Ulam Quaterly – Volume 3, Number 2, 1996

Self-convergence of Weighted Least Squares for Continuous-time ARMAX Model

A.J. Gao[1]

Department of Mathematics
The University of Kansas
Lawrence, KS 66045

Abstract

In this paper we discuss a weighted least squares algorithm for the following continuous-time model:

$$\mathbf{A}(\mathbf{S})y_t = \mathbf{S}\mathbf{B}(\mathbf{S})u_t + \mathbf{C}(\mathbf{S})v_t$$

where $\mathbf{S}$ denotes the integral operator, i.e., $\mathbf{S}y_t = \int_0^t y_s ds$ and $\mathbf{A}(\mathbf{S}), \mathbf{B}(\mathbf{S})$ and $\mathbf{C}(\mathbf{S})$ are matrix polynomials in the integral operator $\mathbf{S}$. Similar to discrete **WLS**([5], [6]), the self-convergence property, the almost sure convergence of the estimate via the continuous **WLS** (**CWLS**) to the true parameter are established. The simulation results for the discrete **WLS** and **ELS**, the **CWLS** and the continuous **ELS** algorithms are given in this paper. The results presented here can be extended to state-space models with continuous time.

1. Introduction

The least squares(**LS**) algorithm and the stochastic gradient algorithm (**SG**) ([1], [2], [3] and [4]) are two important algorithms in the study of recursive identification and adaptive control of discrete-time ARMAX model. Generally, if we focus our attention on the parameter consistency, we choose

[1]The author thanks Prof. B. Pasik-Duncan for suggesting the problem and useful discussions that improved the quality of this paper.

the least squares algorithm because it has better rate of convergence. For adaptive tracking, the stochastic gradient algorithm seems to have a better tracking rate and it is more practical for analyzing computational needs. Recently the extended least squares (**ELS**) algorithm was used to solve the Aström-Wittenmark self-tuning problem. The **ELS** consistency and adaptive tracking optimality results were established. Based on the previous results, the weighted least squares(**WLS**) algorithm was proposed to solve the Aström-Wittenmark adaptive tracking problem([5]). The discrete-time **WLS** has the following advantages: it is almost surely bounded when a weighting sequence is suitably chosen, the **WLS** algorithm can perform as well as the **ELS** for parameter estimation(see the simulation results in Section 4) and it turns out that under some suitable choices of weighting sequence that a better rate can be obtained for adaptive tracking by using the **WLS** than by using the **SG** algorithm. Later in [6], the author proved the self-convergence property of one special kind of **WLS**, which was used to the adaptive pole-placement and LQG control problems. It also will be our further research subject to do similar adaptive control problems by using **CWLS** algorithm.

Although most of the literature on identification and adaptive control of dynamical systems takes place in discrete-time, the study of systems which evolve in continuous-time has recently gained in popularity, due in part to the fact that many physical systems evolve naturally in continuous-time. Continuous-time analysis also can be used to resolve questions which arise in discrete-time models due to very small time steps.

In this paper we analyze the **WLS** algorithm (ref. [11]) for a continuous-time ARMAX model as in [7] or [8]. We prove that the **CWLS** has self-convergence property, the almost sure convergence of the estimate via the **CWLS** to the true parameter. In the simulation section, we show the convergence results for both discrete **ELS**, **WLS** and **CWLS** and **CELS**. Further, we show that the **WLS** converges but the **ELS** diverges for the discrete time example constructed in [10]. Similarly, we show some convergence results for both **CWLS** and **CELS**. We also give one example where **CWLS** converges but **CELS** diverges.

2. Continuous-time WLS algorithm

Consider the system described by the following multi-variable stochastic integral equation:

$$\mathbf{A}(\mathbf{S})y_t = \mathbf{S}\mathbf{B}(\mathbf{S})u_t + \mathbf{C}(\mathbf{S})v_t \tag{1}$$

where $\mathbf{S}$ denotes the integral operator, i.e. $\mathbf{S}y_t = \int_0^t y_s ds$ and $\mathbf{A}(\mathbf{S}), \mathbf{B}(\mathbf{S})$ and $\mathbf{C}(\mathbf{S})$ are matrix polynomials in the integral operator $\mathbf{S}$.

Let $\mathcal{F}_t$ be a family of nondecreasing σ-algebras defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. y_t, u_t adapted to $\mathcal{F}_t$ are the m-dimensional output and the l-dimensional input, respectively. v_t is the system noise that is generated by a known filter $\mathbf{D}^{-1}(\mathbf{S})$, from a standard Wiener process $(w_t, \mathcal{F}_t)$:

$$\mathbf{D}(\mathbf{S})v_t = w_t, \ t \geq 0. \tag{2}$$

Assume that $\mathbf{A}(\mathbf{S}), \mathbf{B}(\mathbf{S})$ and $\mathbf{C}(\mathbf{S})$ are matrix polynomials in $\mathbf{S}$ with unknown coefficients and

$$\mathbf{A}(\mathbf{S}) = \mathbf{I} + \mathbf{A_1 S} + ... + \mathbf{A_p S^p}, \ p \geq 0, \tag{3}$$

$$\mathbf{B}(\mathbf{S}) = \mathbf{B_1} + \mathbf{B_2 S} + ... + \mathbf{B_q S^{q-1}}, \ q \geq 1, \tag{4}$$

$$\mathbf{C}(\mathbf{S}) = \mathbf{I} + \mathbf{C_1 S} + ... + \mathbf{C_r S^r}, \ r \geq 1, \tag{5}$$

$$\mathbf{D}(\mathbf{S}) = \mathbf{I} + \mathbf{D_1 S} + ... + \mathbf{D_r S^r}. \tag{6}$$

