Proceedings of the 33rd
Conference on Decision and Control
Lake Buena Vista, FL - December 1994

Distributed Asynchronous Algorithms with Stochastic Delays for Constrained Optimization Problems with Conditions of Time Drift

Bassem F. Beidas and George P. Papavassilopoulos
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, CA 90089-2536

33rd CDC

Abstract

A distributed asynchronous algorithm for minimizing a function with a nonstationary minimum over a constraint set is considered. The communication delays among the processors are assumed to be stochastic with Markovian character. Conditions which guarantee the mean square and almost sure convergence to the sought solution are presented.

1 Introduction

The recent emphasis on parallel processing is motivated by the compelling need to accelerate computations when solving large dimensional problems in which great memory storage and immense computation capabilities may hinder the performance of centralized algorithms. A number of processors are utilized that operate simultaneously in a collaborative manner on several subproblems decomposed from the original one. To further amend the enhancement of performance, the processors are permitted to communicate asynchronously such that little coordination of communication is maintained. It is shown that dispensing with the synchronization points at the end of each iteration induces improved efficiency, load balancing among processors and reduction of processor idle periods [4, 5, 8].

In this paper, we study asynchronous algorithms with stochastic delays that solve minimization problems with time drifting minimum over a constraint set. The plausibility of the notion of stochastic delays stems from the fact that it models the case of an unpredictable delays in the communication among the processors and therefore addresses various reliability aspects [2]. Constrained optimization problems are prevalent in actual applications, where the nature of the problems solved necessitates imposing natural conditions. Other cases are when the designer often wishes to confine the acceptable values of processors' iterates to lie within a certain region in order to further prevent the processors from straying away from the correct solution.

The paper is organized as follows. In section 2, we describe the model of asynchronous iterations for constrained minimization problems under conditions of time drift. We provide convergence conditions and analysis of the proposed models in section 3. These conditions guarantee also convergence for the case when the problem is time invariant.

2 System Model

We employ a model of n processors working asynchronously on minimizing a function $F(t,x)$ subject to $x \in Q$. The minimum $x^*(t)$ is nonstationary, i.e., changes with time t. We assume that the function $F(t,x)$ is continuously differentiable and that the constraint set Q is nonempty, closed, convex and does not depend on time. Assume that the motion of the minimum is characterized by the known nonlinear function $R(t,x)$ such that

$$x^*(t+1) = R(t, x^*(t)), \tag{1}$$

and the initial value $x^*(0)$ is unknown. Knowledge of the law describing the drift was assumed by Dupac [7] and Tsypkin, Kaplinskiy and Larionov [17]. It, therefore, becomes appropriate to utilize this prediction defined in equation (1) in the formulation of the minimization algorithms.

We assume that the set Q is a Cartesian product of lower dimensional subsets Q_i. This entails that the projection of x on the set Q is equivalent to the projection of x_i on the set Q_i for all i which lends itself naturally to parallelization. Consequently, each processor projects independently on its constraint set. We let $Q_i \subset \mathcal{R}^{n_i}$, where $\sum_{i=1}^{n} n_i = N$ and allow each processor i to update $x_i(t)$. We denote $d_{ji}(t)$ as the delay incurred by transmitting a message from processor j to processor i at time t. We let the communication delays $\{d_{ji}(t)\}$, for all j and i, be stationary Markov chains with state space

$$S = \{1, 2, \ldots, B\},$$

where B is the maximum allowable communication delay for the transmitted messages. We let the probability transition matrix corresponding to $d_{ji}(t)$ be $P_{ji} = (p_{ji}(l,m))$, where

$$p_{ji}(l,m) = \Pr\{d_{ji}(t) = m \mid d_{ji}(t-1) = l\}, \tag{2}$$

for $l, m = 1, 2, \ldots, B$, where here and in the sequel $\Pr\{C\}$ denotes the probability of event C.

