2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Adaptive Identification under Saturated Output Observations with Possibly Correlated and Unbounded Input Signals

Lantian Zhang and Lei Guo, *Fellow, IEEE*

Abstract— **This paper considers adaptive identification and prediction problems for stochastic dynamical systems with saturated output observations, which arise from various problems in science and technology as well as in social and economic systems. A new adaptive algorithm is introduced, which avoids the projection operators used in the related existing work. More importantly, unlike most previous works that require independent and identically distributed conditions as well as bounded conditions on system signals, it is shown that the global convergence of the average regret and strong consistency of the parameter estimates can be established under possibly unbounded, correlated, and non-stationary signal conditions. A numerical example is also given to illustrate the effectiveness of the proposed adaptive algorithm.**

I. INTRODUCTION

The identification of input-output relationships and the prediction of future behaviors of dynamic systems using observation data are fundamental problems in science and technology. Significant advances have been achieved in control systems, signal processing, statistics, machine learning, and related fields. In this work, we focus on the identification and prediction of stochastic dynamical systems with saturated output observation data, which are motivated by wide applications in engineering ([1][3]), economics ([2]-[5]), and even judicial systems ([6]). To be specific, we consider the following nonlinear stochastic system for $k \geq 0$:

$$y_{k+1} = \phi_k^T \theta + e_{k+1}, \quad s_{k+1} = S_k(y_{k+1}), \tag{1}$$

where $\theta \in \mathscr{R}^m$ is an unknown parameter vector to be estimated; $\phi_k \in \mathscr{R}^m$, $y_{k+1} \in \mathscr{R}$, $s_{k+1} \in \mathscr{R}$, $e_{k+1} \in \mathscr{R}$ represent the system stochastic regressor, output, output observation, and random noise, respectively. $S_k(\cdot) : \mathscr{R} \to \mathscr{R}$ is a time-varying saturation function defined as follows:

$$S_k(x) = \begin{cases} L_k & x < l_k \\ x & l_k \leq x \leq u_k \\ U_k & x > u_k \end{cases}, \quad k = 0, 1, \cdots. \tag{2}$$

At each time, the noise-corrupted output can be observed precisely only if its value lies in a certain range $[l_k, u_k]$. However, if the output value exceeds this range, the observation becomes saturated, leading to imprecise information denoted by L_k or U_k.

Unlike the traditional offline identification method in statistics and machine learning, where independent and identically distributed (i.i.d) conditions on regressors are usually needed for the strong consistency of the algorithms ([7]-[9]), a prominent feature of control-oriented adaptive identification is that the signals or data used for parameter estimation are no longer satisfy the i.i.d condition once feedback signals are chosen as the system inputs, thus requiring general data conditions that should include non-independent and non-stationary conditions in general. For traditional linear stochastic systems, the convergence of the classical least squares (LS) and its adaptive prediction performance under nonstationary and non-independent conditions have been extensively studied in a vast literature ([17]-[24]), and successfully applied in the well-known LS-based self-tuning regulators ([22], [25]).

Recently, several identification algorithms have been introduced for stochastic regression models under both saturated observations and general data conditions. For example, the empirical measure approach was applied in [10]-[11], and strong consistency was established under periodic signals. Some stochastic approximation-type algorithms were produced in [12]-[14] under a kind of deterministic persistence of excitation (PE) conditions. In our recent works [15] and [16], we established the global convergence of a projected second-order algorithm under a general and weakest possible non-PE condition, without requiring independence and stationary conditions.

Though the above work has made significant progress in weakening the traditional i.i.d assumptions on the data, most of the results mentioned above require the boundedness condition on the system signals. While the boundedness assumption greatly simplifies the performance analysis of the nonlinear adaptive algorithms, it also limits the application of these algorithms to more general stochastic systems since, for example, the standard Gaussian signals are not bounded in sample paths. In this paper, we introduce a new adaptive algorithm and prove its global convergence for possibly unbounded, nonstationary, and non-independent data. We also prove the global convergence of the averaged adaptive prediction regret without any excitation conditions. This work is a generalization of the recent work [16], where the system signals were required to be bounded in sample paths and the adaptive algorithms are designed by resorting to projection. Hence, this work is a significant extension to the previous works in at least three aspects: it has expanded the applicability condition on the system signals, has weakened the parameter set from convex compact to only boundedness,

*This work was supported by the National Natural Science Foundational of China under Grant No. 12288201.

Lantian Zhang and Lei Guo are with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Mathematical Science, University of Chinese Academy of Sciences, Beijing 100049, China. (e-mails: zhanglantian@amss.ac.cn, Lguo@amss.ac.cn).

2023 62nd IEEE Conference on Decision and Control (CDC) | 979-8-3503-0124-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/CDC49753.2023.10384027

and has simplified the computational complexity of the algorithm.

The rest of the paper is organized as follows: We will first introduce some preliminary notations and assumptions in Section II. The new adaptive algorithm and the global convergence results will be presented in Subsection III. A and Subsection III.B, respectively. The proofs of the main results will be given in Section IV. A numerical example will be provided in Section V. Some concluding remarks will be given in Section VI.

II. NOTATIONS AND ASSUMPTIONS

Notations. We use $\|\cdot\|$ to represent the Euclidean norm for vectors or matrices. The maximum and minimum eigenvalues of a matrix M is denoted by $\lambda_{max}\{M\}$ and $\lambda_{min}\{M\}$, respectively. Additionally, we use *a.s.* to signify "almost surely."

