**RECENT PROGRESS IN PARALLEL
STOCHASTIC APPROXIMATIONS**

By

G. Yin

IMA Preprint Series # 565

August 1989

RECENT PROGRESS IN PARALLEL STOCHASTIC APPROXIMATIONS†

G. Yin‡

Abstract. Parallel recursive algorithms of stochastic approximation type are discussed. The basic idea is to utilize a collection of parallel processors in lieu of using a single one alone as in the classical algorithms. The processors compute and communicate with each other asynchronously. By asynchronous implementations, we mean that each processor updates on its own pace, and uses the most recent received information from other processors. Several procedures are dealt with. These include distributed communicating schemes via convexifications, parallel algorithms with interacting processors and renewal type of computation times, and real time implementable procedures with pipeline arrangements. It appears that from the closed connection between real world applications and mathematical theories, the study of parallel stochastic approximation methods gains not only much of its special appeal, but also much of its intrinsic tension. The main emphasis and interest are the asymptotic properties of such algorithms. The parallel interacting algorithm and its variation are used as representatives for illustrating various limiting results. Convergence and rate of convergence issues are addressed, and robustness analysis is carried out.

Keywords. stochastic approximation, parallel processing, decentralized stochastic optimization, martingale averaging, convergence, rate of convergence, robustness.

AMS (MOS) classification. 60F15, 60F17, 62L20, 93E10, 93E12, 93E25.

1. Introduction

In this work, we survey some of the recent development in parallel stochastic approximation methods. Several classes of recursive algorithms will be reviewed. Problems such as convergence, rate of convergence, robustness etc. will be addressed.

Stochastic approximation (S.A.) methods have a long and illustrious history. The methods were initiated in [RM] and [KW] for locating unknown zeros or maximum values of a nonlinear function with noisy observations. Substantial progress has been made in the past forty years. To date the progress, we mention the work of [Sa], [Lj], [KC], [MP], [Kus]. In fact, the list include merely a few among others. Extensive surveys can be

† This research has been supported in part by the National Science Foundation under grant DMS-8814624, and in part by the Institute for Mathematics and its Applications with fund provided by the National Science Foundation.

‡ Department of Mathematics, Wayne State University, Detroit, MI 48202.

found in [AG], [Wa], [NK], [BMP] etc. The S.A. methods have been employed widely in estimation and detection, Monte Carlo optimization, adaptive control, signal processing, communication theory and many other related fields (cf. [Tsy], [GRC], [LS], [Ch], [Ru], [Pf], [KV], [Ca], [Er] and the references therein).

Comparing with the rapid development in S.A. methods, the study in parallel stochastic approximations is still in its infancy. In 1984, a class of recursive algorithms of stochastic approximation type, with distributed processors and asynchronous communications was first proposed and analyzed in [Ts] and [TBA]. Some asymptotic results were obtained and various potential applications in stochastic control and system identifications were discussed. Such decentralized algorithms soon attract growing interest.

The aforementioned model was studied further in [KY1]. Utilizing the weak convergence and martingale averaging techniques, convergence properties as well as rate of convergence were established under weaker conditions. Moreover, state dependent noise was treated, communication through noisy channels was dealt with and projection procedures were considered. Later, another class of parallel S.A. algorithms was suggested in [KY2]. Such algorithms utilize parallel processing and distributed computations in a natural way. Instead of using a single processor as in the classical setting, a collection of processors is used. Each processor acts on only part of the system vector. These processors computer and communicate with each other interactively and at generalized renewal times. Some interesting asymptotic theorems were obtained. Based on this model, the problem of finding zeros of a nonlinear function through noisy observation via parallel processing methods was treated in [YZ1]. Some robustness issues were considered in [YZ2] and [GZY]. Along another line, seeking for real time implementable procedures via pipelining of communication and computation, algorithms with “delayed” observations were considered in [ZY].

We would like to take a pause and answer the following questions: Why should one be interested in the study of parallel S.A. problems? Why is it useful and important?

- (1) Distributed algorithms exploit the opportunities provided by parallel processing methods for Monte Carlo methods of system optimization or evaluation.
- (2) In various applications, the input data of the computation may be physically distributed (for example, the underlying computation consists of physically separated processors, estimators, trackers, controllers etc.).
- (3) By means of parallel processing, one may be able to take advantage of state space decomposition. As a result, a large dimensional system may be split into small pieces such that each of the sub-system can be handled by one of the parallel processors.

- (4) By employing and utilizing the parallel structure, stochastic approximation type of recursive procedures can be executed through pipelining of computation and communication.

Parallel computation seem to be advantageous in many situations. The list, in fact, is not exhausted. For instance, by means of parallel computation, the desired final result might be obtained much faster, the algorithms may be more efficient than the classical one and more robust with respect to large burst of errors. It seems that from the closed connection between real world applications and mathematical theories, the study of parallel stochastic approximation methods gains not only much of its special appeal, but also much of its intrinsic tension.

The rest of the paper will be arranged as follows: In the next section, several parallel algorithms of stochastic approximation type are reviewed; section 3 contains some convergence and rate of convergence argument for a distributed computation model under the framework of weak convergence; almost sure convergence and robustness problems are dealt with in section 4. Finally, the paper is closed with some concluding remarks in the last section.

To begin with, a word about the notations is in order. Throughout the paper, z' will denote the transpose of z , and $z_x(x)$ denotes the gradient of $z(x)$, similarly for the second derivatives. $|\cdot|$ stands for either an absolute value of a scalar, or a norm of a vector. It will be clear from the context.

2. Several algorithms

We shall describe several parallel stochastic approximation algorithms, and make effort to indicate the salient feature of each algorithm. Although not all details are spelled out, appropriate references will be provided.

2.1. A convexification algorithm

The algorithm described in this section is due to Tsitsiklis Bertsekas and Athans [TBA] although it is a variation of the original version.

A good survey of decentralized and synchronous algorithms for traditional numerical analysis problem can be found in [Mi]. In many applications, *asynchronous* and decentralized algorithms may be more preferable (cf. [Ts] and the references therein).

Motivated by the work of [Be] and [CM], an iterative scheme was proposed in [Ts] and [TBA]. Let there be q parallel processors, each with a state variable of dimension r . Let X_n^i denote the state of processor i at time n and define $X_n = (X_n^1, \dots, X_n^q)'$. We use X

to denote a qr vector with $X = (X^1, \dots, X^q)'$, where each $X^i \in \mathbb{R}^r$. At time n , processor i , $i \leq q$ decides whether or not to communicate the current value of its state to any other processors $j \neq i$, and takes a noisy observation $b_i(X_n^i, \xi_n^i)$. If processor i does not receive any information from j , for all $j \neq i$, then it updates according to

$$X_{n+1}^i = X_n^i + \varepsilon b_i(X_n^i, \xi_n^i)$$

as in the classical procedures. If there are any communications to processor i , then the communicated information must be weighted non-trivially. Let $\{a_{ji}(n)\}$ be a sequence of diagonal $r \times r$ matrices with nonnegative entries and $\sum_j a_{ji}(n) = I$, the identity matrix in $\mathbb{R}^r$. Note that the notation for $a_{ji}(n)$ is not standard. The subscripts of the matrix is of the form (source, destination)=(column, row). The algorithm is given by

$$X_{n+1}^i = \sum_j a_{ji}(n) X_n^j + \varepsilon b_i(X_n^i, \xi_n^i). \quad (2.1)$$

Notice that $a_{ji}(n)$ can be thought of as non-homogeneous Markov transition matrices. Nonetheless, the matrices are also random. Equation (2.1) simply says that the computations are executed by giving suitable weight to each processor through convexifications. Thus, if there is no communications to processor i , then $a_{ji}(n) = \delta_{ji}I$; otherwise $a_{ji}(n) \geq \alpha_0 I$ for some $\alpha_0 > 0$.

Remark: A model with bounded communication delays can be considered. The essential analysis will be the same. In [TBA], the dimensions of X^i are not all the same. However, except that the notations are more involved, such a case can be treated by using the same mathematical framework as presented here. We are using a constant gain in this work, a sequence of decreasing step-size can be incorporated in our formulation (cf. [KY1] section 8).

In the above model, processors do not need to communicate to each other at every time instance. In addition, they do not need to wait for receiving information being sent to them from other processors before doing the actual computations. Due to the asynchronous communication, some processors may compute faster than others while some may remain idle some of the time.

Write

$$A_n = \begin{pmatrix} a_{11}(n) & \dots & a_{q1}(n) \\ \vdots & \ddots & \vdots \\ a_{1q}(n) & \dots & a_{qq}(n) \end{pmatrix}.$$

Let

$$\Phi(n|k) = \begin{cases} A_n \dots A_k, & \text{for } k \leq n; \\ I_{qr}, & \text{for } k = n + 1. \end{cases}$$

Under suitable conditions, it can be shown that $\lim_n \Phi(n|k) = \Phi_k$ exists w.p.1 uniformly in k , and for each $i \leq r$, all rows $i, i+r, \dots, i+qr-r$ of Φ_k are equal. Assume that Φ_k satisfy a law of large numbers type of condition

$$\frac{1}{n} \sum_{k=m}^{n+m-1} E_m \Phi_k \xrightarrow{n,m} \bar{\Phi} \text{ in probability,}$$

where

$$\bar{\Phi} = \begin{pmatrix} \bar{\phi}_1 & \dots & \bar{\phi}_q \\ \vdots & \ddots & \vdots \\ \bar{\phi}_1 & \dots & \bar{\phi}_q \end{pmatrix} = \begin{pmatrix} \hat{\Phi} \\ \vdots \\ \hat{\Phi} \end{pmatrix},$$

$\bar{\phi}_i$ are the diagonal $r \times r$ matrices with diagonal elements $(\bar{\phi}_{i1}, \dots, \bar{\phi}_{ir})$, and $\sum_j \bar{\phi}_{ji} = 1$. In particular, if $r = 1$, then $\bar{\Phi}$ is a matrix with identical rows. This is expected due to the structure of $a_{ji}(n)$ (recall the corresponding results in Markov chains).

To analyze the asymptotic properties, we take a piecewise constant interpolation defined by

$$X(t) = X_n \text{ for } t \in [n\varepsilon, n\varepsilon + \varepsilon).$$

It was shown in [KY1] that under appropriate conditions, $X^\varepsilon(\cdot) \Rightarrow X(\cdot)$ (the symbol $\Rightarrow$ stands for weak convergence), with $X(\cdot) = (x(\cdot), \dots, x(\cdot))'$, such that $x(\cdot)$ satisfies an ordinary differential equation:

$$\dot{x}_i = \sum_{j=1}^q \bar{\phi}_{ji} \bar{b}_j^i(x), \quad i \leq r, \quad (2.2)$$

where $\bar{b}_i(\cdot)$ denotes the average of $b_i(\cdot, \xi)$ and $\bar{b}_i^j(x)$ denotes its j th component.

