

RECURSIVE SYSTEM IDENTIFICATION BY STOCHASTIC APPROXIMATION*

HAN-FU CHEN[†]

Abstract. The convergence theorems for the stochastic approximation (SA) algorithm with expanding truncations are first presented, which the system identification methods discussed in the paper are essentially based on. Then, the recursive identification algorithms are respectively defined for the multivariate errors-in-variables systems, Hammerstein systems, and Wiener systems. All estimates given in the paper are strongly consistent.

Key words: System identification, Hammerstein system, Wiener system, errors-in-variables (EIV), stochastic approximation, recursive algorithm, strong consistency.

1. Introduction. System identification is an important step to control an object when its model is unknown. When the system is parameterized, then the task of system identification is to estimate the unknown parameters contained in the system. One may first collect data with fixed sample size, and then derive the estimates by minimizing some performance index based on the collected data. The minimization may be carried out iteratively in order to use information contained in the data as completely as possible. This type of estimation is usually called as the block algorithm in contrast to the recursive algorithm.

The block algorithm sometimes is not satisfactory, because having obtained the estimate based on the data with sample size N , one has to compute the estimate from beginning if some new data arrive. This makes computation rather time-consuming, and in this case the recursive methods may be more attractive. Recursive system identification consists in obtaining the new estimate by modifying the immediately past estimate by using the new data. In other words, the estimate is updated on-line. This may greatly save the computational time, although the information contained in data may not fully be used.

The least-squares (LS) method [8, 14, 27] probably is most commonly used when estimating parameters by optimizing some index. The LS estimate can also be derived in a recursive way when the identified system is linear. However, the LS method may lead to a biased estimate in a correlated noise case even the system is linear. Further, the LS estimate cannot be calculated recursively for nonlinear systems in general.

Stochastic approximation (SA) pioneered by [31] and further developed by many

*This work is supported by the National Natural Science Foundation of China (Projects 60221301, 60334040, 60474004).

[†]Laboratory of Systems and Control, Institute of Systems Science, AMSS, Chinese Academy of Sciences, Beijing 100080, P. R. China, Fax: 86-10-62587343, E-mail: hfchen@iss.ac.cn

researchers in statistics and in system and control (see, e.g., [6, 7, 9, 17, 24-26, 28-30, 32, 36] among others) is aimed at seeking roots of an unknown function, when the function can be observed at any point in its region of definition, but the observations may be corrupted by noise.

This paper concerns identification of the Hammerstein systems, Wiener systems, and errors-in-variables systems. It is interesting to note that although there are enormous number of papers on this issue, almost all of them are on block algorithms and only a few of them concern a.s. convergence. To fill the gap the author with co-authors for recent years have been working on recursive and a.s. convergent identification algorithms for such kind of systems. We have applied the SA method developed in [9] to identifying various systems and succeeded in providing recursive and strongly consistent identification algorithms. These results published in a set of papers are summarized here with key points of the proof outlined.

The rest of the paper is arranged as follows. The SA algorithm with expanding truncations, which the identification algorithms of the paper are essentially based on, is introduced in Section 2. The recursive identification for multivariate EIV systems is presented in Section 2, while the corresponding results for Hammerstein and Wiener systems in Sections 3 and 4, respectively. Some concluding remarks are given by the end of the paper.

2. SA Algorithms with Expanding Truncations. Let J be the root set of a function $f(\cdot) : R^l \rightarrow R^l$, $J \triangleq \{x \in R^l : f(x) = 0\}$. Assume $f(\cdot)$ is unknown, but it can be observed at any $x \in R^l$, and the observations are noise-corrupted. Let x_k denote the k th approximation to the root set J , and let the observation be carried out at this point, i.e., the $(k+1)$ th observation is

$$(1) \quad y_{k+1} = f(x_k) + \epsilon_{k+1} \quad \text{with} \quad \epsilon_{k+1} \quad \text{being the observation noise.}$$

For the single root case, i.e., for the case where $J = x^0$ the classical Robbins-Monro (RM) algorithm [31] proposes to estimate x^0 by the following recursion

$$x_{k+1} = x_k + a_k y_{k+1} \quad \text{with} \quad a_k > 0, \quad a_k \rightarrow 0, \quad \sum_{k=1}^{\infty} a_k = \infty.$$

However, for the desired convergence $x_k \xrightarrow[k \rightarrow \infty]{} x^0$ a set of restrictive conditions [29, 31] on $f(\cdot)$ and ϵ_{k+1} are needed, which, unfortunately, are hardly to be satisfied for many problems.

To overcome the difficulty we modify the RM algorithm by truncating it at expanding bounds. Let $\{M_k\}$ be a sequence of positive numbers increasingly diverging to infinity, and let x^* be a fixed point in R^l . Arbitrarily fix an initial value x_0 , and

recursively define x_k by the following SA algorithm with expanding truncations:

$$(2) \quad x_{k+1} = (x_k + a_k y_{k+1}) I_{[\|x_k + a_k y_{k+1}\| \leq M_{\sigma_k}]} + x^* I_{[\|x_k + a_k y_{k+1}\| > M_{\sigma_k}]},$$

$$(3) \quad \sigma_k = \sum_{i=1}^{k-1} I_{[\|x_i + a_i y_{i+1}\| > M_{\sigma_i}]}, \quad \sigma_0 = 0,$$

where $I_{[A]}$ is the indicator of an ω -set A : $I_{[A]} = 1$ if $\omega \in A$, and $I_{[A]} = 0$ if $\omega \notin A$. The observation noise ϵ_{k+1} may depend on x_k

Let us list conditions to be used.

A1. $a_k > 0$, $a_k \xrightarrow[k \rightarrow \infty]{} 0$ and $\sum_{k=1}^{\infty} a_k = \infty$.

A2. There exists a continuously differentiable function (not necessarily being nonnegative) $v(\cdot) : R^l \rightarrow R$ such that

$$(4) \quad \sup_{\delta \leq d(x, J) \leq \Delta} f^T(x) v_x(x) < 0$$

for any $\Delta > \delta > 0$, where $d(x, J) \triangleq \inf_y \{\|x - y\| : y \in J\}$ and $v_x(\cdot)$ denotes the gradient of $v(\cdot)$. Further, $v(J) \triangleq \{v(x) : x \in J\}$ is nowhere dense, and x^* used in (2) is such that $v(x^*) < \inf_{\|x\|=c_0} v(x)$ with $\|x^*\| < c_0$ for some $c_0 > 0$.

A3. $f(\cdot)$ is measurable and locally bounded.

Denote by (Ω, \mathcal{F}, P) the probability space. Let $\epsilon_{k+1}(\cdot, \cdot) : (R^l \times \Omega, \mathcal{B}^l \times \mathcal{F}) \rightarrow (R^l \times \mathcal{B}^l)$ be a measurable function defined on the product space, and the noise ϵ_{k+1} be given by

$$\epsilon_{k+1} = \epsilon_{k+1}(x_k, \omega), \quad \omega \in \Omega.$$

A4. For the fixed sample path ω under consideration the following limit takes place:

$$(5) \quad \lim_{T \rightarrow 0} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k, T_k)} a_i \epsilon_{i+1}(x_i(\omega), \omega) \right\| = 0, \quad \forall T_k \in [0, T]$$

along the subscripts $\{n_k\}$ of any convergent subsequences $x_{n_k}(\omega)$, where

$$(6) \quad m(k, T) \triangleq \max \left\{ m : \sum_{i=k}^m a_i \leq T \right\}.$$

The algorithm (1)–(3) is considered for a fixed ω , but ω in $x_i(\omega)$ is often suppressed.