Let $a = (a_t, t \geq 0)$ be a random process adapted to $\mathcal{F}_t$, positive, nonincreasing and $a_t \leq 1$ for $t \geq 0$. The process $a = (a_t)$ will be called a weighting process. Let θ_t be an estimator of θ where

$$\theta^T = [-\mathbf{A}_1, ... - \mathbf{A}_p, \mathbf{B}_1, ...\mathbf{B}_q, \mathbf{C}_1, ...\mathbf{C}_r] \tag{7}$$

In order to estimate θ, we propose the following continuous weighted least squares algorithm(**CWLS**):

$$d\theta_t = a_t \mathbf{R}_t^{-1} \Phi_t \mathbf{D}(\mathbf{S})(dy_t^T - \Phi_t^T \theta_t dt), \, \theta_0 = 0 \tag{8}$$

$$\Phi_t = [y_t^T, \mathbf{S}y_t^T, ...\mathbf{S}^{p-1}y_t^T, u_t^T, \mathbf{S}u_t^T, ...\mathbf{S}^{q-1}u_t^T, \hat{v}_t, \mathbf{S}\hat{v}_t^T, ...\mathbf{S}^{r-1}\hat{v}_t^T]^T \tag{9}$$

$$\Phi_t^0 = [y_t^T, \mathbf{S}y_t^T, ...\mathbf{S}^{p-1}y_t^T, u_t^T, \mathbf{S}u_t^T, ...\mathbf{S}^{q-1}u_t^T, v_t, \mathbf{S}v_t^T, ...\mathbf{S}^{r-1}v_t^T]^T \tag{10}$$

$$d\mathbf{R}_t^{-1} = -a_t \mathbf{R}_t^{-1} \Phi_t \Phi_t^T \mathbf{R}_t^{-1} dt \tag{11}$$

$$\hat{v}_t = y_t - \mathbf{S}\theta_t^T \Phi_t \tag{12}$$

where $\mathbf{R}$ is a positive definite matrix and $\hat{v}_t$ is the predicted value for noise v_t .

3. Main Results

Let

$$f_t = f_t(a) = a_t \Phi_t^T \mathbf{R}_t^{-1} \Phi_t, t \geq 0 \tag{1}$$

$$\Delta = \int_0^\infty a_t f_t dt \tag{2}$$

where $a = (a_t, t \geq 0)$ is a weighting process(see [5]).The process is said to be admissible if Δ is finite. In all the following, we assume a is admissible. It is important to notice that the **CWLS** doesn't include the continuous **ELS** as a special case since the weighting sequence $a_t = 1, t \geq 0$, is not admissible.

Remark. The weighting process does exist. For example, let G be a family of continuous non-increasing functions: $g(.) : R^+ \to R^+$ such that $xg(x) \to 0$ when $x \to \infty$ $\int_c^\infty g(x)dx < \infty$, for $c > 0$. Then we can choose $a_t = g(t)$ or $a_t = g(log(tr(R_t)))$.

Lemma. For the model (1), assume that $\mathbf{D}(\mathbf{S})$ is stable and $\mathbf{D}(\mathbf{S})\mathbf{C}^{-1}(\mathbf{S}) - \mathbf{I}/2$ is strictly positive real. Then the **CWLS** (8) -(12) has the following properties:

$$\| \mathbf{R}_t^{1/2}(\theta_t - \theta) \|^2 \leq M \tag{3}$$

$$\int_0^\infty a_t \| (\theta_t - \theta)^T \Phi_t \|^2 \, dt < \infty \tag{4}$$

$$\int_0^\infty a_t \| \Phi_t - \Phi_t^0 \|^2 \, dt < \infty \tag{5}$$

where $M = k + \Delta$, k is a positive constant and Δ is defined in (2).

Remark. The properties (3) and (4) were presented in [11] already. It is convenient to list all them here together with our other new results.

Proof of Lemma.

For $t \geq 0$, let

$$\bar{\theta_t} = \theta - \theta_t \tag{6}$$

$$\bar{v_t} = v_t - \hat{v}_t \tag{7}$$

$$\bar{\Phi}_t = \Phi_t^0 - \Phi_t. \tag{8}$$

From system equation (1) we know

$$y_t = \mathbf{S}\theta^T\Phi_t^0 + v_t \tag{9}$$

$$\begin{aligned} dy_t &= \theta^T\Phi_t^0 + dv_t = \theta^T(\Phi_t^0 - \Phi_t) + \theta^T\Phi_t + dv_t \\ &= \theta^T\bar{\Phi}_t + \theta^T\Phi_t + dv_t \end{aligned} \tag{10}$$

$$\begin{aligned} \theta^T\bar{\Phi}_t &= dy_t - \theta^T\Phi_t - dv_t = dy_t - \theta_t^T\Phi_t + (\theta_t - \theta)^T\Phi_t dt - dv_t \\ &= d\hat{v}_t - (\theta - \theta_t)^T\Phi_t dt - dv_t(by(12)) \\ &= -d(v_t - \hat{v}_t) - \bar{\theta}_t^T\Phi_t dt = -d\bar{v}_t - \bar{\theta}_t^T\Phi_t dt. \end{aligned} \tag{11}$$

Hence

$$\begin{aligned} \theta^T\bar{\Phi}_t dt + d\bar{v}_t &= -\bar{\theta}_t^T\Phi_t dt \\ \theta^T\bar{\Phi}_t + \frac{d\bar{v}_t}{dt} &= -\bar{\theta}_t^T\Phi_t \end{aligned} \tag{12}$$

