

ON THE FAILURE OF PROPOSED BLIND RECURSIVE IDENTIFICATION SCHEMES BASED ON THE SATO ALGORITHM AND ITS GENERALIZATIONS†

Rodney A. Kennedy‡, Brian D.O. Anderson‡, Zhi Ding*, and C. Richard Johnson Jr.*

† Research supported by ANU Centre for Information Science Research, and the Australian Research Council
‡ Dept. of Systems Eng., RSPPhyS, Australian National University, G.P.O. Box 4, Canberra, A.C.T. 2601, AUSTRALIA
* School of Electrical Engineering, Cornell University, Ithaca, N.Y. 14853, USA.

ABSTRACT: The limitations of certain theoretical studies into the analysis of a broad class of blind recursive identification algorithms is examined here. We consider the seminal group of algorithms developed and analyzed by Benveniste, Goursat and Ruget [2] which generalize the pioneering algorithm due to Sato [1]. In [2] certain ideal blind convergence properties of these algorithms have been established under the assumption that the inverse model to be identified is infinite dimensional. We embellish these results by showing that such ideal convergence properties do not generally hold for practical systems where the equalizer necessarily is finitely parametrized. In short, the existing theoretical works rely fundamentally and implicitly on gross overparametrization (infinite parametrization) to establish their convergence proofs. That such desirable convergence properties fail for simple finitely parametrized practical examples is illustrated.

1 Introduction

The Sato adaptive algorithm was the first widely used recursive identification scheme to perform discrete time system inverse identification based on measuring the output of a system but lacking explicit knowledge (i.e., direct measurements) of its input [1]. The only information concerning the system input utilized by the algorithm was knowledge of its statistical properties, e.g., the input distribution. This feature makes the analysis more difficult than more standard identification schemes which form errors and regressors from both system inputs and system outputs.

The Sato algorithm was intuitively well founded. It was in the seminal work of Benveniste, Goursat and Ruget [2] that a firm theoretical basis and justification of the algorithm was given proving it was an effective means for blind (inverse) identification of non-minimum phase systems. In [2] a broad class of algorithms was considered—which we label the BGR-algorithm—within which the Sato algorithm appears as the simplest manifestation. One of the main results in [2] states that given an infinitely parametrized inverse model and a driving system input which is “subgaussian” in distribution (and i.i.d.) then the gradient based BGR-algorithm has only stable equilibria (convergence points) in parameter space corresponding to ideal identification (up to some inconsequential delay).

This paper establishes that any truncation to a finite number (however large) of parameters in the identified (inverse) model generally destroys these ideal convergence properties—additional local minima are shown to exist which in no sense yield satisfactory identification even approximately. We point to the relevant theoretical difficulties and back up our claims by analyzing a first order example. Our theoretical demonstration will be for the simplest non-trivial system captured by [2]. Explicitly, we treat the Sato algorithm and take as input a sequence of levels which are uniformly distributed (belonging to the class of smooth distributions labelled subgaussian).

Blind identification has major application in digital communications with blind equalization [1-5]. This paper cannot be strictly regarded as being relevant to the equalization problem because digital symbols are discretely distributed and thus cannot be subgaussian. One can, however, regard this work and that in [2] as being pointers or approximations to the equalizer situation. The treatment of discrete driving inputs is treated elsewhere and in most respects is simpler to analyze (these analogous negative results have been established in [4,5]).