To carry out our theoretical analyses, we need the following basic assumptions:

Assumption 1: The stochastic regressor ϕ_k is $\mathscr{F}_k-$measurable for all $k \geq 0$, where $\{\mathscr{F}_k, k \geq 0\}$ is a non-decreasing sequence of $\sigma-$algebras. Besides, the true parameter θ belongs to a known bounded set $D \subseteq \mathbb{R}^m$.

Assumption 2: The thresholds $\{L_k, \mathscr{F}_k\}$, $\{U_k, \mathscr{F}_k\}$, $\{l_k, \mathscr{F}_k\}$, $\{u_k, \mathscr{F}_k\}$ are known adapted stochastic sequences, satisfying

$$l_k - c \leq L_k \leq l_k \leq u_k \leq U_k \leq u_k + c, \quad a.s., \tag{3}$$

where c is a random variable.

Assumption 3: The noise $\{e_k, \mathscr{F}_k\}$ is a martingale difference sequence, and there exists a constant $\eta > 2$, such that

$$\sup_{k\geq 0} \mathbb{E}\left[|e_{k+1}|^{\eta} \mid \mathscr{F}_k\right] < \infty, \ a.s. \tag{4}$$

Besides, the function $G_k(x)$, defined by $G_k(x) = \mathbb{E}\left[S_k(x+e_{k+1}) \mid \mathscr{F}_k\right]$ is twice differentiable and its derivative $G_k'(\cdot)$ satisfies

$$0 < \inf_{|x|\leq M, k\geq 0} G_k'(x) \leq \sup_{|x|\leq M, k\geq 0} G_k'(x) < \infty, \tag{5}$$

for any $M \geq 0$.

Remark 1: It is evident that if the noise e_k is independent with the $\sigma-$algebra $\mathscr{F}_{k-1}$ and follows an identical normal distribution as previously assumed (see, e.g., [13]), then the condition (5) in Assumption 3 will be satisfied.

III. MAIN RESULTS

A. Algorithm

Motivated by the analysis of the classical LS recursive algorithm, a second-order algorithm was introduced in [16] for parameter estimation in stochastic regression models with binary-valued observations. However, due to the inherent challenge of ensuring boundedness in the classical LS, the algorithm presented in [16] incorporates a projection operator to guarantee the boundedness of the estimates. This approach increases the algorithm's complexity and makes the computation speed unsatisfactory in practical calculations. To overcome this limitation, we drew inspiration from the self-convergence property of the weighted least squares algorithm for linear stochastic systems (see, [26]) and designed a new adaptive algorithm without using the projection operator. For simplicity of notation, denote

$$\begin{aligned} \underline{g}_k &= \inf_{|x|\leq \|\phi_k\|(B(D)+\|\hat{\theta}_k\|)} G_k'(x), \\ \overline{g}_k &= \sup_{|x|\leq \|\phi_k\|(B(D)+\|\hat{\theta}_k\|)} G_k'(x), \end{aligned} \tag{6}$$

where $B(D) = \sup_{x\in D}\{\|x\|\}$, D is defined in Assumption 1. Our new adaptive identification algorithm is defined as follows:

Algorithm 1:

$$\hat{\theta}_{k+1} = \hat{\theta}_k + a_k\beta_k P_k\phi_k[s_{k+1} - G_k(\phi_k^{\mathrm{T}}\hat{\theta}_k)], \tag{7}$$

$$P_{k+1} = P_k - \beta_k^2 a_k P_k\phi_k\phi_k^{\mathrm{T}} P_k, \tag{8}$$

$$a_k = \frac{1}{\mu_k + \beta_k^2\phi_k^{\mathrm{T}} P_k\phi_k}, \tag{9}$$

$$\beta_k = \min\{\underline{g}_k, \frac{\mu_k}{2\overline{g}_k\phi_k^T P_k\phi_k + 1}\}, \tag{10}$$

where $\hat{\theta}_k$ is the estimate of θ at time k; G_k is defined in Assumption 3; the initial value $\hat{\theta}_0$ can be chosen arbitrarily in D; $P_0 > 0$ can also be chosen arbitrarily; $\{\mu_k\}$ is the weighting sequence defined by

$$\mu_k = (\log(r_k+1))^{1+\delta}, \ \ r_k = \|P_0^{-1}\| + \sum_{i=0}^{k}\|\phi_i\|^2, \tag{11}$$

where $\delta > 0$ can be chosen arbitrarily.

Furthermore, we have

$$\mathbb{E}(y_{k+1} \mid \mathscr{F}_k) = \theta^T\phi_k, \tag{12}$$

which represents the best prediction for y_{k+1} given $\mathscr{F}_k$ in the mean square sense. By substituting the unknown parameter θ in (12) with its estimate $\hat{\theta}_k$, we can derive a natural adaptive predictor for y_{k+1} as follows:

$$\hat{y}_{k+1} = \hat{\theta}_k^T\phi_k. \tag{13}$$

Typically, the discrepancy between the best prediction and the adaptive prediction can be considered as regret, denoted as R_k, which is given by:

$$R_k = \left[\mathbb{E}(y_{k+1} \mid \mathscr{F}_k) - \hat{y}_{k+1}\right]^2. \tag{14}$$

In this paper, we will prove that the average regret $\frac{1}{n}\sum_{k=1}^{n} R_k$ converges to 0, which will be useful in adaptive control.