By $(6), (7)$ we get, $\bar{\Phi}_t = [0, ...0, 0, ...0, \bar{v}_t{}^T, \mathbf{S}\bar{v}_t{}^T, ...\mathbf{S}^{r-1}\bar{v}_t{}^T]^T$. Therefore (12) becomes

$$\mathbf{C}(\mathbf{S})\frac{d\bar{v}_t}{dt} = -\bar{\theta}_t^T\Phi_t. \tag{13}$$

Define

$$h_t = \bar{\theta}_t^T\Phi_t, g_t = \frac{\mathbf{C}(\mathbf{S}) - \mathbf{D}(\mathbf{S})}{\mathbf{S}}\bar{v}_t + h_t/2 \tag{14}$$

Then (13) becomes

$$\begin{aligned} \mathbf{C}(\mathbf{S})\frac{d\bar{v}_t}{dt} &= -h_t \\ \frac{d\bar{v}_t}{dt} &= -\mathbf{C}^{-1}(\mathbf{S})h_t \end{aligned} \tag{15}$$

$$\begin{aligned} g_t &= \frac{\mathbf{C}(\mathbf{S}) - \mathbf{D}(\mathbf{S})}{\mathbf{S}}\bar{v}_t + h_t/2 = \frac{\mathbf{C}(\mathbf{S}) - \mathbf{D}(\mathbf{S})}{\mathbf{S}}\mathbf{S}(-\mathbf{C}^{-1}(\mathbf{S})h_t) + h_t/2 \\ &= -h_t + \mathbf{D}(\mathbf{S})\mathbf{C}^{-1}(\mathbf{S})h_t + h_t/2 = (\mathbf{D}(\mathbf{S})\mathbf{C}^{-1}(\mathbf{S}) - \mathbf{I}/2)h_t \end{aligned} \tag{16}$$

From algorithms $(8) - (12)$ and (6)

$$\begin{aligned} d\bar{\theta}_t &= -a_t \mathbf{R}_t^{-1} \Phi_t \mathbf{D}(\mathbf{S})(dy_t^T - \Phi_t^T \theta_t dt) - a_t \mathbf{R}_t^{-1} \Phi_t \mathbf{D}(\mathbf{S}) d\hat{v_t}^T \ (by(12)) \\ &= -a_t \mathbf{R}_t^{-1} \Phi_t \mathbf{D}(\mathbf{S})(dv_t - d\bar{v_t})^T = -a_t \mathbf{R}_t^{-1} \Phi_t (\mathbf{D}(\mathbf{S}) dv_t - \mathbf{D}(\mathbf{S}) d\bar{v_t})^T \\ &= -a_t \mathbf{R}_t^{-1} \Phi_t (dw_t - d\bar{v_t} - \mathbf{D}_1 \bar{v_t} dt - \ldots - \mathbf{D}_r \mathbf{S}^{r-1} \bar{v_t} dt)^T. \end{aligned} \tag{17}$$

By (15)

$$\begin{aligned} \mathbf{C}(\mathbf{S}) d\bar{v_t} &= -h_t dt \\ d\bar{v_t} + \mathbf{C}_1 \bar{v_t} dt + \ldots + \mathbf{C}_r \mathbf{S}^{r-1} \bar{v_t} dt &= -h_t dt \\ -d\bar{v_t} = \mathbf{C}_1 \bar{v_t} dt + \ldots + \mathbf{C}_r \mathbf{S}^{r-1} \bar{v_t} dt &+ h_t dt. \end{aligned} \tag{18}$$

Hence from (17) and (18)

$$\begin{aligned} d\bar{\theta}_t &= -a_t \mathbf{R}_t^{-1} \Phi_t (dw_t + \mathbf{C}_1 \bar{v_t} dt + \ldots + \mathbf{C}_r \mathbf{S}^{r-1} \bar{v_t} dt \\ &\quad + h_t dt - \mathbf{D}_1 \bar{v_t} dt - \ldots - \mathbf{D}_r \mathbf{S}^{r-1} \bar{v_t} dt)^T \\ &= -a_t \mathbf{R}_t^{-1} \Phi_t \left(h_t dt + \frac{\mathbf{C}(\mathbf{S}) - \mathbf{D}(\mathbf{S})}{\mathbf{S}} \bar{v_t} dt + dw_t\right)^T \qquad (19) \\ &= -a_t \mathbf{R}_t^{-1} \Phi_t (g_t dt + h_t dt/2 + dw_t)^T \quad ((\text{by } (13)). \end{aligned}$$

Now we want to estimate the differential of $tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t$. By the strictly positive real condition of $\mathbf{D}(\mathbf{S})\mathbf{C}^{-1}(\mathbf{S}) - \mathbf{I}/2$ there exist two positive numbers k_0 and k_1 such that

$$\int_0^t h_s^T (g_s - k_0 h_s) ds + k_1 > 0 \tag{20}$$

By using Ito formula and (19)

$$d(tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t)$$

$$= tr(\frac{\partial(\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t)}{\partial \bar{\theta}_t} d\bar{\theta}_t) + tr(\frac{\partial(\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t)}{\partial t})dt$$

$$+tr\{\frac{\partial^2(\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t)}{\partial^2 \bar{\theta}_t}(-a_t \mathbf{R}_t^{-1}\Phi_t)(-a_t \mathbf{R}_t^{-1}\Phi_t)^T\}/2dt$$

$$= tr\{(\bar{\theta}_t^T \mathbf{R}_t + \mathbf{R}_t \bar{\theta}_t)d\bar{\theta}_t\}$$

$$+tr\{\bar{\theta}_t^T \tfrac{d\mathbf{R}_t}{dt}\bar{\theta}_t\} + tr(a_t^2 \Phi_t^T \mathbf{R}_t^{-1}\Phi_t dt)$$

$$= 2tr(\bar{\theta}_t^T \mathbf{R}_t d\bar{\theta}_t) + tr\bar{\theta}_t^T a_t \Phi_t \Phi_t^T \bar{\theta}_t dt + a_t^2 \Phi_t^T \mathbf{R}_t^{-1}\Phi_t dt.$$

