepsilon}^2 \\ 0.25\widetilde{\varepsilon}^3 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ \widetilde{\varepsilon}/2 \end{bmatrix} \to 0 \text{ as } \widetilde{\varepsilon} \to 0. \end{aligned}$$

6.4. Concluding remarks. We have developed stability of recursive algorithms with a constant stepsize and model mismatch. Sufficient conditions ensuring recurrence of the algorithm have been obtained. Stability in distribution of the algorithm has also been examined.

This paper is devoted to AF algorithms. In future studies we will be interested in extending the results to RLS algorithms. For the RLS algorithms, the main point is to integrate the matrix gain sequence into the analysis. It seems that both stability and recurrence will largely depend on certain matrix products. Another related question involves algorithms that use forgetting factors. Further investigation for tracking algorithms for time-varying parameters is of theoretical and practical value as well, as are problems with hidden Markov models.

REFERENCES

[1] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer, Berlin, 1990.
[2] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York, 1968.
[3] H.-F. Chen and G. Yin, *Asymptotic properties of sign algorithms for adaptive filtering*, IEEE Trans. Automat. Control, 48 (2003), pp. 1545–1556.
[4] S. N. Ethier and T. G. Kurtz, *Markov Processes, Characterization and Convergence*, John Wiley, New York, 1986.
[5] Y. Ephraim and N. Merhav, *Hidden Markov processes*, IEEE Trans. Inform. Theory, 48 (2002), pp. 1518–1569.
[6] E. Eweda and O. Macchi, *Quadratic mean and almost-sure convergence of unbounded stochastic approximation algorithms with correlated observations*, Ann. Henri Poincaré, 19 (1983), pp. 235–255.
[7] B. G. Fitzpatrick, S. McCanless, and Y. Wang, *Bayesian statistical approaches to tracking through turbulence*, in Proceedings of the 2004 American Control Conference, Vol. 2, IEEE, 2004, pp. 1499–1503.
[8] B. G. Fitzpatrick and Y. Wang, *Bayesian methods in actively illuminated laser tracking through deep turbulence*, Proceedings of the 2007 Directed Energy Systems Symposium, Monterey, CA, Directed Energy Professional Society, Albuquerque, NM, 2007.
[9] I. I. Gikhman and A. V. Skorohod, *Introduction to the Theory of Random Processes*, Dover, New York, 1969.
[10] A. Gersho, *Adaptive filtering with binary reinforcement*, IEEE Trans. Inform. Theory, IT-30 (1984), pp. 191–199.
[11] L. Guo, *Stability of recursive stochastic tracking algorithms*, SIAM J. Control Optim., 32 (1994), pp. 1195–1225.
[12] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[13] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[14] R. Z. KHASMINSKII, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Groningen, the Netherlands, 1980.

[15] R .Z. KHASMINSKII AND G. YIN, *Asymptotic behavior of parabolic equations arising from one-dimensional null-recurrent diffusions*, J. Differential Equations, 161 (2000), pp. 154–173.

[16] R. Z. KHASMINSKII, C. ZHU, AND G. YIN, *Stability of regime-switching diffusions*, Stochastic Process. Appl., 117 (2007), pp. 1037–1051.

[17] V. KRISHNAMURTHY, X. WANG, AND G. YIN, *Spreading code optimization and adaptation in CDMA via discrete stochastic approximation*, IEEE Trans. Inform. Theory, 50 (2004), pp. 1927–1949.

[18] H. J. KUSHNER, *Stochastic approximation with discontiuous dynamics and state dependent noise: w.p.1 and weak convergence*, J. Math. Anal. Appl., 82 (1982), pp. 524–542.

[19] H. J. KUSHNER, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.

[20] H. J. KUSHNER, *Scheduling and control of multi-node mobile communications systems with randomly-varying channels by stability methods*, in Wireless Communications, P. Agrawal, D.M. Andrews, P.J. Fleming, G. Yin, and L. Zhang, eds., Springer, New York, 2006, pp. 177–197.

[21] H. J. KUSHNER AND A. SHWARTZ, *Weak convergence and asymptotic properties of adaptive filters with constant gains*, IEEE Trans. Inform. Theory, IT-30 (1984), pp. 177–182.

