Article

An Introduction to Development of Centralized and Distributed Stochastic Approximation Algorithm with Expanding Truncations

Wenxiao Zhao [1,2]

1   Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; wxzhao@amss.ac.cn; Tel.: +86-10-8254-1271
2   School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: The stochastic approximation algorithm (SAA), starting from the pioneer work by Robbins and Monro in 1950s, has been successfully applied in systems and control, statistics, machine learning, and so forth. In this paper, we will review the development of SAA in China, to be specific, the stochastic approximation algorithm with expanding truncations (SAAWET) developed by Han-Fu Chen and his colleagues during the past 35 years. We first review the historical development for the centralized algorithm including the probabilistic method (PM) and the ordinary differential equation (ODE) method for SAA and the trajectory-subsequence method for SAAWET. Then, we will give an application example of SAAWET to the recursive principal component analysis. We will also introduce the recent progress on SAAWET in a networked and distributed setting, named the distributed SAAWET (DSAAWET).

1. Introduction

SAA was first introduced by Robbins and Monro in 1950s [1] and is also called the RM algorithm in the literature. Since the 1950s, SAA has found successful applications in systems and control, statistics, machine learning, the stochastic arithmetic, and the CESTAC method for numerical calculation [2,3], energy internet [4], and so forth, and there are plenty of studies on various theoretical properties of SAA, including the sample path convergence [1,5–7], the weak convergence [8], convergence of time-varying root-searching functions [9–12], robustness of SAA [13], stable and unstable limit points [14], convergence rate and asymptotic normality [7,15–20], and so forth. In this paper, due to space limitations, we will not give an overview of all aspects of SAA, but focus on the sample path analysis of this kind of algorithm, and particularly introduce the centralized and distributed SAAWET developed by Han-Fu Chen and his colleagues during the past 35 years.

At the early stage of SAA, to guarantee the convergence of the algorithm, it usually requires restrictive conditions on the noise $\{\varepsilon_k\}_{k \geq 0}$ and the function $f(\cdot)$, for example, $\{\varepsilon_k\}_{k \geq 0}$ being a martingale difference sequence and $f(\cdot)$ being bounded by a linear function. Then, by tools from the probability theory, such as the martingale theory, the almost sure convergence of SAA can be established and such an analysis method is called the probabilistic method (PM) [1,20]. However, in many situations, the noise $\{\varepsilon_k\}_{k \geq 0}$ may not be a martingale difference sequence, and it may depend on the past estimates generated by SAA itself, which we call the state-dependent noise, and in such cases, PM fails. Towards relaxing the restrictive conditions in PM, in the 1970s a so-called ordinary differential equation (ODE) method was introduced [7,21], which transforms the convergence analysis of SAA into the stability of the equilibrium point of an associated ODE and forms another representative direction for theoretical analysis of SAA. For the ODE method, it *a priori*

assumes that the estimate sequence generated by the algorithm is bounded, which is not easy to be verified in practice. For many theoretical problems, due to the complexity of the system noise and less *a priori* information on the estimate sequence, neither PM nor the ODE method can be applied.

In order to essentially relax the assumptions required by PM and the ODE method, in the past 35 years an expanding truncation technique was introduced into SAA by Han-Fu Chen and his colleagues, and a so-called stochastic approximation algorithm with expanding truncations (SAAWET) has been developed [22,23]. The key idea of the expanding truncation technique lies in that if the magnitude of the estimate exceeds the truncation bound, then the estimate is reset to the initial value but with a decreased stepsize, and, at the same time, an enlarged truncation bound will be used for the next recursion, while if the magnitude of estimate is smaller than the truncation bound, then SAAWET evolves as the classical SAA in the next recursion and the truncation bound remains unchanged. The expanding truncation technique introduces a way of adaptively choosing the step-sizes of the algorithm and essentially relaxes the conditions on the noises and the root-searching function. Furthermore, to guarantee the convergence of SAAWET, one only needs to verify the noise condition along the index sequence $\{n_k\}_{k\geq 0}$ of any convergent subsequence $\{x_{n_k}\}_{k\geq 0}$ rather than along the whole sequence $\{x_k\}_{k\geq 0}$, which greatly facilitates the application of such an algorithm. The convergence analysis method of SAAWET is accordingly named the trajectory-subsequence (TS) method. In the past 35 years, SAAWET has successfully been applied to recursive identification of a multivariate ARMAX system [24], recursive identification of nonlinear systems [25,26], adaptive regulation of a nonlinear system [27], consensus control of multiagent systems [28], convergence of a distributed randomized PageRank algorithm [29], and so forth. In this paper, we first review the probabilistic method (PM) and the ordinary differential equation (ODE) method for SAA and the trajectory-subsequence method for SAAWET. Then, we will give an application example of SAAWET to the recursive principal component analysis.

All the above algorithms and methods are applied in a centralized situation. In recent years, distributed algorithms have been extensively investigated—for example, the consensus problem [30,31], distributed estimation [32,33], and so on. The distributed algorithms work in a network situation, where there is no central agent collecting and processing the data over the whole network but the agents cooperatively accomplish a global objective by using their local observations and information obtained from communication with their adjacent neighbours [34–38]. Recently, many distributed problems have been solved by SAA-based algorithms, such as distributed estimation [32,39], and distributed convex optimization [34,35]. As a result, the distributed algorithm for SAA has naturally drawn much attention from researchers. Distributed SAA (DSAA) was proposed in [40–43] and so forth, in which agents cooperatively find roots of a function, being a sum of local functions associated with agents connected in a network. It is noticed that almost all the above DSAA algorithms require restrictive conditions to guarantee convergence; for example, in [41] it requires the global Lipschitz continuity of local functions and the observation noise as a martingale difference sequence. However, these assumptions may not hold for many problems, including the distributed principal component analysis and the distributed gradient-free optimization, and so forth. Aiming at solving the distributed root-seeking problem under weaker conditions compared with those used in [41], a distributed version of SAAWET (DSAAWET) was recently developed [44]. In this paper, we will introduce DSAAWET and its convergence properties and also give some application examples of DSAAWET.

The rest of the paper is arranged as follows. In Section 2, the centralized SAA and PM, as well as the ODE method for its convergence analysis are reviewed. Then, SAAWET is introduced with the conditions for its convergence analysis and its general convergence theorem given. In this paper, SAAWET is also applied to solve the recursive principal component analysis problem. In Section 3, DSAAWET and its convergence properties

are introduced, and some application examples are given as well. In Section 4, some concluding remarks are addressed. The notations used in this paper are listed in Table 1.

Table 1. Notations.