In addition to getting the ODE limit, we showed that

$$X^\varepsilon(t_\varepsilon + \cdot) \Rightarrow \Theta = (\theta, \dots, \theta)', \text{ as } \varepsilon \rightarrow 0, t_\varepsilon \rightarrow \infty,$$

where θ is a unique asymptotic stable point of the limiting ODE.

Next, denote $\bar{B}(x) = (\bar{b}_1(x), \dots, \bar{b}_q(x))'$. Consider the normalized error $U_n = X_n - \Theta/\sqrt{\varepsilon}$. Let $U^\varepsilon(\cdot)$ be a suitably interpolated sequence of U_n . We can show that $U^\varepsilon(\cdot) \Rightarrow U(\cdot) = (u(\cdot), \dots, u(\cdot))'$ a Gauss-Markov process with $u(\cdot)$ satisfying the following stochastic differential equation

$$du = \hat{\Phi} \bar{B}_x(\Theta) u dt + dw, \quad (2.3)$$

where $w(\cdot)$ is an r -dimensional Brownian motion. Interested readers are referred to [KY1] for details. In view of the above discussion, these results conform that all processors asymptotically come together, and reach an agreement (cf. [WT]).

The essence of this algorithm is that a collection of parallel processors iterates on the same (or essentially the same) system, with perhaps different dynamics and noise processes. The processors share information and occasionally communicate with each other through convexifications. Some applications of such algorithm for on-line identifications were discussed in [Ts].

2.2. An interacting algorithm

The algorithm discussed in this section was suggested in [KY2]. It is quite natural and suitable for parallel processing. Instead of using a single processor alone as in the traditional approach, a collection of processors are used interactively.

Let X be an rp -vector. Let there be r -processors, such that the i th one controls X^i , with $X^i \in \mathbb{R}^p$. Without loss of generality, we may assume that $p = 1$, i.e., each processor controls one component of the state vector. In our model, the computations are done asynchronously and at random times.

To be more precise, for each $i \leq r$, let processor i takes Y_n^i units of time to complete its n th iteration. For notational simplicity, we assume Y_n^i to be integer-valued random variables. Define $\tau_n^i = \sum_{j=1}^n Y_j^i$ and $\tau_0^i = 0$. These are “generalized renewal” type of random computation times. It is “generalized”, since we do not require Y_n^i to be i.i.d. sequences. Let $\tilde{\xi}_n^i$ be the noise incurred in the n th iteration for processor i , and $X_{\tau_n^i}^i$ denote the value of the i th component of the state at the end of n th iteration of processor i , and $X_{\tau_n^i} = (X_{\tau_n^1}^1, \dots, X_{\tau_n^r}^r)'$. The algorithm is given by:

$$X_{\tau_{n+1}^i}^i = X_{\tau_n^i}^i + \varepsilon b_i(X_{\tau_n^i}^i, \tilde{\xi}_n^i), \quad i \leq r. \quad (2.4)$$

To proceed, we need the following working definitions. Define the process $\{\xi_n^i\}$ by

$$\xi_n^i = \tilde{\xi}_j^i \text{ for } n \in [\tau_j^i, \tau_{j+1}^i),$$

and define

$$N_i(n) = \sup\{j; \tau_j^i \leq n\}, \quad \Delta_n^i = n - \tau_{N_i(n)}^i.$$

$\{\xi_n^i\}$ is a piecewise constant interpolation of $\{\tilde{\xi}_n^i\}$ in the “renewal interval” $[\tau_j^i, \tau_{j+1}^i)$. As is the case of the usual renewal process, for each i , $\{N_i(n)\}$ is a counting processes. It counts the number of computations being completed up to time n . Similarly, Δ_n^i are those “current life” processes. Define augmented vectors $\tilde{X}_n$ by

$$\tilde{X}_n = (\tilde{X}_n^1, \dots, \tilde{X}_n^r)',$$

where $\tilde{X}_n^i = (X_{\tau_{N_i(n)}^1}^1, \dots, X_{\tau_{N_i(n)}^r}^r)'$. The notation appears to be pretty heavy, but it seems to be unavoidable. We have to keep track of all the state values used in the r -processors at each time n , these in general are different for different processors. Although each component $b_i(\cdot)$ still depends on a r -vector, such vectors are in general different from processor to processor due to asynchronous computation times.

Now, put

$$X_n = \left(X_{\tau_{N_1(n)}^1}^1, \dots, X_{\tau_{N_r(n)}^r}^r \right), \text{ and } I_n^i = I_{\{\Delta_n^i=0\}}.$$

In view of the above definitions, equation (2.4) can also be written as

$$X_{n+1}^i = X_n^i + \varepsilon b_i(\tilde{X}_n^i, \xi_n^i) I_{n+1}^i. \quad (2.5)$$

With slight abuse of notations, we write

$$b(\tilde{X}_n, \xi_n) = (b_1(\tilde{X}_n^1, \xi_n^1), \dots, b_r(\tilde{X}_n^r, \xi_n^r))',$$

and

$$I_n = \begin{pmatrix} I_n^1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & I_n^r \end{pmatrix}.$$

Thus, (2.5) can in turn be written as

$$X_{n+1} = X_n + \varepsilon I_{n+1} b(\tilde{X}_n, \xi_n). \quad (2.6)$$

In the above algorithm, each processor takes a random time to complete its iteration. For different processors, the “renewal” times are different. This reflects the salient nature of the current algorithm, but it also brings about much of the difficulties. Due to the asynchronization, the iteration number can no longer be used as a common index of time indicator for all the processors. We thus have to use the elapsed processing time as the time indicator.

For more detailed motivations and discussions, we refer to the paper of [KY2]. This algorithm and its specialized version, which is described in section 2.3, will be used as archetype examples to illustrate various asymptotic approaches in the following sections. The analysis of asymptotic properties of the algorithm formulated in this section will be discussed in section 3 by using weak convergence methods; the algorithm described in section 2.3 will be studied in section 4.

2.3. A parallel RM type of algorithm

We concern ourselves with a specialized version of the algorithm given in section 2.2. The objective here is to locate the zeros of a function with random noise corrupted observations by RM type of algorithms with many processors.

The use of parallel processors come from the following consideration. Suppose that the dimension of the underlying problem is large. Then, enormous memory and extensive computation would be needed. Instead of using the classical RM algorithms, we use a number of processors to solve the problem in a cooperative way. By means of state space decomposition, a large dimensional system can be partitioned into sub-systems with relatively small dimensions. Each processor iterates on only a small portion of the state vector. As a result, micro-computers may be used to carry out the computations.

As in the previous section, let $X \in \mathbb{R}^r$, and let there be r processors such that each processor controls one component of the vector X . Let $b(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^r$ be a function with $b(\cdot) = (b_1(\cdot), \dots, b_r(\cdot))'$, such that $b_i(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}$. The basic algorithm is given by

$$X_{\tau_{n+1}}^i = X_{\tau_n}^i + a_{\tau_n} (b_i(X_{\tau_n}) + \tilde{\xi}_n^i), \quad i \leq r. \quad (2.7)$$

Similarly, we can derive other equations like (2.5) and (2.6). Starting from an arbitrary initial value, X^i is computed recursively based on the most recently determined values of X^j , $j \leq r$. The newly computed values are passed to all other processors. For simplicity, we assume that the passage of information does not involve any time delay. This does not seem to be restrictive. In fact, time delays can be incorporated into our model as long as they are bounded.

In order to get the almost sure convergent result, we shall consider a modified version of the algorithm. In proving w.p.1 convergence theorems of stochastic approximation type, normally, one assumes that either some compact set in the domain of attraction of an averaged ODE (ordinary differential equation) to be entered infinitely often, or the function $b(\cdot)$ has certain growth rate. To ensure the boundedness condition to be satisfied, one often uses a projection or truncation scheme, which in turn requires some *a priori* information on the bounded region. To relax these conditions, a procedure with randomly varying truncations was suggested in [CZ]. Their result was further extended to the non-additive noise case in [YZ3]. We now describe the randomly truncated version of (2.7) as follows. Let $\{M(n)\}$ be a sequence of monotone increasing positive real numbers, such that $M(n) \xrightarrow{n} \infty$. Define σ_n as:

$$\sigma_0 = 0, \quad \sigma_{n+1} = \sigma_n + I_{\{|X_n + a_n I_{n+1}(b(\tilde{X}_n) + \xi_n)| > M(\sigma_n)\}},$$

and write

$$J_n = I_{\{|X_n + a_n I_{n+1}(b(\tilde{X}_n) + \xi_n)| \leq M(\sigma_n)\}}$$

$$J_n^c = I_{\{|X_n + a_n I_{n+1}(b(\tilde{X}_n) + \xi_n)| > M(\sigma_n)\}},$$

where

$$b(\tilde{X}_n) = (b_1(\tilde{X}_n^1), \dots, b_r(\tilde{X}_n^r))',$$

and I_n, ξ_n are similar to the definition as in section 2.2. The algorithm with random truncation bounds is:

$$X_{n+1} = \left(X_n + a_n I_{n+1}(b(\tilde{X}_n) + \xi_n) \right) J_n + \hat{x} J_n^c, \quad (2.8)$$

where $\hat{x} \in \mathbb{R}^r$ is a fixed but otherwise arbitrary point, and $|\hat{x}| < M$ for some $M > 0$.

Remark: It can be easily seen that the truncations are so designed that they force the iterates returning to a bounded region if they ever exit from there. The main advantage of such truncations is that the usual restriction on the growth rate of the function $b(\cdot)$ can be discarded completely. As opposed to the traditional projection or truncation methods, no *a priori* information on the projection region is needed.

2.4. Pipelining of computation and communication

A real time implementable algorithm is discussed in this section. We start with the classical RM algorithm

$$X_{n+1} = X_n + a_n O_n, \quad O_n = b(X_n) + \xi_n, \quad (2.9)$$

where $X, \xi \in \mathbb{R}^r$ and $b(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^r$. The above procedure can be thought of as a two phase computations. The first phase is taking observations O_n , and the second phase is forming new approximations by adding X_n to $a_n O_n$.