2 Problem Formulation

The overall system under study is shown in Fig.1. To the left we have a linear system h (not necessarily minimum phase) whose “inverse” we wish to identify. The driving input is denoted $\{a_k\}$ and is assumed to be i.i.d., with symmetric distribution $\nu_a(\cdot)$. The system output is denoted x_k . Connected in series with the system we have a finitely parametrized model which by convention we call the equalizer (slightly out of context) with tap weights given by

$$\hat{\theta}(k) \triangleq [\hat{\theta}_0(k) \hat{\theta}_1(k) \dots \hat{\theta}_m(k)]'. \quad (2.1)$$

As the notation suggests the parameters are time varying, actually adjusted according to some adaptive law to be defined later. What

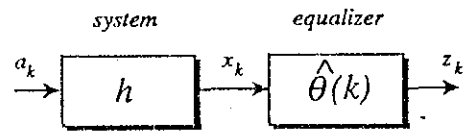


Figure 1: System and Equalizer

distinguishes the type of identification under study here is that during adaptation only $\{x_k\}$ can be measured and incorporated for learning. Explicit knowledge of the input $\{a_k\}$ is lacking. Finally, in Fig.1, z_k denotes the equalizer output which also certainly is measurable. When the linear system h is minimum phase the objective of ideal adaptation is to achieve $z_k = a_k \forall k$ by adjusting $\hat{\theta}(k)$. In all other cases the ideal objective of adaptation is to adjust $\hat{\theta}(k)$ to achieve

$$z_k = +a_{k-\delta} \text{ or } z_k = -a_{k-\delta} \quad \forall k, \delta \in \mathbb{Z} \quad (2.2)$$

where δ is some fixed delay. Note that due to the symmetry of $\nu_a(\cdot)$ the ambiguity in the sign in (2.2) is unavoidable unless additional a priori knowledge of the system is available. The ideal objective must often be relaxed in practice, to allow approximate equality in (2.2). Issues regarding the ability of a finite impulse response (FIR) filter to approximate a non-minimum phase inverse (with delay δ) are not crucial for our analysis. We are simply analyzing existing algorithms and studying their behavior. The development and justification of these algorithms and the precise qualification of conditions like (2.2) are left to other works [2,3]. The question we pose and solve here is:

Problem: Does the BGR-algorithm with a subgaussian $\nu_a(\cdot)$ distributed driving sequence $\{a_k\}$ always ensure convergence of a finite dimensional equalizer to a parameter setting yielding one of the (desirable) possibilities in (2.2) at least in approximation?

Recall, in [2] when the number of equalizer parameters in (2.1) is infinite, i.e., $m = \infty$, the answer is yes. Unfortunately, here for the finite case, we show the answer is no, at least for a class of simple systems (we believe the same conclusion holds for the general situation).

3 Blind Algorithm and its Convergence Analysis

3.1 Algorithm

The BGR-algorithm [2] takes the simple form

$$\hat{\theta}(k+1) = \hat{\theta}(k) - \mu \psi(z_k) X_k, \quad (3.1)$$

where μ is a small stepsize, and

$$X_k \triangleq [x_k \ x_{k-1} \ \dots \ x_{k-m}]' \quad (3.2)$$

is the regressor composed of system outputs. The scalar $\psi(z_k)$ is a prediction error generated from the equalizer output z_k . In [1]

Sato gave the following form for the memoryless non-linearity $\psi(\cdot)$ generating this error:

$$\psi(z) \triangleq z - \gamma \operatorname{sgn}(z) \quad (3.3)$$

where the dispersion constant is given by:

$$\gamma \triangleq \frac{E\{a_k^2\}}{E\{|a_k|\}} \equiv \frac{\int_{-\infty}^{+\infty} u^2 \nu_a(du)}{\int_{-\infty}^{+\infty} |u| \nu_a(du)} \quad (3.4)$$

where $E\{\cdot\}$ denotes expectation. The generalization given in [2], but probably developed independently, takes the form:

$$\psi(z) \triangleq \tilde{\psi}(z) - \xi \operatorname{sgn}(z) \quad (3.5)$$

where the pair $(\xi, \tilde{\psi}(\cdot))$ satisfy

$$\xi \triangleq \frac{E\{a_k \tilde{\psi}(a_k)\}}{E\{|a_k|\}} \equiv \frac{\int_{-\infty}^{+\infty} u \tilde{\psi}(u) \nu_a(du)}{\int_{-\infty}^{+\infty} |u| \nu_a(du)} \quad (3.6)$$

such that $\psi''(\cdot) \geq 0$ on $(0, \infty)$, $\psi(\cdot)$ is odd and $\psi(O_+) < 0$. We recover the Sato algorithm with the simple choice of $\tilde{\psi}(z) = z$. In what follows we focus on the Sato algorithm as representative of the BGR-algorithm.