As was done in Beidas and Papavassilopoulos [2], we utilize the vector $y^i(t)$ that summarizes the information available to each processor due to the presence of the communication delays, i.e.,

$$y^i(t) = \begin{bmatrix} x_1(t+1-d_{1i}(t)) \\ \vdots \\ x_n(t+1-d_{ni}(t)) \end{bmatrix}. \tag{3}$$

Assume that instead of computing the gradient, the processors are only able to obtain a noise corrupted version of it. The asynchronous gradient projection algorithm proceeds as follows. Processor i evaluates a gradient iteration, projects back onto the set Q_i using the unique closest Euclidean distance and assumes this value as its new update.

$$\begin{aligned} \tilde{x}_i(t+1) &= R_i(t, y^i(t)) - \gamma(t)(\nabla_i F(t, R(t, y^i(t))) + \zeta_i(t)) \\ x_i(t+1) &= \Pi_i[\tilde{x}_i(t+1)], \end{aligned} \tag{4}$$

where

$$\Pi_i[x] = \min_{z \in Q_i} \|z - x\|. \quad (5)$$

Namely, the projection operator when applied to the point x is the point in the constraint set that minimizes the Euclidean distance to x. Convergence is studied with the use of the Lyapunov function defined as the squared norm of the distance away from the desired minimum, i.e.,

$$V(t, x) = \frac{1}{2}\|x - x^*(t)\|^2, \quad (6)$$

where here and in the sequel $\|\cdot\|$ is the Euclidean norm. Let $\mathcal{I}_t$ define the previous information of the algorithm until time t such that

$$\mathcal{I}_t = \{ d_{ji}(\tau), \zeta_i(\tau), \text{ for } \tau < t \text{ and } j, i = 1, \ldots, n \}. \quad (7)$$

$\mathcal{I}_0$ includes the initial condition information. We note that $x_i(t)$ is uniquely determined by the random variables defined by $\mathcal{I}_t$.

The basic assumptions are introduced, the form of which is expressed in terms of $y^i(t)$ which is the information available to each processor. This permits the individual verification of the basic assumptions by the various processors. It is important to note that the ability of such verification is an intrinsic property of asynchronous iterations.

Basic assumptions:

1. There exists deterministic positive $K_1(t)$ such that for all i, we have

$$\begin{aligned} &E\left[(y_i^i(t) - x_i^*(t))'(\nabla_i F(t, y^i(t)) + \zeta_i(t)) \mid \mathcal{I}_t \right] \\ &\geq K_1(t) E\left[\|y^i(t) - x^*(t)\|^2 \mid \mathcal{I}_t \right]. \end{aligned} \quad (8)$$

2. There exist deterministic nonnegative $K_2(t)$ and $K_3(t)$ such that for all i, we have

$$\begin{aligned} &E\left[\|\nabla_i F(t, y^i(t)) + \zeta_i(t)\|^2 \mid \mathcal{I}_t \right] \\ &\leq K_2(t) + K_3(t) E\left[\|y^i(t) - x^*(t+1)\|^2 \mid \mathcal{I}_t \right]. \end{aligned} \quad (9)$$

3. There exist nonnegative $\alpha(t)$ and $\beta(t)$ such that

$$\begin{aligned} &(1 + \alpha(t))\|x(t) - x^*(t)\| \\ &\leq \|R(t, x(t)) - R(t, x^*(t))\| \\ &\qquad \leq (1 + \beta(t))\|x(t) - x^*(t)\|. \end{aligned} \quad (10)$$

4. For the initial approximation, we have

$$E\|x(0) - x^*(0)\|^2 < \infty \text{ and } \|x^*(0)\| < \infty. \quad (11)$$

Given the previous history of the algorithm, inequality (8) requires that the expected direction of $-\nabla F(t, y^i(t))$ is one of decrease with respect to the Lyapunov function $V(t, y^i(t))$. Inequality (9) imposes growth conditions on the update $\nabla_i F(t, y^i(t))$ and the noise. The inclusion of $K_2(t)$ in inequality (9) is indicative of the presence of additive noise with variance that is not necessarily finite.