B. Global convergence results

To give the main theorems, we first establish the following lemma:

Lemma 1: Let Assumptions 1-3 be satisfied. Then the parameter estimate given by Algorithm 1 has the following property as n $\to \infty$:

$$\tilde{\theta}_{n+1}^T P_{n+1}^{-1}\tilde{\theta}_{n+1} + \sum_{k=0}^{n}\frac{\beta_k^2}{\mu_k}(\tilde{\theta}_k^T\phi_k)^2 = O(1). \tag{15}$$

Theorem 1: Consider the Algorithm 1 under Assumptions 1-3. If there exists a $\alpha > 2$ such that

$$\sum_{k=1}^{n}\|\phi_k\|^{\alpha} = O(n), \ \ a.s., \tag{16}$$

then the sample paths of the average regrets will have the following property as $n \to \infty$:

$$\sum_{k=0}^{n} R_k = o(n), \ \ a.s. \tag{17}$$

where R_k is defined by $R_k = (\mathbb{E}[y_{k+1} \mid \mathscr{F}_k] - \hat{y}_{k+1})^2$.

Remark 2: We remark that from Theorem 1, the sample paths of the average regrets converge to 0 without any excitation conditions on system signals and this result does not require the boundedness of system signals, making it convenient to apply in feedback control systems.

Theorem 2: Under Assumptions 1-3, the estimator $\hat{\theta}_k$ given by the Algorithm 1 has the following upper bound almost surely as $k \to \infty$:

$$\|\tilde{\theta}_{k+1}\|^2 = O\left(\frac{1}{\lambda_{min}\{P_{k+1}^{-1}\}}\right), \ \ a.s. \tag{18}$$

where $\tilde{\theta}_k = \theta - \hat{\theta}_k$.

Remark 3: Since $\mu_k = (\log r_k)^{1+\delta}$ for some $\delta > 0$, we then have

$$\lambda_{\min}\{P_0^{-1} + \sum_{i=1}^{n} \frac{\beta_i^2 \phi_i \phi_i^T}{\mu_i}\} \geq \frac{1}{\mu_n} \lambda_{\min}\{P_0^{-1} + \sum_{i=1}^{n} \beta_i^2 \phi_i \phi_i^T\}. \tag{19}$$

Hence by Theorem 1, we have

$$\left\|\tilde{\theta}_{n+1}\right\|^2 = O\left(\frac{(\log r_n)^{1+\delta}}{\lambda_{\min}\left\{P_0^{-1} + \sum_{i=1}^{n} \beta_i^2 \phi_i \phi_i^{\tau}\right\}}\right), \ \text{a.s}, \ \ \delta > 0. \tag{20}$$

This error bound is weaker than the result in [16], where (20) holds with $\delta = 0$. However, by equation (18) and the fact that $\lambda_{\min}\{P_{k+1}^{-1}\}$ is increasing, the current algorithm naturally guarantees the boundedness of the parameter estimates, a property that is important for the convergence analysis of the adaptive regret later. In [16], the boundedness of the estimation is guaranteed by a projection operator. Thus for some large-scale systems, the current algorithm avoids the complex projection operation in [16].

Corollary 1: Let the conditions of Theorem 1 be satisfied. If the PE condition is satisfied, i.e.

$$n = O(\lambda_{\min}\{P_0^{-1} + \sum_{i=0}^{n} \phi_i \phi_i^T\}), \ \ a.s.,$$

then we have

$$\|\tilde{\theta}_n\|^2 = O(\frac{\log^{1+\delta} n}{n}), \ \ a.s. \tag{21}$$

IV. PROOF OF MAIN RESULTS

To prove the main results, we need the following lemma.

Lemma 2: ([24]). Let $\{f_n, \mathscr{F}_n\}$ be an adapted sequence and $\{w_n, \mathscr{F}_n\}$ a martingale difference sequence. If

$$\sup_n \mathbb{E}[|w_{n+1}|^{\alpha} \mid \mathscr{F}_n] < \infty \ \ a.s. \tag{22}$$

for some $\alpha \in (0, 2]$, then as $n \to \infty$:

$$\sum_{i=0}^{n} f_i w_{i+1} = O(s_n(\alpha) \log^{\frac{1}{\alpha}+\eta}(s_n^{\alpha}(\alpha) + e)) \ a.s., \forall \eta > 0, \tag{23}$$

where

$$s_n(\alpha) = \left(\sum_{i=0}^{n} |f_i|^{\alpha}\right)^{\frac{1}{\alpha}}. \tag{24}$$

A. Proof of Lemma 1.

For convenience, let

$$\begin{aligned} \psi_k &= G_k(\phi_k^T \theta) - G_k(\phi_k^T \hat{\theta}_k), \\ w_{k+1} &= s_{k+1} - G_k(\phi_k^T \theta). \end{aligned} \tag{25}$$

By Assumption 3, we have $\mathbb{E}[w_{k+1} \mid \mathscr{F}_k] = 0$ for any $k \geq 0$, and we will prove that

$$\sup_{k \geq 0} \mathbb{E}[|w_{k+1}|^{2+\eta} \mid \mathscr{F}_k] < \infty. \tag{26}$$

