AVERAGING THEORY FOR SIGN-SIGN LMS

B. D. O. Anderson[1], I. M. Y. Mareels[2], W. A. Sethares[4] and C. R. Johnson[2]

ABSTRACT

This paper examines the behavior of the "sign-sign" algorithm, in which the sign (or signum) function is applied to both the regressor vector and the error signal. Despite the discontinuity of the sign function, we show that averaging theory can be applied when the algorithm is driven by ergodic random inputs. The analysis defines an averaged equation around which the trajectories of the sign-sign algorithm cluster. The averaged equation is shown to converge when the algorithm is excited by Gaussian inputs. Simulations show that other input distributions cause the averaged equation to diverge.

1. INTRODUCTION

The parameter of the Least Mean Square (LMS) algorithm are updated according to [1]

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \mu X_k e_k \tag{1.1}$$

where μ is a small scalar gain, X_k is the regressor vector containing time shifted versions of the inputs, and e_k is the measured error. Implementation of (1.1) imposes certain hardware demands, including a need for multipliers. To lessen such demands, several variants of (1.1) have been proposed, which replace the update kernel $X_k e_k$ by $\mathrm{sgn}(X_k)e_k$, $X_k\mathrm{sgn}(e_k)$ or $\mathrm{sgn}(X_k)\mathrm{sgn}(e_k)$. These are referred to as the signed regressor algorithm, the signed error algorithm and the sign-sign algorithms, respectively. If θ^* represents the true parameterization, then the evolution of the parameter error $\theta_k = \theta^* - \theta_k$ for the sign-sign algorithm is defined by

$$\theta_{k+1} = \theta_k - \mu\, \mathrm{sgn}(X_k)\mathrm{sgn}(X_k'\theta_k) \ . \tag{1.2}$$

With stochastic inputs, LMS is well studied, the signed regressor has recently been analyzed [4] and algorithms of the form of the signed error are amenable to analysis as in [10]. This paper therefore focuses on the sign-sign algorithm and shows by example that nontrivial stochastic inputs can cause instability. In view of this possibility of divergence (for

---

[1]Department of Systems Engineering, Research School of Physical Sciences, P.O. Box 4, Australian National University, Canberra, A.C.T. 2601, Australia.

[2]School of Electrical Engineering, Phillips Hall, Cornell University, Ithaca, NY, 14853-5401, U.S.A. Supported by NSF grant, ECS-85-06149 and INT-8513400.

[3]Department of Electrical and Computer Engineering, University of Newcastle, New South Wales, Australia.

[4]Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI.

excitation sequences which pose no problem for the LMS algorithm) it becomes important to classify situations leading to acceptable behavior and to unacceptable behavior. This paper contributes to this classification by proving the stability of the sign-sign algorithm when excited by a correlated Gaussian process. In general, the application of averaging theory to an equation like

$$\theta_{k+1} = \theta_k - \mu f(X_k, \theta_k) \tag{1.3}$$

requires $f(.,.)$ to satisfy a Lipschitz continuity condition in its second argument, see e.g. [7]. In case the X_k sequence is independent, an averaging result for $f(.,.)$ discontinuous in θ_k is available in [9], while [10] offers an averaging result for dependent X_k and discontinuous $f(.,.)$. This result takes some work to obtain, and, reasonably, requires side conditions (stationarity) on the X_k sequence. The result is applied in [10] to the signed error algorithm.

The method of [10] leads to results governing the probability distribution of the supremum over a finite time interval of the difference between the solution of the unaveraged equation and an averaged differential equation. Here we present a result for the supremum of the expectation of this difference, again over a finite time interval. This is of course a similar result; it is obtained with much simpler machinery. In the process, the meaning of small μ is clarified. Conventional averaging requires that μ be small. Here it only makes sense to say μ is small with respect to the magnitude of $\|\theta_k\|$. For if $\{\theta_k\}$ is a solution trajectory of the sign-sign algorithm for some μ, then $\{\lambda\theta_k\}$ is a solution trajectory when μ is replaced by $\lambda\mu$, for any positive λ, and the X_k sequence is left unaltered. The real point is that one can average and secure insights while μ is small <u>relative</u> to $\|\theta_k\|$. Averaging theory is thus inapplicable if θ_k tends to the vicinity of the origin. On the other hand $\|\theta_k\|$ small means that the estimate θ_k is close to the desired parameterization θ^*.