In a wide range of applications, most of the computation times are spent on the phase 1 operation. The second phase is normally less time consuming. In particular, this is the case when we have very large dimensional systems, or the places where the data are collected are far away from the data processing center. This feature cannot be seen from (2.9) owing to the fact that n is the iteration number rather than the real time elapsed in one iteration.

Taking the aforementioned fact into account, we shall design another algorithm in this section. A number of processors is lined up as on a production line. The processors update asynchronously. After one iteration is completed, the available data is passed to the next processor on line.

To give some motivations, we consider the following example first. Suppose that the time required for taking a measurement is 3 units of real time, and the time needed for the addition of X_n and $a_n O_n$ is 1 unit. Instead of using a single processor alone to implement the computation, three parallel processors will be employed in a pipeline fashion. All processors perform the same kind of computation, but with different starting times. Let n denote discrete real time. At a given instance n , suppose that processor 1 has completed its measurement taken three time units ago, i.e., O_{n-3} has been obtained, and suppose that X_n is available which was communicated by processor 3. Then, step 2 is executed for processor 1, i.e.,

$$X_{n+1} = X_n + a_{n-3} O_{n-3}.$$

At the next time instance $n + 1$, processor 1 sends its content X_{n+1} to processor 2. For simplicity, we assume that the passage of information is instantaneous, i.e., there is no time delay (again, bounded delays can be considered in the model). By now, processor 2 has completed its measurement started at $n - 2$. Then,

$$X_{n+2} = X_{n+1} + a_{n-2} O_{n-2}$$

is obtained, and the result X_{n+2} is passed to processor 3, etc.

Now, we are in a position to design the algorithm in a more general setting. In view of the above illustration, the overall system consists of a number of parallel processors connected through asynchronous communications. A sequence of estimates is generated by using a number of processors cyclically and periodically. The length of a cycle is equal to the number of processors participating in the computation, which is also equal to the time required for a single processor to complete its phase one operation.

To be more specific, let d be a positive integer, which is the unit of time for a single processor to complete one step of iteration. d identical processors are selected. All processors work on the same system vector X asynchronously, and communicate with each other through pipeline structure. In spite of different time indices, the scheme is the same for all the processors taking part in the computation. Thus, it suffices to use a single formula to describe the procedure, i.e., all d processors iterate according to the same format, but with different starting time. The initial conditions $X_0, X_2, \dots, X_{d-1}$ are given. In place of (2.9), we consider

$$X_{n+1} = X_n + a_{n-d} O_{n-d} \tag{2.10}$$

with $O_n = b(X_n) + \xi_n$. Equivalently, we can write

$$X_{n+1} = X_n + a_{n-r} b(X_{n-d}) + a_{n-d} \xi_{n-d}.$$

One of the crucial questions is that if the algorithm actually works, and do the convergence properties and asymptotic normality still hold? These questions were answered in [ZY]. In that paper, we showed that X_n converges almost surely to the set $E = \{x \in \mathbb{R}^r; b(x) = 0\}$. Assuming further that $E = \{\theta\}$ a singleton set, $a_n = \frac{1}{n}$, and $b(x) = H(x - \theta) + o(|x - \theta|)$, we then established the asymptotic normality, i.e., $\sqrt{n}(X_n - \theta)$ converges in distribution to a normal random variable. This is done, by virtue of an functional invariant theorem in [EK] and examining the interpolation

$$M_n(t) = \frac{[nt]}{\sqrt{n}} \sum_{k=1}^{[nt]} \frac{1}{k} A_{[nt],k} \xi_{k-d}, \text{ for } t \in [0, 1],$$

where $[z]$ denotes the integral part of z , and

$$A_{jk} = \begin{cases} \prod_{l=k+1}^j (I + \frac{1}{l}H), & \text{if } j > k; \\ I, & \text{otherwise.} \end{cases}$$

To some extend, the algorithm discussed in this sub-section can be viewed as an aggregation method. It uses extensive concurrency by means of pipelining of computation and communication for each individual processor. The computing time (for both phase 1 and phase 2) is uniformly across processors because we have to keep load balancing in our mind (cf. [BT]). Moreover, the fact that when one processor transfers its partial results to another processor, all other processors, including the one being communicated, are still in operation, makes the communication penalty relatively insignificant. As a result, the communication penalty will not prevent the concurrent utilization of a large number of processors in parallel when the underlying computing task is large. If the classical procedure is used with a single processor, then a certain amount of time must be spent on the waiting for the required data to become available. In the newly developed algorithm, at any given time, there is always one processor doing the phase 2 computation (addition), and all the others are in phase 1—the data collection mode. Proceeding in this way, the amount of real time required for iteration is reduced from d units to one unit only. It seems that the regular pipeline and the “delayed” observation structure provides a way to speed up the underlying computation.

3. Study of asymptotic properties via weak convergence

In this section, we review some asymptotic results by means of weak convergence argument. We shall concentrate on the algorithm described in 2.2. This algorithm is served as a representative for illustrating the weak convergence approach. In fact, we shall try to make the main idea clear without giving much of the technical details. For verbatim proofs, we refer to our paper [KY1], [KY2].

3.1. ODE limit

To study the convergence properties of algorithm (2.2), we define the continuous time interpolation $X^\varepsilon(\cdot)$ by $X^\varepsilon(t) = X_n$ on $[n\varepsilon, n\varepsilon + \varepsilon)$. In this section, we treat the weak convergence of $\{X^\varepsilon(\cdot)\}$ as $\varepsilon \rightarrow 0$. The weak convergence methods have been found to be very useful to study various recursive stochastic algorithms (cf. [KH], [KS], [Kus] and the references therein). We shall use various notion of weak convergence freely. Consult the work of [Bi], [Kus] and [EK] among others, for reference of definitions, notations, basic facts and theorems.

The weak convergence arguments usually require that first tightness of the underlying sequence be established, and then the limit process be identified or characterized. Roughly speaking, tightness is equivalent to sequential compactness in any complete separable metric space.

For simplicity, we shall consider the case that the noise processes $\{\tilde{\xi}_n^i\}$, $i \leq r$ are “exogenous” in this paper. For more general state dependent noise cases, see the treatment of [KY2] and [KS].

We assume that:

(A.1.1) For each $i \leq r$, $b_i(\cdot, \xi)$ is continuous in x , uniformly on each bounded (x, ξ) set. For each $N < \infty$, $\sup_{|x| \leq N} \{|b_i(x, \tilde{\xi}_n^i)|; i, n\}$ is uniformly integrable.

(A.1.2) $\{Y_n^i\}$, $i \leq r$ are bounded, and there are $\tilde{\mu}_i(x, \xi)$, continuous in x uniformly in each bounded (x, ξ) set, such that $E_{\tau_n^i}^\varepsilon Y_n^i = \tilde{\mu}_i(X_{\tau_n^i}, \tilde{\xi}_n^i)$, where $E_{\tau_n^i}^\varepsilon$ denotes the conditioning on the σ -algebra generated by $\mathcal{F}_{\tau_n^i}^\varepsilon$.

(A.1.3) For each $i \leq r$, there exist continuous functions $\bar{b}_i(\cdot)$ and $\mu_i(\cdot)$, such that for each x ,

$$\frac{1}{n} \sum_{j=m}^{m+n} \tilde{E}_m^i b_i(x, \tilde{\xi}_j^i) \xrightarrow{n, m} \bar{b}_i(x) \text{ in probability,} \quad (3.1)$$

and

$$\frac{1}{n} \sum_{j=m}^{m+n} \tilde{E}_m^i \tilde{\mu}_i(x, \tilde{\xi}_j^i) \xrightarrow{n, m} \mu_i(x) \text{ in probability.} \quad (3.2)$$

Remark: The uniform integrability in (A.1.1) is a standard condition in weak convergence methods. The continuity of $b_i(\cdot, \xi)$ can be replaced by “weak” continuity (cf. [KS], [Kus] and [KY2]). (A.1.2) allows the renewal interval to depend on both the state variable and the noise. (A.1.3) is the main averaging assumption. Eq. (3.1) concerns about the average on the noise and Eq. (3.2) takes care of the average on the random computation time.

From the definition of the interpolated process $\{X^\varepsilon(\cdot)\}$, $X^\varepsilon(\cdot) \in D^r[0, \infty)$ the space of functions that are right continuous and have left-hand limits endowed with the Skorohod topology (cf. [Bi], [Kus], and [EK] and the references therein).

Theorem 3.1. *Suppose that (A.1.1) to (A.1.3) are satisfied and suppose the ODE (3.3) has a unique solution on $[0, \infty)$ for each initial condition $x = (x_1, \dots, x_r)'$,*

$$\dot{x}_i = \frac{\bar{b}_i(x)}{\mu_i(x)}, \quad i \leq r. \quad (3.3)$$

If $X^\varepsilon(0) \Rightarrow x(0)$, then $\{X^\varepsilon(\cdot)\}$ is tight, and the limit of any convergent subsequence satisfies Eq. (3.3).

Idea of proof: Without loss of generality, we may assume that X_n^i are bounded, otherwise, we can always use a truncation device as in [Kus]. It can be shown that $\{X^\varepsilon(\cdot)\}$ is tight and that all limits are uniformly Lipschitz continuous. We thus need to characterize the limiting process. This can be divided into the following steps:

(1) We show that a limit theorem holds for the random computation time. For each $i \leq r$, define Z_{ni}^ε , $Z_i^\varepsilon(\cdot)$, B_{ni}^ε and $B_i^\varepsilon(\cdot)$ by

$$Z_{ni}^\varepsilon = \varepsilon \sum_{j=1}^n Y_j^i, \quad Z_i^\varepsilon(t) = Z_{ni}^\varepsilon \text{ on } [n\varepsilon, n\varepsilon + \varepsilon)$$

$$B_{ni}^\varepsilon = \varepsilon \sum_{j=1}^n b_i(X_{\tau_j^i}, \tilde{\xi}_j^i), \quad B_i^\varepsilon(t) = B_{ni}^\varepsilon \text{ on } [n\varepsilon, n\varepsilon + \varepsilon).$$

The set $\{X_i^\varepsilon(\cdot), Z_i^\varepsilon(\cdot), B_i^\varepsilon(\cdot), i \leq r\}$ is tight in the space $D^{3r}[0, \infty)$ and all limits are uniformly Lipschitz continuous. Extract a convergent subsequence, also indexed by ε and with limit $(x_i(\cdot), Z_i(\cdot), B_i(\cdot)), i \leq r$.