3 Convergence Analysis

We formulate the main convergence results of process (4). Let us denote

$$\begin{aligned} q(t) &= (1 + \beta(t))^2 + \gamma^2(t) K_3(t)(1 + \beta(t))^2 \\ &\quad - 2\gamma(t) K_1(t)(1 + \alpha(t))^2. \end{aligned} \quad (12)$$

For a closed and convex set Q_i, the projection operator has the following properties,

$$\begin{aligned} &(x - \Pi_i[x])'(y - \Pi_i[x]) \leq 0 \quad \text{for all } y \in Q_i \\ &\|\Pi_i[x] - \Pi_i[y]\| \leq \|x - y\| \quad \text{for any } x,\ y \end{aligned} \quad (13)$$

The analysis is carried out by utilizing the second property of the projection operator $\Pi_i[\cdot]$ which allows the Lyapunov function defined by equation (6) to maintain a supermartingale and hence the convergence is retained as in the nonconstrained case established by Beidas and Papavassilopoulos [2]. Therefore, since $x_i(t+1)$ is an orthogonal projection of $\bar{x}_i(t+1)$ on Q_i and since $x_i^*(t+1) \in Q_i$ then

$$\begin{aligned} &\|x_i(t+1) - x_i^*(t+1)\| \\ &\leq \|R_i(t, y^i(t)) - R_i(t, x^*(t)) \\ &\qquad - \gamma(t)(\nabla_i F(t, R(t, y^i(t)) + \zeta_i(t))\| \end{aligned} \quad (14)$$

This enables the usual Lyapunov argument to be exploited to reduce the error defined in equation (6) and shaping this error equation to fit the form of an easily manageable vector inequality.

We also note that a sequence $\nu(t)$ of random variables converges to a random variable ν almost surely if

$$\Pr\left\{ \lim_{t \to \infty} \nu(t) = \nu \right\} = 1. \quad (15)$$

Theorem 1 *Consider the sequence $\{x_i(t)\}$ generated by equation (4). Suppose that the cost function $F(t, x)$ has a unique minimum at $x = x^*(t) \in Q$ for any t. Let the basic assumptions (1) - (4) be satisfied. In addition, assume that*

1. $\sum_{t=0}^{\infty} q(t) < \infty$, $q(t) \geq 0$,
2. $\sum_{t=0}^{\infty} \beta(t) < \infty$,
3. $\sum_{t=0}^{\infty} \|R(t, 0)\|^2 < \infty$,
4. $\sum_{t=0}^{\infty} \gamma^2(t) K_2(t) < \infty$.

Then for every initial condition the sequence $\{x_i(t)\}$ converges to $x_i^*(t) \in Q_i$ in the mean square and almost surely for each i.

Proof: Subtracting $x_i^*(t+1)$ from equation (4) and taking norms, we write

$$\begin{aligned} &\|x_i(t+1) - x_i^*(t+1)\|^2 \\ &= \|\Pi_i[R_i(t, y^i(t)) - \gamma(t)(\nabla_i F(t, R(t, y^i(t)) + \zeta_i(t))] \\ &\qquad - R_i(t, x^*(t))\|^2 \end{aligned} \quad (16)$$

Using the properties of the projection operator and recalling the fact that $x_i^*(t) \in Q_i$, we write

$$\begin{aligned} &\|x_i(t+1) - x_i^*(t+1)\|^2 \\ &\leq \|R_i(t, y^i(t)) - R_i(t, x^*(t))\|^2 \\ &\quad + \gamma^2(t)\|\nabla_i F(t, r(t, y^i(t)) + \zeta_i(t)\|^2 \\ &\quad - 2\gamma(t)[R_i(t, y^i(t)) \\ &\quad - R_i(t, x^*(t))]'(\nabla_i F(t, R(t, y^i(t)) + \zeta_i(t)). \quad (17) \end{aligned}$$

Here, we notice that applying steps similar to those of the proofs contained in Beidas and Papavassilopoulos [2] yields the required result.

Q.E.D.

It is worthy of mention that convergence in the mean square requires a weaker version of condition *4.* which can be replaced by $\lim_{t\to\infty} \gamma^2(t)K_2(t) = 0$.

Next we cover different cases of the projection operator as the constraint set Q_i takes more specific forms.