$\|v\|, \|A\|$	L_2 norm of vector v, matrix A
$\|v\|_1, \|A\|_1$	L_1 norm of vector v, matrix A
$\mathbf{I}_m$	$m \times m$ identity matrix
$\mathbf{1}$	Vector or matrix with all entries equal to 1
$\mathbf{0}$	Vector or matrix with all entries equal to 0
X^T	Transpose of matrix X
$\text{col}\{x_1, \ldots, x_m\}$	$\text{col}\{x_1, \ldots, x_m\} \triangleq \{x_1^T, \ldots, x_m^T\}^T$, stacks of vectors or matrices $\{x_1, \ldots, x_m\}$
$\mathbb{I}_A(x)$	Indicator function, $\mathbb{I}_A(x) = 1$ if $x \in A$, $\mathbb{I}_A(x) = 0$ otherwise
$\mathbb{E}[\cdot]$	Expectation operator
$m(k, T)$	$m(k, T) \triangleq \max\{m : \sum_{i=k}^{m} a_i \le T\}$
$\sigma_{i,k}$	Truncation number of agent i at time k
$\hat{\sigma}_{i,k}$	$\hat{\sigma}_{i,k} \triangleq \max_{j \in N_i(k)} \sigma_{j,k}$, where $N_i(k)$ is the set of neighbors of agent i at time k
σ_k	Largest truncation number of all agents at time k, that is, $\sigma_k = \max_{i \in \mathcal{V}} \sigma_{i,k} = \max_{i \in \mathcal{V}} \hat{\sigma}_{i,k}$

2. Centralized Stochastic Approximation Algorithm

2.1. Probabilistic Martingale Method

We first give the detailed problem formulation of SAA.

Assume that $f(\cdot) : \mathbb{R}^l \to \mathbb{R}^l$ is an unknown function with root x^0, that is, $f(x^0) = 0$. At time k, assume that x_k is an estimate of x^0 and the measurement of $f(\cdot)$ at x_k is y_{k+1}, that is,

$$y_{k+1} = f(x_k) + \varepsilon_{k+1}, \quad k \ge 0, \tag{1}$$

where ε_{k+1} is the observation noise.

With an arbitrary initial value x_0, in [1] a recursive algorithm for searching the root x^0 of $f(\cdot)$ is introduced

$$x_{k+1} = x_k + a_k y_{k+1}, \quad k \ge 0, \tag{2}$$

where $\{a_k\}_{k \ge 1}$ is the stepsize sequence which satisfies

$$a_k > 0, \quad a_k \xrightarrow[k \to \infty]{} 0, \quad \sum_{k=0}^{\infty} a_k = \infty. \tag{3}$$

Assumption (3) requires that the stepsize sequence tends to zero, but the sum of the stepsizes goes to infinity. This indicates that the noise effect can be asymptotically reduced while the algorithm has the ability to search in the whole feasible domain.

Example 1. *Assume that at time k, the estimate x_k is very close to the root x^0. Then there exists an $\varepsilon > 0$ small enough, such that*

$$\|x_k - x^0 + a_k f(x_k)\| < \varepsilon. \tag{4}$$

Then,

$$\|x_{k+1} - x^0\| = \|x_k - x^0 + a_k f(x_k) + a_k \varepsilon_{k+1}\| \geq \|a_k \varepsilon_{k+1}\| - \varepsilon. \tag{5}$$

From (5) it can be seen that if the stepsize sequence $\{a_k\}_{k\geq 1}$ has a positive lower bound, then for the unbounded noises, such as Gaussian variables, the estimation error will not converge to zero. Thus, for the stepsize sequence $\{a_k\}_{k\geq 1}$, $a_k \xrightarrow[k\to\infty]{} 0$ is a necessary condition for convergence of SAA.

Example 2. *Consider the case $\varepsilon_{k+1} \equiv 0$. If $\sum_{k=0}^{\infty} a_k < \infty$ and $f(\cdot)$ is bounded, then*

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| \leq \sum_{k=0}^{\infty} a_k \|f(x_k)\| < \infty. \tag{6}$$

If the initial value x_0 satisfies $\|x_0 - x^0\| - \sum_{k=0}^{\infty} a_k \|f(x_k)\| > 0$, then from (6) it follows that

$$\|x_{k+1} - x^0\| = \|x_{k+1} - x_0 + x_0 - x^0\| \geq \|x_0 - x^0\| - \sum_{k=0}^{\infty} \|x_{k+1} - x_k\| > 0, \tag{7}$$

and $\{x_k\}_{k\geq 0}$ will not converge to x^0. Thus, $\sum_{k=0}^{\infty} a_k = \infty$ is another necessary condition for convergence of SAA.

We make the following assumptions.

A2.1 The stepsize sequence $\{a_k\}_{k\geq 0}$ satisfies

$$a_k > 0, \quad \sum_{k=0}^{\infty} a_k = \infty, \quad \sum_{k=0}^{\infty} a_k^2 < \infty. \tag{8}$$

A2.2 There exists a twice-differentiable function $v(\cdot) : \mathbb{R}^l \to \mathbb{R}$, such that:

   (i)   The Hessian matrix of $v(\cdot)$ is bounded.
   (ii) $v(x^0) = 0$, $v(x) > 0$, $\forall\, x \neq x^0$ and $v(x) \to \infty$ as $\|x\| \to \infty$.
   (iii) For any $\varepsilon > 0$, there exists an $\beta_\varepsilon > 0$ such that

$$\sup_{\|x - x^0\| > \varepsilon} \nabla v(x)^T f(x) = -\beta_\varepsilon < 0. \tag{9}$$

A2.3 $\{\varepsilon_k, \mathcal{F}_k\}_{k\geq 0}$ is an m.d.s., that is, $E[\varepsilon_{k+1}|\mathcal{F}_k] = 0$, $k \geq 0$ and further, $\sup_{k\geq 0} E\|\varepsilon_k\|^2 < \infty$.
A2.4 There exists a constant $c > 0$ such that

$$\|f(x)\|^2 + E[\|\varepsilon_{k+1}\|^2|\mathcal{F}_k] < c(1 + v(x)), \quad k \geq 0. \tag{10}$$

Theorem 1. *([20,45,46]) Assume that A2.1–A2.4 hold. Then, with an arbitrary initial value x_0, estimates generated from (2) converge to the root x^0 of $f(\cdot)$ almost surely, that is,*

$$x_k \xrightarrow[k\to\infty]{} x^0 \quad \text{a.s.} \tag{11}$$

Proof. Here, we list the sketch of the proof.
By using the function $v(\cdot)$ given in A2.2, define v_{k+1}:

$$v_{k+1} \triangleq (1 + v(x_{k+1})) \prod_{i=k+1}^{\infty} (1 + a_i^2), \tag{12}$$

By (9), we can obtain

$$
\begin{aligned}
0 \leq &E(v_{k+1}|\mathcal{F}_k) \\
\leq &v_k + a_k \nabla v(x_k)^T f(x_k) \prod_{i=k+1}^{\infty}(1 + a_i^2) \\
\leq &v_k + a_k \nabla v(x_k)^T f(x_k) \leq v_k.
\end{aligned}
\tag{13}
$$

By the properties of a nonnegative supermartingale, we have that $\{v_k\}_{k \geq 0}$ converges almost surely. Noting (8), we further obtain $\prod_{i=k+1}^{\infty}(1 + a_i^2) \underset{k \to \infty}{\longrightarrow} 1$ and thus, $\{v(x_k)\}_{k \geq 0}$ converges almost surely.