It is worth noting that the LMS algorithm, the signed regressor, signed error and the sign-sign algorithms do not have identical convergence properties. For example, the sign-sign algorithm takes steps of magnitude $\pm\mu$, so convergence to zero is generically impossible. More importantly though, there exist periodic deterministic input sequences for which the LMS algorithm converges (with suitably small μ), and which led to divergence of the signed regressor and sign-sign algorithms, see [6] and [4]. There are also deterministic inputs for which LMS and sign-sign algorithms converge but signed-regressor diverges, and others for which LMS and signed-regressor converge but sign-sign diverges.

Section 2 defines an averaged function $f_{av}(\theta)$ (analogous to $f(.,.)$ of (1.3) and an averaged error system with error θ^a, and then shows (under appropriate hypothesis) that the error system for the sign-sign algorithm is stable whenever the averaged error system is stable. The result is in the form of a bound on the expected difference between θ_k and θ^a.

Section 3 demonstrates that θ^a is uniformly stable when the input is a Gaussian process. Combined with the result of section 2, this shows that the estimates $\hat{\theta}_k$ of the sign-sign algorithm are convergent to a μ-region about θ^*.

Section 4 presents two sets of simulations. The first simply supports the preceding analysis by example (i.e., that θ^a stable implies θ_k). The second set shows that θ^a is not always well behaved. By choosing certain non-Gaussian random inputs, we shown that θ^a and the sign-sign algorithm diverge together. This shows that the relationship between θ^a and the sign-sign algorithm is tighter than we were able to prove since we were unable to prove that unstable θ^a implies unstable θ. It also shows that the stability of the algorithm is highly dependent on the properties of the excitation sequence.

2. STOCHASTIC AVERAGING FOR THE SIGN-SIGN ALGORITHM

Suppose that X_k is a n-dimensional random process $X_k = (x_k, x_{k-1}, \ldots x_{k-n-1})$, with x_k ergodic. Define the averaged update term as

$$f_{av}(\theta) = E\{\text{sgn } (X_k) \text{ sgn } (X_k'\theta)\} \ . \tag{2.1}$$

This section relates the behavior of the sign-sign error system (1.2) to the behavior of the averaged equation

$$\theta^a_{k+1} = \theta^a_k - \mu f_{av}(\theta^a_k) \tag{2.2}$$

with the additional assumption that (1.2) and (2.2) are initialized identically, that is,

$$\theta_0 = \theta^a_0 \ . \tag{2.3}$$

The initial value θ_0, if regarded as a random variable, is independent of the X_k sequence.

For $n > 1$ we will make several assumptions concerning f_{av} and the properties of the input process X_k. The first requires Lipschitz continuity of the averaged update $f_{av}(\theta)$.

Assumption 1: Assume that $f_{av}(\theta)$ exists for all $\theta \in R^n$ and that f_{av} is Lipschitz continuous in D, an open subset of $R^n/\{0\}$.

Notice that it is impossible for f_{av} to be Lipschitz continuous at the origin unless f_{av} is identically zero. An important special case is provided by a Gaussian, zero mean sequence $\{x_k\}$. Let

$$E[X_k X_k'] = R = (r_{ij}) \ , \tag{2.5}$$

let r_i' be the rows of R, and suppose R is positive definite. Then for a fixed but arbitrary $\bar\theta$, the i^{th} component of the vector $f_{av}(\bar\theta)$ can be written (see for example [11, p. 484])

$$E(\text{sgn } X_k \text{ sgn } X_k'\bar\theta)_i = \frac{2}{\pi}\sin^{-1}\left[\frac{r_i'\,\bar\theta}{(r_{ii})^{1/2}(\bar\theta' R\bar\theta)^{1/2}}\right] = [f_{av}(\bar\theta)]_i \tag{2.6}$$