Paying attention to (19)

$$d(tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t)$$

$$= 2tr\{\bar{\theta}_t^T \mathbf{R}_t(-a_t \mathbf{R}_t^{-1}\Phi_t(g_t dt + h_t/2dt + dw_t)^T)\}$$

$$+tr\bar{\theta}_t^T a_t \Phi_t \Phi_t^T \bar{\theta}_t dt + a_t^2 \Phi_t^T \mathbf{R}_t^{-1}\Phi_t dt$$

$$= -2tra_t \bar{\theta}_t^T \Phi_t g_t^T dt - tra_t \bar{\theta}_t^T \Phi_t h_t^T dt$$

$$-2tra_t \bar{\theta}_t^T \Phi_t dw_t^T + tra_t h_t h_t^T dt + a_t^2 \Phi_t^T \mathbf{R}_t^{-1}\Phi_t dt$$

$$= -2a_t h_t^T g_t dt - 2a_t h_t^T dw_t + a_t^2 \Phi_t^T \mathbf{R}_t^{-1}\Phi_t dt.$$

By integrating both sides of the above equation and the definition of f_t we get

$$0 \le tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t = \bar{\theta}_0^T \mathbf{R}_0 \bar{\theta}_0 + \int_0^t a_s^2 \Phi_s^T \mathbf{R}_s^{-1}\Phi_s ds - 2\int_0^t a_s h_s^T g_s ds - 2\int_0^t a_s h_s^T dw_s$$

$$= \bar{\theta}_0^T \mathbf{R}_0 \bar{\theta}_0 + \int_0^t a_s^2 f_s ds - 2\int_0^t a_s h_s^T g_s ds - 2\int_0^t a_s h_s^T dw_s. \tag{21}$$

Denote

$$k_2 = \bar{\theta}_0^T \mathbf{R}_0 \bar{\theta}_0,\ M_t = \int_0^t a_s^2 \Phi_s^T \mathbf{R}_s^{-1}\Phi_s ds$$

where k_2 is a constant number and in the following it will be used to denote any constant. Further by admissible property of the process $a = (a_t, t \ge 0)$, it is obvious $0 \leq M_t \leq \Delta$, $t \geq 0$, for some random variable M.Also by (20) finally we get

$$\begin{aligned} 0 &\leq tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t \\ &\leq k_2 + \Delta - 2\int_0^t a_s h_s^T g_s ds - 2\int_0^t a_s h_s^T dw_s \\ &\leq k_2 + \Delta + k_1 - 2k_0 \int_0^t a_s h_s^T h_s ds - 2\int_0^t a_s h_s^T dw_s \\ &\leq k_2 + \Delta - 2k_0 \int_0^t a_s h_s^T h_s ds - 2\int_0^t a_s h_s^T dw_s. \end{aligned} \tag{22}$$

Applying the following Itô integral estimation ([9], Lemma 4):

$$\int_0^t \xi_s dw_s = O(1) + o((\int_0^t \| \xi_s \|^2 ds)^{1/2+k}) a.s., k > 0$$

for any predictable process $(\xi_s, \mathcal{F}_t)$. Let $k = 1/2$, we get

$$\begin{aligned} &0 \leq tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t \\ &\leq k_2 + \Delta - 2k_0 \int_0^t a_s \| h_s \|^2 ds + O(1) + o(\int_0^t \| a_s h_s \|^2 ds) \\ &\leq k_2 + M + O(1) \end{aligned} \tag{23}$$

because $-2k_0 \int_0^t a_s \| h_s \|^2 ds + o(\int_0^t \| a_s h_s \|^2 ds)$ is negative. Now we use M to denote the new random variable $k_2 + \Delta + O(1)$, hence

$$\| \mathbf{R}_t^{1/2}(\theta_t - \theta) \|^2 \leq M. \tag{24}$$

where M is a random variable, (3) is proved. Further, by (23), its obvious

$$tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t \leq k_2 + M - (2k_0 - \epsilon) \int_0^t a_s \| h_s \|^2 ds \tag{25}$$

for sufficient large t and sufficient small $\epsilon \geq 0$, hence

$$\int_0^t a_s \| h_s \|^2 ds \leq \frac{tr\bar{\theta}_t^T \mathbf{R}_t \bar{\theta}_t - (k_2 + \Delta)}{2k_0 - \epsilon} \tag{26}$$

(4) is proved.

By the strictly positive real condition of $\mathbf{D}(\mathbf{S})\mathbf{C}^{-1}(\mathbf{S}) - \mathbf{I}/2$ and stability of $\mathbf{D}(\mathbf{S})$, we know $\mathbf{D}(\mathbf{S})\mathbf{C}^{-1}(\mathbf{S})$ and $\mathbf{C}(\mathbf{S})$ are stable. Therefore by (26) and Lemma 1 in [7], we have

$$\int_0^\infty a_t \parallel \mathbf{C^{-1}}(\mathbf{S})h_t \parallel^2 dt < \infty \tag{27}$$

From (12) and ((15) we get

$$\theta^T \bar{\Phi}_t + \bar{\theta}_t^T \Phi_t = \mathbf{C^{-1}}(\mathbf{S})h_t. \tag{28}$$

Hence

$$\theta^T \bar{\Phi}_t = \mathbf{C^{-1}}(\mathbf{S})h_t - \bar{\theta}_t^T \Phi_t \tag{29}$$

and

$$\int_0^t a_s \parallel \theta^T \bar{\Phi}_s \parallel^2 ds \leq 2(\int_0^t a_s \parallel \mathbf{C^{-1}}(\mathbf{S})h_t \parallel^2 ds + \int_0^t a_s \parallel \theta_s^T \Phi_s \parallel^2 ds) \tag{30}$$

and (5) is proved.