With an arbitrary $\varepsilon > 0$, define

$$
G_\varepsilon \triangleq \{x : \|x - x^0\| > \varepsilon\}
\tag{14}
$$

and a stopping sequence $\{\sigma_\varepsilon^i\}_{i \geq 0}$:

$$
\sigma_\varepsilon^i = \min\{t : t > \sigma_\varepsilon^{i-1}, \; x_t \in G_\varepsilon^c\}.
\tag{15}
$$

By (13) and (14) and noting (9), we have

$$
E[v_{k+1}|\mathcal{F}_k] \leq v_k - \beta_\varepsilon a_k I_{[\sigma_\varepsilon^i > k]}.
\tag{16}
$$

From (16) and by the properties of the supermartingale, we have

$$
\sum_{k=1}^{\infty} a_k I_{[\sigma_\varepsilon^i > k]} < \infty \quad \text{a.s.}
\tag{17}
$$

By (17) and (8), we can prove that σ_ε^i is finite almost surely, and there exists a subsequence $\{x_{k_i}\}_{i \geq 1}$ which satisfies $\|x_{k_i} - x^0\| \leq \varepsilon$. Since $\varepsilon > 0$ can be arbitrarily small, we know that there exists a subsequence of $\{x_{k_i}\}_{i \geq 0}$ which converges to x^0. By the convergence of $\{v(x_k)\}_{k \geq 0}$, we can finally prove (11). □

The above proof is based on the convergence of the nonnegative supermartingale, and the analysis method is accordingly named the probabilistic method (PM).

Remark 1. *If A2.2 is changed to*

$$
\inf_{\|x-x^0\|>\varepsilon} \nabla v(x)^T f(x) = \beta_\varepsilon > 0,
\tag{18}
$$

then we can prove the almost sure convergence of the following SAA

$$
x_{k+1} = x_k - a_k y_{k+1}, \; k \geq 0.
\tag{19}
$$

Remark 2. *From A2.3 we can see that it requires the observation noise being an m.d.s. In many problems, the observation noise may contain complicated state-dependent uncertainties, for example, $\varepsilon_{k+1} = \varepsilon_{k+1}(x_k)$. On the other hand, if we choose $v(x) = \|x - x^0\|^2$, then from A2.4 we have*

$$
\|f(x)\|^2 \leq c(1 + \|x - x^0\|)^2,
\tag{20}
$$

which indicates that $f(\cdot)$ is bounded by a linear function. These are restrictions for PM.

2.2. Ordinary Differential Equation Method

We first introduce two technical lemmas.

Lemma 1. *(Arzelà-Ascoli, [47]) Let $\{f_\lambda(t),\ t \in [0, +\infty),\ \lambda \in \Lambda\}$ be a function class satisfying the condition of uniform boundedness and equi-continuity. Then there exists a continuous function $f(t)$ and a subsequence $\{f_{\lambda_k}(t)\}_{k \geq 1}$ such that $\{f_{\lambda_k}(t)\}_{k \geq 1}$ converges to $f(t)$ uniformly on any fixed bounded interval in $[0, +\infty)$.*

Remark 3. *For the function class $\{f_\lambda(t),\ t \in [0, +\infty),\ \lambda \in \Lambda\}$, the equi-continuity means that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $|t - s| < \delta$,*

$$if_\lambda(t) - f_\lambda(s)| < \varepsilon, \quad \forall \lambda \in \Lambda. \tag{21}$$

The following result is well-known for the Lyapunov stability of ODE.

Lemma 2. *Consider the ODE*

$$\dot{x}(t) = f(x(t)),\ t \geq 0 \tag{22}$$

with an equilibrium point x^0, that is, $f(x^0) = 0$. If there exists a continuously differentiable $v(x)$ such that

(i)　$v(x^0) = 0$;

(ii)　$v(x) > 0,\ \forall\ x \neq x^0; v(x) \underset{\|x\| \to \infty}{\longrightarrow} \infty$;

(iii)　$\nabla v(x)^T f(x) < 0,\ \forall\ x \neq x^0$;

then x^0 is the globally stable solution of (22).

Before introducing the conditions and the results for the ODE method, we first introduce a crucial notation to be used in this paper. For any $T > 0$, define

$$m(n, T) \triangleq \max\left\{ m : \sum_{i=n}^{m} a_i \leq T \right\}. \tag{23}$$

From the definition in (23) we can see that $m(n, T)$ is the maximal number of steps starting from n with the sum of stepsizes not exceeding T. Noting that $a_k \to 0$ as $k \to \infty$, we have that $m(n, T) \to \infty$ as $n \to \infty$.

We make the following assumptions.

A3.1 $a_k > 0,\ a_k \longrightarrow_{k \to \infty} 0,\ \sum_{k=0}^{\infty} a_k = \infty$.

A3.2 There exists a twice-differentiable function $v(\cdot) : \mathbb{R}^l \to \mathbb{R}$ such that

(i)　$v(x^0) = 0$;

(ii)　$v(x) > 0,\ \forall\ x \neq x^0$ and $v(x) \to \infty$ as $\|x\| \to \infty$;

(iii)　$\nabla v(x)^T f(x) < 0,\ \forall\ x \neq x^0$.

A3.3 The noise $\{\varepsilon_k\}_{k \geq 0}$ satisfies

$$\lim_{T \to 0} \limsup_{n \to \infty} \frac{1}{T} \left\| \sum_{i=n}^{m(n,T)} a_i \varepsilon_{i+1} \right\| = 0, \quad a_k \varepsilon_{k+1} \underset{k \to \infty}{\longrightarrow} 0. \tag{24}$$

A3.4 The function $f(\cdot)$ is continuous.

Remark 4. *If the stepsizes further satisfy $\sum_{k=0}^{\infty} a_k^2 < \infty$ and the noise $\{\varepsilon_k, \mathcal{F}_k\}_{k \geq 0}$ is an m.d.s. and $\sup_{k \geq 0} E[\|\varepsilon_{k+1}\|^2|\mathcal{F}_k] < \infty$ a.s., then it can be directly verified that $\sum_{k=0}^{\infty} a_k \varepsilon_{k+1} < \infty$, and hence, A3.3 holds. If the noise is not stochastic but satisfies $\varepsilon_k \underset{k \to \infty}{\longrightarrow} 0$, then A3.3 also holds. In this regard, (24) is more general compared with A2.3.*

We give the following result for the ODE method.

Theorem 2. *([6,7,21]) Consider the sample path ω where A3.3 takes place. If A3.1, A3.2, and A3.4 hold on the sample path ω and $\{x_k(\omega)\}_{k\geq 0}$ is bounded, then*

$$x_k(\omega) \xrightarrow[k\to\infty]{} x^0. \tag{25}$$

Proof. We sketch the proof.

Define $t_k \triangleq \sum_{i=0}^{k} a_i$ with $t_0 = 0$. By interpolating the estimates generated from (2), we have a sequence of continuous functions

$$x_t^0 \triangleq \frac{t_k - t}{a_k} x_{k-1} + \frac{t - t_{k-1}}{a_k} x_k, \quad t \in [t_{k-1}, t_k], \tag{26}$$

$$x_k(t) = x_{t+k}^0, \quad k \geq 1, \tag{27}$$

which can be proved to be bounded by the boundedness of $\{x_k\}_{k\geq 0}$ and uniform continuity by A3.3.