(2) We claim that for each $i \leq r$,

$$Z_i(t) = \int_0^t \mu_i(x(Z_i(u))) du. \quad (3.4)$$

To verify this, we need only show that

$$M_i(t) = Z_i(t) - \int_0^t \mu_i(x(Z_i(u))) du$$

is a martingale. Since $M_i(\cdot)$ is Lipschitz continuous, the martingale property will imply that $M_i(t) = \text{constant}$. However, $M_i(0) = 0$, as a result $M_i(t) \equiv 0$ and (3.4) holds.

In order to verify the martingale property, we show that for each bounded and continuous function $h(\cdot)$, arbitrary integer ν and $t_j < t < t + s$,

$$Eh(x(Z_i(t_j)), Z_i(t_j), j \leq \nu) \left(Z_i(t + s) - Z_i(t) - \int_t^{t+s} \mu_i(x(Z_i(u))) du \right) = 0.$$

To begin with, we examine the process $Z_i^\varepsilon(\cdot)$ and show that the limit satisfies the above equation.

Choose a sequence $n_\varepsilon \rightarrow \infty$ as $\varepsilon \rightarrow 0$, such that $\varepsilon n_\varepsilon = \delta_\varepsilon \rightarrow 0$. It can be shown that

$$0 = \lim_{\varepsilon} Eh(X^\varepsilon(Z_i^\varepsilon(t_j)), j \leq \nu) \\ \times \left(Z_i^\varepsilon(t + s) - Z_i^\varepsilon(t) - \delta_\varepsilon \sum_{l=t/\delta_\varepsilon}^{(t+s)/\delta_\varepsilon} \frac{1}{n_\varepsilon} E_{\tau_{ln_\varepsilon}}^\varepsilon \sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \tilde{\mu}_i(X^\varepsilon(Z_i^\varepsilon(\varepsilon j)), \tilde{\xi}_j^i) \right).$$

Letting $l\delta_\varepsilon \rightarrow u$, and by virtue of the weak convergence of $Z_i^\varepsilon(\cdot)$ and the Skorohod imbedding, we need only show

$$\frac{1}{n_\varepsilon} E_{\tau_{ln_\varepsilon}}^\varepsilon \sum_{j=ln_\varepsilon}^{ln_\varepsilon+n_\varepsilon-1} \tilde{\mu}_i(X^\varepsilon(Z_i^\varepsilon(\varepsilon j)), \tilde{\xi}_j^i) \xrightarrow{\varepsilon} \mu_i(x(Z_i(u))) \text{ in probability.}$$

The details can be found in [KY2].

(3) By the same token, we can show

$$B_i(t) = \int_0^t \bar{b}_i(x(Z_i(u))) du.$$

(4) Owing to the definition of the interpolations $X_i^\varepsilon(\cdot)$,

$$X_i^\varepsilon(t) - X_i^\varepsilon(0) = \varepsilon \sum_{j=0}^{N_i(t/\varepsilon)-1} b_i(X_{\tau_j^i}, \tilde{\xi}_j^i) = \varepsilon \sum_{\varepsilon j=0}^{Z_i^{\varepsilon,-1}(t)} b_i(X^\varepsilon(Z_i^\varepsilon(\varepsilon j)), \tilde{\xi}_j^i),$$

where $Z^{\varepsilon,-1}(\cdot)$ and later on $Z_i^{-1}(\cdot)$ denote the inverse of $Z^\varepsilon(\cdot)$ and $Z_i(\cdot)$, respectively (the existence of these inverses is guaranteed by the conditions of Y_n^i and the definition of $Z_i^\varepsilon(\cdot)$). Combining the above results (1)-(3), we have the following

$$x_i(t) = x_i(0) + \int_0^{Z_i^{-1}(t)} \bar{b}_i(x(Z_i(u))) du.$$

This yields the desired result. A few details are omitted.

3.2. Some order of magnitude estimates

To proceed, we shall assume that there is a point θ which is globally attracting for the ordinary differential equation (3.3) and is stable in the sense of Liapunov.

Owing to the Skorohod topology of $D^r[0, \infty)$, the weak convergence provides information on arbitrarily large but bounded intervals. For many applications, it is very useful to have a convergence result which is uniform in t . To proceed further, the following asymptotic estimates are of particular importance.

Suppose that there is a twice continuously differentiable Liapunov function $V(\cdot)$ for (3.3), such that $V(\cdot) > 0$ for $x \neq \theta$, $V(\theta) = 0$; $V(x) \rightarrow \infty$ as $|x| \rightarrow \infty$; $V_{xx}(\cdot)$ is bounded;

$$\sum V_{x_i}(x) \bar{b}_i(x) / \mu_i(x) < -\lambda V(x), \text{ for some } \lambda > 0.$$

Suppose that some “appropriate” conditions are satisfied for the functions $b_i(\cdot)$, the noise processes etc. Then, we have

- (a) $\lim_n EV(X_n) = O(\varepsilon)$;
- (b) The set $\{X_n; n < \infty, \varepsilon > 0\}$ is tight;
- (c) The set $\{(X_n - \theta) / \sqrt{\varepsilon}; n \geq N_\varepsilon, \varepsilon > 0\}$ is tight for $0 < N_\varepsilon \xrightarrow{\varepsilon} \infty$.

To avoid excessive notations, we are being pretty vague about the “appropriate” conditions. The precise statements can be found in [KY2]. The proof of the argument uses a technique known as perturbed Liapunov functions (cf. [Kus], [KH], [BP], [PSV] among others).

3.3. Further limit results

Theorem 3.2. *Let the conditions of Theorem 3.1 be satisfied, M_m^i be a sequence of stopping times and Eqs. (3.1) (3.2) hold with m replaced by M_m^i . If (b) in the above subsection holds, then*

$$X^\varepsilon(t_\varepsilon + \cdot) \Rightarrow \theta \text{ for any sequence } t_\varepsilon \xrightarrow{\varepsilon} \infty.$$

We give indications on how this theorem can be proved. Define

$$\begin{aligned} \hat{Z}_{ni}^\varepsilon &= \varepsilon \sum_{j=N_i(t_\varepsilon/\varepsilon)}^{N_i(t_\varepsilon/\varepsilon)+n-1} Y_j^i, \\ \hat{X}_{ni}^\varepsilon &= \varepsilon \sum_{j=N_i(t_\varepsilon/\varepsilon)}^{N_i(t_\varepsilon/\varepsilon)+n-1} b_i(X_{\tau_j^i}, \tilde{\xi}_j^i) + X_{t_\varepsilon/\varepsilon}^i, \\ \hat{X}_i^\varepsilon(t) &= \hat{X}_{ni}^\varepsilon, \quad \hat{Z}_i^\varepsilon(t) = \hat{Z}_{ni}^\varepsilon \text{ on } [n\varepsilon, n\varepsilon + \varepsilon), \\ \hat{X}_n^\varepsilon &= (\hat{X}_{n1}^\varepsilon, \dots, \hat{X}_{nr}^\varepsilon), \quad \hat{Z}_n^\varepsilon = (\hat{Z}_{n1}^\varepsilon, \dots, \hat{Z}_{nr}^\varepsilon). \end{aligned}$$

In the above, $t_\varepsilon/\varepsilon$ is interpreted as a positive integer. This is no loss in generality since we can always take the integral part of it.

It can be easily seen that $\hat{X}^\varepsilon(\cdot) = X^\varepsilon(t_\varepsilon + \cdot)$ and $\{\hat{X}^\varepsilon(\cdot)\}$ is tight. Extract a weakly convergent subsequence, and denote it (with no loss of generality) by $\{\hat{X}^\varepsilon(\cdot)\}$ with limit $\hat{x}(\cdot)$. The proof is quite similar to that of Theorem 3.1. In fact, we only need to show $\hat{x}(0) = \theta$. To this end, for any $T > 0$,

$$\{\hat{X}^\varepsilon(\cdot), \hat{Z}_i^\varepsilon(\cdot), \hat{X}^\varepsilon(\cdot - T), \hat{Z}_i^\varepsilon(\cdot - T)\} = \{X^\varepsilon(t_\varepsilon + \cdot), Z^\varepsilon(t_\varepsilon + \cdot), X^\varepsilon(t_\varepsilon + \cdot - T), Z^\varepsilon(t_\varepsilon + \cdot - T)\}$$

with limit denoted by $(\hat{x}(\cdot), \hat{Z}(\cdot), \hat{x}_T(\cdot), \hat{Z}_T(\cdot))$, we have $\hat{x}(0) = \hat{x}_T(T)$, and all possible values of $\hat{x}_T(0)$ belong to a set which is tight. For any $\eta > 0$, there is a T_η large enough, such that for all $T \geq T_\eta$,

$$P\{|\hat{x}_T(T) - \theta| > \eta\} < \eta.$$

The theorem thus follows.

Remark: To elaborate further on the above theorem, we consider the special case $b_i(x, \xi^i) = b_i(x) + \xi^i$. In this case, θ is the unique zero of the function $b(x)$. The problem reduces to the Robbins-Monro setting with parallel processors.

Next, we discuss the rate of convergence issue. Define

$$U_n = \frac{X_n - \theta}{\sqrt{\varepsilon}}.$$

We show that the interpolated process of U_n converges to a Gauss-Markov process. The covariance of this process provides a measure of the asymptotic dependence of $X_n - \theta$ on ε . Thus, the study on the normalized error provides a way to analyze rates of convergence.

By virtue of a truncated Taylor expansion,

$$\begin{aligned} U_{n+1}^i &= U_n^i + \varepsilon b'_{ix}(\theta, \xi_n^i) U_n I_{n+1}^i \\ &\quad + \sqrt{\varepsilon} b_i(\theta, \xi_n^i) I_{n+1}^i + o(\varepsilon) O(|U_n|^2). \end{aligned} \tag{3.5}$$

Define the continuous time interpolation of U_n by

$$U^\varepsilon(t) = U_n \text{ on } [\varepsilon n - \varepsilon N_\varepsilon, \varepsilon n - \varepsilon N_\varepsilon + \varepsilon).$$

For any n , let

$$\begin{aligned} z_{ni}^\varepsilon &= \varepsilon \sum_{j=N_\varepsilon}^{N_\varepsilon+n-1} Y_j^i, \quad z_i^\varepsilon(t) = z_{ni}^\varepsilon \text{ on } [n\varepsilon, n\varepsilon + \varepsilon), \\ W_{ni}^\varepsilon &= \sqrt{\varepsilon} \sum_{j=N_\varepsilon}^{N_\varepsilon+n-1} b_i(\theta, \xi_j^i) I_{j+1}^i, \quad W_i^\varepsilon(t) = W_{ni}^\varepsilon \text{ on } [n\varepsilon, n\varepsilon + \varepsilon), \\ \tilde{W}_{ni}^\varepsilon &= \sqrt{\varepsilon} \sum_{j=N_\varepsilon}^{N_\varepsilon+n-1} b_i(\theta, \tilde{\xi}_j) \text{ and } \tilde{W}_i^\varepsilon(t) = \tilde{W}_{ni}^\varepsilon \text{ on } [n\varepsilon, n\varepsilon + \varepsilon). \end{aligned}$$

We assume:

(A.3.1) $b_i(\cdot, \cdot)$ are bounded and continuously differentiable in x (with bounded derivatives) uniformly in ξ , and $E b_i(\theta, \xi_n^i) = 0$.