We shall move onto a statement of the theorem found in [2] after we provide a little background on concepts relevant to the analysis of the convergence properties.

3.2 Convergence Analysis—Equilibria

In this section we introduce two important notions which are standard. The first notion is that of a mean cost surface for which (3.1) may be viewed as a gradient descent strategy. The second notion is that of an equilibrium which is a parameter setting that, if frozen, gives an average update in (3.1) of zero (i.e., expectation of zero but not necessarily requiring the instantaneous update in (3.1) to be zero).

In connection with the Sato algorithm (3.3) we introduce a scalar cost function defined as follows (this is denoted $\Psi(\cdot)$ in [2]):

$$J(z) \triangleq \frac{1}{2} (|z| - \gamma)^2 \quad (3.7)$$

Then (3.1) and (3.3) can be seen as a gradient descent algorithm when it is rewritten using (3.7) as

$$\hat{\theta}(k+1) = \hat{\theta}(k) - \mu \frac{\partial J(z_k)}{\partial \hat{\theta}(k)} \quad (3.8)$$

where $z_k = X_k' \hat{\theta}(k)$. The average or mean cost surface, as a function of the parameters, is then defined by

$$\begin{aligned} \mathcal{J}(\hat{\theta}(k)) &\triangleq E\{J(X_k' \hat{\theta}(k))\} \\ &= \frac{1}{2} E\{(|X_k' \hat{\theta}(k)| - \gamma)^2\} \\ &= \frac{1}{2} \hat{\theta}(k)' E\{X_k X_k'\} \hat{\theta}(k) - \gamma E\{|X_k' \hat{\theta}(k)|\} + \frac{\gamma^2}{2}. \end{aligned} \quad (3.9)$$

In this last expression the computation of the middle term is generally difficult and one of the causes of analytical difficulties. These problem are accentuated further when attempting to tackle the general case (3.5)-(3.6). We will return to (3.9) later. We now move on to define equilibria of the algorithm.

The stationary points on the mean cost surface (3.9) are defined by

$$\frac{\partial}{\partial \hat{\theta}(k)} \mathcal{J}(\hat{\theta}(k)) \equiv E\left\{\frac{\partial J(z_k)}{\partial \hat{\theta}(k)}\right\} = 0 \quad (3.10)$$

in which (the commutativity) follows by the continuity of the cost (3.7). In turn it is clear how the stationary points defined by (3.10) can be interpreted as being the parameter settings for which the average update in (3.8) is the zero vector.

3.3 Stability

Determining equilibria according to (3.10) is only half the story. Of greatest interest is determining stable equilibria. Testing the hessian to be positive definite is a sufficient condition for stability (but not necessary):

$$\frac{\partial}{\partial \hat{\theta}(k)} E\left\{\frac{\partial J(z_k)}{\partial \hat{\theta}(k)}\right\} > 0. \quad (3.11)$$

One cannot in this expression commute the differentiation and expectation operators because the argument of the expectation is discontinuous at the origin.

We now summarize our conceptual approach to describing the convergence properties of the BGR-algorithm: (i) determine the mean cost surface (3.9) which describes succinctly the global adaptation tendencies of the adaptation update; and (ii) locate and describe the set of stable (attractive) equilibria (3.10) in the equalizer parameter space and ascertain if these parameter values fulfill the desirable objective of forming an (approximate) inverse (2.2).