General Convergence Theorem (GCT). Let $\{x_k\}$ be given by (1)–(3) with a given initial value x_0 . Assume A1–A3 hold. Then, the distance between x_k and J , $d(x_k, J) \xrightarrow[k \rightarrow \infty]{} 0$ for any sample paths (ω) for which A4 holds.

For the proof of the theorem we refer to [9]. The idea of expanding truncation was originally proposed in [17]. The method has further been developed incorporating with the trajectory subsequence (TS) analysis method, which makes it possible to verify A4 only along n_k of any convergent subsequences $\{x_{n_k}\}$ rather than along the whole sequence $\{x_k\}$. Due to such a relaxation, SA algorithms with expanding truncations with TS analysis method have successfully solved a series of problems arising from systems and control [10-13, 15].

Remark 1. GCT remains true, if the condition $v(x^*) < \inf_{\|x\|=c_0} v(x)$ in A2 is replaced by

$$(7) \quad v(x^*) < \inf_{\{k:\|x_k\|\geq c_0\}} v(x_k)$$

for the sample path under consideration, where $\{x_k\}$ is recursively given by (2),(3).

Remark 2. If $\sum_{k=1}^{\infty} a_k \epsilon_{k+1} < \infty$, then A4 is fulfilled.

Remark 3. The selection of $\{M_k\}$ defines the tolerated divergence rate of $\{x_k\}$. This is because by (2) and (3)

$$\|x_{k+1}\| \leq \|x^*\| \vee M_{\sigma_k} \leq \|x^*\| \vee M_{k-1} \leq M_{k-1} \quad \text{for sufficiently large } k,$$

where $a \vee b$ means the maximum of a and b .

Remark 4. Under the conditions of GCT $\{x_k\}$ given by (1)–(3) is bounded, and hence the truncation in (2) ceases in a finite number of steps, and the asymptotic behavior of $\{x_k\}$ is the same as that given by the RM algorithm.

Example. We now give an example showing that the RM algorithm fast diverges even though the observations are free of noise, while the SA algorithm with expanding truncations is convergent. Let $f(x) = -(x - 10)^3$, $a_k = \frac{1}{k+1}$, and $y_{k+1} = f(x_k)$. According to the RM algorithm $x_{k+1} = x_k - \frac{1}{k} y_{k+1}$, the computation gives $x_0 = 0$, $x_1 = -1000$, $x_2 = 515149400, \dots$. In contrast to this, the SA algorithm with expanding truncations converges to the root 10 of $-(x - 10)^3$ even in the noise environment: $y_{k+1} = f(x_k) + \epsilon_{k+1}$, $\epsilon_{k+1} - 0.9\epsilon_k = w_{k+1} + 0.5w_k$, $w_k \in \mathcal{N}(0, 0.1)$. With $M_k = 2^{k+1}$, $x^* = 0.5$, the computation gives $x_0 = 0$, $x_{100} = 9.26$, $x_{400} = 9.61, \dots$

Consider the linear function case $f(x) = -Fx + b$, where F is an arbitrary $n \times m$ -matrix and b is an n -vector. It is clear that $f(x) = 0$ is solvable if and only if $FF^+b = b$, where F^+ denotes the pseudo-inverse of F . In this case for both $-Fx + b$ and $-F^T Fx + F^T b$ the root set $J = \{F^+b + (I - F^+F)c, \forall c \in R^m\}$ is the same. Since at the right-hand side of the algorithm (2) the sign “+” is for a_k , it is convenient to deal with $-F^T Fx + F^T b$ rather than $-Fx + b$ by noticing that $-F^T F \leq 0$.

Assume $y_{k+1} = -F^T Fx_k + F^T b + \epsilon_{k+1}$. Let us check in the linear case if we can simplify conditions used in GCT.

As $v(x)$ required in A2 we may take

$$(8) \quad v(x) = (-F^T Fx + F^T b)^T (-F^T Fx + F^T b).$$

It is clear that (4) holds and $v(J) = 0$. Condition A2 is automatically satisfied if $F^+ F = I$. If $F^+ F \neq I$, then J is a connected but unbounded set in R^m , and hence the condition $v(x^*) < \inf_{\|x\|=c_0} v(x)$ may not be satisfied however large c_0 is taken.

Let $V \triangleq I - F^+ F$. Then $Vy_{k+1} = V\epsilon_{k+1}$.

By (2),(3) it follows that

$$(9) \quad Vx_{k+1} = (Vx_k + a_k V\epsilon_{k+1})I_{[\|x_k + a_k y_{k+1}\| \leq M\sigma_k]} + Vx^* I_{[\|x_k + a_k y_{k+1}\| > M\sigma_k]}.$$

From (9) it is seen that $\{Vx_k\}$ is uniformly bounded for those sample paths for which the following condition S5 holds.

$$\text{A5. } \sum_{k=1}^{\infty} a_k V\epsilon_{k+1} < \infty.$$

Consequently, (7) is satisfied for those ω for which A5 holds.

It is worth noting that the following A6 implies both A4 and A5:

$$\text{A6. } \sum_{k=1}^{\infty} a_k \epsilon_{k+1} < \infty.$$

GCT for Linear Functions. In the case $f(x) = -(F^T Fx + F^T b)$, assume A1 holds. Then $\{x_k\}$ recursively given by (1)-(3) is uniformly bounded and $d(x_k, J) \xrightarrow[k \rightarrow \infty]{} 0$ for those ω for which A4 and A5 hold, where $J = \{F^+ b + (I - F^+ F)c \mid c \in R^m\}$, the root set of $f(\cdot)$. Moreover, if for some ω A6 holds, then x_k converges to some point in J as k tends to infinity.

This theorem follows from GCT as explained above except the last assertion, which is given in [18].

3. Multivariate EIV Systems. Consider the multi-input multi-output (MIMO) system

$$(10) \quad A(z)y_k^0 = B(z)u_k^0,$$

where z is the backward-shift operator, and u_k^0 and y_k^0 are m -input and n -output, respectively. They are observed with additive noises ξ_k and η_k :

$$(11) \quad y_k = y_k^0 + \xi_k, \quad u_k = u_k^0 + \eta_k.$$

The problem is to estimate the coefficients contained in the matrix polynomials

$$(12) \quad A(z) = I + A_1 z + \cdots + A_{s_a} z^{s_a},$$

and

$$(13) \quad B(z) = B_1 z + B_2 z^2 + \cdots + B_{s_b} z^{s_b}$$

on the basis of the noisy observations $\{u_k\}$ and $\{y_k\}$.