One can verify that $f_{av}(\theta)$ is Lipschitz continuous in the set

$$D = \{\theta \in R^n, \ \|\theta\| \geq \varepsilon\}$$

where ε is an arbitrary, positive constant, with a global Lipschitz constant inversely proportional to ε: notice that

$$[f_{av}(\theta)]_i = \frac{2}{\pi} \sin^{-1} \frac{(r_i' R^{-1/2})(R^{1/2}\theta)}{(\theta' R\theta)^{1/2} r_{ii}^{1/2}}$$

$$= \frac{2}{\pi} [\frac{\pi}{2} - \cos^{-1} \frac{(r_i' R^{-1/2})(R^{1/2}\theta)}{(\theta' R\theta)^{1/2} r_{ii}^{1/2}}]$$

$$= 1 - \frac{2}{\pi} \begin{pmatrix}\text{angle between vectors} \\ R^{-1/2} \; r_i \;\text{ and }\; R^{1/2}\,\theta\end{pmatrix} \tag{2.7}$$

In comparing (1.2) and (2.2), we shall also make assumptions on the character of the process X_k. The first concerns the direction in which the vector X_k can point, essentially, requiring that no direction is infinitely likely, in the sense that the probability density associated with the X_k direction contains no delta function.

Assumption 2: Let dS be an elemental area on the unit sphere. Then there exists $\beta > 0$ (independent of the location of dS) such that

$$Pr[X_k \| X_k\|^{-1} \in dS] \leq \beta dS \quad \text{for every } k \tag{2.8}$$

Remark: Assumption 2 is a technical assumption which insures that θ is not orthogonal to X_k infinitely often, and so prevents a "lock up" of the algorithm in some particular direction in parameter space. The only practically important stochastic processes which violate this assumption, at least in a communication or control engineering context, are those whose range consists of a finite number of points.

Another assumption on the X_k is a mixing condition. This essentially requires that information about the process at time k becomes less significant in determining the value at time $k+m$ as m becomes larger. Mixing is a way of exploiting analytically the idea that the auto correlation of a dependent process decreases with time.

Assumption 3: The following mixing properties hold:

$$\lim_{K\to\infty} E \{[\| \frac{1}{K} \sum_{k=mK}^{(m+1)K-1} \mathrm{sgn}X_k[\mathrm{sgn}(X_k'\theta_k)-\mathrm{sgn}(X_k'\theta_{mK})]\|]|\theta_{mK} = \theta^+\}$$

$$= \lim_{K\to\infty} E \{[\| \frac{1}{K} \sum_{k=0}^{K-1} \mathrm{sgn}X_k[X_k'\theta_k)-\mathrm{sgn}(X_k'\theta_0)\|]|\theta_0 = \theta^+\} \tag{2.10}$$

We note that we shall also be using the following particular ergodic property:

<u>Assumption 4</u>: In the set D, there holds

$$\lim_{K\to\infty} E\|\left[\frac{1}{K} \sum_{k=mK}^{(m+1)K-1} \operatorname{sgn}X_k \operatorname{sgn}(X_k'\theta^a_{mK}) - f_{av}(\theta^a_{mK})\right]\| = 0 \tag{2.11}$$

In the following result, we shall establish conditions under which $E[\|\theta_k - \theta^a_k\|]$ remains small. One of these conditions is that $\|\theta^a_k\|$ does not get too small, which should be no surprise. The fact that the Lipschitz constant associated with $f_{av}(\bar{\theta})$ becomes infinite as one seeks to include arbitrarily small $\|\bar{\theta}\|$ in the domain of definition suggests that one might expect difficulties in obtaining a result which permitted arbitrarily small $\|\theta^a_k\|$. Also, as mentioned in the introduction, averaging suggests that we may need μ small relative to $\|\theta^a_k\|$.

The first connection between (1.2) and (2.2) is as follows.

<u>Proposition 1</u>: Consider equations (1.2) and (2.2) with the same initial condition as per (2.3), and with Assumptions 1 through 4 in force. Let $\varepsilon > 0$ be arbitrary. With K sufficiently large, with μK sufficiently small and with r such that for all $m \le r$ there holds

$$\|\theta^a_{mK}\| \ge \sigma > 2\mu K/n \text{ for some } \sigma$$

and for all $k \le rK$ there holds $\theta^a_k \in D$, then

$$\sup_{0\le k\le rK} E\|\theta_k - \theta^a_k\| \le \mu K \varepsilon M \tag{2.12}$$

with

$$M = 3r\{1 + (1 + \mu KL_1)[(1 + \mu KL_1)^r - 1]\} \tag{2.13}$$

and L independent of μ, K, r.