(A.3.2) The following holds.

(1) For any sequence $N_\varepsilon \rightarrow \infty$,

$$\frac{1}{n} \sum_{j=N_i(n)}^{N_i(N_\varepsilon)+n-1} b_{ix}(\theta, \tilde{\xi}_j^i) \xrightarrow{n} \bar{b}_{ix}(\theta) \text{ in probability.} \quad (3.6)$$

(2) (3.1) (3.2) hold with m replaced by $N_i(N_\varepsilon)$.

(A.3.3) The sets $\{\xi_n^i; n < \infty\}$, $i \leq r$ are mutually independent. For each $i \leq r$, $\tilde{W}_i^\varepsilon(\cdot) \Rightarrow \tilde{W}_i(\cdot)$ a Brownian motion with covariance $\sigma_i^2 t$, for any sequence of stopping times $\{N_\varepsilon\}$ with respect to the σ -algebra $\mathcal{F}_n^\varepsilon$, and $\{\xi_n^i\}$ is uniformly mixing with rate $\phi_k \xrightarrow{k} 0$.

Theorem 3.3. *Under (A.3.1) to (A.3.3), the Brownian motions $\tilde{W}_i(\cdot)$, $i \leq r$ are mutually independent. Suppose the tightness condition (c) of section 3.2 is satisfied, $U^\varepsilon(0) \Rightarrow U(0)$, and the equation (3.7) has a unique solution (in the sense of distribution) for each initial condition $U(0)$. Then, $\{U_i^\varepsilon(\cdot)\}$ $i \leq r$ are tight and the limits satisfy*

$$dU_i = \frac{\bar{b}'_{ix}(\theta)}{\mu_i(\theta)} U dt + \frac{\sigma_i}{\sqrt{\mu_i(\theta)}} d\hat{W}_i, \quad i \leq r, \quad (3.7)$$

where $\hat{W}_i(\cdot)$, $i \leq r$ are standard mutually independent Brownian motions.

We omit the proof here. The computed asymptotic covariance can be used for comparing different algorithms. For some numerical comparisons of the asymptotic performance of this algorithm and the classical procedures, see [KY2].

4. With probability one convergence and robustness

In this section, we review the results obtained in [YZ1] and [GZY]. The computation model was given in section 2.3. The approach we are taking is the random truncation methods (cf. [CZ]). In the first subsection, we state the global convergence, and in the second section we deal with the problem of joint robustness with respect to the Liapunov function and noise processes.

4.1. Convergence property

We let the set of zeros of $b(\cdot)$ be denoted by $E = \{x \in \mathbb{R}^r; b(x) = 0\}$. For simplicity, we take the step-size to be of the simplest form $a_n = \frac{1}{n}$. All the subsequent developments remain to be valid for a_n satisfy conditions of the form: $\limsup_n n a_n < \infty$, and

$$1 - \alpha_1/(n+1) \leq \frac{a_{n+1}}{a_n} \leq 1 - \alpha_2/(n+1),$$

for some $0 < \alpha_2 < \alpha_1 < 2$. Of cause, some modifications are necessary.

The following assumptions are needed in the sequel.

(B.1.1) For each $i \leq r$, $\{Y_n^i\}$ is bounded. There exist a sequence of random variables μ_n^i and a $\mu_i > 0$, such that

$$E(Y_n^i - \mu_n^i | Y_l^i - \mu_l^i) = 0, \quad \text{for } l < n, \quad \mu_n^i \xrightarrow{n} \mu_i \text{ w.p.1.}$$

Moreover, for some $0 < \alpha < 1$,

$$\sum_{n=1}^{\infty} n^{\frac{1}{2}(1-\alpha)} (\mu_n^i - \mu_{n-1}^i) \text{ converges w.p.1.}$$

(B.1.2) For each $i \leq r$,

$$\xi_n^i = \tilde{\phi}_n^i + \tilde{\psi}_n^i,$$

with $\tilde{\psi}_n^i \xrightarrow{n} 0$ w.p.1 and $\sum a_{r_j} \tilde{\phi}_j^i$ converges w.p.1.

Remark: Similar to the definition of ξ_n^i , we can define ϕ_n^i and ψ_n^i in an obvious manner.

(B.1.3) There are a twice continuously differentiable Liapunov function $V(\cdot)$, a real number $M > 0$ (as in the definition of the truncations), such that

$$V(\hat{x}) < \inf\{V(x); |x| = M\} = d, \quad [V(\hat{x}), d] \cap V(E) \neq [V(\hat{x}), d]$$

$$V(x) \neq V(y), \quad V'_x(x) \mu^{-1} b(x) < 0, \quad \forall x \notin E, y \in E,$$

where

$$\mu^{-1} = \begin{pmatrix} \frac{1}{\mu^1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\mu^r} \end{pmatrix}.$$

Theorem 4.1. Under (B.1.1) to (B.1.3), for any initial condition X_1 , algorithm (2.5) is strongly convergent, i.e., $\lim_{n \rightarrow \infty} d(X_n, E) = 0$ w.p.1, where $d(\cdot, \cdot)$ denotes the distant function.

The proof of this theorem consists of four steps.

Step 1. We prove a strong law of large numbers type of lemma for the random computation times. This is stated as:

Lemma 4.1. Under (B.1.1), $\frac{1}{n}\tau_n^i - \mu_i = o\left(\frac{1}{n^{\frac{1}{2}(1-\alpha)}}\right)$.

In view of the above lemma,

$$\frac{\tau_{N_i(n)}^i}{N_i(n)} \leq \frac{n}{N_i(n)} \leq \frac{\tau_{N_i(n)+1}^i}{N_i(n)+1} \frac{N_i(n)+1}{N_i(n)}.$$

As a consequence, w.p.1,

$$\frac{N_i(n)}{n} \xrightarrow{n} \frac{1}{\mu_i}, \quad \frac{N_i(n)}{n} - \frac{1}{\mu_i} = o\left(\frac{1}{n^{\frac{1}{2}(1-\alpha)}}\right). \quad (4.1)$$

In fact, this lemma should be compared with the well-known renewal theorem. It has certain resemblance to the conventional renewal theorem. Owing to (4.1), not only convergence, but also rate of convergence are given.

Step 2. To proceed, we observe that since σ_n is an increasing sequence, either $\sigma_n \rightarrow \sigma < \infty$, or $\sigma = \infty$. The finiteness of σ actually imply the boundedness of the algorithm. If $\sigma < \infty$ w.p.1, then for sufficiently large n ,

$$|X_{n+1}| = \left| X_n + \frac{1}{n} I_{n+1}(b(\tilde{X}_n) + \xi_n) \right| \leq M(\sigma).$$

Hence,

$$X_{n+1} = X_n + \frac{1}{n} I_{n+1}(b(\tilde{X}_n) + \xi_n),$$

i.e., truncations are employed for only finitely many times. Thus, we need only prove the following lemma.

Lemma 4.2. Under (B.1.1) to (B.1.3), $\sigma < \infty$ w.p.1.

Denote $m(n, \eta) = \max\{m; \sum_{j=n}^m \frac{1}{j} \leq \eta\}$. Let $\{X_{n_k}\}$ be a convergent subsequence of $\{X_n\}$. Suppose that $\{\tilde{X}_n\}$ is bounded whenever $\{X_n\}$ is. We show that there is a $\delta > 0$, for any $0 < \eta < \delta$, and some $K_i > 0$ $i = 1, 2$, $K_\eta > 0$, and any $k \geq K_\eta$,

$$\left| \sum_{j=n_k}^{m(n_k, \eta)} \frac{1}{j} I_{j+1}(b(\tilde{X}_j) + \xi_j) \right| \leq K_1, \quad \forall m \in [n_k, m(n_k, \eta) + 1], \quad (4.2)$$

$$|X_m - X_{n_k}| \leq K_2 \eta, \quad \forall m \in [n_k, m(n_k, \eta) + 1].$$

Then, we use these estimate to prove Lemma 4.2. The verification is mainly a proof by contradiction. The approach of the truncation methods is closely related to that of [CZ]. Nevertheless, in our problem, the averaging procedure is in fact tow folds. One is the

average of the noise processes and another one is the average of the computation times. As in the approach of weak convergence discussed in section 3, the two averaging procedures are intertwined, and hence the proofs are more involved and complicated. Let $\{X_k\}$ denote a convergent subsequence with limit $\bar{x}$ (this can be shown to exist). A typical argument in the averaging involves dealing with a term of the form

$$\sum_{j=N_i(k)}^{N_i(m(k,\eta))} \frac{1}{\tau_j^i} b_i(X_{\tau_j^i}).$$

We refer to the paper [YZ1] for detailed discussions.

Step 3. Put

$$t_n = \sum_{j=1}^n a_j = \sum_{j=1}^n \frac{1}{j}, \quad m(t) = \sup\{n; t_n \leq t\}.$$

Let $x^0(\cdot)$ be the piecewise linear interpolation of X_n , such that

$$x^0(t_n) = X_n, \quad x^0(t) = \frac{t_{n+1} - t}{a_n} X_n + \frac{t - t_n}{a_n} X_{n+1}, \quad t \in [t_n, t_{n+1}),$$

and let $\bar{x}(\cdot)$ be the piecewise constant interpolation of X_n with $\bar{x}(t) = X_n$, for $t \in [t_n, t_{n+1})$, and define $x^n(\cdot) = x^0(\cdot + t_n)$. We first show that $\{x^n(\cdot)\}$ is uniformly bounded and equicontinuous. Then, by virtue of the Ascoli-Arzelà's Lemma, we can extract a convergent subsequence $\{x^k(\cdot)\}$ with limit $x(\cdot)$. Note that for each $i \leq r$,

$$x^{k,i}(t) - x^{k,i}(0) = \sum_{j=N_i(k)}^{N_i(m(t+t_k))-1} \frac{1}{\tau_j^i} b_i(X_{\tau_j^i}) + \sum_{j=N_i(k)}^{N_i(m(t+t_k))-1} \frac{1}{\tau_j^i} \tilde{\xi}_j^i + O\left(\frac{1}{k}\right).$$

Using this representation and similar kind of argument as in step 2, we can show that x^i , $i \leq r$ satisfy the ODE $\dot{x}^i = \frac{1}{\mu_i} b_i(x)$, where $x = (x^1, \dots, x^r)'$.