4 Theoretical Skim

4.1 Existing Theory on Convergence

One major result established in [2] can be paraphrased as follows (modulo certain estoteric details): When the input distribution $\nu_a(\cdot)$ is subgaussian, i.e., satisfies

$$\nu_a(du) = K e^{-g(u)} du \quad (4.1)$$

where K is a constant, $g(u)$ is an even function such that $g(u)$ and $g'(u)/u$ are strictly increasing on \mathbb{R}^+ , then the only stable equilibria of the BGR-algorithm (i.e., ones satisfying (3.10)-(3.11)) are those which achieve (2.2). This establishes a very desirable property of these blind algorithms.

One limiting case of a subgaussian distribution is the uniform distribution and for definiteness in our later demonstration it is with this choice that we rely.

4.2 Finite Parametrization Qualification

An implicit qualification in the establishment of the above convergence result in [2] is that the equalizer be infinite dimensional, i.e., $m = \infty$ in (2.1). The precise manner in which one makes rigorous this necessity can be found in [4] (in a slightly different context) and will not be expounded here. Rather our approach here will be to show that for a finitely parametrized equalizer stable undesirable equilibria exist, violating the spirit of the positive results found in [2].

One may argue that choosing too few equalizer parameters may readily lead to multi-minima and that such difficulties are vacuous; in effect the problem being ill-posed. This is not the mechanism that we rely on. In the following section we deliberately choose a first order auto-recursive system (thus minimum phase). Then a finite number of equalizer parameters may achieve perfect equalization—in no sense is the example underparametrized. Worst still deliberately or unwittingly overparametrizing the equalizer still shows undesirable local minima, as do equalizers having an arbitrarily high but finite number of taps. Yet in the appropriate limit an infinite dimensional equalizer has no such minima, as can be concluded from [2]. These results are not contradictory but reflect the lack of physical relevance of the theoretical results in [2] to a certain degree. We now move onto the demonstration of our claims.

5 Example and Corresponding Theory

5.1 First Order Minimum Phase Example

Let our system h (in Fig.1) be determined by the auto-recursion

$$x_k + \beta x_{k-1} = a_k, \quad |\beta| < 1 \quad (5.1)$$

which may be alternatively expressed as a rational function in the delay operator

$$H(q^{-1}) \triangleq \frac{1}{1 + \beta q^{-1}}, \quad (5.2)$$

This is a first order, minimum phase system. Note a zeroth order equalizer (i.e., one which scales the system output) cannot satisfy

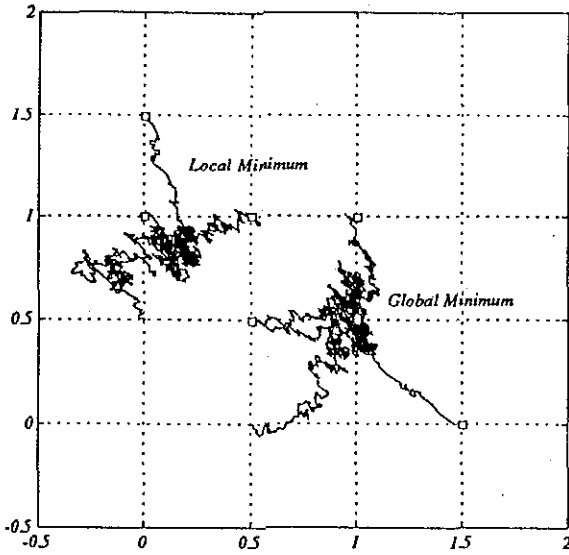


Figure 2: Simulation of Sato Algorithm with Uniform Input

the ideal objective (2.2), even approximately unless β is very small. In contrast the two tap equalizer ($m = 1$)

$$\Theta(q^{-1}) \triangleq \hat{\theta}_0(k) + \hat{\theta}_1(k)q^{-1} \quad (5.3)$$

can achieve the ideal objective of $z_k = a_k \forall k$ by setting $\hat{\theta}_0(k) = 1$ and $\hat{\theta}_1(k) = \beta$ (therefore the problem is well-posed).

We choose the input distribution to be uniform in the interval $[-d, +d]$. This distribution is subgaussian and along with the Sato algorithm forms a system which is captured under the theoretical considerations of [2] (apart from the equalizer being finite dimensional in our case).