We note that by Assumption 2,

$$|S_k(\phi_k^T \theta + e_{k+1}) - S_k(\phi_k^T \theta)| \leq |e_{k+1}| + 2c. \tag{27}$$

From (27), we have

$$\begin{aligned} &\mathbb{E}[|S_k(\phi_k^T \theta + e_{k+1}) - \mathbb{E}[S_k(\phi_k^T \theta + e_{k+1}) \mid \mathscr{F}_k]|^{2+\eta} \mid \mathscr{F}_k] \\ =&O(\mathbb{E}[|S_k(\phi_k^T \theta + e_{k+1}) - S_k(\phi_k^T \theta)|^{2+\eta} \mid \mathscr{F}_k]) \\ &+ O(\mathbb{E}[|S_k(\phi_k^T \theta) - \mathbb{E}[S_k(\phi_k^T \theta + e_{k+1}) \mid \mathscr{F}_k]|^{2+\eta} \mid \mathscr{F}_k]) \\ =&O(\mathbb{E}[|e_{k+1}|^{2+\eta} \mid \mathscr{F}_k]). \end{aligned} \tag{28}$$

Thus, by Assumption 3, (26) is obtained.

Now, drawing inspiration from the analysis of classical least-squares in linear stochastic regression models (see e.g., [18], [19], [22]), we consider the following stochastic Lyapunov function: $V_k = \tilde{\theta}_k^T P_k^{-1} \tilde{\theta}_k$. By Algorithm 1 , we know that

$$\begin{aligned} V_{k+1} =& \tilde{\theta}_k^T P_{k+1}^{-1} \tilde{\theta}_k - 2a_k \beta_k \tilde{\theta}_k^T P_{k+1}^{-1} P_k \phi_k \psi_k \\ &+ a_k^2 \beta_k^2 \psi_k^T \phi_k^T P_k P_{k+1}^{-1} P_k \phi_k \psi_k \\ &+ 2a_k^2 \beta_k^2 \psi_k \phi_k^T P_k P_{k+1}^{-1} P_k \phi_k w_{k+1} \\ &- 2a_k \beta_k \phi_k^T P_k P_{k+1}^{-1} \tilde{\theta}_k w_{k+1} \\ &+ a_k^2 \beta_k^2 w_{k+1}^T \phi_k^T P_k P_{k+1}^{-1} P_k \phi_k w_{k+1}. \end{aligned} \tag{29}$$

Let us now analyze the right-hand-side (RHS) of (29) term by term. From (8)-(9), we know that

$$\tilde{\theta}_k^T P_{k+1}^{-1} \tilde{\theta}_k = \tilde{\theta}_k^T P_k^{-1} \tilde{\theta}_k + \frac{\beta_k^2}{\mu_k} \tilde{\theta}_k^T \phi_k \phi_k^T \tilde{\theta}_k. \tag{30}$$

Moreover, by (8) again, we know that

$$\begin{aligned} &a_k P_{k+1}^{-1} P_k \phi_k \\ =&a_k \left(I + \frac{\beta_k^2}{\mu_k} \phi_k \phi_k^T P_k\right) \phi_k \\ =&a_k \phi_k \left(1 + \frac{\beta_k^2}{\mu_k} \phi_k^T P_k \phi_k\right) = \frac{\phi_k}{\mu_k}. \end{aligned} \tag{31}$$

Hence, we have

$$\begin{aligned} &a_k \beta_k \tilde{\theta}_k^T P_{k+1}^{-1} P_k \phi_k \psi_k \\ =& \frac{\beta_k}{\mu_k} \tilde{\theta}_k^T \phi_k \psi_k \geq \frac{\beta_k^2}{\mu_k} (\tilde{\theta}_k^T \phi_k)^2, \end{aligned} \tag{32}$$

Similarly, by (31),

$$\begin{aligned}&a_k^2\beta_k^2\phi_k^T P_k P_{k+1}^{-1} P_k \phi_k \psi_k^2\\=&a_k\frac{\beta_k^2}{\mu_k}\phi_k^T P_k\phi_k\psi_k^2 \le a_k\frac{\beta_k^2}{\mu_k}\overline{g}_k\phi_k^T P_k\phi_k(\tilde{\theta}_k^T\phi_k)\psi_k\end{aligned}\tag{33}$$

where $\overline{g}_k$ is defined in (6). Now, substituting (30), (32) and (33) into (29) we get

$$\begin{aligned}V_{k+1}\le&\tilde{\theta}_k^T P_k^{-1}\tilde{\theta}_k-(\frac{\beta_k}{\mu_k}-a_k\frac{\beta_k^2}{\mu_k}\overline{g}_k\phi_k^T P_k\phi_k)\tilde{\theta}_k^T\phi_k\psi_k\\&+2a_k\frac{\beta_k^2}{\mu_k}\psi_k\phi_k^T P_k\phi_k w_{k+1}-2\frac{\beta_k}{\mu_k}\phi_k^T\tilde{\theta}_k w_{k+1}\\&+a_k\frac{\beta_k^2}{\mu_k}\phi_k^T P_k\phi_k w_{k+1}^2\end{aligned}\tag{34}$$