Step 4. By virtue of the well-known stability results of ordinary differential equations and condition (B.1.3),

$$\lim_{t \rightarrow \infty} d(x(t), E) = 0,$$

where $x(\cdot)$ is the limit of $x^k(\cdot)$. Now, by using an argument as in the proof of Lemma 4.2, we can derive a contradiction. This in turn completes the proof of the theorem.

Remark: We point out that the convergence property can also be established by the martingale averaging techniques described in [Ku]. There are certain advantages in each approach. Using that approach, functions with discontinuity can be treated, but some conditions on the growth rate of the function needs to be imposed. The treatment

discussed in this section can deal with functions having arbitrary nonlinearity, however, the continuity seems to be crucial.

4.2. Robustness

Recently, the robustness issue has been brought much attention. It was proved in [Cl] that a strong law of large numbers type of condition for measurement errors with 0 limit is a necessary and sufficient condition for with probability one convergence of the Robbins-Monro method. That result was extended further in [CGG] under much relaxed conditions. Moreover, the robustness issue for stochastic approximations was brought up.

In a broad informal sense, robustness analysis for S.A. algorithms gives an account on the allowable tolerance and relates deviations from idealized assumptions. In other words, it intends to answer the question: “If the idealized model is only approximately true, can one still get a reasonably good estimates for the true parameter within certain bounds?” It is a common practice to prove convergence of stochastic approximation algorithms by using Liapunov function approach. As a result, the following questions are of particular interest. What kind of measurement errors can be tolerated? What kind of deviations can be allowed for the corresponding Liapunov function.

It seems that the analysis of robustness plays an important role in organizing information about the algorithms to a manageable form, and appears to be quite informative and helpful for various applications.

The approach which we are taking is some what different from that of [YZ1]. In lieu of the ODE (ordinary differential equation) method (cf. [Lj], [KC]), elementary inequalities will be employed to establish the desire results. With the assumption that a strong law of large numbers with non-zero limit holds for the noise processes, and the Liapunov function evaluated at θ (the true parameter which we are seeking), is different from zero, we obtain asymptotic upper bounds for the absolute deviations. It turns out that the upper bound depends jointly on the local behavior of the Liapunov function (near the true parameter), and the large time properties of the measurement noise. In this sense, the result obtained is interpreted as joint robustness. In addition, the marginal or individual robustness with respect to either the noise or the Liapunov function is also derived as a direct consequence. Finally, the almost sure convergence as well as the necessary and sufficient condition for convergence are obtained by virtue of the robustness property.

The following assumptions are needed in the subsequent development.

(B.2.1) For each $i \leq r$, $|Y_n^i| \leq \nu$ a.s. for some $\nu > 0$, and there is a $\mu_i > 0$, such that

$$a_n \sum_{j=1}^n Y_j^i \xrightarrow{n} \mu_i \text{ a.s.}$$

Denote $\mu = \max_{i \leq r} \mu_i$.

(B.2.2) For each $i \leq r$, there exist a twice continuously differentiable function $V_i(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ and a positive real number ε_1 , such that

$$|V_i(\theta^i)| = \bar{\varepsilon}^i < \varepsilon_1, \quad |V_{i,x^i}(\theta^i)| = \hat{\varepsilon}^i < \varepsilon_1$$

$$\sup_{|x-\theta| > \varepsilon_1} V_{i,x^i}(x) b_i(x) < 0$$

where $V_{i,x^i}(\cdot)$ denotes the partial derivative of $V_i(\cdot)$ with respect to x^i . Let $V(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}$, be given by

$$V(x) = \sum_{i=1}^r V_i(x^i).$$

Suppose

$$V(x) > 0, \text{ for all } |x - \theta| > \varepsilon_1 \text{ and } \lim_{|x| \rightarrow \infty} V(x) = \infty.$$

(B.2.3) For each $i \leq r$, there is a $\varepsilon_2^i \geq 0$, such that

$$\limsup_{n \rightarrow \infty} a_{\tau_n^i} \left| \sum_{j=1}^n \tilde{\xi}_j^i \right| = \varepsilon_2^i \text{ a.s.}$$

Denote $\varepsilon_2 = \max_{i \leq r} \varepsilon_2^i$, and

$$d_0 = V(x^*) < \inf\{V(x); |x| > M\} = d$$

$$\frac{d - d_0}{2} > \sup\{V(x); |x - \theta| < \varepsilon_1\}$$

where $M > 0$ is as given in section 2.3. In fact, the existence of M is guaranteed by the above condition (B.2.2).

(B.2.4) The sequence $\{a_n\}$ satisfies: $0 < a_n \xrightarrow{n} 0$, $\sum_{n=1}^{\infty} a_n = \infty$, $a_n \geq a_{n+1}$, and assume (without loss of generality) that there is a $0 < b < 1$, such that

$$a_n \leq a_{n+1}(1 + ba_{n+1}).$$

Remark: The above assumptions are the conditions on the random computation times, Liapunov functions and noise processes, respectively. (B.2.1) is essentially a law of large numbers type of condition. It is satisfied for example, if $\sum_{j=1}^n Y_j^i$ are renewal processes and $a_n = \frac{1}{n}$. It also holds for certain stationary processes, and many other processes.

(B.2.3) is a law of large numbers type of condition on the measurement noise. Such assumptions are needed in the robustness analysis. For more discussion on this point and various examples, we refer to [CGG].

(B.2.4) gives conditions on the gain or step size sequence. The regularity requirement on a_n includes a wide variety of sequences. In particular, (B.2.4) is satisfied by the usual choice $a_n = \frac{A}{n}$ for some $A > 0$.

In most of the previous convergence analyses, the noise processes were assumed to satisfy a law of large numbers type (or the like) condition with limit 0, and $V(\theta) = 0$. The robustness analysis deals with the situation where such conditions are violated or disturbed. In our assumption, the Liapunov function takes positive values when it is away from the true parameter. Whenever it is close to the true parameter, a certain amount of variation is allowed. As for the noise processes, the limit of the law of large number type can be different from zero.

The robustness theorems are given as follows.

Theorem 4.2.

- (1) Under assumption (B.2.1)-(B.2.4), there exist $\varepsilon_1^* > 0$, $0 < \varepsilon_2^* \leq 1$, and a function $g(\cdot, \cdot) : [0, \infty) \times [0, \infty) \mapsto [0, \infty)$, which is monotone non-decreasing and left continuous with respect to each variable t_1, t_2 , and

$$\begin{aligned} \lim_{t_2 \rightarrow 0} g(t_1, t_2) &= g(t_1, 0) \\ \lim_{t_1 \rightarrow 0} g(t_1, t_2) &= g(0, t_2) \\ \lim_{\substack{t_1 \rightarrow 0 \\ t_2 \rightarrow 0}} g(t_1, t_2) &= 0; \end{aligned}$$

for all $\varepsilon_1 \in (0, \varepsilon_1^*]$, $\varepsilon_2 \in (0, \varepsilon_2^*]$, and X_n given by (2.8),

$$\limsup_{n \rightarrow \infty} |X_n - \theta| \leq g(\varepsilon_1, \varepsilon_2). \quad (4.3)$$

- (2) If in addition, for each $i \leq r$, $\bar{\varepsilon}^i, \hat{\varepsilon}^i$ given in (B.2.2) satisfy $\bar{\varepsilon}^i = \hat{\varepsilon}^i = 0$, then for all $\varepsilon_2 \in (0, \varepsilon_2^*]$,

$$\limsup_{n \rightarrow \infty} |X_n - \theta| \leq g_2(\varepsilon_2) = g(0, \varepsilon_2). \quad (4.4)$$

- (3) If the conditions are fulfilled with $\varepsilon_2 = 0$, then for all $\varepsilon_1 \in (0, \varepsilon_1^*]$,

$$\limsup_{n \rightarrow \infty} |X_n - \theta| \leq g_1(\varepsilon_1) = g(\varepsilon_1, 0). \quad (4.5)$$

(1) gives an account on the joint robustness; (2) and (3) are the corresponding marginal robustness. $g(\cdot, \cdot)$ is a measure on the joint robustness property. Similarly, $g_1(\cdot)$ and $g_2(\cdot)$ are the marginal measures.

Further more, we can prove the following theorem.

Theorem 4.3. Assume that (B.2.1)-(B.2.4) are satisfied with the following modifications:

- (1) (B.2.2) holds for any $\varepsilon_1 > 0$, i.e., $\bar{\varepsilon}^i = \hat{\varepsilon}^i = 0$, $i \leq r$; (B.2.3) holds with $\varepsilon_2 = 0$. Then, algorithm is strongly consistent, i.e., $X_n \xrightarrow{n} \theta$ a.s.
- (2) Suppose assumptions (B.2.1)-(B.2.4) are satisfied with an additional condition on the step-size:

$$\limsup_{n \rightarrow \infty} na_n < \infty$$

$$1 - \frac{\beta_1}{n+1} \leq \frac{a_{n+1}}{a_n} \leq 1 - \frac{\alpha_1}{n+1}$$

where $0 < \alpha_1 < \beta_1 < 2$. If $X_n \xrightarrow{n} \theta$ a.s., then for each $i \leq r$,

$$a_{\tau_n^i} \sum_{j=1}^n \tilde{\xi}_j^i \xrightarrow{n} 0 \text{ a.s.} \quad (4.6)$$

Idea of proofs: As in the previous section, to prove Theorem 4.3, we have to show the number of truncations is finite first. This is done by establishing the following facts in succession.