Then by Lemma 1, we can prove that there exists a subsequence $\{x_{n_k}(t)\}_{k\geq 0}$ converging a continuous function $x(t)$ satisfying the following ODE

$$\dot{x}(t) = f(x(t)). \tag{28}$$

By A3.2 and Lemma 2 we can further prove that the equilibrium point x^0 is globally stable and the estimates $\{x_k\}_{k\geq 0}$ converge to x^0 on the sample path ω. □

The essential idea of the above proof is to transform the convergence of a recursive algorithm to the stability of an associated ODE, and thus the analysis method is called the ODE method. From assumption A3.3 we can see that the ODE method has a wider application potential than the PM method. However, the boundedness assumption on the estimate sequence is still restrictive. Aiming at removing this condition and further relaxing the technical assumptions on the noise and the root-seeking function, in the next section we will introduce the SAA with expanding truncations (SAAWET) and its general convergence theorem.

2.3. SAAWET and Trajectory-Subsequence Method

We first give an example.

Example 3. *([5]) Consider the root-seeking of the function $f(x) = -(x - 10)^3$ and assume that the observation noise $\varepsilon_k \equiv 0$.*

(i) *If we choose the initial value $x_0 = 0$ and the stepsize $a_k = \frac{1}{k+1}$, it can be directly verified that the estimates generated from (2)*

$$x_0 = 0, \quad x_1 = 1000, \quad x_2 = -485,148,500, \quad x_3 \approx 3.8 \times 10^{25}, \cdots$$

(ii) *If we choose the initial value $x_0 = 0$ and the stepsize $a_k = \frac{1}{k+10^3}$, it can be directly verified that the estimates generated from (2) is convergent, that is, $x_k \xrightarrow[k\to\infty]{} 10$.*

For Example 3, neither the growth rate condition on $f(\cdot)$ used by PM nor the boundedness condition on the estimate sequence used by the ODE method take place. On the other hand, if we can choose the stepsize in an adaptive manner, estimates generated from SAA may still converge. The core idea of SAAWET consists in adaptively choosing the stepsizes. Let us describe it.

Denote by J the root set of an unknown function $f(\cdot) : \mathbb{R}^l \to \mathbb{R}^l$, that is, $f(x^0) = 0$ $\forall x^0 \in J$. Choose $\{M_k\}_{k\geq 0}$ a positive sequence increasingly diverging to infinity, $M_k \xrightarrow[k\to\infty]{} \infty$.

With an arbitrary initial value $x_0 \in \mathbb{R}^l$, the estimate sequence $\{x_k\}_{k\geq 1}$ is generated by the following SAAWET:

$$x_{k+1} = (x_k + a_k y_{k+1}) I_{[\|x_k + a_k y_{k+1}\| \leq M_{\sigma_k}]} + x^* I_{[\|x_k + a_k y_{k+1}\| > M_{\sigma_k}]}, \tag{29}$$

$$\sigma_k = \sigma_{k-1} + I_{[\|x_{k-1} + a_{k-1} y_k\| > M_{\sigma_{k-1}}]}, \quad \sigma_0 = 0, \tag{30}$$

where

$$y_{k+1} = f(x_k) + \varepsilon_{k+1}, \tag{31}$$

y_{k+1} is the observation of $f(\cdot)$ at x_k, ε_{k+1} is the observation noise, a_k is the stepsize and x_k is the estimate for the root set of $f(\cdot)$ at time k.

From (29) and (30), we can see that if the magnitude of $x_k + a_k y_{k+1}$ is located in the truncation bound, then the algorithm evolves as the classical SAA (2), while if the magnitude of $x_k + a_k y_{k+1}$ exceeds the truncation bound, then x_{k+1} is pulled back to x^* with the truncation bound being enlarged and the stepsize being reduced for the next recursion.

We first list conditions for theoretical analysis.

A4.1 $a_k > 0$, $a_k \xrightarrow[k\to\infty]{} 0$, $\sum\limits_{k=1}^{\infty} a_k = \infty$;

A4.2 There exists a continuously differentiable function $v(\cdot): \mathbb{R}^l \to \mathbb{R}$ such that

(i) $\sup\limits_{\delta \leq d(x,J) \leq \Delta} \nabla v(x)^T f(x) < 0, \quad \forall\, 0 < \delta < \Delta,$

(ii) The set $v(J) \triangleq \{v(x), x \in J\}$ is nowhere dense,

(iii) For $x^* \in \mathbb{R}^l$ in (29), there is a $c_0 > 0$ such that $\|x^*\| < c_0$ and $v(x^*) < \inf\limits_{\|x\| = c_0} v(x)$,

A4.3 On the sample path ω, for any convergent subsequence $\{x_{n_k}\}_{k\geq 1}$ generated from (29) and (30), it holds that

$$\lim_{T\to 0} \limsup_{k\to\infty} \frac{1}{T} \left\| \sum_{i=n_k}^{m(n_k,T_k)} a_i \varepsilon_{i+1} \right\| = 0, \quad \forall\, T_k \in (0, T], \tag{32}$$

with $m(n_k, T_k) \triangleq \max\left\{ m : \sum\limits_{i=n_k}^{m} a_i \leq T_k \right\}$;

A4.4 The function $f(\cdot)$ is measurable and locally bounded.

Remark 5. *A set being nowhere dense means that the set of the interior points of its closure is empty. It is clear that a set with a single point is nowhere dense. Note that for the noise condition A3.3, it only requires to verify (32) on a fixed sample path and along the index of any convergent subsequence, which compared with (24) is much easier to be verified in practice. In this regard, the analysis method for SAAWET is called the trajectory-subsequence (TS) method.*

Theorem 3. *([5]) For SAAWET (29) and (30), if A4.1, A4.2 and A4.4 hold, then on the sample path ω where A4.3 takes place, it holds that*

$$d(x_k, J) \xrightarrow[k\to\infty]{} 0, \tag{33}$$

where $d(x_k, J) \triangleq \inf\limits_{y\in J} \|x_k - y\|$. Denote by $\bar{J}$ the closure of J and by J^* the set of the limit points of $\{x_k\}$. It holds that

$$d(x_k, J^*) \xrightarrow[k\to\infty]{} 0, \tag{34}$$

and J^* is a closed connected subset of $\bar{J}$.

It is direct to verify that if the root set of $f(x) = 0$ is a singleton, then under the conditions of Theorem 3 we have that $d(x_k, x^0) \underset{k \to \infty}{\longrightarrow} 0$ and $J^* = J = \{x^0\}$.

Proof. We outline the proof.

Denote by $\{x_{n_k}\}_{k \geq 1}$ a convergent subsequence generated by (29) and (30).

Step 1. By A4.3, prove that for all k large enough and $t > 0$ small enough, there is no truncation for x_l, $l \in \{n_k, \cdots, m(n_k, t)\}$, that is,

$$x_{l+1} = x_l + a_l y_{l+1}, \quad l \in \{n_k, \cdots, m(n_k, t)\}, \tag{35}$$

and

$$\|x_{l+1} - x_l\| \leq ct, \quad l \in \{n_k, \cdots, m(n_k, t)\}. \tag{36}$$

Step 2. By the stability condition A4.2 and (35) and (36), prove that for SAAWET, the number of truncations is finite, that is, algorithms (29) and (30) evolve as the classical SAA after a finite number of steps.