5.2 Simulation Evidence

One direct and meaning way of investigating local minima is through simulation, although in no real sense does this convincingly prove the existence of local minima. Figure 2 show the trajectories of the Sato algorithm (3.1) on the AR(1) channel (5.1) for $\beta = 0.55$ for various initializations with an i.i.d., uniform $[-1.5, +1.5]$ system input. The trajectories appear to hang in the vicinity of the vertical axis symptomatic of a local minimum (other trajectories converge onto the desired parametrization $(1, 0.55)'$ which can be proven to be a global minimum). This simulation evidence for a local (potentially non-global) minimum guides our theoretical investigations to follow.

5.3 Local Equilibria

Local equilibria (3.10) are those points $\bar{\theta}$ which (using (3.7)) satisfy

$$E \left\{ \frac{\partial J(z_k)}{\partial \theta(k)} \right\} \Big|_{\theta(k)=\bar{\theta}} = E \left\{ (X_k' \bar{\theta} - \frac{2d}{3}) \text{sgn}(X_k' \bar{\theta}) X_k \right\} = 0 \quad (5.4)$$

where we have determined the dispersion from (3.4) to be $\gamma = 2d/3$ (the input is uniformly distributed on $[-d, +d]$). Based on the simulation evidence we wish to verify that there is indeed a locally stable equilibrium on the positive $\theta_1(k)$ axis. Therefore we constrain our search for local minima of the form

$$\bar{\theta} \triangleq [0, \bar{\theta}_1]' \quad (5.5)$$

which means $X_k' \bar{\theta} = x_{k-1} \bar{\theta}_1$ in (5.4). Then with the aid of the recursion (5.1) we can deduce that indeed there is such an equilibrium, when $\bar{\theta}_1$ satisfies

$$E \left\{ (|x_{k-1}| |\bar{\theta}_1| - \frac{2d}{3}) \text{sgn}(\bar{\theta}_1 x_{k-1}) x_{k-1} \right\} = 0 \quad (5.6)$$

which can be rearranged to obtain (there are two solutions; we just take the positive one)

$$\bar{\theta}_1 = + \frac{2d}{3} \frac{E\{|x_k|\}}{E\{|x_k|^2\}}. \quad (5.7)$$

When we attempt to evaluate (5.7) we have difficulties determining the term $E\{|x_k|\}$ explicitly. A closed form expression eludes us but this is of little concern as it can be evaluated numerically. More importantly we need to establish that (5.5) with (5.7) is a stable equilibrium rather than a saddle point.

5.4 Checking Stability

The following condition from (3.11) and (3.7) if positive definite implies stability:

$$\frac{\partial}{\partial \theta(k)} E \left\{ \frac{\partial J(z_k)}{\partial \theta(k)} \right\} = \underbrace{\frac{\partial}{\partial \theta(k)} E\{X_k X_k'\}}_{E\{X_k X_k'\}} - \frac{2d}{3} \frac{\partial}{\partial \theta(k)} E\{\text{sgn}(z_k) X_k\} \quad (5.8)$$

when evaluated at our candidate equilibrium $[0, \bar{\theta}_1]'$. In (5.8) the second term is highly non-trivial to evaluate. Clearly the first term is positive definite but the second term has the potential to destroy this property in the overall hessian.

As establishing the existence of local minima has serious consequences for the utility of existing (past) theoretical demonstrations we will take some care to prove our claims.