Based on our Lemma, we have the following main results.

Theorem 1. Under the same conditions in Lemma, the **CWLS** described by (8)-(12) has self-convergence property, i.e., θ_t converges almost surely to a finite random vector $\bar{\theta}_t$.

Proof of Theorem 1. From (8) we get

$$\begin{aligned} d\theta_t &= - \quad a_t \mathbf{R}_t^{-1} \Phi_t \mathbf{D}(\mathbf{S})(dy_t^T - \Phi_t^T \theta_t dt) \\ &= - \quad a_t \mathbf{R}_t^{-1} \Phi_t \mathbf{D}(\mathbf{S})(\Phi_t^{0T} \theta dt + dv_t^T - \Phi_t^T \theta_t dt) \\ &= - \quad a_t \mathbf{R}_t^{-1} \Phi_t \mathbf{D}(\mathbf{S})(\bar{\Phi}_t^T \theta dt + \Phi_t^T \bar{\theta}_t dt + dv_t^T). \end{aligned} \tag{31}$$

By taking integration on both sides we get

$$\begin{aligned} \theta_0 - \theta_t &= \int_0^t a_s \mathbf{R}_s^{-1} \Phi_s \mathbf{D}(\mathbf{S}) \bar{\Phi}_s^T \theta ds \\ &+ \int_0^t a_s \mathbf{R}_s^{-1} \Phi_s \mathbf{D}(\mathbf{S}) \Phi_s^T \bar{\theta}_s ds \\ &+ \int_0^t a_s \mathbf{R}_s^{-1} \Phi_s \mathbf{D}(\mathbf{S}) dv_s^T. \end{aligned} \tag{32}$$

By Lemma and Schwarz inequality, we can estimate all three terms on the right side of (32) as following.

$$\begin{aligned}
&\int_0^t \| a_s \mathbf{R}_s^{-1} \Phi_s \mathbf{D}(\mathbf{S}) \bar{\Phi}_s^T \theta \| \, ds \\
&\leq (\int_0^t \| a_s \mathbf{R}_s^{-1} \Phi_s \Phi_s^T \mathbf{R}_s^{-1} \| \, ds \int_0^t \| a_s \mathbf{D}(\mathbf{S}) \bar{\Phi}_s^T \theta \|^2 \, ds)^{1/2} \\
&\leq (tr(\int_0^t a_s \mathbf{R}_s^{-1} \Phi_s \Phi_s^T \mathbf{R}_s^{-1} ds) \int_0^t \| a_s \mathbf{D}(\mathbf{S}) \bar{\Phi}_s^T \theta \|^2 \, ds)^{1/2} \qquad (33) \\
&\leq ((tr(\mathbf{R}_0^{-1}) - tr(\mathbf{R}_t^{-1})) \int_0^t \| a_s \mathbf{D}(\mathbf{S}) \bar{\Phi}_s^T \theta \|^2 \, ds)^{1/2} \\
&\leq (tr(\mathbf{R}_0^{-1}) \int_0^t \| a_s \mathbf{D}(\mathbf{S}) \bar{\Phi}_s^T \theta \|^2 \, ds)^{1/2}
\end{aligned}$$

together with (5) and the stability of $\mathbf{D}(\mathbf{S})$, it follows that the first term on the right side of (32) is convergent.

Similarly,

$$\begin{aligned}
&\int_0^t \| a_s \mathbf{R}_s^{-1} \Phi_s \mathbf{D}(\mathbf{S}) \Phi_s^T \bar{\theta}_s \| \, ds \\
&\leq (\int_0^t \| a_s \mathbf{R}_s^{-1} \Phi_s \Phi_s^T \mathbf{R}_s^{-1} \| \, ds \int_0^t \| a_s \mathbf{D}(\mathbf{S}) \Phi_s^T \bar{\theta}_s \|^2)^{1/2} \qquad (34) \\
&\leq (tr(\mathbf{R}_0^{-1}) \int_0^t \| a_s \mathbf{D}(\mathbf{S}) \Phi_s^T \bar{\theta}_s \|^2 \, ds)^{1/2}
\end{aligned}$$

and together with (4) and stability of $\mathbf{D}(\mathbf{S})$, we know the second term on the right side of (32) is convergent. Finally the last term on the right side of (32) converges because of (2) and

$$\int_0^t \| a_s^2 \mathbf{R}_s^{-1} \Phi_s \Phi_s^T \mathbf{R}_s^{-1} \| \, ds < tr(\mathbf{R}_0^{-1}).$$

Therefore the self-convergence property of **CWLS** has been proved.

Furthermore, by Lemma it follows that the **CWLS** is almost sure consistency to the true parameter.

Theorem 2. Under the same conditions as in Lemma, the **CWLS** is strongly consistent on $I = \{\lim_{t\to\infty} \lambda_{min} \mathbf{R}_t = \infty\}$ to the true parameters and

$$\| \theta_t - \theta \|^2 = O((\lambda_{min} \mathbf{R}_t)^{-1}) \ a.s. \qquad (35)$$

Theorem 2 follows directly from (3).