Identification of EIV systems has been attracting a considerable attention from researchers of control and statistics, e.g., [1, 2, 16, 33, 34] among others, but in the existing literature almost all estimates are for SISO systems and generated by block algorithms. We intend to give recursive and strongly consistent estimates for the matrix coefficients in (12) (13). For this we first list conditions to be used.

B1. The input $\{u_k^0\}$ is an ARMA process

$$(14) \quad P(z)u_k^0 = Q(z)\varepsilon_k$$

with

$$(15) \quad P(z) = I + P_1z + \cdots + P_{s_p}z^{s_p}, \quad Q(z) = I + Q_1z + \cdots + Q_{s_q}z^{s_q}.$$

B2. $A(z)$ and $P(z)$ are stable, i.e., all roots of $\det A(z)$ and $\det P(z)$ are outside the closed unit disk.

B3. $\Delta_k \triangleq [\xi_k^T, \eta_k^T, \varepsilon_k^T]^T$ is a sequence of iid random vectors with $E\Delta_k = 0$ and

$$E\Delta_k\Delta_k^T \triangleq R_\Delta \triangleq \begin{bmatrix} R_\xi & R_{\xi\eta} & R_{\xi\varepsilon} \\ R_{\eta\xi} & R_\eta & R_{\eta\varepsilon} \\ R_{\varepsilon\xi} & R_{\varepsilon\eta} & R_\varepsilon \end{bmatrix}.$$

Let us present the $n + m$ - observation process $z_k = [y_k^T, u_k^T]^T$ in the ARMA form. Set $\lambda \triangleq \max(s_a, s_b, s_p, s_q)$ and $A_i \triangleq 0, B_j \triangleq 0, P_s \triangleq 0, Q_t \triangleq 0$ respectively for $i > s_a, j > s_b, s > s_p, t > s_q$, and set

$$(16) \quad G(z) \triangleq \begin{bmatrix} A(z) & -B(z) \\ 0 & P(z) \end{bmatrix} \triangleq I + G_1z + \cdots + G_\lambda z^\lambda \quad \text{with} \quad G_i = \begin{bmatrix} A_i & -B_i \\ 0 & P_i \end{bmatrix},$$

$$(17) \quad S(z) \triangleq \begin{bmatrix} A(z) & -B(z) & 0 \\ 0 & P(z) & Q(z) \end{bmatrix} = S_0 + S_1z + \cdots + S_\lambda z^\lambda,$$

where $A(z), B(z), P(z)$, and $Q(z)$ are given by (12),(13),(14), and (15), respectively. Then the observation process satisfies the following equation:

$$(18) \quad G(z)z_k = S(z)\Delta_k.$$

We now derive the multivariate Yule-Walker equation [35, 37] of the observation process. Noticing that the dimension of Δ_k is $n + 2m$ while the dimension of z_k is $n + m$, we see that (18) is not a standard multivariate ARMA process.

Since $G(z)$ is stable by B2 and Δ_k is iid with $E\|\Delta_k\|^2 < \infty$ by B3, $\{z_k\}$ is a stationary and ergodic process with $E\|z_k\|^2 < \infty$. Hence,

$$(19) \quad Ez_k z_{k-i}^T \triangleq R_i = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N z_k z_{k-i}^T.$$

Set

$$(20) \quad \phi_{k-1}^T \triangleq [z_{k-1}^T, \dots, z_{k-\lambda}^T], \quad \psi_{k-\lambda-1}^T \triangleq [z_{k-\lambda-1}^T, z_{k-\lambda-2}^T, \dots, z_{k-\lambda-\mu}^T],$$

$$(21) \quad \chi_{k-1}^T \triangleq [\zeta_{k-1}^T, \dots, \zeta_{k-\lambda}^T], \quad G \triangleq [G_1, \dots, G_\lambda],$$

where $\mu = \lambda(n+m)$ and $\zeta_k^T \triangleq [\xi_k^T, \eta_k^T]$.

Then by (10),(11), and (14) we have

$$(22) \quad z_k = -G\phi_{k-1} + G\chi_{k-1} + [\xi_k^T, \eta_k^T + (Q(z)\varepsilon_k)^T]^T.$$

By (19) it follows that

$$(23) \quad \Gamma \triangleq E\phi_{k-1}\psi_{k-1-\lambda}^T = \begin{bmatrix} R_\lambda & R_{\lambda+1} & \cdots & R_{\lambda+\mu-1} \\ \vdots & \vdots & & \vdots \\ R_1 & R_2 & \cdots & R_\mu \end{bmatrix}$$

$$(24) \quad Ez_k\psi_{k-\lambda-1}^T = [R_{\lambda+1}, \dots, R_{\lambda+\mu}] \triangleq W.$$

By B3 and by noticing $s_q \leq \lambda$ we have $E\chi_{k-1}\psi_{k-\lambda-1}^T = 0$, and $E[\xi_k^T, \eta_k^T + (Q(z)\varepsilon_k)^T]^T \psi_{k-\lambda-1}^T = 0$.

Consequently, by (22)-(24) we derive the multivariate Yule-Walker equation:

$$(25) \quad G\Gamma + W = 0.$$

Remark 5. The Yule-Walker equation (25) is equivalent to

$$(26) \quad \Gamma\Gamma^T G^T + \Gamma W^T = 0,$$

and the solution to (26) composes a set:

$$(27) \quad J \triangleq \{G = -W\Gamma^+ + G_0(I - \Gamma\Gamma^+), \forall G_0 \in R^{(n+m) \times \lambda(n+m)}\}.$$

If $\Gamma\Gamma^+ = I$, then G is uniquely defined: $G = -W\Gamma^+$.

In order to estimate coefficients of $A(z)$, $B(z)$, and $P(z)$, it suffices to estimate G , since $G_i = \begin{bmatrix} A_i & -B_i \\ 0 & P_i \end{bmatrix}$.

We estimate G^T by Θ_k given by the SA algorithm with expanding truncations as follows. Let $M_k = k^{\frac{1}{2}-\delta}$ with any $\delta \in (0, \frac{1}{2})$. Recursively define

$$(28) \quad \Gamma_k = (1 - \frac{1}{k})\Gamma_{k-1} + \frac{1}{k}\phi_{k-1}\psi_{k-\lambda-1}^T$$

$$(29) \quad W_k = (1 - \frac{1}{k})W_{k-1} + \frac{1}{k}z_k\psi_{k-\lambda-1}^T$$

$$(30) \quad \Theta_k = (\Theta_{k-1} - \frac{1}{k}\Gamma_k(\Gamma_k^T\Theta_{k-1} + W_k^T))I_{[(\Theta_{k-1} - \frac{1}{k}\Gamma_k(\Gamma_k^T\Theta_{k-1} + W_k^T)) \leq M_{\sigma_k}]}$$

$$(31) \quad \sigma_k = \sum_{i=1}^{k-1} I_{[(\Theta_{i-1} - \frac{1}{i}\Gamma_i(\Gamma_i^T\Theta_{i-1} + W_i^T)) > M_{\sigma_i}]}$$

with arbitrary initial values Γ_0, W_0, Θ_0 , and $\sigma_0 = 0$.