In the appendix we prove the following crucial relationship

$$\frac{\partial}{\partial \theta(k)} E\{\text{sgn}(z_k) X_k\} = \begin{bmatrix} \frac{2d^2 p(0)}{3\bar{\theta}_1} & 0 \\ 0 & 0 \end{bmatrix} \quad (5.9)$$

where $p(0)$ is the probability density of the system output x_k at zero which is shown to satisfy the following useful inequality (see the Appendix)

$$p(0) \leq \frac{1}{2d}. \quad (5.10)$$

Assembling the pieces we can write the second term in (5.8) as

$$-\frac{2d}{3} \begin{bmatrix} \frac{2d^2 p(0)}{3} & \frac{3}{2d} \frac{E\{|x_k|^2\}}{E\{|x_k|\}} & 0 \\ 0 & 0 & 0 \end{bmatrix} = -E\{|x_k|^2\} \begin{bmatrix} \frac{2d^2 p(0)}{3E\{|x_k|\}} & 0 \\ 0 & 0 \end{bmatrix} \quad (5.11)$$

using (5.7) and (5.9).

Now we focus on the first portion of (5.8) where we require the relationship $E\{x_k x_{k-1}\} = -\beta E\{|x_k|^2\}$ which comes from (5.1). Therefore we obtain

$$E\{X_k X_k'\} = E\{|x_k|^2\} \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix}. \quad (5.12)$$

So the positive definiteness of (5.8) hinges on

$$1 - \frac{2d^2 p(0)}{3E\{|x_k|\}} > \beta^2 \quad (5.13)$$

which is the determinant positivity condition obtained by combining (5.11) and (5.12).

This expression needs to be evaluated and given our earlier difficulty in computing $E\{|x_k|\}$ we will content ourselves with a lower bound (and this can be coupled with the upper bound in (5.10)). Write using Bayes' rule ($p(\cdot)$ denotes probability density)

$$\begin{aligned} E\{|x_k|\} &= \int_{-d}^{+d} p(a_k) E\{|x_k| | a_k\} da_k \\ &= \frac{1}{2d} \int_{-d}^{+d} E\{|x_k| | a_k\} da_k. \end{aligned} \quad (5.14)$$

Now we note that from (5.1) and the independence of $\{a_k\}$

$$\begin{aligned} E\{|x_k| | a_k = \eta\} &= E\{|\eta - \beta a_{k-1} + \beta^2 a_{k-2} - \dots|\} \\ &\geq |E\{\eta - \beta a_{k-1} + \beta^2 a_{k-2} - \dots\}| \\ &= |\eta| \end{aligned} \quad (5.15)$$

where we have used Jensen's inequality. Combining (5.14) and (5.15) gives

$$E\{|x_k|\} \geq \frac{1}{2d} \int_{-d}^{+d} |a_k| da_k = \frac{d}{2}. \quad (5.16)$$

By using our approximate bounds on $p(0)$ and $E\{|x_k|\}$ we can see whether (5.13) holds nonetheless. We have using (5.10) and (5.16) that

$$\frac{2d^2 p(0)}{3E\{|x_k|\}} \leq \frac{2}{3} \quad (5.17)$$

showing (5.13) is valid for (at least)

$$|\beta| < \frac{1}{\sqrt{3}} \cong 0.5774. \quad (5.18)$$

There are ways of tightening (5.15) to prove that for an even broader range of β (e.g., all β for which the system is stable) the equilibrium in (5.5) with (5.7) is indeed a minimum.

6 Discussion

So this completes our demonstration. We have constructed an example based on a simple AR(1) system for which a BGR-algorithm with subgaussian input (actually a Sato algorithm with a uniform input) exhibits a local stable undesirable equilibrium (not corresponding to the desired system inverse nor even a reasonable approximation of the same). This property which has serious consequences as far as practical implementation is concerned was for a finitely parametrized equalizer (neither an overparametrization nor an underparametrization). It can be demonstrated that for the same example any finite overparametrization (more than two adapted parameters; $1 < m < \infty$) will also have undesirable local minima. It is only when the equalizer is infinite dimensional (and therefore unrealizable; not to be confused with IIR) does the theory in [2] apply and prove there are no local undesirable minima (as mentioned before our results and those in [2] are not contradictory).