Notice that by the definition of β_k, we have $a_k\beta_k\overline{g}_k\phi_k^T P_k\phi_k\le\frac{1}{2}$, thus we can obtain

$$\begin{aligned}&(\frac{\beta_k}{\mu_k}-a_k\frac{\beta_k^2}{\mu_k}\overline{g}_k\phi_k^T P_k\phi_k)\tilde{\theta}_k^T\phi_k\psi_k\\\ge&\frac{\beta_k}{2\mu_k}(\tilde{\theta}_k^T\phi_k)\psi_k\ge\frac{\beta_k^2}{2\mu_k}(\tilde{\theta}_k^T\phi_k)^2.\end{aligned}\tag{35}$$

By (35) and summing up both sides of (34) from 0 to n, we have

$$\begin{aligned}V_{n+1}\le&\tilde{\theta}_0^T P_0^{-1}\tilde{\theta}_0-\sum_{k=0}^n\frac{\beta_k^2}{2\mu_k}(\tilde{\theta}_k^T\phi_k)^2\\&-\sum_{k=0}^n 2\frac{\beta_k}{\mu_k}\phi_k^T\tilde{\theta}_k w_{k+1}+\sum_{k=0}^n 2a_k\frac{\beta_k^2}{\mu_k}\psi_k\phi_k^T P_k\phi_k w_{k+1}\\&+\sum_{k=0}^n a_k\frac{\beta_k^2}{\mu_k}\phi_k^T P_k\phi_k w_{k+1}^2.\end{aligned}\tag{36}$$

We now analyze the last three terms in (36) which are related to the martingale difference sequence $\{w_k,\mathscr{F}_k\}$.

Denote

$$\tilde{S}_n=\sqrt{\sum_{k=0}^n\left(a_k\frac{\beta_k^2}{\mu_k}\phi_k^T P_k\phi_k\right)^2\|\psi_k\|^2}.\tag{37}$$

By (31) and Lemma 2, we have

$$\begin{aligned}&\|\sum_{k=0}^n 2a_k\frac{\beta_k^2}{\mu_k}\psi_k\phi_k^T P_k\phi_k w_{k+1}\|\\=&O\left(\tilde{S}_n\log^{\frac{1}{2}+\varepsilon}\left(\tilde{S}_n^2+e\right)\right)\\=&o\left(\sum_{k=0}^n\frac{\beta_k^2}{\mu_k}(\tilde{\theta}_k^T\phi_k)^2\right)+O(1),\quad\text{a.s.}\quad\forall\varepsilon>0.\end{aligned}\tag{38}$$

where we have used the fact that $a_k\frac{\beta_k^2}{\mu_k}\overline{g}_k\phi_k^T P_k\phi_k\le\frac{\beta_k}{2\mu_k}$.

Also, by Lemma 2 again, we know that

$$\begin{aligned}&\sum_{k=0}^n\frac{\beta_k}{\mu_k}\phi_k^T\tilde{\theta}_k w_{k+1}\\=&O\left(\sum_{k=0}^n\frac{\beta_k^2}{\mu_k^2}(\tilde{\theta}_k^T\phi_k)^2\right)^{\frac{1}{2}+\varepsilon}\\=&o\left(\sum_{k=0}^n\frac{\beta_k^2}{\mu_k}(\tilde{\theta}_k^T\phi_k)^2\right)+O(1)\quad\text{a.s.}\quad\forall\varepsilon>0\end{aligned}\tag{39}$$

As for the last term of right side of (36), a similar analysis as Remark 2 in [26] gives

$$\sum_{k=0}^n a_k\frac{\beta_k}{\mu_k}\phi_k^T P_k\phi_k=O(1).\tag{40}$$

Moreover, from C_r-inequality and Lyapunov inequality, we have for every $\delta\in(2,\min(\eta,4)]$,

$$\sup_k\mathbb{E}\left[\left|w_{k+1}^2-\mathbb{E}[w_{k+1}^2\mid\mathscr{F}_k]\right|^{\frac{\delta}{2}}\mid\mathscr{F}_k\right]<\infty,\quad\text{a.s}~.\tag{41}$$

Denote $\Lambda_n=(\sum_{k=0}^n\left(a_k\frac{\beta_k}{\mu_k}\phi_k^T P_k\phi_k\right)^{\frac{\delta}{2}})^{\frac{2}{\delta}}$, by Lemma 2 and letting $\alpha=2$, we get

$$\begin{aligned}&\sum_{k=0}^n a_k\frac{\beta_k}{\mu_k}\phi_k^T P_k\phi_k\left\{w_{k+1}^2-\mathbb{E}[w_{k+1}^2\mid\mathscr{F}_k]\right\}\\=&O\left(\Lambda_n\log^{\frac{2}{\delta}+\varepsilon}(\Lambda_n+e)\right)=O(1),\quad\text{a.s.}~\forall\varepsilon>0\end{aligned}\tag{42}$$

where the last equality is from (40). Hence, from (40) and (42)

$$\begin{aligned}&\sum_{k=0}^n a_k\frac{\beta_k}{\mu_k}\phi_k^T P_k\phi_k w_{k+1}^2\\\le&\sum_{k=0}^n a_k\frac{\beta_k}{\mu_k}\phi_k^T P_k\phi_k\left(w_{k+1}^2-\mathbb{E}[w_{k+1}^2\mid\mathscr{F}_k]\right)\\&+\sum_{k=0}^n a_k\frac{\beta_k}{\mu_k}\phi_k^T P_k\phi_k\mathbb{E}[w_{k+1}^2\mid\mathscr{F}_k]=O(1),\quad\text{a.s.}\end{aligned}\tag{43}$$

Substituting (38), (39) and (43) into (36) we finally obtain (15).