Example 1

Let the set Q_i consist of simple constraints such that

$$Q_i = \{x_i : x_i \geq 0\}.$$

In this case, the asynchronous gradient projection algorithm is described as

$$x_i(t+1) = \max\{0, R_i(t, y^i(t)) - \gamma(t)(\nabla_i F(t, R(t, y^i(t))) + \zeta_i(t))\}. \quad (18)$$

Example 2

Let the set Q_i consist of upper and lower bounds such that

$$Q_i = \{x_i : a_i \leq x_i \leq b_i\},$$

where a_i and $b_i \in \mathcal{R}^{n_i}$. In this case, the asynchronous gradient projection algorithm takes the form of

$$x_i(t+1) = \begin{cases} a_i & \tilde{x}_i(t+1) < a_i \\ b_i & \tilde{x}_i(t+1) > b_i \\ R_i(t, y^i(t)) - \gamma(t) \cdot (\nabla_i F(t, R(t, y^i(t))) + \zeta_i(t)) & \text{otherwise} \end{cases} \quad (19)$$

4 Conclusion

We studied the behavior of distributed asynchronous iterations with stochastic delays that solve optimization problems with nonstationary minimum over a constrained set. For the purpose of confining the iterates to the constraint set, each processor evaluates a gradient iteration and then projects back its iterate independently of the other processors. This procedure guarantees that each iterate generated by the algorithm is contained in the constraint set. The analysis that establishes the sufficiency conditions required to guarantee mean square and almost sure convergence is based upon utilizing a Lyapunov function given by equation (6) and using properties of the projection that maintain its supermartingale property and, finally, showing that the adverse effects possibly inflicted by the communication delays are negligible.

References

[1] G. M. Baudet, Asynchronous Iterative Methods for Multiprocessors, *J. ACM* vol. 25, 1978 226-244.

[2] B. F. Beidas and G. P. Papavassilopoulos, Asynchronous Implementation with Stochastic Delays of Optimization Algorithms Under Conditions of Drift, submitted to *IEEE Trans. Auto. Cont.*

[3] B. F. Beidas and G. P. Papavassilopoulos, Convergence Analysis of Linear Asynchronous Iterations with Stochastic Delays, to appear in *Parallel Computing.*

[4] D. P. Bertsekas, Distributed Dynamic Programming, *IEEE Trans. Auto. Cont.*, vol. AC-27 (1982) 610-616.

[5] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation, Numerical Methods*, (Prentice Hall, New Jersey, 1989).

[6] D. Chazan and W. Miranaker, Chaotic Relaxation, *Linear Algebra Appl.* vol. 2 (1969) 199-222.

[7] V. A. Dupac, A Dynamic Stochastic Approximation Method, *Ann. Math. Statistics* vol. 36 (1965) 1695-1702.

[8] H. T. Kung, Synchronized and Asynchronous Parallel Algorithms for Multiprocessors, in: J. F. Traub, ed., *Algorithms and Complexity: New Directions and Recent Results* (Academic Press, 1976) 153-200.

[9] H. J. Kushner and G. Yin, Asymptotic Properties of Distributed and Communicating Stochastic Approximation Algorithms, *SIAM J. Cont. and Optimiz.* vol. 25 (1987) 1266-1290.

[10] E. Kaszkurewicz, A. Bhaya and D. D. Šiljak, On the Convergence of Parallel Asynchronous Block-Iterative Computations, *Linear Algebra Appl.* vol. 131 (1990) 139-160.

[11] D. G. Luenberger, *Linear and Nonlinear Programming* (Addison-Wesley Publishing Co., 1984).

[12] B. T. Polyak, Convergence and Convergence Rate of Iterative Algorithms, Part I: General Case, *Automation Remote Control* vol. 37 (1976) 83-94.

[13] B. T. Polyak and Ya. Z. Tsypkin, Pseudogradient Adaptation and Training Algorithms, *Automation Remote Control* vol. 34 (1973) 45-68.

[14] D. D. Šiljak, *Large-Scale Dynamic systems* (North-Holland, New York, 1978).

[15] P. Tseng, D. P. Bertsekas and J. N. Tsitsiklis, Partially Asynchronous, Parallel Algorithms for Network Flow and Other Problems, *SIAM J. Cont. and Optimiz.* vol. 28 (1990) 678-710.

[16] J. N. Tsitsiklis, D. M. Bertsekas and M. Athans, Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms, *IEEE Trans. Auto. Cont.* vol. AC-31 (1986) 803-812.

[17] Ya. Z. Tsypkin, A. I. Kaplinskiy and K. A. Larionov, Adaptation and Learning Algorithms under Nonstationary Conditions, *Engineering Cybernetics* vol. 5 (1970) 829-840.

[18] A. Üresin and M. Dubois, Parallel Asynchronous Algorithms for Discrete Data, *J. ACM* vol. 37 (1990) 588-606.