THEOREM 1. *Assume B1-B3 hold. Then $\{\Theta_k^T\}$ given by (30) converges to a matrix belonging to J given by (27), the solution set of the Yule-Walker equation (26) a.s. as $k \rightarrow \infty$.*

Proof. We only outline the key points, for the detailed proof we refer to [11]. First of all, (28) and (29) present the time averages of $\phi_{k-1}\psi_{k-\lambda-1}^T$ and $z_k\psi_{k-\lambda-1}^T$, respectively. By ergodicity we have

$$(32) \quad \Gamma_k \xrightarrow[k \rightarrow \infty]{} \Gamma, \quad \text{and} \quad W_k \xrightarrow[k \rightarrow \infty]{} W,$$

where Γ and W are given by (23) and (24), respectively.

Comparing (30) with (2) we find that x^* and y_{k+1} in (2) correspond to 0 and $-\Gamma_k(\Gamma_k^T\Theta_{k-1} + W_k^T)$, respectively. Therefore, the function under consideration $f(\Theta) \triangleq -\Gamma\Gamma^T\Theta + \Gamma W^T$ is linear, and

$$(33) \quad y_k \triangleq -\Gamma_k(\Gamma_k^T\Theta_{k-1} + W_k^T) = -(\Gamma\Gamma^T\Theta_{k-1} + \Gamma W^T) + \epsilon_k,$$

where

$$(34) \quad \epsilon_k = -(\Gamma_k\Gamma_k^T\Theta_{k-1} + \Gamma_k W_k^T) + (\Gamma\Gamma^T\Theta_{k-1} + \Gamma W^T).$$

Thus, the conclusion of the theorem follows from GCT for Linear Functions given in Section 2 if it can be shown that

$$(35) \quad \sum_{k=1}^{\infty} \frac{1}{k} [-(\Gamma_k\Gamma_k^T\Theta_{k-1} + \Gamma_k W_k^T) + (\Gamma\Gamma^T\Theta_{k-1} + \Gamma W^T)] < \infty.$$

Introducing

$$F \triangleq \begin{bmatrix} -G_1 & I & 0 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \\ \vdots & & & \ddots & 0 & 0 \\ -G_\lambda & 0 & \cdots & 0 & I & \\ 0 & & & 0 & 0 & 0 \end{bmatrix}, \quad L \triangleq \begin{bmatrix} S_0 \\ \vdots \\ S_\lambda \end{bmatrix},$$

we express z_k given by (18) in the state space form

$$z_k = H\pi_k, \quad \pi_k = F\pi_{k-1} + L\Delta_k, \quad H = \overbrace{[I, 0, \dots, 0]}^{(\lambda+1)(n+m)}(n+m).$$

By stability of F and B3 it can be shown that

$$(36) \quad \left\| \frac{1}{n} \sum_{k=0}^n \pi_{k+j+1} \pi_{k+1}^T - \sum_{i=0}^{\infty} F^{i+j} L R_{\Delta} L^T F^{Ti} \right\| = O(n^{-(\frac{1}{2}-\beta)}) \quad \text{for any } \beta > 0.$$

Since $z_k = H\pi_k$, from (36) it follows that

$$(37) \quad \left\| \frac{1}{n} \sum_{k=1}^n z_{k+j} z_k^T - R_j \right\| = O(n^{-(\frac{1}{2}-\beta)})$$

with $R_j = H \sum_{i=0}^{\infty} F^{i+j} L R_{\Delta} L^T F^{Ti} H^T$ for any $\beta > 0$.

As pointed out in Remark 3 by (30) $\|\Theta_k\| \leq M_{\sigma_k} \leq M_{k-1} < k^{\frac{1}{2}-\delta}$, which incorporating with (37) yields

$$\sum_{k=1}^{\infty} \frac{1}{k} (\Gamma_k \Gamma_k^T - \Gamma \Gamma^T) \Theta_{k-1} < \infty,$$

which in turn implies (35). □

By Theorem 1 Θ_k given by (30) is a strongly consistent estimate for G^T whenever $\Gamma \Gamma^T$ is nondegenerate. We now give conditions guaranteeing the required nondegeneracy.

B4. $\det A_{\lambda} \neq 0, \quad \det P_{\lambda} \neq 0$.

B5. $\det \Psi(z) \det \Phi(z)$ is coprime with $\det A(z) \det P(z)$, where

$$(38) \quad \Psi(z) \triangleq (P(z)R_{\eta} + Q(z)R_{\epsilon\eta})z^{\lambda}P^T(z^{-1}) + (Q(z)R_{\epsilon} + P(z)R_{\eta\epsilon})z^{\lambda}Q^T(z^{-1})$$

$$\Phi(z) \triangleq A(z)R_{\xi}z^{\lambda}A^T(z^{-1}) + B(z)R_{\eta}z^{\lambda}B^T(z^{-1})$$

$$- B(z)R_{\eta\xi}z^{\lambda}A^T(z^{-1}) - A(z)R_{\xi\eta}z^{\lambda}B^T(z^{-1})$$

$$- [(A(z)R_{\xi\eta} - B(z)R_{\eta})z^{\lambda}P^T(z^{-1}) + (A(z)R_{\xi\epsilon} - B(z)R_{\eta\epsilon})z^{\lambda}Q^T(z^{-1})]$$

$$\cdot [(P(z)R_{\eta} + Q(z)R_{\epsilon\eta})z^{\lambda}P^T(z^{-1}) + (Q(z)R_{\epsilon} + P(z)R_{\eta\epsilon})z^{\lambda}Q^T(z^{-1})]^{-1}$$

$$(39) \quad \cdot [(P(z)R_{\eta\xi} + Q(z)R_{\epsilon\xi})z^{\lambda}A^T(z^{-1}) - (P(z)R_{\eta} + Q(z)R_{\epsilon\eta})z^{\lambda}B^T(z^{-1})].$$

THEOREM 2. *Assume B1-B5 hold and $R_{\Delta} > 0$. Then Γ is of row-full-rank, and Θ_k given by (30) converges to G^T a.s.*

Proof. Denote by ζ_k the $(n+m)$ -dimensional process $S(z)\Delta_k$, i.e.,

$$(40) \quad \zeta_k = S(z)\Delta_k.$$

By using the innovation representation [3] it can be shown that ζ_k can be represented as

$$(41) \quad \zeta_k = D(z)w_k, \quad D(z) \triangleq I + D_1z + \cdots + D_\lambda z^\lambda$$

with $Ew_k = 0$, $Ew_k w_k^T \triangleq R_w > 0 \quad \forall k \geq 0$, $Ew_k w_s^T = 0$ if $k \neq s$, where $D(z)$ is a stable $(n + m) \times (n + m)$ -matrix polynomial. Then, we have

$$(42) \quad G(z)z_k = D(z)w_k.$$

By [37] the rank of Γ is $\lambda(n + m)$ if and only if $G(z)$ and $D(z)$ have no common left-factor and $\text{rank}[G_\lambda : D_\lambda] = n + m$. The latter is guaranteed by B4.