Step 3. To establish the convergence of $\{x_k\}_{k \geq 0}$.　□

Remark 6. *If it is a priori known that estimates $\{x_k\}_{k \geq 0}$ generated from (29) and (30) are in a subspace $S(\subset \mathbb{R}^l)$, then for the convergence of $\{x_k\}_{k \geq 0}$, we only need to verify the stability condition in $J \cap S$ and in such a case assumption A4.2 is formulated as follows:*

A4.2' There exists a continuously differentiable function $v(\cdot) : \mathbb{R}^l \to \mathbb{R}$ such that

(i) $\displaystyle\sup_{\delta \leq d(x, J \cap S) \leq \Delta, \, x \in S} \nabla v(x)^T f(x) < 0, \quad \forall \, 0 < \delta < \Delta,$

(ii) The set $v(J \cap S) \triangleq \{v(x), x \in J\}$ is nowhere dense,

(iii) For $x^* \in S$ in (29), there exists a constant $c_0 > 0$ such that $\|x^*\| < c_0$ and $v(x^*) < \inf_{\|x\| = c_0} v(x)$.

Example 4. *Consider the root-seeking problem in Example 3. Assume that $y_{k+1} = f(x_k) + \varepsilon_{k+1}$ with $\varepsilon_{k+1} - 0.9\varepsilon_k = w_{k+1} + 0.5 w_k$ and $\{w_k\}_{k \geq 1}$ being a sequence of iid Gaussian variables $w_k \sim N(0, 0.1^2)$. In (29) and (30), choose $M_k = 2^{k+1}$, $x^* = 0.5$. By a direct calculation, it obtains that*

$$x_{10} = 0.5 \quad x_{50} = 7.76 \quad x_{100} = 9.26$$
$$x_{200} = 9.46 \quad x_{300} = 9.52 \quad x_{400} = 9.61.$$

From the above sections, we can see that SAAWET does not require that the observation noise is purely stochastic or the estimate sequence is bounded. In fact, it can be shown that conditions for convergence of SAAWET are the weakest possible in a certain sense [5]. In the next section, we will apply SAAWET to give a recursive algorithm for principal component analysis.

2.4. Recursive Principal Component Analysis Algorithm

Before introducing the recursive algorithm, we first make the following assumption.

A5.1 $\{A, A_k, k \geq 1\}$ are $n \times n$ symmetric matrices. A is unknown and $\{A_k\}_{k \geq 1}$ is the observation sequence of A satisfying $A_k \underset{k \to \infty}{\longrightarrow} A$.

The principal component analysis algorithm considered in this paper aims at estimating the eigenvalues and eigenvectors based on the observation sequence $\{A_k\}_{k \geq 1}$. Since A is unknown, if we perform SVD or QR decomposition for each A_k, $k \geq 1$, it would be rather time-consuming. In the following, we will introduce a SAAWET-based recursive algorithm for solving the problem.

The eigenvectors of matrix A are recursively estimated as follows:

Choose the stepsize $a_k = \frac{1}{k}$ and any initial values $u_0^{(1)} \in \mathbb{R}^n$ with unit modular and define

$$\widetilde{u}_{k+1}^{(1)} = u_k^{(1)} - a_k A_k u_k^{(1)}. \tag{37}$$

If $\|\widetilde{u}_{k+1}^{(1)}\| \neq 0$, then

$$u_{k+1}^{(1)} = \widetilde{u}_{k+1}^{(1)} / \|\widetilde{u}_{k+1}^{(1)}\|. \tag{38}$$

Otherwise, $\|\widetilde{u}_{k+1}^{(1)}\| = 0$ and reset $u_{k+1}^{(1)}$ as another vector with a unit modular.

Assume that $u_k^{(i)}$, $i = 1, \cdots, j$ have been well-defined. Next, we inductively define the estimation algorithm for $u_k^{(j+1)}$.

Define the $n \times j$-matrix

$$V_k^{(j)} \triangleq \begin{bmatrix} u_k^{(1)} & P_k^{(1)} u_k^{(2)} & \cdots & P_k^{(j-1)} u_k^{(j)} \end{bmatrix},$$

where $P_k^{(i)} \triangleq I - V_k^{(i)} V_k^{(i)+}$, $i = 1, \cdots, j - 1$ and $V_k^{(i)+}$ is the pseudo inverse of $V_k^{(i)}$.

With an arbitrary initial vector $u_0^{(j+1)}$ with a unit modular, define

$$\widetilde{u}_{k+1}^{(j+1)} = P_k^{(j)} u_k^{(j+1)} - a_k P_k^{(j)} A_k P_k^{(j)} \cdot u_k^{(j+1)}. \tag{39}$$

If $\|\widetilde{u}_{k+1}^{(j+1)}\| \neq 0$, then

$$u_{k+1}^{(j+1)} = \widetilde{u}_{k+1}^{(j+1)} / \|\widetilde{u}_{k+1}^{(j+1)}\|. \tag{40}$$

Otherwise, $\|\widetilde{u}_{k+1}^{(j+1)}\| = 0$ and reset $u_{k+1}^{(j+1)}$ as another vector denoted still by $u_{k+1}^{(j+1)}$ satisfying $\|u_{k+1}^{(j+1)}\| = 1$ and $V_k^{(j)\tau} u_{k+1}^{(j+1)} = 0$.

$\{u_k^{(1)}, \cdots, u_k^{(n)}\}$ serve as the estimates for eigenvectors of matrix A.

Based on $\{u_k^{(1)}, \cdots, u_k^{(n)}\}$, we introduce the recursive algorithms for estimating eigenvalues of matrix A.

Choose a positive sequence increasingly diverging to infinity, $M_{k+1} > M_k > 0$, $M_k \underset{k \to \infty}{\longrightarrow} \infty$. With arbitrary initial values $\lambda_0^{(j)}$, $\{\lambda_k^{(j)}\}_{k \geq 1}$, $j = 1, \cdots, n$ are generated by the following algorithms:

$$\lambda_{k+1}^{(j)} = \left[\lambda_k^{(j)} - a_k \left(\lambda_k^{(j)} - u_k^{(j)T} A_k u_k^{(j)} \right) \right] \cdot I_{\left[\left| \lambda_k^{(j)} - a_k \left(\lambda_k^{(j)} - u_k^{(j)T} A_k u_k^{(j)} \right) \right| \leq M_{\sigma_k(j)} \right]}, \tag{41}$$

$$\sigma_k(j) = \sum_{i=1}^{k-1} I_{\left[\left| \lambda_i^{(j)} - a_i \left(\lambda_i^{(j)} - u_i^{(j)T} A_i u_i^{(j)} \right) \right| > M_{\sigma_i(j)} \right]}, \qquad \sigma_0(j) = 0, \tag{42}$$

where $\lambda_k^{(j)}$ is the estimate for the eigenvalue corresponding to the eigenvector which $u_k^{(j)}$ estimates at time k.

Denote S the unit sphere in $\mathbb{R}^n$ and J the set of all eigenvectors with unit modular of matrix A. Denote the set of all eigenvalues of matrix A by $V(J)$, that is, $V(J) \triangleq \{\lambda^{(1)}, \cdots, \lambda^{(n)}\}$.

For algorithms (39)–(42), the following result takes place.