Remark: The minimum value which K can take depends on both the ergodic and mixing properties of $\{X_k\}$ and on ε, with smaller ε requiring a large value of K. The maximum value which μK can take depends on ε, with smaller ε requiring a smaller maximum value. It is most convenient to think of ε being chosen first, then (large enough) K, then (small enough) μ. Then r is chosen in effect so that over $[o, rK]$, the solution of the averaged equation stays in D and does not get too small.

<u>Proof:</u> See [16]

Proposition 1 demonstrates conditions under which the sign-sign error system (1.2) and the averaged equation (2.2) behave alike, assuming fixed μ. Note also that the solution of (2.2) can be related to the solution of a differential equation:

$$\dot{\theta}^d = -f_{av}(\theta^d) \qquad \theta^d(0) = \theta^d_o \tag{2.22}$$

so that as $\mu \to 0$, with μK constant, $\theta^a{}_{mK} \to \theta^d(\mu mK)$. This limiting version shows that as $\mu \to 0$, $\theta_k \to \theta^a{}_k$ in a certain sense over an interval proportional to $\varepsilon\mu^{-1}$. Assume that if ε is chosen first, and K is chosen to exceed its allowed lower bound, then μ can be chosen arbitrarily small. Consider in fact ε, K fixed, and μ, r variable, with μrK fixed so that $\|\theta^a_{\mu rK}\| \approx \|\theta^d(\mu rK)\| \geq \sigma$, and indeed $\|\theta^d(t)\| \geq \sigma$ for all $t < \mu rK$ and such that $\theta^a_k \in D$, $K \leq rK$. Then as $\mu \to 0$, it is easily checked that

$$\mu K \varepsilon M + \varepsilon(\mu Kr)\, e^{\mu Kr} . \tag{2.23}$$

Since ε is arbitrary, this shows that

$$\lim_{\mu\to 0} \sup_{0\leq k\leq rK} E\|\theta_k - \theta^a_k\| = 0 .$$

when μrK remains constant. We thus have

<u>Proposition 2</u>: Consider equations (1.2) and (2.2) with the same initial condition as per (2.3), and with Assumptions 1 through 4 in force. Suppose that the solution of (2.22) obeys $\|\theta^d(t)\| \geq \sigma > 0$ and remains in D for all $t \in [0,T]$. Then,

$$\lim_{\mu\to 0} \sup_{0 \leq k \leq \frac{T}{\mu}} E\|\theta_k - \theta^a_k\| = 0 .$$

Proposition 2 is essentially a consistency result, showing that the sign-sign algorithm and the averaged equation behave the same way for vanishing stepsizes. The result was easy to derive once the fixed stepsize case was shown in proposition 1.

3. BEHAVIOR OF THE AVERAGED EQUATION WITH GAUSSIAN INPUTS

In the previous section, we described how the solutions of the stochastic equation and the averaged equation are related. This section shows that the averaged equation converges to a region about the origin when the input is a Gaussian process. The averaged equation from (2.2) and (2.6) is

$$\theta^a_{k+1} = \theta^a_k - \mu \frac{2}{\pi} \operatorname{vec}\left[\sin^{-1} \frac{r_i{}'\theta^a_k}{(r_{ii})^{\frac{1}{2}}(\theta^a_k{}'R\theta^a_k)^{\frac{1}{2}}}\right] \tag{3.1}$$

Now observe that because $(2/\pi)\theta \leq \sin\theta \leq \theta$, for $\theta \in [0,\pi/2]$, we have $z \leq \sin^{-1}z \leq (\pi/2)z$ for $z\in[0,1]$, while $z \geq \sin^{-1}z \geq (\pi/2)z$ for $z \in [-1,0]$. This means $\sin^{-1}z = \nu(z)z$ where $\nu(z) \in [1,\pi/2]$ for every $z \in [-1,1]$. Accordingly, (3.1) can be replaced by

$$\theta^a_{k+1} = \theta^a_k - \mu \frac{2}{\pi} \operatorname{diag}[\nu(z_1),\ldots,\nu(z_n)] \operatorname{vec}\left[\frac{r_i{}'\theta^a_k}{(r_{ii})^{\frac{1}{2}}(\theta^a_k{}'R\theta^a_k)^{\frac{1}{2}}}\right] \tag{3.2}$$

with $\nu(z_i) \in [1, \pi/2]$, and $z_i = \dfrac{r_i{}'\theta^a_k}{(r_{ii})^{\frac{1}{2}}(\theta^a_k{}'R\theta^a_k)^{\frac{1}{2}}}$