Noticing that

$$D(z)R_w D^T(z^{-1}) = S(z)R_\Delta S^T(z^{-1}), \quad \det G(z) = \det A(z)P(z),$$

$$\det S(z)R_\Delta z^\lambda S^T(z^{-1}) = \det \Psi(z)\det \Phi(z),$$

by B5 we find $G(z)$ and $D(z)$ have no common left-factor, and hence Γ is of row-full-rank. This proves the theorem, for details we refer to [11]. □

4. Hammerstein Systems. We now consider identification of the SISO Hammerstein system, which consists of a static nonlinearity $f(\cdot)$ followed by an ARMA type linear subsystem. Here we restricted to the MA type linear subsystem. The Hammerstein system and the Wiener system to be discussed in the next section are important in practice, and their identification issue has attracted much attention from both researchers and practitioners, e.g., [4, 5, 19-23, 38, 39] among others. The block diagram of Hammerstein system is presented in Figure 1.

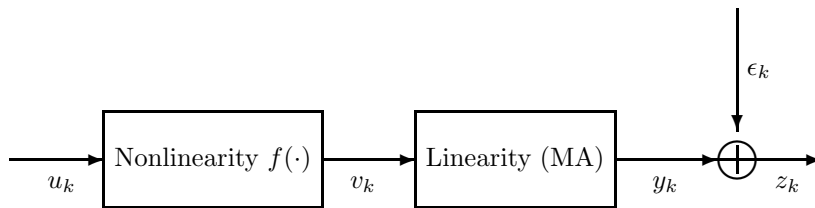


Fig. 1. Hammerstein system.

The system input, output, and observation are respectively denoted by u_k , y_k , and z_k :

$$(43) \quad y_{k+1} = \sum_{j=0}^r d_j v_{k-j}, \quad d_0 = 1, \quad v_j = f(u_j)$$

$$(44) \quad z_k = y_k + \epsilon_k,$$

where ϵ_k is the observation noise.

The coefficients d_i , $i = 1, \dots, r$ of the linear subsystem and the function $f(\cdot)$ are unknown and to be estimated. Since $f(\cdot)$ is not parameterized, we intend to estimate $f(u)$ at any given u .

Reference [20] probably is the first attempt to identify Hammerstein systems by SA method, but the obtained there estimates still contain some unknown parameters and thus the system cannot be identified completely. We now apply the SA algorithm with expanding truncations to solve the problem.

Fix a u , where $f(u)$ is to be estimated. Let the system input $\{u_k\}$ be a sequence of bounded independent and identically distributed (iid) random variables $|u_k| < c_1$, $\forall i = 1, 2, \dots$, with $Eu_k = 0$ and with density $p(\cdot)$, where $c_1 > 0$ is a constant $c_1 \neq |u|$ and $p(\cdot)$ is continuous at u with $p(u) > 0$. Let $\{u_k\}$ be also independent of the observation noise $\{\epsilon_k\}$. The conditions to be used are as follows.

C1. The nonlinear function $f(\cdot)$ is measurable, locally bounded, and continuous at u , where $f(u)$ is estimated.

C2. The observation noise $\{\epsilon_k\}$ is a sequence of mutually independent random variables with $E\epsilon_k \xrightarrow[k \rightarrow \infty]{} 0$ and $\sup_k E\epsilon_k^2 < \infty$.

To consistently estimate $\{d_i\}$ Conditions C1 and C2 are sufficient. It is worth noting that no assumption is made on the structure of $f(\cdot)$.

In order to uniquely define $f(u)$ and Ev_k we need a condition to guarantee that the response of the linear subsystem to a nonzero constant input is nonzero:

$$\text{C3. } \sum_{j=0}^r d_j \neq 0.$$

Let $a_k = \frac{1}{k}$ and let $\{M_k\}$ be a sequence of increasing real numbers diverging to infinity:

$$M_k > 0, \quad M_{k+1} > M_k \quad \forall k, \quad \text{and} \quad M_k \xrightarrow[k \rightarrow \infty]{} \infty.$$

Define

(45)

$$\theta_{k+1}(i) = \begin{cases} \theta_k(i) - a_k(\theta_k(i) - u_k z_{k+1+i}), & \text{if } |\theta_k(i) - a_k(\theta_k(i) - u_k z_{k+1+i})| \leq M_{\sigma_k(i)}, \\ 0, & \text{otherwise,} \end{cases}$$

(46)

$$\begin{aligned} \sigma_k(i) &= \sum_{j=1}^{k-1} I_{[|\theta_j(i) - a_j(\theta_j(i) - u_j z_{j+1+i})| > M_{\sigma_j(i)}]}, \\ \sigma_0(i) &= 0 \end{aligned}$$

with an initial value $\theta_0(i)$, $i = 0, 1, \dots, r$, where $\theta_0(0)$ is used to estimate $\rho \triangleq Eu_1 v_1$, while $\theta_k(i)$, $i = 1, \dots, r$ are used to estimate ρd_i , $i = 1, \dots, r$, respectively.

For estimating $Ef(u_1)$ and $f(u)$ we define

$$(47) \quad \gamma_{k+1} = \begin{cases} \gamma_k - a_k(\gamma_k - z_{k+1}), & \text{if } |\gamma_k - a_k(\gamma_k - z_{k+1})| \leq M_{\nu_k}, \\ 0, & \text{otherwise,} \end{cases}$$

$$(48) \quad \nu_k = \sum_{j=1}^{k-1} I_{[|\gamma_j - a_j(\gamma_j - z_{j+1})| > M_{\nu_j}]}, \quad \nu_0 = 0$$

with an initial value γ_0 , and

$$(49) \quad \mu_{k+1}(u) = \begin{cases} \mu_k(u) - a_k w_k(\mu_k(u) - z_{k+1}), & \text{if } |\mu_k(u) - a_k w_k(\mu_k(u) - z_{k+1})| \leq M_{\lambda_k(u)}, \\ 0, & \text{otherwise,} \end{cases}$$

$$(50) \quad \begin{aligned} \lambda_k(u) &= \sum_{j=1}^{k-1} I_{[|\mu_j(u) - a_j w_j(\mu_j(u) - z_{j+1})| > M_{\lambda_j(u)}]}, \\ \lambda_0(u) &= 0 \end{aligned}$$

with an initial value $\mu_0(u)$, where w_k is a kernel function

$$(51) \quad w_k \triangleq \frac{1}{b_k} e^{-\left(\frac{u_k - u}{b_k}\right)^2},$$

where $b_k = \frac{1}{k^\delta}$ with $\delta \in (0, \frac{1}{2})$.