Theorem 4. *Assume that A5.1 holds.*

(i)　There exists a closed-subset J_j^* of J such that

$$d\left(u_k^{(j)}, J_j^*\right) \underset{k\to\infty}{\longrightarrow} 0. \tag{43}$$

(ii)　There exists $\lambda^{(j)} \in V(J)$ such that

$$d\left(u_k^{(j)T} A u_k^{(j)}, \lambda^{(j)}\right) \underset{k\to\infty}{\longrightarrow} 0,$$

and for any $v^{(j)} \in J_j^*$

$$A v^{(j)} = \lambda^{(j)} v^{(j)}.$$

(iii)　$\lambda_k^{(j)}$ converges to the eigenvalue $\lambda^{(j)}$ of matrix A and the limit set of $\lambda_k^{(j)}$, $j = 1, \cdots, n$ coincides with $V(J)$.

Proof. Noting that $\{u_k^{(j)}\}_{k\geq 0}$, $j =, \cdots, n$ are with a unit modular, algorithms (39)–(42) are SAAWET. The proof can be obtained by applying Theorem 3 and Remark 5. Due to space limitations, the detailed proof is omitted. □

3. Distributed Stochastic Approximation Algorithm

In this section, we will introduce the distributed version of SAAWET (DSAAWET). The key difference from the existing distributed SAA, see for example, [40], lies in the expanding truncation mechanism that adaptively defines enclosing bounds. The theoretical results of DSAAWET guarantee that estimates generated by DSAAWET for all agents converge almost surely to a consensus set, which is contained in the root set of the sum function.

We first introduce the problem formulation.

Consider the case where all agents in a network cooperatively search the root of the sum function given by

$$f(\cdot) = \frac{1}{N} \sum_{i=1}^N f_i(\cdot), \tag{44}$$

where $f_i(\cdot) : \mathbb{R}^l \to \mathbb{R}^l$ is the local objective function which can only be observed by agent i. Denote the root set of $f(\cdot)$ by $J \triangleq \{x \in \mathbb{R}^l : f(x) = 0\}$.

For any $i \in \mathcal{V}$, denote by $x_{i,k} \in \mathbb{R}^l$ the estimate for the root of $f(\cdot)$ given by agent i at time k. Since at time $k+1$, agent i only has its local noisy observation

$$O_{i,k+1} = f_i(x_{i,k}) + \varepsilon_{i,k+1}, \tag{45}$$

where $\varepsilon_{i,k+1}$ is the observation noise, to estimate the root of $f(\cdot)$, all agents need to exchange information with adjacent agents via the topology of the network.

The topology of the network at time k is described by a digraph $\mathcal{G}(k) = \{\mathcal{V}, \mathcal{E}(k)\}$, where $\mathcal{V} = \{1, \cdots, N\}$ is the index set of all agents, and $\mathcal{E}(k) \subset \mathcal{V} \times \mathcal{V}$ is the edge set with $(j, i) \in \mathcal{E}(k)$ representing the information flow from agent j to agent i at time k. Denote the adjacency matrix of the network at time k by $W(k) = [\omega_{ij}(k)]_{i,j=1}^N$, where $\omega_{ij}(k) > 0$ if, and only if $(j, i) \in \mathcal{E}(k)$, and $\omega_{ij}(k) = 0$, otherwise. Denote by $N_i(k) = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}(k)\}$ the set of neighbors of agent i at time k.

We apply the idea of expanding truncation to the distributed estimation. Denote by $x_{i,k}$ the estimate for agent i at time k and by $\sigma_{i,k}$ the number of truncations for agent i up to time k. DSAAWET is given by (46)–(49) with any initial values $x_{i,0}$, where $O_{i,k+1}$ is defined by (45), $\{\gamma_k\}$ is the stepsize, $x^* \in \mathbb{R}^l$ is a vector known to all agents, $\{M_k\}_{k\geq 0}$ is a positive sequence increasingly diverging to infinity with $M_0 \geq \|x^*\|$.

$$\sigma_{i,0} = 0, \quad \hat{\sigma}_{i,k} \overset{\Delta}{=} \max_{j \in N_i(k)} \sigma_{j,k}, \tag{46}$$

$$x'_{i,k+1} = \left(\sum_{j \in N_i(k)} \omega_{ij}(k)(x_{j,k} I_{[\sigma_{j,k} = \hat{\sigma}_{i,k}]} + x^* I_{[\sigma_{j,k} < \hat{\sigma}_{i,k}]}) + \gamma_k O_{i,k+1} \right) I_{[\sigma_{i,k} = \hat{\sigma}_{i,k}]} + x^* I_{[\sigma_{i,k} < \hat{\sigma}_{i,k}]}, \tag{47}$$

$$x_{i,k+1} = x^* I_{[\|x'_{i,k+1}\| > M_{\hat{\sigma}_{i,k}}]} + x'_{i,k+1} I_{[\|x'_{i,k+1}\| \leq M_{\hat{\sigma}_{i,k}}]}, \tag{48}$$

$$\sigma_{i,k+1} = \hat{\sigma}_{i,k} + I_{\left[\|x'_{i,k+1}\| > M_{\hat{\sigma}_{i,k}}\right]}, \tag{49}$$

Denote by

$$\sigma_k \overset{\Delta}{=} \max_{i \in \mathcal{V}} \sigma_{i,k} \tag{50}$$

the largest truncation number among all agents at time k. If $N = 1$, then by denoting $\hat{\sigma}_{i,k} = \sigma_{i,k} \overset{\Delta}{=} \sigma_k$, $x_{i,k} \overset{\Delta}{=} x_k$, $x'_{i,k} \overset{\Delta}{=} x'_k$, $O_{i,k} \overset{\Delta}{=} O_k$, (46)–(49) becomes

$$x'_{k+1} = x_k + \gamma_k O_{k+1},$$
$$x_{k+1} = x^* I_{\left[\|x'_{k+1}\| > M_{\sigma_k}\right]} + x'_{k+1} I_{\left[\|x'_{k+1}\| \leq M_{\sigma_k}\right]},$$
$$\sigma_{k+1} = \sigma_k + I_{\left[\|x'_{k+1}\| > M_{\sigma_k}\right]},$$

which is the SAAWET in the centralized setting.

We first introduce assumptions to be used.

A5.1 $\gamma_k > 0, \gamma_k \underset{k \to \infty}{\longrightarrow} 0$, and $\sum_{k=1}^{\infty} \gamma_k = \infty$.

A5.2 There exists a continuously differentiable function $v(\cdot) : \mathbb{R}^l \to \mathbb{R}$ such that

(a)
$$\sup_{\delta \leq d(x,J) \leq \Delta} f^T(x) v_x(x) < 0 \tag{51}$$

for any $\Delta > \delta > 0$, where $v_x(\cdot)$ denotes the gradient of $v(\cdot)$ and $d(x, J) = \min_y \{ \| x - y \| : y \in J \}$,

(b) $v(J) \overset{\Delta}{=} \{ v(x) : x \in J \}$ is nowhere dense,

(c) $\|x^*\| < c_0$ and $v(x^*) < \inf_{\|x\| = c_0} v(x)$ for some positive constant c_0, where x^* is used in (47) and (48).