Equation (3.2) may be rewritten as follows:

$$\theta^a_{k+1} = \theta^a_k - \mu L_k R \theta^a_k (\theta^a_k{}' R \theta^a_k)^{-\frac{1}{2}} \qquad (3.3)$$

where the matrix L_k is defined as:

$$L_k = \frac{2}{\pi} \operatorname{diag}(\nu(z_i)) \operatorname{diag}((r_{ii})^{-\frac{1}{2}}) \qquad (3.4)$$

To demonstrate that the averaged equation (3.1) is uniformly stable, consider the following candidate Lyapunov function:

$$V(\theta) = \theta' R \theta \qquad (3.5)$$

along the trajectories of (3.3) the difference

$$\begin{aligned} & V(\theta^a_{k+1}) - V(\theta^a_k) \\ &= -\mu\, \theta^a_k{}'(R L_k R + R L_k R)\theta^a_k (\theta^a_k{}' R \theta^a_k)^{-\frac{1}{2}} \\ &+ \mu^2\, \theta^a_k{}' R L_k\, R L_k R \theta^a_k (\theta^a_k{}' R \theta^a_k)^{-1} \end{aligned}$$

using the definition of L_k in (3.4) and the property that $\nu(z_i) \in [1, \pi/2]$, can be over bounded as

$$V(\theta^a_{k+1}) - V(\theta^a_k) \leq -\mu\alpha_1 V(\theta^a_k)^{\frac{1}{2}} + \mu^2 \alpha_2 \qquad (3.6)$$

where α_1 and α_2 are given by:

$$\begin{aligned} \alpha_1 &= \frac{2}{\pi} \min_i [r_{ii}^{-\frac{1}{2}}] \lambda_{min}(R) \\ \alpha_2 &= \{\max_i [r_{ii}^{-\frac{1}{2}}] \lambda_{max}(R)\}^2 . \end{aligned}$$

This demonstrates that

$$V(\theta^a_{k+1}) - V(\theta^a_k) < 0 \quad \text{if} \quad V(\theta^a_k) > (\mu\, \alpha_2/\alpha_1)^2$$

Thus the averaged equation (3.3) is uniformly stable and $\|\theta^a_k\|$ is asymptotically of the order of μ.

From proposition 1 it follows that the trajectories of the averaged equation are on average good approximations to the trajectories of equation (2.2) over the interval $[0,N]$ provided $\|\theta^a_k\| > \sigma$, $k \in [0,N]$ and $\theta^a_k \in D$. In the Gaussian case, these two conditions are identical, since one can always choose D to be $\{\theta | \theta \in R^n, \|\theta\| \geq \sigma, \|\theta\| \leq \lambda^{1/2}_{max}(R) \lambda^{-1/2}_{min}(R) \|\theta_o\|\}$. From proposition 1, $\sigma > 2(\mu K)/\sqrt{n}$, and so the sign-sign algorithm drive by a Gaussian input sequence $\{x_k\}$ on average exhibits convergent behavior for all initial conditions $\|\theta_0\| \geq \mu C$, (for some constant $C > 0$ independent of μ) and asymptotically $\|\theta_k\|$ becomes of the order of μ.

4. SIMULATION RESULTS

This section presents two sets of simulations. The first set illustrates the validity of the averaging results of the previous sections and the second set shows that the sign-sign algorithm tracks the averaged equation even when the averaged equation diverges.