THEOREM 3. *Assume C1 and C2 hold. Then $\theta_k(i)$, $i = 0, 1, \dots, r$, defined by (45) (46) are strongly consistent:*

$$(52) \quad \theta_k(0) \xrightarrow[k \rightarrow \infty]{} \rho \triangleq E(u_1 f(u_1)) (= E u_1 v_1) \text{ a.s.}$$

and

$$(53) \quad \theta_k(i) \xrightarrow[k \rightarrow \infty]{} \rho d_i \text{ a.s., } \quad i = 1, \dots, r.$$

Further, if in addition, C3 holds and $\rho \neq 0$, then

$$(54) \quad \theta_k(0) \gamma_k \left(\sum_{i=0}^r \theta_k(i) \right)^{-1} \xrightarrow[k \rightarrow \infty]{} Ef(u_1) (= E v_1) \text{ a.s.,}$$

$$(55) \quad \mu_k(u) \xrightarrow[k \rightarrow \infty]{} \mu(u) \triangleq f(u) + \sum_{j=1}^r d_j Ef(u_1) \text{ a.s.,}$$

and

$$(56) \quad \mu_k(u) - \frac{\sum_{j=1}^r \theta_k(j) \gamma_k}{\sum_{j=0}^r \theta_k(j)} \xrightarrow[k \rightarrow \infty]{} f(u) \text{ a.s.,}$$

where $\theta_k(i)$, γ_k , and $\mu_k(u)$ are defined by (45)(46), (47)(48), and (49)(50), respectively.

Proof. We outline the key points of the proof.

Rewrite (45) as follows:

$$(57) \quad \theta_{k+1}(i) = \begin{cases} \theta_k(i) - a_k(\theta_k(i) - d_i\rho) - a_k\epsilon_{k+1}(i), \\ \quad \text{if } |\theta_k(i) - a_k(\theta_k(i) - d_i\rho) - a_k\epsilon_{k+1}(i)| \leq M_{\sigma_k}(i), \\ 0, \quad \text{otherwise,} \end{cases}$$

where

$$(58) \quad \epsilon_{k+1}(i) = -u_k z_{k+1+i} + d_i\rho, \quad i = 0, 1, \dots, r.$$

The algorithm (57) is for the linear function $-x + d_i\rho$, which corresponds to $-Fx + b$ in GCT for Linear Functions with $F = 1$ and $b = d_i\rho$. Consequently, $V = 0$ and A5 is automatically satisfied. The root set J in the present case consists of a singleton $d_i\rho$. Therefore, by GCT for Linear Functions, for (53) it suffices to verify A4 for $\epsilon_k(i)$, $i = 0, \dots, r$ given by (58). This can be done by using the convergence theorem for martingale difference sequences.

Next, rewrite (47) (48) as

$$(59) \quad \gamma_{k+1} = \begin{cases} \gamma_k - a_k(\gamma_k - \sum_{j=0}^r d_j Ef(u_1)) + a_k\delta_{k+1}, \\ \quad \text{if } |\gamma_k - a_k(\gamma_k - \sum_{j=0}^r d_j Ef(u_1) - \delta_{k+1})| \leq M_{\nu_k}, \\ 0, \quad \text{otherwise,} \end{cases}$$

where

$$(60) \quad \delta_{k+1} = z_{k+1} - \sum_{j=0}^r d_j Ef(u_1).$$

Again, by verifying that A4 holds with ϵ_{k+1} replaced by δ_{k+1} , we apply GCT for Linear Functions to (59)(60) and conclude (54).

Finally, for (55) we rewrite (49) as

$$(61) \quad \mu_{k+1}(u) = \begin{cases} \mu_k(u) - a_k\sqrt{\pi}p(u)(\mu_k(u) - f(u) - \sum_{j=1}^r d_j Ef(u_1)) - a_k e_{k+1}(u), \\ \quad \text{if } |\mu_k(u) - a_k\sqrt{\pi}p(u)(\mu_k(u) - f(u) - \sum_{j=1}^r d_j Ef(u_1) - a_k e_{k+1}(u))| \\ \quad \leq M_{\lambda_k(u)}, \\ 0, \quad \text{otherwise,} \end{cases}$$

where

$$(62) \quad e_{k+1}(u) = w_k(\mu_k(u) - z_{k+1}) - \sqrt{\pi}p(u)(\mu_k(u) - f(u) - \sum_{j=1}^r d_j E f(u_1)).$$

It is clear that the algorithm (61) is for the linear function

$$(63) \quad \sqrt{\pi}p(u)(x - f(u) - \sum_{j=1}^r d_j E f(u_1))$$

Similar to the argument given above for (55) it suffices to verify A4 for $e_{k+1}(u)$, then the conclusion follows from GCT for Linear Functions.

However, it is worth noting that A4 may be verified for $\epsilon_{k+1}(i)$ and δ_{k+1} along the whole sequence of subscripts k , while the direct verification of A4 for $e_{k+1}(u)$ is feasible only along subscripts n_k for which x_{n_k} converges. This is because only along convergent subsequences it can be shown that for all large enough k and sufficiently small $T > 0$

$$(64) \quad \mu_{s+1}(u) = \mu_s(u) - a_s w_s(\mu_s(u) - z_{s+1}),$$

and

$$(65) \quad \|\mu_{s+1}(u) - \mu_{n_k}(u)\| \leq cT, \quad s = n_k, n_k + 1, \dots, m(n_k, T),$$

where c is a constant independent of k but may depend on sample path ω . For verifying A4 with ϵ_k replaced by $e_k(u)$ it is important to have (64)(65), which, roughly speaking, mean that for s not far from n_k $\mu_s(u)$ is close to $\mu_{n_k}(u)$ and the algorithm suffers from no truncation. For details we refer to [10]. \square

Remark 6. The linear subsystem is not necessarily restricted to be an MA process. The results may be extended to ARMA systems.

5. Wiener Systems. We now consider identification of the SISO Wiener system, which is also a linear system cascaded with a static nonlinearity similar to the Hammerstein system but in the reverse order. The block diagram of Wiener systems is presented in Figure 2.

As in Section 4 the input, output, and observation are respectively denoted by u_k, y_k , and z_k :

$$(66) \quad z_k = y_k + \epsilon_k, \quad y_k = f(v_k),$$

$$(67) \quad v_{k+1} = \sum_{j=0}^r d_j u_{k-j}, \quad d_0 = 1.$$

The problem is to recursively estimate the coefficients d_i , $i = 1, \dots, r$ of the linear subsystem and the value $f(v)$ at any v on the basis of the observations $\{z_k\}$ and appropriately designed inputs $\{u_k\}$.

We now define the system input $\{u_k\}$. Let $\{\eta_k\}$ be a sequence of iid Gaussian random variables $\eta_k \in \mathcal{N}(0, 1)$ independent of the observation noise $\{\epsilon_k\}$. Define functions $T_k(\cdot)$:

$$(68) \quad T_k(x) = \begin{cases} x, & x \in S(k) \triangleq (-k^\delta, k^\delta), \\ k^\delta \text{sign}(x), & x \in S^c(k). \end{cases}$$

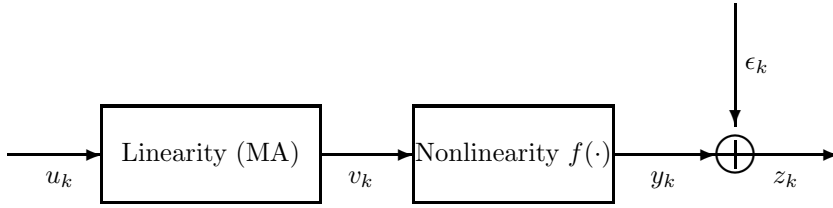


Fig. 2. Wiener system

It is clear that $|T_k(x)| \leq k^\delta$ and the range of $T_k(x)$ unboundedly increases as $k \rightarrow \infty$.