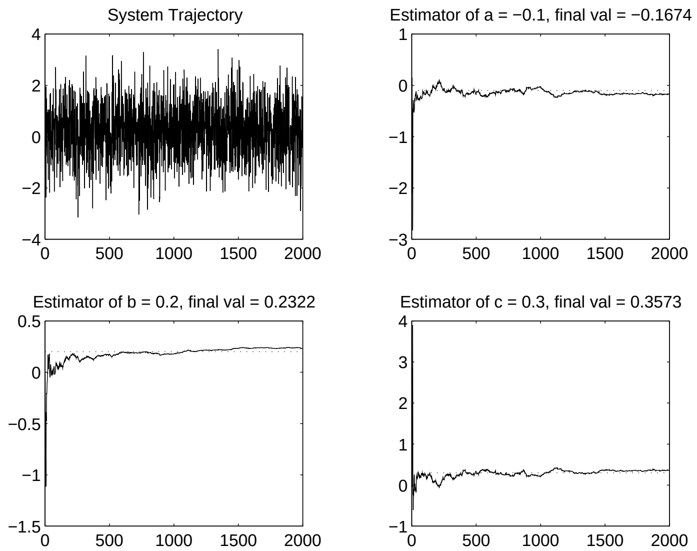

Fig. 1: Discrete-time ELS algorithm

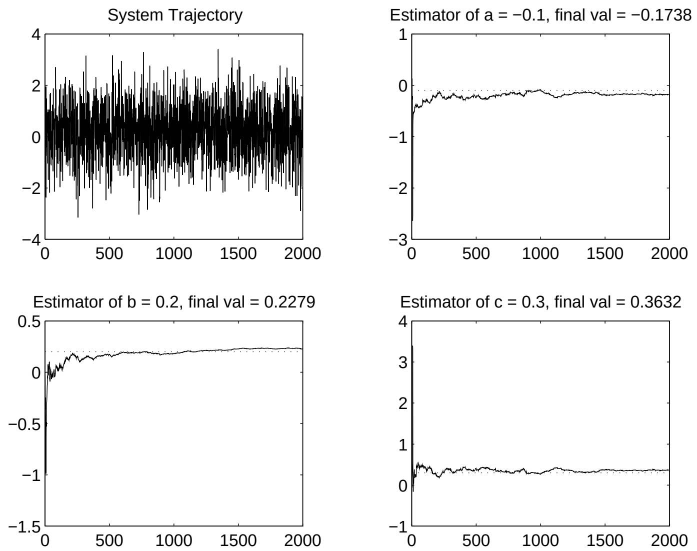

Fig. 2: Discrete-time WLS algorithm

4. Simulation

In this section we simulate both the **WLS** and the **CWLS** algorithms and also compare them with discrete **ELS** and **CELS** algorithms.

For a discrete-time model, we simulate the **ELS algorithm** in [1] and the **WLS** in [5] for $p = q = r = 1, a = -0.1, b = 0.2, c = 0.3, \theta_0 = 0$.The weighting sequence for **WLS** is $a_n = (log(trace(S_n)))^{-1.1}$. The results are shown in figure 1 and 2 for **ELS** and **WLS**,respectively. Then we choose one example $y_{n+1} = a + bu_n + w_{n+1}$ which was constructed in [10]. Figure 3 and 4 show that under the same control law **ELS** diverges but **WLS** converges for this example. This is the self-convergence property of **WLS** which **ELS** is lack of.

For a continuous-time model, we simulate the **CELS** (see [7] and [8]) and our **CWLS** for $p = q = 1, r = 0, \mathbf{A}(\mathbf{S}) = 1 + a\mathbf{S}, \mathbf{B}(\mathbf{S}) = b\mathbf{S}$, $\mathbf{C}(\mathbf{S}) = \mathbf{D}(\mathbf{S}) = 1$ and $\theta_0 = 0$. The weighting process for **CWLS** is chosen as $a_t = (log(trace(R_t))^{-1.1}$. The convergence results are shown in figure 5 and 6 for continuous **ELS** and **CWLS** respectively. Similar to the discrete case, we also show one example in Figure 7 and Figure 8 where **CWLS** converges and **CELS** diverges.