[22] H. J. KUSHNER AND G. YIN, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed., Springer, New York, 2003.

[23] C. P. KWONG, *Dual signal algorithm for adaptive filtering*, IEEE Trans. Comm, COM-34 (1986), pp. 1272–1275.

[24] J. P. LASALLE AND S. LEFSCHITZ, *The Stability by Lyapunov Direct Method*, Academic Press, New York, 1961.

[25] L. LJUNG, *Analysis of stochastic gradient algorithms for linear regression problems*, IEEE Trans. Inform. Theory, IT-30 (1984), pp. 151–160.

[26] L. LJUNG, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[27] O. MACCHI AND E. EWEDA, *Convergence analysis of self-adaptive equalizers*, IEEE Trans. Inform. Theory, IT-30 (1984), pp. 161–176.

[28] P. K. ORZECHOWSKI, N. CHEN, S. GIBSON, AND T. C. TSAO, *Optimal suppression of laser beam jitter by high-order RLS adaptive control*, IEEE Trans. Control Syst. Technol., 16 (2008), pp. 255–267.

[29] N.O. PÉREZ, N. CHEN, S. GIBSON, AND T.-C. TSAO, *Variable-order adaptive control of a MEMS sterring mirro for suppression of laser beam jitter*, Opt. Engrg., 45 (2006), pp. 104206-1–104206-12.

[30] H. V. POOR AND X. WANG, *Code-aided interference suppression for DS/CDMA communications—Part I: Interference suppression capability*, IEEE Trans. Comm., 45 (1997), pp. 1101–1111.

[31] H. V. POOR AND X. WANG, *Code-aided interference suppression for DS/CDMA communications—Part II: Parallel blind adaptive implementations*, IEEE Trans. Comm., 45 (1997), pp. 1112–1122.

[32] S. R. VENKATESH AND M. A. DAHLEH, *Identification in the presence of classes of unmodelled dynamics and noise*, IEEE Trans. Automat. Control, 42 (1997), pp. 1620–1635.

[33] L. Y. WANG AND G. YIN, *Persistent identification of systems with unmodeled dynamics and exogenous disturbances*, IEEE Trans. Automat. Control, 45 (2000), pp. 1246–1256.

[34] L. Y. WANG AND G. YIN, *Closed-loop persistent identification of linear systems with unmodeled dynamics and stochastic disturbances*, Automatica J. IFAC, 38 (2002), pp. 1463–1474.

[35] L. Y. WANG, G. YIN, J.-F. ZHANG, AND Y. L. ZHAO, *System Identification with Quantized Observations*, Birkhäuser Boston, Cambridge, MA, 2010.

[36] B. WIDROW AND S. D. STEARNS, *Adaptive Signal Processing*, Prentice-Hall, Englewood, Cliffs, NJ, 1985.

[37] W. M. WONHAM, *Lyapunov criteria for weak stochastic stability*, J. Differential Equations, 2 (1966), pp. 195–207.

[38] G. YIN, *Asymptotic properties of an adaptive beam former algorithm*, IEEE Trans. Inform. Theory, IT-35 (1989), pp. 859–867.

[39] G. YIN, S. KAN, L. Y. WANG, AND C. Z. XU, *Identification of systems with regime switching and unmodelled dynamics*, IEEE Trans. Automat. Control, 54 (2009), pp. 34–47.

[40] G. Yin and V. Krishnamurthy, *Least mean square algorithms with Markov regime switching limit*, IEEE Trans. Automat. Control, 50 (2005), pp. 577–593.

[41] G. Yin, V. Krishnamurthy, and C. Ion, *Iterate-averaging sign algorithms for adaptive filtering with applications to blind multiuser detection*, IEEE Trans. Inform. Theory, 49 (2003), pp. 657–671.

[42] G. Yin and Y. M. Zhu, *Averaging procedures in adaptive filtering: An efficient approach*, IEEE Trans. Automat. Control, AC-37 (1992), pp. 466–475.

[43] G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications*, Springer, New York, 2010.