Figure 1 shows simulations in which the two dimensional sign-sign algorithm was excited by i.i.d. (pseudorandom) Gaussian inputs. In each run, the two dimensional sign-sign algorithm was initialized at (2,5) in parameter error space. Each converges to a region about the origin (parameter error equal to zero) in a few hundred iterations. The steady center line is an approximation to the averaged equation θ^a, which was computed at each θ by $f_{av}(\theta) - \frac{1}{K} \sum_{i=1}^{K} \operatorname{sgn} X_i \operatorname{sign} (X_i'\theta)$ where X_i is a pseudorandum input sequence. The value $K = 10{,}000$ was found empirically to give an appropriately smooth averaging window for this calculation. The simulation shows that individual runs of the sign-sign algorithm tend to cluster around the averaged equation, as expected from proportions 1 and 2. The stepsize was set at $\mu = 0.1$.

For the second set of simulations, a non-Gaussian, dependent stochastic process was chosen. The process alternates between two nonzero mean (but independent) Gaussian random variables S1 and S2 in a Markov like manner. Let S1 be normally distributed with mean 1.5 and variance 0.1, and let S2 be normal with mean - 0.1 and variance .01. If the value of the process at time k is determined by S1, then with probability α, the value of the process at time k+1 will be determined by S1, and with probability $1-\alpha$, the value will be determined by S2. Similarly for S2 with transition probabilities β and $1-\beta$. For the simulation, $\alpha = .2$ and $\beta = .8$ were chosen. This particular process was used because it fulfills the Assumptions, (a simple Markov chain would fail Assumption 2 since X_k would point infinitely often in certain directions), and because the averaged equation diverges in simulation. It is an interesting (and nontrivial) question whether the divergence of θ^a could actually be proven.

This process was used to excite a six dimensional sign-sign algorithm. The trajectories of the averaged equation (which is computed as before) are shown in figure 2. The graph is a two dimensional slice of the six dimensional parameter error space and the trajectories of θ^a are shown for seven different starting values. In each case, the trajectory moves near the origin, and then proceeds to diverge along one of the two 45° lines. Since the function θ^a fails to be Lipschitz at the origin, it is not surprising that the solution fails to be well-defined near the origin. In the present case, there are two solutions, corresponding to the two unstable directions. It should be noted that the initial convergence to the origin in figure 2 is not a generic behavior, that is, other input sequences, other plants (values of θ^*), and other 2-D slices of the 6-D parameter error space would not necessarily show convergence to the origin before diverging. What is generic, however, it that multiple solutions of θ^a at the origin can lead to diverging "arms" in parameter error space. Moreover, this example is not singular, in the sense that a continuum of values for S1, S2, α and β lead to similar divergent behavior. It is also not hard to find multi-state Markov-like processes for which θ^a diverges in simulation.

The behavior of θ^a is only a useful guide if individual runs of the algorithm behave like the averaged equation. Figure 3 shows a number of simulations all initialized at the same location (-2, -6) in parameter error space. The individual runs tend to cluster around the averaged equation until they approach the origin, where averaging fails to hold. Then, the trajectories move out along one or the other of the unstable branches, apparently forever. These simulations indicate that the assumptions made in the theory are not restrictive, in that the behavior of the averaged equation is a good representation of the actual response, in both stable and unstable situations.

5. CONCLUSION

We have established two key facts about the sign-sign algorithm. The random trajectories of the unaveraged equation cluster around those of the averaged equation, at least until the solution of the latter becomes small, (under conditions of ergodicity, mixing, small μ etc). Second, trajectories of the averaged equation excited by Gaussian inputs converge to a ball around the origin. Combining these facts, the trajectories of the unaveraged random equation with Gaussian inputs must also converge to a region in the vicinity of the origin.

In addition, we have investigated (through simulation) the behavior of the sign-sign algorithm and the related averaged equation when excited by certain non-Gaussian processes. In all cases investigated, stability/instability of the averaged equation was coupled with stability/instability of the actual algorithm. In particular, a simple class of input processes was found for which the sign-sign algorithm diverges.

The averaged equation provides a tool which may lead to a further understanding of what types of input excitation lead to stability/instability of the sign-sign algorithm, though more insights are necessary in order to unambiguously classify all suitable and unsuitable inputs.

REFERENCES

[1] B. Widrow and S. D. Stearns, Adaptive Signal Processing, Prentice Hall, Inc., Englewood Cliffs, NJ, 1985.