B. Proof of Theorem 2.

We note that

$$\lambda_{min}\left\{P_{n+1}^{-1}\right\}\|\tilde{\theta}_{n+1}\|^2\le\tilde{\theta}_{n+1}^T P_{n+1}^{-1}\tilde{\theta}_{n+1},\tag{44}$$

then Theorem 2 follows from Lemma 1 immediately.

C. Proof of Theorem 1.

From Theorem 2, $\{\|\hat{\theta}_k\|,k\ge0\}$ is bounded almost surely. Let $\bar{B}$ be the upper bound of $\|\hat{\theta}_k\|$, i.e., $\|\hat{\theta}_k\|\le\bar{B}$ for every $k\ge0$. For every $t\in\mathbb{R}^+$, define the function $\underline{g}(t)=\inf_{|x|\le t(\bar{B}+B(D)),k\ge0}G_k'(x)$, $\overline{g}(t)=\sup_{|x|\le t(\bar{B}+B(D)),k\ge0}G_k'(x)$, and $f(t)=\min(\underline{g}(t)~,~\frac{\mu_0}{2\|P_0\|\overline{g}(t)t^2+1})$. From Assumption 3, we can easily find $f(\cdot)$ is a non-increasing and positive function. Besides, we have for every $k\ge0$,

$$f(\|\phi_k\|)\le\beta_k.\tag{45}$$

From (45) and (15), we have

$$\sum_{k=0}^{n}(f(\|\phi_k\|))^2(\phi_k^T\tilde{\theta}_k)^2 = O(\mu_k) = O(\log^{1+\delta}k),\ a.s. \quad (46)$$

If $\lim\limits_{t\to\infty} f(t) \ge \gamma > 0$, we will have

$$\sum_{k=0}^{n}R_k = \sum_{k=0}^{n}(\phi_k^T\tilde{\theta}_k)^2 = O(\log^{1+\delta}n),\ a.s., \quad (47)$$

otherwise, we will have

$$\lim_{t\to\infty} f(t) = 0. \quad (48)$$

In this case, let

$$h(s) = \inf\{t \ge 0 : f(t) = s\},\ s \in \{f(t) : t \ge 0\}, \quad (49)$$

we also have $h(\cdot)$ is a non-increasing function. Besides, by (48) and the fact that $f(\cdot)$ is positive, we have $\lim\limits_{k\to\infty} h(\frac{f_0}{\sqrt{k^\gamma}}) = \infty$, *a.s.*, where f_0 is a positive number such that $h(f_0) > 0$. Thus, for every $0 < \gamma < 1$, we can obtain

$$\begin{aligned}&\sum_{k=1}^{n}(\phi_k^T\tilde{\theta}_k)^2 I_{\{(f(\|\phi_k\|))^2 < f_0^2 k^{-\gamma}\}}\\ \le&\sum_{k=1}^{n}(\phi_k^T\tilde{\theta}_k)^2 I_{\{\|\phi_k\| \ge h(\frac{f_0}{\sqrt{k^\gamma}})\}} \le \sum_{k=1}^{n}(\phi_k^T\tilde{\theta}_k)^2\left(\frac{\|\phi_k\|}{h(\frac{f_0}{\sqrt{k^\gamma}})}\right)^{\varepsilon}\\ =&O\left(\left[\sum_{k=1}^{n}\|\phi_k\|^\alpha\right]^{\frac{2+\varepsilon}{\alpha}}\left[\sum_{k=1}^{n}\left(\frac{1}{h(\frac{f_0}{\sqrt{k^\gamma}})}\right)^{\frac{\alpha\varepsilon}{\alpha-(2+\varepsilon)}}\right]^{1-\frac{2+\varepsilon}{\alpha}}\right)\\ =&O(n^{\frac{2+\varepsilon}{\alpha}})o(n^{\frac{\alpha-2-\varepsilon}{\alpha}}) = o(n),\ a.s.,\ \forall 0 < \varepsilon < \alpha - 2,\end{aligned} \quad (50)$$

where $I_{\{\cdot\}}$ is the indicator function, defined by

$$I_A = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in A^c \end{cases}. \quad (51)$$

Moreover,

$$\begin{aligned}&\sum_{k=0}^{n}(\phi_k^T\tilde{\theta}_k)^2 I_{\{(f(\|\phi_k\|))^2 \ge f_0^2 k^{-\gamma}\}}\\ \le&\sum_{k=0}^{n}(f(\|\phi_k\|))^2(\phi_k^T\tilde{\theta}_k)^2\frac{k^\gamma}{f_0^2} I_{\{(f(\|\phi_k\|))^2 \ge k^{-\gamma}\}}\\ \le&[\sum_{k=0}^{n}(f(\|\phi_k\|))^2(\phi_k^T\tilde{\theta}_k)^2]\frac{n^\gamma}{f_0^2} = o(n),\quad a.s.\end{aligned} \quad (52)$$

Consequently, we have

$$\sum_{k=0}^{n}R_k = \sum_{k=0}^{n}(\phi_k^T\tilde{\theta}_k)^2 = o(n),\ a.s. \quad (53)$$