Define the system input u_k as

$$(69) \quad u_k = T_k(\eta_k).$$

Let us define sequences of real numbers to be used in the algorithms:

$$(70) \quad a_k = \frac{1}{k}, \quad b_k = \frac{1}{k^\alpha}, \quad M_k = M_0 + k^\beta, \quad k = 1, \dots,$$

where $\alpha > 0$, $\beta > 0$, $M_0 > 0$, $3\alpha + \beta + \delta < 1/2$, where $\delta > 0$ is the one given in (68).

For estimating the linear subsystem as in Section 4 we apply the SA algorithms with expanding truncations:

$$(71) \quad \theta_{k+1}(i) = [\theta_k(i) - a_k(\theta_k(i) - u_k z_{k+i+1})] I_{[|\theta_k(i) - a_k(\theta_k(i) - u_k z_{k+i+1})| \leq M_{\sigma_k(i)}]},$$

$$(72) \quad \sigma_k(i) = \sum_{j=1}^{k-1} I_{[|\theta_j(i) - a_j(\theta_j(i) - u_j z_{j+i+1})| > M_{\sigma_j(i)}]}, \quad \sigma_0(i) = 0,$$

with any initial values $\theta_0(i)$, $i = 0, 1, \dots, r$.

Set

$$(73) \quad \sigma \triangleq \left(\sum_{i=0}^r d_i^2 \right)^{1/2} \quad \text{and} \quad \rho \triangleq \frac{1}{\sqrt{2\pi}\sigma^3} \int_R t f(t) e^{-\frac{t^2}{2\sigma^2}} dt.$$

Note that $\theta_k(0)$ is used for estimating ρ , and $\frac{\theta_k(i)}{\theta_k(0)}$ for d_i , $i = 1, \dots, q$, wherever $\rho \neq 0$.

Denote by d_{ik} the k -th estimate for d_i . Then the output v_k and its standard deviation of the linear subsystem are respectively estimated by

$$(74) \quad \hat{v}_k \triangleq u_{k-1} + d_{1k}u_{k-2} + \dots + d_{rk}u_{k-r-1}, \quad \hat{\sigma}_k \triangleq \left(\sum_{i=0}^r d_{ik}^2 \right)^{1/2}.$$

In order to estimate $f(v)$ we introduce the following kernel function

$$(75) \quad w_k \triangleq \frac{\sqrt{2\pi}\sigma}{b_k} e^{-\left(\frac{v_k-v}{b_k}\right)^2 + \frac{v_k^2}{2\sigma^2}}.$$

By (74), w_k is naturally estimated by

$$(76) \quad \hat{w}_k \triangleq \frac{\sqrt{2\pi}\hat{\sigma}_k}{b_k} e^{-\left(\frac{\hat{v}_k-v}{b_k}\right)^2 + \frac{\hat{v}_k^2}{2\hat{\sigma}_k^2}}.$$

With any initial value $\mu_0(v)$, $f(v)$ is estimated by $\mu_k(v)$, which is recursively calculated according to the following SA algorithm with expanding truncations:

$$(77) \quad \mu_{k+1}(v) = [\mu_k(v) - a_k \hat{w}_k(\mu_k(v) - z_k)] I_{[|\mu_k(v) - a_k \hat{w}_k(\mu_k(v) - z_k)| \leq M_{\nu_k(v)}]},$$

$$(78) \quad \nu_k(v) = \sum_{j=1}^{k-1} I_{[|\mu_j(v) - a_j \hat{w}_j(\mu_j(v) - y_j)| > M_{\nu_j(v)}]}, \quad \nu_0(v) = 0.$$

Let us list conditions to be used.

D1. The static nonlinearity $f(\cdot)$ is a bounded measurable function and continuous at v , where $f(v)$ is estimated.

D2. The observation noise $\{\epsilon_k\}$ is a sequence of mutually independent random variables with $E\epsilon_k = 0$ and $\sup_k E\epsilon_k^2 < \infty$, and is independent of $\{\eta_k\}$.

THEOREM 4. Assume D1 and D2 hold. Then

$$(79) \quad \theta_k(0) \xrightarrow[k \rightarrow \infty]{} \rho, \quad \theta_k(i) \xrightarrow[k \rightarrow \infty]{} \rho d_i \quad a.s. \quad i = 1, \dots, r.$$

Further, if, in addition, $\rho \neq 0$, then

$$(80) \quad \mu_k(v) \xrightarrow[k \rightarrow \infty]{} f(v) \quad a.s.$$

Proof. We outline the key points of the proof. For estimating the linear subsystem (71) is rewritten as

$$(81) \quad \theta_{k+1}(i) = \begin{cases} \theta_k(i) - a_k(\theta_k(i) - \rho d_i) - a_k \bar{\epsilon}_{k+1}(i), & \text{if } |\theta_k(i) - a_k(\theta_k(i) - \rho d_i) - a_k \bar{\epsilon}_{k+1}(i)| \leq M_{\sigma_k}(i), \\ 0, & \text{otherwise,} \end{cases}$$

where

$$(82) \quad \bar{e}_{k+1}(i) = -u_k z_{k+i+1} + \rho d_i, \quad i = 0, 1, \dots, r.$$

By verifying A4 satisfied by $\bar{e}_{k+1}(i)$, (79) follows from GCT for Linear Functions.

Further, rewrite (77) as

$$(83) \quad \mu_{k+1}(v) = \begin{cases} \mu_k(v) - a_k \sqrt{\pi}(\mu_k(v) - f(v)) + a_k e_k(v), \\ \quad \text{if } |\mu_k(v) - a_k \sqrt{\pi}(\mu_k(v) - f(v)) + a_k e_k(v)| \leq M_{\nu_k(v)}, \\ 0, \quad \text{otherwise,} \end{cases}$$

where

$$(84) \quad e_k(v) = \sqrt{\pi}(\mu_k(v) - f(v)) - \hat{w}_k(\mu_k(v) - z_k).$$

Noticing that $f(v)$ is the root of $\sqrt{\pi}(x - f(v))$, by GCT for Linear Functions, for (80) it suffices to verify A4 for $e_k(v)$. It is important to note that A4 is verifiable along subscripts n_k of any convergent subsequences, but not along the whole sequence of subscripts k . For details we refer to [23]. □

Remark 7. Theorem 4 can be extended to the case where the linear subsystem is an ARMA process and $f(\cdot)$ is unbounded but with some growth rate restriction.

6. Concluding Remarks. The SA algorithms with expanding truncations incorporating with TS method was successfully applied to solving a set of problems arising from systems and control. This paper demonstrates that the method is also a powerful tool for system identification. By this method the recursive and strongly consistent estimates are given for multivariate EIV systems, Hammerstein systems and Wiener systems with non-parameterized nonlinearity $f(\cdot)$. By the way, the Hammerstein and Wiener systems with $f(\cdot)$ being a piece-wise linear function have also been recursively identified recently by the author with co-authors giving the strongly consistent estimates for all unknown system parameters.