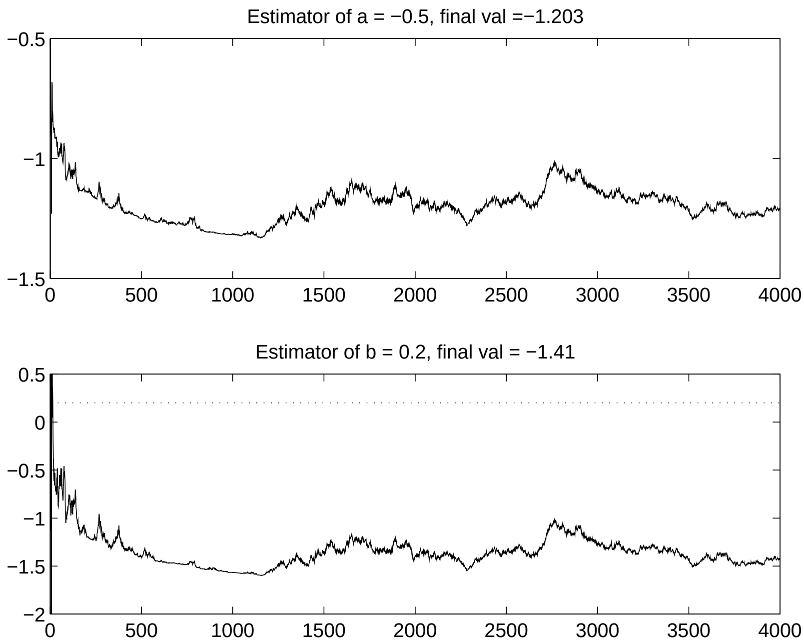

Fig. 3: Divergence Discrete -time ELS

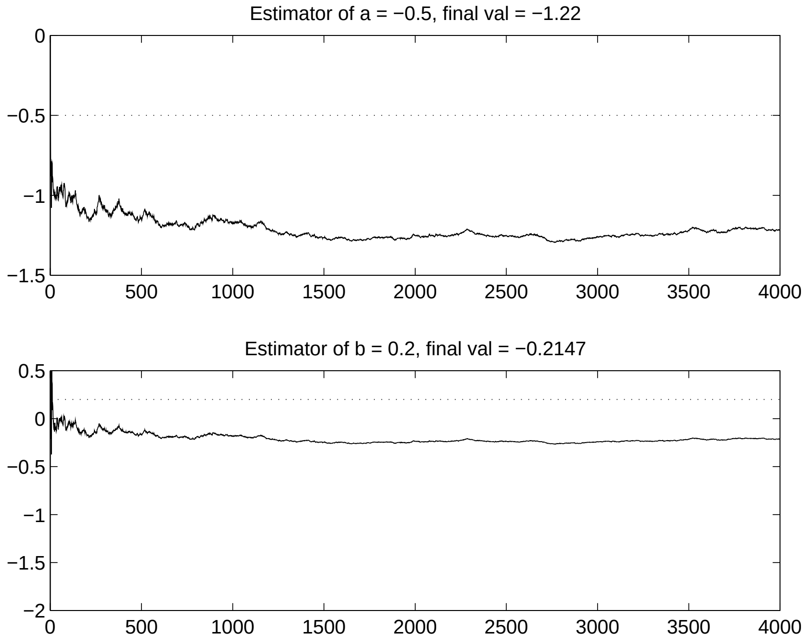

Fig. 4: Convergence Discrete-time WLS

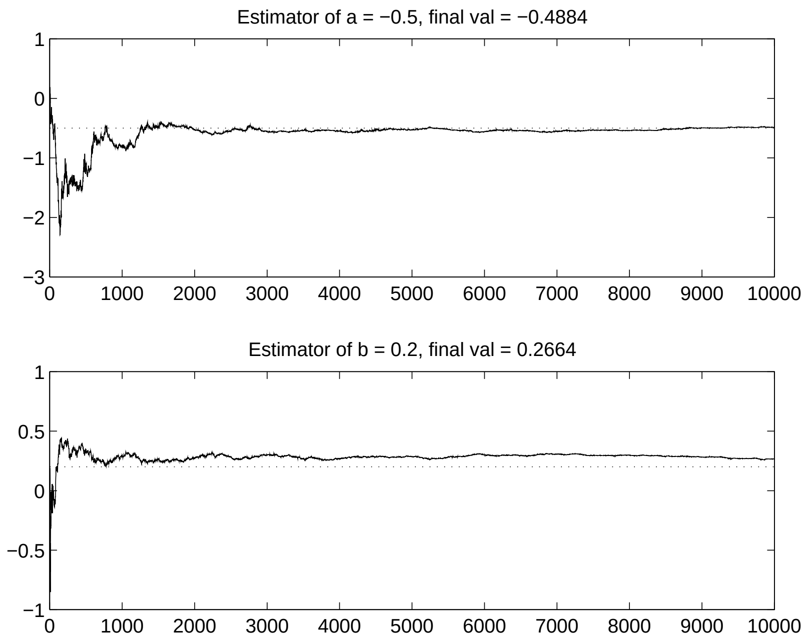

Fig. 5: Continuous-time ELS algorithm

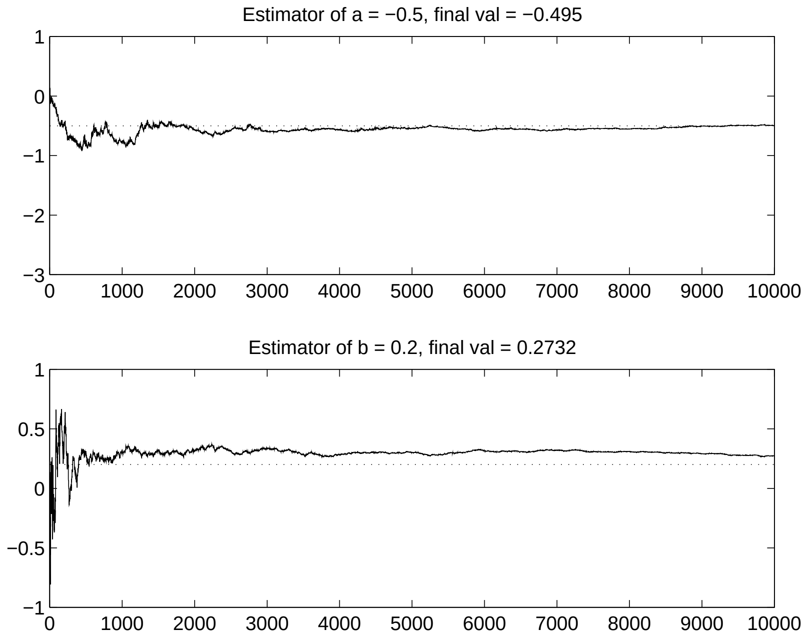

Fig. 6: Continuous-time WLS algorithm

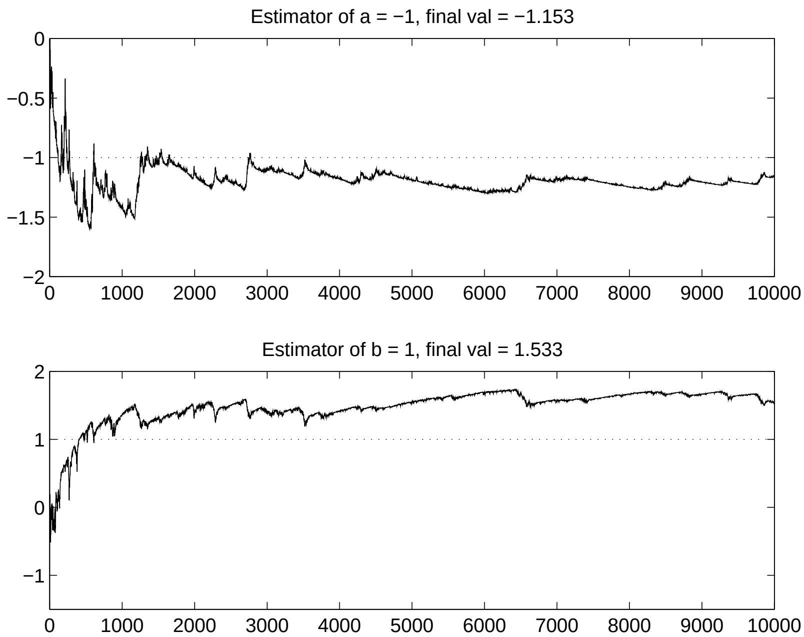

Fig. 7: Divergence Continuous-time ELS

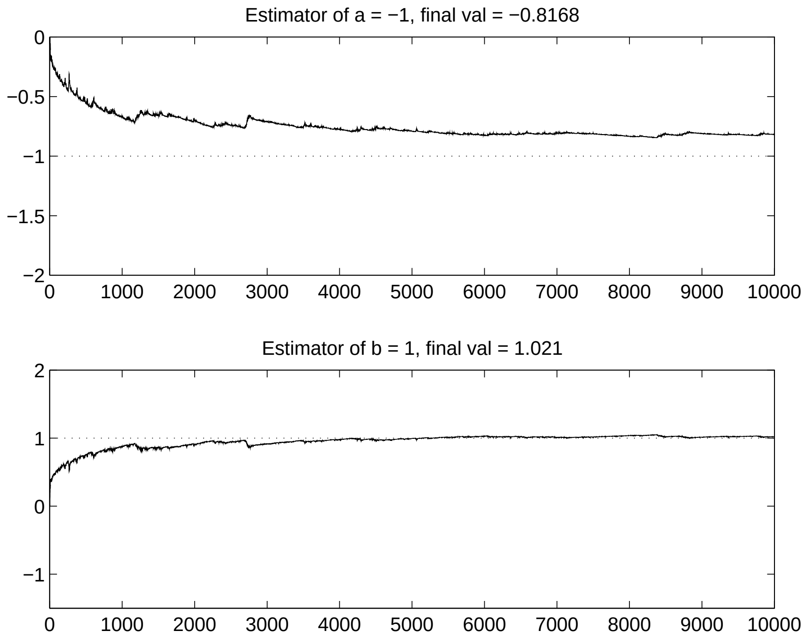

Fig. 8: Convergence Continuous-time WLS

References

[1] H. F. Chen, and L. Guo, *Convergence Rate of Least Squares r, Identification and Adaptive control for Stochastic Systems*, Int. J. Control, Vol.44, pp. 1459-1476, (1986).

[2] L. Guo, H.F. Chen, *the Åström-Witlenmark Self-tuning Regulator Revisited and ELS-based Adaptive Trackers*, IEEE Trans. Automat. Control, Vol.36, pp. 802-812, (1991).

[3] G. C. Goodwin, P. J. Ramadge and P. E. Caines, *Discrete Time Stochastic Adaptive Control*, SIAM J. Control and Optimization, Vol. 19, pp.829-853, (1981).

[4] H. F. Chen, and L. Guo, *Asymptotically Optimal Adaptive Control with Consistent Parameter Estimates*, SIAM J. Control and Optimization, Vol. 25, pp.558-575, (1987).