A5.3 The local functions $f_i(\cdot) \ \forall i \in \mathcal{V}$ are continuous.

A5.4 (a) $W(k) \ \forall k \geq 0$ are doubly stochastic matrices (A matrix is said to be doubly stochastic if all entries being nonnegative and the sum of entries in each row and each column being 1.);

(b) There exists a constant $0 < \eta < 1$ such that

$$\omega_{ij}(k) \geq \eta \quad \forall j \in \mathcal{N}_i(k) \ \forall i \in \mathcal{V} \ \forall k \geq 0;$$

(c) The digraph $\mathcal{G}_\infty = \{\mathcal{V}, \mathcal{E}_\infty\}$ is strongly connected, where $\mathcal{E}_\infty$ is the set of edges (j, i) such that j is a neighbor of i which communicates with i infinitely often, that is,

$$\mathcal{E}_\infty = \{(j, i) : (j, i) \in \mathcal{E}(k) \text{ for infinitely many indices } k\};$$

(d) There exists a positive integer B such that for every $(j, i) \in \mathcal{E}_\infty$, agent j sends information to the neighbor i at least once every B consecutive time slots, that is,

$$(j, i) \in \mathcal{E}(k) \cup \mathcal{E}(k+1) \cup \cdots \cup \mathcal{E}(k+B-1)$$

for all $(j, i) \in \mathcal{E}_\infty$ and any $k \geq 0$.

Remark 7. *Assumptions A5.1–A5.3 are similar to those conditions used in the centralized SAAWET. Since the DSAAWET is in a distributed setting, assumption A5.4 is a commonly used condition describing the topology of the network [34].*

Next, we introduce the condition on the observation noises for each agent.

A5.5 For the sample path ω under consideration and for each agent i,

(a) $\gamma_k \varepsilon_{i,k+1} \underset{k \to \infty}{\longrightarrow} 0$, and

(b) $\lim_{T \to 0} \limsup_{k \to \infty} \frac{1}{T} \| \sum_{m=n_k}^{m(n_k, t_k)} \gamma_m \varepsilon_{i,m+1} I_{[\|x_{i,m}\| \leq K]} \| = 0 \ \forall t_k \in [0, T]$ along indices $\{n_k\}$ whenever $\{x_{i,n_k}\}$ converges.

We now give the main results of DSAAWET.

Define the vectors $X_k \overset{\triangle}{=} \mathrm{col}\{x_{1,k}, \cdots, x_{N,k}\}$, $\varepsilon_k \overset{\triangle}{=} \mathrm{col}\{\varepsilon_{1,k}, \cdots, \varepsilon_{N,k}\}$, $F(X_k) \overset{\triangle}{=} \mathrm{col}\{f_1(x_{1,k}), \cdots, f_N(x_{N,k})\}$. Denote by $X_{\perp,k} \overset{\triangle}{=} D_\perp X_k$ the disagreement vector of X_k with $D_\perp \overset{\triangle}{=} (\mathbf{I}_N - \frac{\mathbf{11}^T}{N}) \otimes \mathbf{I}_l$, and by $x_k = \frac{1}{N} \sum_{i=1}^N x_{i,k}$ the average of the estimates derived at all agents at time k.

Theorem 5. *([44]) Let $\{x_{i,k}\}$ be produced by (46)–(49) with an arbitrary initial value $x_{i,0}$. Assume A5.1–A5.4 hold. Then on the sample path ω where A5.5 holds, the following assertions take place:*
(i) $\{x_{i,k}\}$ is bounded and there exists a positive integer k_0 possibly depending on ω such that

$$x_{i,k+1} = \sum_{j \in N_i(k)} \omega_{ij}(k) x_{j,k} + \gamma_k O_{i,k+1} \ \forall k \geq k_0, \tag{52}$$

or in the compact form:

$$X_{k+1} = (W(k) \otimes \mathbf{I}_l) X_k + \gamma_k (F(X_k) + \varepsilon_{k+1}) \ \forall k \geq k_0; \tag{53}$$

(ii)$X_{\perp,k} \underset{k \to \infty}{\longrightarrow} \mathbf{0}$ and $d(x_k, J) \underset{k \to \infty}{\longrightarrow} 0$; $\tag{54}$

(iii) There exists a connected subset $J^* \subset J$ such that

$$d(x_k, J^*) \underset{k \to \infty}{\longrightarrow} 0. \tag{55}$$

Proof. The proof can be obtained by first proving the finite number of truncations, then the consensus of estimates, and finally, the convergence of estimates. The detailed proof can be found in [44]. □

DSAAWET provides a likely tool for solving the distributed problems over networks. In fact, DSAAWET has been successfully applied in the following problems:

- Distributed identification of linear systems [48],
- Distributed blind identification of communication channels [49],
- Output consensus of networked Hammerstein and Wiener systems [38].

4. Concluding Remarks

In this paper, we briefly reviewed the historical development of SAA, and in particular, introduced SAAWET and its distributed version (DSAAWET) developed by Han-Fu Chen and his colleagues during the past 35 years. SAAWET and DSAAWET establish general convergence results for the root-seeking problems with noise observations. Since many problems, such as identification, estimation, adaptive control, and optimization can be transformed into the root-seeking problem, SAAWET and DSAAWET can hopefully provide powerful tools for solving such problems.