(i) Suppose (B.2.3) holds for some $\varepsilon_2 \leq \varepsilon_2^* \leq 1$, and suppose that for each $i \leq r$ and K with $\tau_{N_i(K)}^i > K$, there exists a $n \geq K$, $|\tilde{X}_n^i| \leq M$. Then, there is a T such that for any $\hat{T} \in [0, T]$, and any m satisfying

$$n \leq m \leq m(n, \hat{T}) \equiv \max \left\{ m; \sum_{j=n}^m a_j \leq \hat{T} \right\}$$

$$\left| \sum_{j=n}^{m+1} a_j I_{j+1}(b(\tilde{x}_j) + \xi_j) \right| \leq 6\sqrt{r}\varepsilon_2^* + 2.$$

(ii) Under the conditions of (i), for each $\hat{T} \in [0, T]$, and all $m \in [n, m(n, T) + 1]$,

$$|X_m - X_n| \leq 3\sqrt{r}\varepsilon_2^* + \hat{T}(b_1 + 2\sqrt{r}\varepsilon_2^*).$$

(iii) Under (B.2.1)-(B.2.4) with $\varepsilon_1 \in (0, \varepsilon_1^*]$ and $\varepsilon_2 \in (0, \varepsilon_2^*]$, there is an integer-valued random variable σ , such that

$$\sup_n \sigma_n \leq \sigma < \infty \text{ a.s.}$$

After (iii) is proved, we then carry out analysis as in [CGG], although the averaging is more involved due to the renewal times. The difficulties also stem from the fact that the iteration number is no longer a common index of all the processors.

As for the proof of Theorem 4.3, we note that (1) in Theorem 4.3 can be proved by letting $\varepsilon \rightarrow 0$ in Theorem 4.2, whereas the proof of (2) depends on an observation that for sufficiently large n , the truncations stop, i.e., there is a $\tilde{N}$, such that for all $n \geq \tilde{N}$,

$$X_{n+1} = X_n + a_n I_{n+1}(b(\tilde{X}_n) + \xi_n), \forall n \geq \tilde{N}.$$

Consequently,

$$\begin{aligned} & (X_{n+1} - \theta) + a_n \sum_{j=\tilde{N}-1}^{n-1} \left(\frac{1}{a_j} - \frac{1}{a_{j+1}} \right) (X_{j+1} - \theta) \\ & - \frac{a_n}{a_{\tilde{N}-1}} (X_{\tilde{N}} - \theta) - a_n \sum_{j=\tilde{N}}^n I_{j+1} b(\tilde{X}_j) = a_n \sum_{j=\tilde{N}}^n I_{j+1} \xi_j. \end{aligned}$$

The desired result follows from this expression.

5. Concluding remarks

A survey on some of the recent development in parallel stochastic approximations is made in this work. The purpose is to promote understanding of the potential of using parallel and distributed computation for stochastic iterative schemes, which arise naturally in Monte Carlo optimization and many other related fields. We have focused on those algorithms that are suited for large scale parallelization and represent the hope for solving problems which are much larger than what can be solved at present.

Several algorithms have been discussed here. The asymptotic properties of the algorithms are of our main concerns. We believe that a thorough understanding of an algorithm (by that we mean not only the understanding of its structure and procedure, but also the asymptotic behavior) is essential for applications to various problems. Based on this point of view, we presented some limit results for a class of parallel S.A. algorithms and their variations.

As was mentioned at the beginning, the research in parallel S.A. is quite new. There are still many interesting and challenging problems need to be solved. To be a little bit more specific, the following items are worthwhile considering.

(1) In implementing any recursive algorithm, one needs a good stopping criterion. There is no exception for the parallel S.A. methods. One of the possible approaches is to use the construction of sequentially determined confidence ellipsoids (cf. [Yi] and the references therein). We can design a stopping rule for each individual processor first by constructing appropriate ellipsoidal regions, and then take the maximum value of those

individual stopping times as the stopping rule for the overall system. Among other things, the Bonferroni's inequality is needed in establishing the limit theorem.

(2) Recently, progress has been made for analyzing S.A. type of recursive algorithms via large deviation approach (cf. [DK] and the references therein). It seems that similar approach can be taken to treat the parallel S.A. problems.

(3) An interesting problem is the actual implementation of parallel algorithms, especially the VLSI type of implementations. They will help us to get deeper insight about the algorithms and lead to further improvement.

(4) We have made some numerical comparisons of asymptotic performance for different algorithms in [KY1] and [KY2]. Further investigation in optimal design can be taken. It is clear that designing efficient algorithms is not only theoretically necessary but also practically important.

Parallel and decentralized S.A. algorithms offer promise because they make the utilization of computing power on many problems possible. Our study is actually just started. It is conceivable that parallel stochastic approximations will play more and more important role in the future, especially those in the areas such as systems theory and optimization.

References

- [AG] A.E. Albert and L.A. Gardner, *Stochastic Approximation and Nonlinear Regression*, MIT Press, Cambridge, MA, 1967.
- [BMP] A. Benveniste, M. Metivier and P. Priouret, *Algorithmes Adaptatifs et Approximations Stochastiques*, Masson, Paris, 1987.
- [Be] D.P. Bertsekas, Distributed computation of fixed points, *Mathematical Programming*, 27(1983), 107-120.
- [BT] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computing*, Prentice-Hall, New Jersey, 1989.
- [Bi] P. Billingsley, *Convergence of Probability Measures*, J. Wiley & Sons, New York, 1968.
- [BP] G.B. Blankenship and G.C. Papanicolaou, Stability and control of stochastic systems with wide band noise, *SIAM J. Appl. Math.*, 34(1978), 437-476.
- [Ca] P. E. Caines, *Linear Stochastic Systems*, J. Wiley, New York, 1988.
- [Ch] H.F. Chen, *Recursive Estimation and Control for Stochastic Systems*, J. Wiley, New York, 1985.
- [CGG] H.F. Chen, L. Guo and A.J. Gao, Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds, *Stochastic Process. Appl.*, 27(1988), 217-231.
- [CZ] H.F. Chen and Y.M. Zhu, Stochastic approximation procedures with randomly varying truncations, *Scientia Sinica (series A)*, vol. XXIX No. 9(1986), 914-926.

- [CM] D. Chazan and W. Miranker, Chaotic relaxation, *Linear Algebra and Appl.*, 2(1969), 199-222.
- [Cl] D.S. Clark, Necessary and sufficient conditions for the Robbins-Monro method, *Stochastic Process. Appl.*, 17(1984), 359-367.
- [DK] P. Dupuis and H.J. Kushner, Stochastic approximation and large deviations: upper bounds and w.p.1 convergence, to appear in *SIAM J. Control Optim.*
- [Er] Yu. Ermoliev, Stochastic quasigradient Methods and their applications to system optimization, *Stochastics*, 9(1983), 1-36.
- [EK] S. Ethier and T.G. Kurtz, *Markov Processes*, J. Wiley & Sons, New York, 1986.
- [GRC] G. Goodwin, P. Ramadge and P. Caines, Discrete time stochastic adaptive control, *SIAM J. Control Optim.*, 19(1981), 829-853.
- [GZY] A.J. Gao, Y.M. Zhu and G. Yin, Joint robustness on noise and Liapunov function for parallel stochastic approximation algorithms, Preprint, 1989.
- [KW] E. Kiefer and J. Wolfowitz, Stochastic estimation of the maximum of a regression function, *Ann. Math. Statist.* 23(1952), 462-466.
- [KV] P.R. Kumar and P.P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
- [Ku] H.J. Kushner, Stochastic approximation with discontinuous dynamics and state dependent noise: w.p.1 and weak convergence, *J. Math. Anal. Appl.*, 82(1981), 527-542.
- [Kus] H.J. Kushner, *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.
- [KC] H.J. Kushner and D.S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems*, Springer-Verlag, Berlin, 1978.
- [KH] H.J. Kushner and H. Huang, Asymptotic properties of stochastic approximations with constant coefficients, *SIAM J. Control Optim.*, 19(1981), 86-105.
- [KS] H.J. Kushner and A. Shwartz, An invariant measure approach to the convergence of stochastic approximations with state-dependent noise, *SIAM J. Control Optim.*, 22(1984), 13-27.
- [KY1] H.J. Kushner and G. Yin, Asymptotic properties of distributed and communicating stochastic approximation algorithms, *SIAM J. Control Optim.*, 25(1987), 1266-1290.
- [KY2] H.J. Kushner and G. Yin, Stochastic approximation algorithms for parallel and distributed processing, *Stochastics*, 22(1987), 219-250.
- [LB] S. Li and T. Basar, Asymptotic agreement and convergence of stochastic algorithm, *IEEE Tran. Automat. Control*, AC-32 (1987), 612-618.
- [Lj] L. Ljung, Analysis of recursive stochastic algorithms, *IEEE Trans. Automat. Control* AC-22, 551-575.
- [LS] L. Ljung and T. Söderström, *Recursive identification*, MIT Press, Cambridge, MA, 1983.
- [MP] M. Metivier and P. Priouret, Applications of a Kushner and Clark Lemma to general classes of stochastic algorithms, *IEEE Trans. Inform. Theory*, IT-30(1984), 140-151.
- [Mi] W.L. Miranker, A survey of parallelism in numerical analysis, *SIAM Review*, 13(1971), 524-547.

- [NK] M.B. Nevel'son and R.Z. Khas'minskiĭ, *Stochastic Approximation and Recursive Estimation*, Translation of Math. Monographs, v47, AMS, Providence, 1973.
- [Pf] G. Ch. Pflug, Stochastic minimization with constant step-size: asymptotic laws, *SIAM J. Control Optim.*, 24(1986), 655-666.
- [PSV] G.C. Papanicolaou, D. Stroock and S.R.S. Varadhan, Martingale approach to some limit theorems, *Proc. of the 1976 Duke Univ. Conference on Turbulence*, M. Reed ed., Duke University Mathematics Series, vol. 3, Durham, NC, 1977.
- [Qu] M.J. Quinn, *Designing Efficient Algorithms for Parallel Computers*, McGraw-Hill, New York, 1987.
- [RM] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Statist.* 22(1951), 400-407.
- [Ru] R.Y. Rubinstein, *Monte Carlo Optimization and Sensitivity of Queueing Networks*, J. Wiley & Sons, New York, 1986.
- [Sa] J. Sacks, Asymptotic distribution of stochastic approximation procedures, *Ann. Math. Statist.*, 29(1958), 159-167.
- [Ts] J. Tsitsiklis, Problem in decentralized decision making and computation, Ph.D thesis, MIT, Cambridge, MA, 1984.
- [TBA] J.N. Tsitsiklis, D.P. Bertsekas and M. Athans, Distributed asynchronous deterministic and stochastic gradient optimization algorithms, *IEEE Trans. Automat. Control*, AC-31(1986), 803-812.
- [Tsy] Ya. Z. Tsypkin, *Adaptation and Learning in Automatic Systems*, Academic Press, New York, 1971.
- [Wa] M.T. Wasan, *Stochastic Approximation*, Cambridge Press, London, 1969.
- [WT] R.B. Washburn and D. Teneketzis, Asymptotic agreement among communicating decision makers, *Stochastics*, 13(1984), 103-129.
- [Yi] G. Yin, A stopping rule for the Robbins-Monro method, to appear in *J. Optim. Theory. Appl.*
- [YZ1] G. Yin and Y.M. Zhu, On w.p.1 convergence of a parallel stochastic approximation algorithm, *Probab. Eng. Inform. Sci.*, 3(1989), 55-75.
- [YZ2] G. Yin and Y.M. Zhu, On robustness of the Robbins-Monro method for parallel processing, *Syst. Control Lett.*, 12(1989), 77-86.
- [YZ3] G. Yin and Y.M. Zhu, Almost sure convergence of stochastic approximation algorithms with non-additive noise, *Int'l J. Control*, 49(1989), 1361-1376.
- [ZY] Y.M. Zhu and G. Yin, Stochastic approximation in real time: pipelining computation and communication, Preprint, 1989.