D. Proof of Corollary 1.

For every $\|v\| = 1$, $M > 0$, we have

$$\begin{aligned}&v^T\frac{1}{n}(\sum_{k=0}^{n}\phi_k\phi_k^T I_{\{\|\phi_k\| > M\}})v\\ \le&\frac{1}{n}\sum_{k=0}^{n}\|\phi_k\|^2 I_{\{\|\phi_k\| > M\}} \le \frac{1}{n}\sum_{k=0}^{n}\frac{\|\phi_k\|^\alpha}{M^{\alpha-2}}.\end{aligned} \quad (54)$$

By (45) and (54), we can obtain

$$\begin{aligned}&v^T\frac{1}{n}(\sum_{k=1}^{n}\beta_k^2\phi_k\phi_k^T I_{\{\|\phi_k\| \le M\}} + P_0^{-1})v\\ \ge&v^T\frac{1}{n}(\sum_{k=1}^{n}f^2(\|\phi_k\|)\phi_k\phi_k^T I_{\{\|\phi_k\| \le M\}} + P_0^{-1})v\\ \ge&f^2(M)[v^T\frac{1}{n}(\sum_{k=1}^{n}\phi_k\phi_k^T + \frac{P_0^{-1}}{f^2(M)})v - \frac{1}{n}\sum_{k=1}^{n}v^T\phi_k\phi_k^T v I_{\{\|\phi_k\| > M\}}]\\ \ge&f^2(M)[\frac{1}{n}\lambda_{\min}\{\frac{P_0^{-1}}{f^2(M)} + \sum_{k=1}^{n}\phi_k\phi_k^T\} - \frac{1}{M^{\alpha-2}}\frac{1}{n}\sum_{k=1}^{n}\|\phi_k\|^\alpha].\end{aligned} \quad (55)$$

Let M be sufficiently large, we can easily obtain that

$$n = O(\lambda_{\min}\{\sum_{k=1}^{n}\beta_k^2\phi_k\phi_k^T + P_0^{-1}\}),\ a.s. \quad (56)$$

Hence, (21) is obtained from (56) and (20).

V. NUMERICAL SIMULATION

In this section, we give an example to illustrate and demonstrate the theoretical results obtained in this paper. Consider the model $(1)-(2)$ with $l_k = L_k = 0$ and $u_k = U_k = 15$ for $k \ge 0$. The regressors $\{\phi_k\}(m = 5)$ and observations $\{s_{k+1}\}$ are generated by the following dynamical system model:

$$\begin{cases} \phi_{k+1} &= A\phi_k + u_k \\ s_{k+1} &= S_k(\phi_k^T\theta + e_{k+1}) \end{cases}. \quad (57)$$

The input $u_k = (u_k^{(1)}, \cdots, u_k^{(5)})$, where $u_k^{(j)}$ are independent with the distribution $u_k^{(i)} \sim N(0,1)$ for any $k \ge 0$ and $i = 1, \cdots, 5$. The state matrix $A = diag[0.3, 0.2, 0.1, 0.4, 0.7]$, the parameter $\theta = [-1.2, 0.5, 1, -0.5, 1.5]^T$, and the noise sequence $\{e_{k+1}\}$ is i.i.d with normal distribution $N(0,1)$. Let $\phi_0 = 0$, it can be verified that the independence of regressors is not satisfied. Besides, the boundedness of regressor ϕ_k is also not satisfied since u_k is unbounded. Moreover, we can easily verify that $n = O(\lambda_{\min}\{\sum\limits_{k=0}^{n}\beta_k^2\phi_k\phi_k^T\})$. Fig. 1 shows the trajectory of $\frac{1}{n}\sum\limits_{k=0}^{n}R_k$, which converges to 0 by Theorem 1. For the parameter estimation, the estimate error $\|\tilde{\theta}_n\|^2$ will convergent to 0 with the convergence rate $O\left(\frac{\log^2 n}{n}\right)$ by Corollary 1,which is verified by the boundedness of the trajectory of $\frac{n\|\tilde{\theta}_n\|^2}{\log^2 n}$ in Fig. 2.

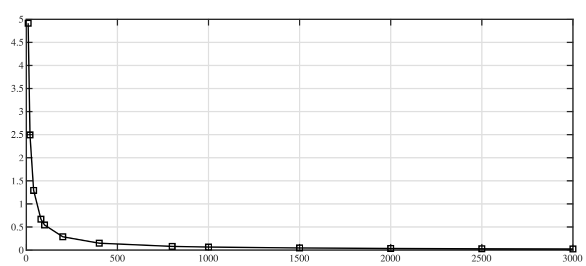

Fig. 1. A trajectory of $\frac{1}{n}\sum_{k=1}^{n}R_k$

Fig. 2. A trajectory of $\frac{n\|\tilde{\theta}_n\|^2}{\log^2 n}$

VI. CONCLUSIONS

In this paper, we have introduced a new adaptive algorithm for stochastic systems with saturated output observations. The global convergence of the parameter estimates has been established under possibly unbounded, non-independent, and nonstationary conditions on regression signals. Moreover, the averaged regret of adaptive predictors has also been shown to converge to 0 for possibly unbounded regressors without requiring any excitation conditions. These results make it possible for our theory to be applicable to feedback control systems and lay a foundation for possible generalization to related identification problems. For further exploration, several challenges still need to be addressed, for example, how to solve adapted control problems with saturated observations in stochastic dynamical control systems, and how to establish global convergence for adaptive estimation algorithms in more complex stochastic nonlinear models such as multi-layer neural networks, among others.