References

1. Robbins, H.; Monro, S. A stochastic approximation method. *Annu. Math. Stat.* **1951**, *22*, 400–407. [CrossRef]
2. Abadie, J.; Dekhli, F. A variant of the CESTAC method and its application to constrained optimization. *Math. Comput. Simul.* **1988**, *30*, 519–529. [CrossRef]
3. Dickson, J.A.; McLeod, R.D.; Card,H.C. Stochastic arithmetic implementations of neural networks with in situ learning. *IEEE Int. Conf. Neural Netw.* **1993**, 711–716. [CrossRef]
4. Hua, H.; Wei, Z.; Qin, Y.; Wang, T.; Li, L.; Cao, J. A review of distributed control and optimization in energy internet: From traditional methods to artificial intelligence-based methods. *Iet-Cyber-Phys. Syst. Theory Appl.* **2021**, 1–17. [CrossRef]
5. Chen, H.-F. *Stochastic Approximation and Its Applications*; Kluwer: Dordrecht, The Netherland, 2002.
6. Kushner, H.J.; Clark, D.S. *Stochastic Approximation for Constrained and Unconstrained Systems*; Springer: New York, NY, USA, 1978.
7. Kushner, H.J.; Yin, G. *Stochastic Approximation Algorithms and Applications*; Springer: New York, NY, USA, 1997.
8. Kushner, H.J. *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*; MIT Press: Cambridge, MA, USA, 1984.
9. Benaim, M. A dynamical systems approach to stochastic approximation. *Siam Control. Optim.* **1996**, *34*, 437–472. [CrossRef]
10. Chen, H.-F.; Uosaki, R. Convergence analysis of dynamic stochastic approximation. *Syst. Control. Lett.* **1998**, *35*, 309–315. [CrossRef]
11. Dupac, V. A dynamic stochastic method. *Annu. Math. Stat.* **1965**, *36*, 1695–1702. [CrossRef]
12. Uosaki, K. Some generalizations of dynamic stochastic approximation processes. *Annu. Stat.* **1974**, *2*, 1042–1048. [CrossRef]
13. Chen, H.-F.; Guo, L.; Gao, A.J. Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds. *Stoch. Process. Their Appl.* **1988**, *27*, 217–231. [CrossRef]
14. Fang, H.-T.; Chen, H.-F. Stability and instability of limit points of stochastic approximation algorithms. *IEEE Trans. Autom. Control.* **2000**, *45*, 413–420. [CrossRef]
15. Chen, H.-F.; Zhu, Y.M. *Stochastic Approximation*; Scientific and Technological Publisher: Shanghai, China, 1996. (In Chinese)
16. Chen, H.-F. Convergence rate of stochastic approximation algorithms in the degenerate case. *Siam Control. Optim.* **1998**, *36*, 100–114. [CrossRef]
17. Fabian, V. On asymptotically efficient recursive estimation. *Annu. Stat.* **1978**, *6*, 854–856. [CrossRef]
18. Fang, H.-T.; Chen, H.-F. Sharp convergence rates of stochastic approximation for degenerate roots. *Sci. China* **1998**, *41*, 383–392. [CrossRef]
19. Ljung, L.; Pflug, G.; Walk, H. *Stochastic Approximation and Optimization of Random Processes*; Birkhäuser: Basel, Switzerland, 1992.
20. Nevelson, M. B.; Khasminskii, R.Z. Stochastic approximation and recursive estimation. In *Translation of Math. Monographs*; American Mathematical Society: Providence, RI, USA, 1976; Volume 47.
21. Ljung, L. On positive real transfer functions and the convergence of some recursive schemes. *IEEE Trans. Autom. Control.* **1977**, *22*, 539–551. [CrossRef]
22. Chen, H.-F.; Zhu, Y.M. Stochastic approximation procedures with randomly varying truncations. *Sci. Sin.* **1986**, *29*, 914–926.
23. Chen, H.-F. Stochastic approximation and its new applications. In Proceedings of the 1994 Hong Kong International Workshop on New Directions of Control and Manufacturing, Hong Kong, China, 7–9 November 1994; pp. 2–12.
24. Chen, H.-F. New approach to recursive identification for ARMA systems. *IEEE Trans. Autom. Control.* **2010**, *55*, 868–879. [CrossRef]
25. Zhao, W. X.; Chen, H.-F. Identification of Wiener, Hammerstein, and NARX systems as Markov chains with improved estimates for their nonlinearities. *Syst. Control. Lett.* **2012**, *61*, 1175–1186. [CrossRef]
26. Zhao, W. X.; Chen, H.-F. Markov chain approach to identifying Wiener systems. *Sci. China Inf. Sci.* **2012**, *55*, 1201–1217. [CrossRef]
27. Chen, H.-F. Adaptive Regulator for Hammerstein and Wiener systems with noisy observations. *IEEE Trans. Autom. Control.* **2007**, *4*, 703–709. [CrossRef]
28. Fang, H.-T.; Chen, H.-F.; Wen, L. On control of strong consensus for networked agents with noisy observations. *J. Syst. Sci. Complex.* **2012**, *25*, 1–12. [CrossRef]
29. Zhao, W. X.; Chen, H.-F.; Fang, H.-T. Convergence of distributed randomized PageRank algorithms. *IEEE Trans. Autom. Control.* **2013**, *58*, 3255–3259. [CrossRef]
30. Ren, W.; Beard, R.W. Consensus seeking in multiagent systems under dynamically changong interaction topologies. *IEEE Trans. Autom. Control* **2005**, *50*, 655–661. [CrossRef]
31. Olfati-Saber, R.; Murray, R.M. Consuensus problems in networks of agnets with switching topology ans time-delays. *IEEE Trans. Autom. Control* **2004**, *49*, 1520–1533. [CrossRef]

32. Zhang, Q.; Zhang, J.F. Distributed parameter estimation over unreliable networks with Markovian switching topologies. *IEEE Trans. Autom. Control* **2012**, *57*, 2545–2560. [CrossRef]
33. Lopes, C.G.; Sayed, A.H. Diffusion least-mean square over adaptive networks: Formulation and performance analysis. *IEEE Trans. Singnal Process.* **2008**, *56*, 3122–3136. [CrossRef]
34. Nedic, A.; Ozdaglar, A. Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Autom. Control.* **2009**, *54*, 48–61. [CrossRef]
35. Nedic, A.; Ozdaglar, A.; Parrilo, P.A. Constrained consensus and optimization in multi-agent networks. *IEEE Trans. Autom. Control.* **2010**, *55*, 922–938. [CrossRef]
36. Simonetto, A.; Leus, G. Distributed asynchronous time-varying constrained optimization. In Proceedings of the 48th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2–5 November 2014; pp. 2142–2146.
37. Yi, P.; Hong, Y.; Liu, F. Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. *Automatica* **2016**, *74*, 259–269. [CrossRef]
38. Feng, W.; Chen, H.-F. Output Consensus of Networked Hammerstein and Wiener Systems. *SIAM J. Control Optim.* **2019**, *57*, 1230–1254. [CrossRef]
39. lei, J.; Chen, H.-F. Distributed randomized PageRank algorithm based on stochastic approximation. *IEEE Trans. Autom. Control.* **2015**, *60*, 1641–1646. [CrossRef]
40. Bianchi, P.; Fort, G.; Hachem, W. Performance of a distributed stochastic approximation algorithm. *IEEE Trans. Inf. Theory* **2013**, *59*, 7405–7418. [CrossRef]
41. Morral, G.; Bianchi, P.; Fort, G.; Jakubowicz, J. Distributed stochastic approximation: The price of non-double stochasticity. *Proc. Asilomoar* **2012**, 1473–1477.
42. Bokar, V.S. Asynchronous stochastic approximation. *Siam J. Control Optim.* **1998**, *36*, 840–851. [CrossRef]
43. Lei, J.; Fang, H.T.; Chen, H.-F. Asymptotic Properties of Primal-Dual Algorithm for Distributed Stochastic Optimization over Random Networks with Imperfect Communications. *Siam J. Control. Optim.* **2018**, *56*, 2159–2188. [CrossRef]
44. Lei, J.; Chen, H.-F. Distributed stochastic approximation algorithm with expanding truncations. *IEEE Trans. Autom. Control.* **2020**, *65*, 664–679. [CrossRef]
45. Blum, J.R. Multidimensional stochastic approximation. *Annu. Math. Stat.* **1954**, *9*, 737–744. [CrossRef]
46. Gladyshev, E.G. On stochastic approximation. *Theory Probab. Appl.* **1965**, *10*, 275–278. (In Russian) [CrossRef]
47. Dunford, N.; Schwartz, J.T. *Linear Operators, Part I: General Theory*; Wiley-Interscience: New York, NY, USA, 1966.
48. Lei, J.; Chen, H.-F. Distributed estimation for parameter in heterogeneous linear time-varying models with observations at network sensors. *Commun. Inf. Syst.* **2015**, *15*, 423–451. [CrossRef]
49. Liu, R.; Chen, H.-F. Distributed and Recursive Blind Channel Identification to Sensor Networks. *Control Theory Technol.* **2017**, *15*, 274–287. [CrossRef]