A set of examples have been computed according to the algorithms given in the paper for various systems, and the numerical simulation results are consistent with theoretical analysis. The general picture is that some fluctuations appear at the first steps and then estimates approach to the true values.

For further research it is of interest to consider more general inputs u_k , more complicated nonlinearities, and possibly the adaptive control problems.

REFERENCES

[1] B. D. O. ANDERSON, *Identification of scalar errors-in-variables models with dynamics*, Automatica, 21(1985), pp. 709-716.

- [2] B. D. O. ANDERSON & M. DEISTLER, *Identifiability in dynamic errors-in-variables models*, Journal of Time Series Analysis, 5(1984), pp. 1-13.
- [3] B. D. O. ANDERSON & J. B. MOORE, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [4] E. W. BAI, *Frequency domain identification of Hammerstein models*, IEEE Trans. Autom. Control, 48:4(2003), pp. 530-542.
- [5] E. W. BAI, *Convergence of the iterative Hammerstein system identification algorithm*, IEEE Trans. Automat. Contr., 49:11(2004), pp. 1929-1940.
- [6] A. BENVENISTE, M. METIVIER, & P. PRIOURET, *Adaptive Algorithms and Stochastic Approximation*, Springer-Verlag, New York, 1990.
- [7] V. S. BOKAR, *Asynchronous stochastic approximation*, SIAM J. Control & Optimization, 36(1998), pp. 840-851.
- [8] P. E. CAINES, *Linear Stochastic Systems*, Wiley, New York, 1988.
- [9] H. F. CHEN, *Stochastic Approximation and Its Applications*, Kluwer, Dordrecht, the Netherlands, 2002.
- [10] H. F. CHEN, *Pathwise convergence of recursive identification algorithms for Hammerstein systems*, IEEE Trans. Autom. Control, 49:10(2004), pp. 1641-1649.
- [11] H. F. CHEN, *Recursive Identification for Multivariate Errors-in-Variables Systems*, to appear in Automatica, 2007.
- [12] H. F. CHEN & X. R. CAO, *Controllability is not necessary for adaptive pole placement control*, IEEE Trans. Autom. Control, 42:9(1997), pp. 1222-1229.
- [13] H. F. CHEN, T. DUNCAN, & B. PASIK-DUNCAN, *A Kiefer-Wolfowitz algorithm with randomized differences*, IEEE Trans. Autom. Control, 44:3(1999), pp. 442-453.
- [14] H. F. CHEN & L. GUO, *Identification and Stochastic Adaptive Control*, Birkhäuser, Boston, 1991.
- [15] H. F. CHEN & G. YIN, *Asymptotic properties of sign algorithms for adaptive filtering*, IEEE Trans. Autom. Control, 48:9(2003), pp. 1545-1556.
- [16] H. F. CHEN & J. M. YANG, *Strongly consistent coefficient estimate for errors-in-variables models*, Automatica, 41 (2005), pp. 1025-1033.
- [17] CHEN, H. F. & Y. M. ZHU, *Stochastic approximation procedures with randomly varying truncations*, Scientia Sinica (Series A), 29:9(1986), pp. 914-926.
- [18] H. T. FANG & H. F. CHEN, *Stability and instability of limit points for stochastic approximation algorithms*, IEEE Trans. Autom. Control, 45:3(2000), pp. 413-420.
- [19] W. GREBLICKI, *Recursive identification of Wiener system*, Applied Mathematics and computer science, 11:4(2001), pp. 977-991.

- [20] W. GREBLICKI, *Stochastic approximation in nonparametric identification of Hammerstein systems*, IEEE Trans. Autom. Control, 47:11(2002), pp. 1800–1810.
- [21] W. GREBLICKI AND M. PAWLAK, *Nonparametric identification of Hammerstein systems*, IEEE Trans. Inform. Theory, 35:3(1989), pp. 409–418.
- [22] X. L. HU & H. F. CHEN, *Strong consistency of recursive identification for Wiener systems*, Automatica, 41:11(2005), pp. 1905–1916.
- [23] X. L. HU & H. F. CHEN, *Recursive identification for Wiener systems using truncated Gaussian inputs*, submitted for publication.
- [24] J. KORONASKI, *Random-seeking methods for the stochastic unconstrained optimization*, Int. J. Control, 21(1975), pp. 517–527.
- [25] J. KUSHNER & D. S. CLARK, *Stochastic Approximation for Constrained and Unconstrained Systems*, Springer-Verlag, New York, 1978.
- [26] H. J. KUSHNER & G. YIN, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York, 2003.
- [27] L. LJUNG, *System Identification*, Prentice-Hall, Englewood Cliffs, N-J., 1987.
- [28] L. LJUNG, *Analysis of recursive stochastic algorithms*, IEEE Trans. Autom. Control, 22(1977), pp. 551–575.
- [29] M. B. NEVELSON & R. Z. KHASHMINSKII, *Stochastic Approximation and Recursive Estimation*, Amer. Math. Soc., Providence, RI, Translation of Math. Monographs, Vol. 47, 1976.
- [30] B. T. POLYAK & A. B. JUDISKY, *Acceleration of stochastic approximation by averaging*, SIAM J. Control & Optimization, 30(1992), pp. 838–855.
- [31] H. ROBBINS & S. MONRO, *A stochastic approximation method*, Ann. Math. Statist., 22(1951), pp. 400–407.
- [32] D. RUPPERT, *Stochastic Approximation*, In B. K. Ghosh & P. K. Sen, Editors, Handbook in Sequential Analysis, Marcel Dekker, New York, 1991, pp 503–529.
- [33] T. SÖDERSTRÖM, *Errors-in-variables in system identification*, Plenary talk at the 14th IFAC Symposium on System Identification, March 29–31, 2006, Newcastle, Australia.
- [34] T. SÖDERSTRÖM, K. MAHATA, & U. SOVERINI, *Identification of dynamic errors-in-variables models: Approaches based on two-dimensional ARMA modeling of the data*, Automatica, 39(2003), pp. 929–935.
- [35] T. SÖDERSTRÖM & P. STOICA, *System Identification*, Prentice Hall International, Hemel Hempstead, UK, 1989.
- [36] J. C. SPALL, *Multivariate stochastic approximation using a simultaneous perturbation gradient approximation*, IEEE Trans. Autom. Control, 37(1992),

pp. 331-341.

- [37] P. STOICA, *Generalized Yule-Walker equations and testing the orders of multivariate time series*, Int. J. Control, 37:5(1983), pp. 1159-1166.
- [38] J. VÖRÖS, *Parameter identification of Wiener systems with discontinuous nonlinearities*, Systems and Control Letters, 44(2001), pp. 363-372.
- [39] T. WIGREN, *Convergence analysis of recursive algorithms based on the nonlinear Wiener model*, IEEE Trans. Automat. Contr., 39(1994), pp. 2191-2206.