#	Author/s	Title
495	Lucas Hsu, Niky Kamran and Peter J. Olver	Equivalence of Higher Order Lagrangians II. The Cartan Form for Particle Lagrangians
496	D.J. Kaup and Peter J. Olver	Quantization of BiHamiltonian Systems
497	Metin Arik, Fahrünisa Neyzi, Yavuz Nutku, Peter J. Olver and John M. Verosky	Multi-Hamiltonian Structure of the Born-Infeld Equation
498	David H. Wagner	Detonation Waves and Deflagration Waves in the One Dimensional ZND Model for High Mach Number Combustion
499	Jerrold R. Griggs and Daniel J. Kleitman	Minimum Cutsets for an Element of a Boolean Lattice
500	Dieter Jungnickel	On Affine Difference Sets
501	Pierre Leroux	Reduced Matrices and q-log Concavity Properties of q-Stirling Numbers
502	A. Narain and Y. Kizilyalli	The Flow of Pure Vapor Undergoing Film Condensation Between Parallel Plates
503	Donald A. French	On the Convergence of Finite Element Approximations of a Relaxed Variational Problem
504	Yisong Yang	Computation, Dimensionality, and Zero Dissipation Limit of the Ginzburg-Landau Wave Equation
505	Jürgen Sprekels	One-Dimensional Thermomechanical Phase Transitions with Non-Convex Potentials of Ginzburg-Landau Type
506	Yisong Yang	A Note On Nonabelian Vortices
507	Yisong Yang	On the Abelian Higgs Models with Sources
508	Chjan. C. Lim	Existence of Kam Tori in the Phase Space of Vortex Systems
509	John Weiss	Bäcklund Transformations and the Painlevé Property
510	Pu Fu-cho and D.H. Sattinger	The Yang-Baxter Equation for Integrable Systems
511	E. Bruce Pitman and David G. Schaeffer	Instability and Ill-Posedness in Granular Flow
512	Brian A. Coomes	Polynomial Flows on $\mathbb{C}^{n*}$
513	Bernardo Cockburn, Suchung Hou and Chi-Wang Shu	The Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws IV: The Multidimensional Case
514	Peter J. Olver	Invariant Theory, Equivalence Problems, and the Calculus of Variations
515	Daniel D. Joseph and Thomas S. Lundgren with an appendix by R. Jackson and D.A. Saville	Ensemble Averaged and Mixture Theory Equations
516	P. Singh, Ph. Caussignac, A. Fortes, D.D. Joseph and T. Lundgren	Stability of Periodic Arrays of Cylinders Across the Stream by Direct Simulation
517	Daniel D. Joseph	Generalization of the Foscolo-Gibilario Analysis of Dynamic Waves
518	A. Narain and D.D. Joseph	Note on the Balance of Energy at a Phase Change Interface
519	Daniel D. Joseph	Remarks on inertial radii, persistent normal stresses, secondary motions, and non-elastic extensional viscosities
520	D. D. Joseph	Mathematical Problems Associated with the Elasticity of Liquids
521	Henry C. Simpson and Scott J. Spector	Some Necessary Conditions at an Internal Boundary for Minimizers in Finite Elasticity
522	Peter Gritzmann and Victor Klee	On the 0-1 Maximization of Positive Definite Quadratic Forms
523	Fu-Cho Pu and D.H. Sattinger	The Yang-Baxter Equations and Differential Identities
524	Avner Friedman and Fernando Reitich	A Hyperbolic Inverse Problem Arising in the Evolution of Combustion Aerosol
525	E.G. Kalnins, Raphael D. Levine and Willard Miller, Jr.	Conformal Symmetries and Generalized Recurrences for Heat and Schrödinger Equations in One Spatial Dimension
526	Wang Jinghua and Gerald Warnecke	On Entropy Consistency of Large Time Step Godunov and Glimm Schemes
527	C. Guillopé and J.C. Saut	Existence Results for the Flow of Viscoelastic Fluids with a Differential Constitutive Law
528	H.L. Bodlaender, P. Gritzmann, V. Klee and J. Van Leeuwen	Computational Complexity of Norm-Maximization
529	Li Ta-tsien (Li Da-qian) and Yu Xin	Life-Span of Classical Solutions to Fully Nonlinear Wave Equations
530	Jong-Shenq Guo	A Variational Inequality Associated with a Lubrication Problem
531	Jong-Shenq Guo	On the Semilinear Elliptic Equation $\Delta w - \frac{1}{2}y \cdot \nabla w + \lambda w - w^{-\beta} = 0$ in R^n
532	Andrew E. Yagle	Inversion of the Bloch transform in magnetic resonance imaging using asymmetric two-component inverse scattering

Recent IMA Preprints (Continued)

#	Author/s	Title
533	Bei Hu , A Fiber Tapering Problem	
534	Peter J. Olver , Canonical Variables for BiHamiltonian Systems	
535	Michael Renardy , A Well-Posed Boundary Value Problem for Supercritical Flow of Viscoelastic Fluids of Maxwell Type	
536	Michael Renardy , Ill-Posedness Resulting from Slip As a Possible Explanation of Melt Fracture	
537	Michael Renardy , Compatibility Conditions at Corners Between Walls and Inflow Boundaries for Fluids of Maxwell Type	
538	Rolf Rees , The Spectrum of Restricted Resolvable Designs with $r = 2$	
539	D. Lewis and J.C. Simo , Nonlinear stability of rotating pseudo-rigid bodies	
540	Robert Hardt and David Kinderlehrer , Variational Principles with Linear Growth	
541	San Yih Lin and Yisong Yang , Computation of Superconductivity in Thin Films	
542	A. Narain , Pressure Driven Flow of Pure Vapor Undergoing Laminar Film Condensation Between Parallel Plates	
543	P.J. Vassiliou , On Local Equivalence for Vector Field Systems	
544	Brian A. Coomes , On Conditions Sufficient for Injectivity of Maps	
545	Yanchun Zhao , A Class of Global Smooth Solutions of the One Dimensional Gas Dynamics System	
546	H. Holden, L. Holden and N.H. Risebro , Some Qualitative Properties of 2×2 Systems of Conservation Laws of Mixed Type	
547	M. Slemrod , Dynamics of Measured Valued Solutions to a Backward-Forward Heat Equation	
548	Avner Friedman and Jürgen Sprekels , Steady States of Austenitic–Martensitic–Domains in the Ginzburg–Landau Theory of Shape Memory Alloys	
549	Avner Friedman and Bei Hu , Degenerate Hamilton–Jacobi–Bellman Equations in a Bounded Domain	
550	E.G. Kalnins, Willard Miller, Jr., and M.V. Tratnik , Families of Orthogonal and Biorthogonal Polynomials on the N-Sphere	
551	Heinrich Freistühler , On Compact Linear Degeneracy	
552	Matthew Witten , Quantifying the Concepts of Rate and Acceleration/Deceleration of Aging	
553	J.P. Albert and J.L. Bona , Total Positivity and the Stability of Internal Waves in Stratified Fluids of Finite Depth	
554	Brian Coomes and Victor Zurkowski , Linearization of Polynomial Flows and Spectra of Derivations	
555	Yuriko Renardy , A Couette-Poiseuille Flow of Two Fluids in a Channel	
556	Michael Renardy , Short wave instabilities resulting from memory slip	
557	Daniel D. Joseph and Michael Renardy , Stokes' first problem for linear viscoelastic fluids with finite memory	
558	Xiaxi Ding , Superlinear Conservation Law with Viscosity	
559	J.L. Ericksen , Liquid Crystals with Variable Degree of Orientation	
560	F. Robert Ore, Jr. and Xinfu Chen , Electro-Optic Modulation in an Arbitrary Cross-Section Waveguide	
561	M.V. Tratnik , Multivariable biorthogonal continuous-discrete Wilson and Racah polynomials	
562	Yisong Yang , Existence of Solutions for a Generalized Yang-Mills Theory	
563	Peter Gritzmann, Laurent Habsieger and Victor Klee , Good and Bad Radii of Convex Polygons	
564	Martin Golubitsky, Martin Krupa and Chjan. C. Lim , Time-Reversibility and Particle Sedimentation	
565	G. Yin , Recent Progress in Parallel Stochastic Approximations	
566	G. Yin , On H-Valued SA: Finite Dimensional Approximations	
567	Chien-Cheng Chang , Accurate Evaluation of the Effect of Diffusion and Conductivity in Certain Equations	
568	Chien-Cheng Chang and Ruey-Ling Chern , The Effect of Viscous Diffusion in Discrete Vortex Dynamics for Slightly Viscous Flows	
569	Li Ta-Tsien (Li Da-qian) and Zhao Yan-Chun , Global Existence of Classical Solutions to the Typical Free Boundary Problem for General Quasilinear Hyperbolic Systems and its Applications	
570	Thierry Cazenave and Fred B. Weissler , The Structure of Solutions to the Pseudo-Conformally Invariant Nonlinear Schrödinger Equation	
571	Marshall Slemrod and Athanasios E. Tzavaras , A Limiting Viscosity Approach for the Riemann Problem in Isentropic Gas Dynamics	
572	Richard D. James and Scott J. Spector , The Formation of Filamentary Voids in Solids	
573	P.J. Vassiliou , On the Geometry of Semi-Linear Hyperbolic Partial Differential Equations in the Plane Integrable by the Method of Darboux	
574	Jerome V. Moloney and Alan C. Newell , Nonlinear Optics	
575	Keti Tenenblat , A Note on Solutions for the Intrinsic Generalized Wave and Sine-Gordon Equations	
576	P. Szmolyan , Heteroclinic Orbits in Singularly Perturbed Differential Equations	
577	Wenxiong Liu , A Parabolic System Arising In Film Development	