[5] B. Bercu, *Weighted Estimation and Tracking for ARMAX Models*, SIAM J. Control and Optimization, Vol. 33(1), pp.89-106, (1995).

[6] L. Guo, *Self-Convergence of Weighted Least-Squares with Applications to Stochastic Adaptive Control*, IEEE Trans. Automatic Control, Vol.41(1), pp.79-89, (1994).

[7] H. F. Chen, and J. B. Moore, *Convergence rates of Continuous Time Stochastic ELS Parameter Estimation*, IEEE Transactions on Automatic Control, Vol. AC-32, No.3, pp. 267-269, (1987).

[8] H. F. Chen, and L. Guo, *Continuous Time Stochastic Adaptive Control Tracking Robustness and Asymptotic properties.* SIAM J. Control and Optimization, Vol. 28, No. 3, pp. 513-527, (1990).

[9] N. Christopeit, *Quasi-least-squares estimation in semimartingale regression models*, Stochastics, 16, pp255-278, (1986).

[10] N.T. Karim, and R. Wei, *On sufficient conditions for strong consistency of least-squares estimates* (Preprint).

[11] A.J. Gao, *Weighted Least Squares Algorithm for Continuous-time Model*, Proceedings of the 33rd IEEE Conference on Dicision and Control, Lake Buena Vista, Florida, Vol.1, pp263-264, (1994).