REFERENCES

[1] J. Sun, Y. W. Kim and L. W, Aftertreatment control and adaptation for automotive lean burn engines with HEGO sensors, *International Journal of Adaptive Control and Signal Processing*, vol. 18, no. 2, pp. 145-166, Mar. 2004.

[2] J. Tobin, Estimation of relationships for limited dependent variables, *Econometrica: Journal of the Econometric Society,* vol. 26, no. 1, pp. 24-36, Jan. 1958.

[3] H. F. Grip, L. Imsland, T. A. Johansen, J. C. Kalkkuhl, and A. Suissa, Vehicle sideslip estimation, *IEEE Control Systems Magazine*, vol. 29, no. 5, pp. 36-52, Oct. 2009.

[4] M. S. Jeon and J. H. Lee, Estimation of willingness-to-pay for premium economy class by type of service, *Journal of Air Transport Management,* vol. 84, May. 2020.

[5] A. Bykhovskaya, Time series approach to the evolution of networks: prediction and estimation, *Journal of Business* & *Economic Statistics,* vol. 41, pp. 170-183, 2023.

[6] F. Wang, L. Zhang, and L. Guo, Applications of nonlinear recursive identification theory in sentencing data analyses, *SCIENTIA SINICA Informationis,* vol. 52, no. 10, 2022, https://doi.org/10.1360/SSI-2022-0325.

[7] B. W. Turnbull, The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society: Series B,* vol. The Annals of Statistics, vol. 38, no. 3, pp. 290-295, 1976.

[8] J. L. Powell, Least absolute deviations estimation for the censored regression model, *Journal of Econometrics,* vol. 25, no. 3, pp. 303-325, Jul. 1984.

[9] S. Mei, Y. Bai, and A. Montanari, The Landscape of empirical risk for nonconvex losses, *The Annals of Statistics,* vol. 46, no. 6A, pp. 2747-2774, Dec. 2018.

[10] L. Y. Wang, J. F. Zhang, and G. G. Yin, System identification using binary sensors, *IEEE Transactions on Automatic Control,* vol. 48, no. 11, pp. 1892-1907, Nov. 2003.

[11] L. Y. Wang and G. G. Yin, Asymptotically efficient parameter estimation using quantized output observations, *Automatica,* vol. 43, no. 7, pp. 1178-1191, Jul. 2007.

[12] H. Zhang, T. Wang, and Y. Zhao, Asymptotically efficient recursive identification of FIR systems with binary-valued observations, *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 51, pp. 2687-2700, 2019.

[13] J. Guo, and Y. Zhao, Recursive projection algorithm on FIR system identification with binary-valued observations, *Automatica,* vol. 49, pp. 3396–3401, 2013.

[14] Y. Wang, Y. Zhao, J. F. Zhang, and J. Guo, A unified identification algorithm of FIR systems based on binary observations with time-varying thresholds, *Automatica,* vol. 135, 2022.

[15] L. Zhang, Y. Zhao, and L. Guo, Identification and adaptation with binary-valued observations under non-persistent excitation, *Automatica,* vol. 138, Apr. 2022.

[16] L. Zhang and L. Guo, Adaptation identification with guaranteed performance under saturated-observation and non-persistent excitation, 2022, Arxiv: 2207.02422.

[17] L. Ljung, Consistency of the least-squares identification method, *IEEE Transactions on Automatic Control,* vol. 21, pp. 779-781, 1976.

[18] J. B. Moore, On strong consistency of least squares identification algorithm, *Automatica,* vol. 14, no. 5, pp. 505–509, Sep. 1978.

[19] T. L. Lai and C. Z. Wei, Least squares estimates in stochastic regression models with applications to identification and control dynamic systems, *Annals of Statistics,* vol. 10, no. 1, pp. 154–166, March 1982.

[20] H. F. Chen, Strong consistency and convergence rate of least squares identification, *Sci. Sinica, Ser. A,* vol. 25, no. 7, pp. 771–784, 1982.

[21] H. F. Chen and L. Guo, Convergence rate of least-squares identification and adaptive control for stochastic systems, *int. J. Control,* vol. 44, no. 5, pp. 1459–1476, Nov. 1986.

[22] L. Guo, Convergence and logarithm laws of self-tuning regulators, *Automatica,* vol. 31, no. 3, pp. 435-450, Mar. 1995.

[23] T. L. Lai and C. Z. Wei, Extended least squares and their applications to adaptive control and prediction in linear systems, *IEEE Transactions on Automatic Control,* vol. 31, no. 10, pp. 898–906, Oct. 1986.

[24] H. F. Chen and L. Guo, *Identification and Stochastic Adaptive Control,* Birkhäsuser, Boston, 1991.

[25] L. Guo and H. F. Chen, The Astrom-Wittenmark self-tuning regulator revisited and ELS-based adaptive trackers, *IEEE Transactions on Automatic Control,* vol. 36, pp. 802-812, 1991.

[26] L. Guo, Self-convergence of weighted least-squares with applications to stochastic adaptive control, *IEEE Transactions on Automatic Control,* vol. 41, pp. 79-89, 1996.