[2] J. L. Moschner, 'Adaptive filter with clipped input data,' Tech. Dept. 6796-1, Information Systems Laboratory, Stanford University, 1970.

[3] R. Kumar and J. B. Moore, 'Adaptive equalization via fast quantized - stable methods,' IEEE Trans. Communications, Vol. COM-29, October 1981, pp. 492-1501.

[4] W. A. Sethares, I. M. Y. Mareels, B. D. O. Anderson, C. R. Johnson, Jr., and R. R. Bitmead, 'Excitation conditions for signed regressor least mean squares adaptation,' IEEE Trans. Circuits and Systems, June 1988.

[5] R. W. Lucky, 'Techniques for adaptive equalization of digital communication systems,' Bell Syst. Tech. J., Vol. 45, February, pp. 255-266.

[6] S. Dasgupta and C. R. Johnson, Jr., 'Some comments on the behavior of sign-sign adaptive identifiers,' Syst. Control Letters, Vol. 7, April 1986, pp. 75-82.

[7] R. Z. Khas'minskii, 'On stochastic processes defined by differential equations with a small parameter,' Theory of Probability and Its Applications, Vol. XI, 1966, pp. 211-228.

[8] R. R. Bitmead, B. D. O. Anderson and T. S. Ng, 'Convergence rate determination for gradient-based adaptive estimators,' Automatica, Vol. 22, March 1986, pp. 185-191.
[9] D. P. Derevitskii and A. L. Fradkov, 'Two models for analysing the dynamics of adaptation algorithms,' Autom. and Remote Control, Vol. 35, January 1974, pp. 59-67.
[10] A. Benveniste, M. Goursat and G. Rugt, 'Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications,' IEEE Trans. Auto. Control, Vol. AC-25, December 1980, pp. 1042-1058.
[11] A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill Book Co., New York, 1965.
[12] C. A. Desoer and M. Vidyasagar, Feedback Systems: Input-Output Properties, Academic Press, New York, 1975.
[13] A. Gersho, 'Adaptive Filtering with Binary Reinforcement,' IEEE Trans. on Info. Thy, Vol. IT-30, March 1984, pp. 191-198.
[14] L. Ljung and T. Soderstrom, theory and Practice of Recursive Identification, MIT Press, Cambridge, MA, 1983.
[15] N. S. Jayant and P. Noll, Digital Coding of Waveforms, Prentice Hall, 1984.
[16] B. D. O. Anderson, IMY Mareels, W. A. Sethares, C. R. Johnson, Jr., 'Averaging Theory for Sign-Sign LMS,' technical report Univ. of Wisconsin.

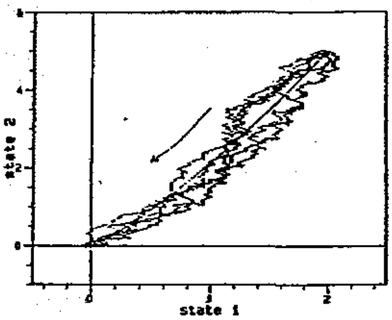

Figure 1: Trajectories and averaged trajectories of two dimensional algorithm with gaussian regressor

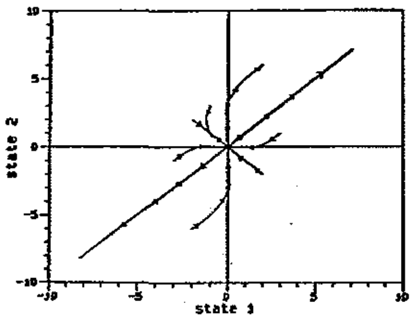

Figure 2: Two-dimensional cross section of six-dimensional algorithm exhibiting instability

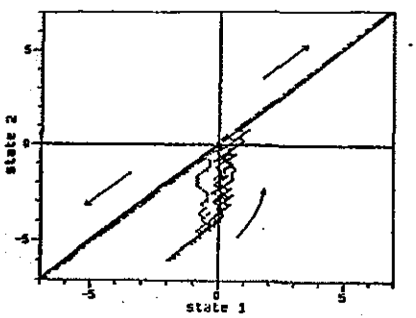

Figure 3: Unstable algorithm showing ultimate divergence