References

- [1] Y. Sato, "A method of self-recovering equalization for multi-level amplitude modulation," *IEEE Trans. on Commun.*, vol. COM-23, pp. 679-682, June 1975.
- [2] A. Benveniste M. Goursat and G. Ruget, "Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communications", *IEEE Trans. on Auto. Control*, vol. AC-25, pp. 385-399, June 1980.
- [3] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. on Commun.*, vol. COM-28, pp. 1867-1875, November 1980.
- [4] Z. Ding, C.R. Johnson Jr., and R.A. Kennedy, "On the (non-)existence and undesirable equilibria of Godard blind equalizers", *IEEE Trans. Acoust., Speech, Signal Processing.* (under review).
- [5] Z. Ding, R.A. Kennedy, B.D.O. Anderson, and C.R. Johnson Jr., "Ill-convergence of Godard blind equalizers in data communications", *IEEE Trans. Commun.* (accepted).

Appendix: Hessian Calculations

Here we wish to prove (5.9) when at equilibrium (5.5) with (5.7). Observe that for ϵ smaller than $|\bar{\theta}_1|$, there holds

$$\text{sgn}(\bar{\theta}_1 x_{k-1}) = \text{sgn}((\bar{\theta}_1 + \epsilon)x_{k-1})$$

and so at $\hat{\theta}(k) = \bar{\theta} = \{0, \bar{\theta}_1\}$

$$\frac{\partial}{\partial \hat{\theta}_1(k)} E\{\text{sgn}(z_k) X_k\} = \lim_{\epsilon \rightarrow 0} \epsilon^{-1} \left(E\{\text{sgn}((\bar{\theta}_1 + \epsilon)x_{k-1}) X_k\} - E\{\text{sgn}(\bar{\theta}_1 x_{k-1}) X_k\} \right) = 0.$$

Thus the second column of (5.9) is indeed zero.

Now we consider the derivatives with respect to $\hat{\theta}_0$. Towards this end define

$$f(\epsilon) \triangleq E\{x_{k-1} \text{sgn}(cx_k + \bar{\theta}_1 x_{k-1})\} \\ = E\{x_{k-1} \text{sgn}((\bar{\theta}_1 - \beta\epsilon)x_{k-1} + ca_k)\}$$

using the recursion (5.1). We are interested in computing $f'(\epsilon)$ at $\epsilon = 0$, but we need to incorporate our knowledge of the distribution of $\{a_k\}$ which is uniform on $[-d, +d]$ into the analysis. Let $\epsilon \geq 0$ be sufficiently small such that

$$k(\epsilon) \triangleq \frac{\epsilon}{\bar{\theta}_1 - \beta\epsilon} \geq 0,$$

and write using Bayes' rule

$$f(\epsilon) = E\{g(\epsilon, \eta)\} \triangleq \frac{1}{2d} \int_0^d (g(\epsilon, \eta) + g(\epsilon, -\eta)) d\eta$$

where

$$g(\epsilon, \eta) \triangleq E\{x_{k-1} \text{sgn}(x_{k-1} + k(\epsilon)a_k) \mid a_k = \eta\} \\ = \left\{ \int_{-k(\epsilon)\eta}^{+\infty} - \int_{-\infty}^{-k(\epsilon)\eta} \right\} x_{k-1} p(x_{k-1}) dx_{k-1}$$

noting a_k and x_{k-1} are independent random variables. Then

$$h(\epsilon, \eta) \triangleq g(\epsilon, \eta) + g(\epsilon, -\eta) \\ = \left\{ \int_{-k(\epsilon)\eta}^{+\infty} - \int_{-\infty}^{-k(\epsilon)\eta} + \int_{+k(\epsilon)\eta}^{+\infty} - \int_{-\infty}^{+k(\epsilon)\eta} \right\} xp(x) dx \\ = 4 \int_{+k(\epsilon)\eta}^{+\infty} x_{k-1} p(x_{k-1}) dx_{k-1}$$

once various symmetries are observed. Note that for any $\eta \in (0, d]$, $h(\epsilon, \eta)$ has a local maximum at $\epsilon = 0$. Hence $f(\epsilon)$ has a local maximum at $\epsilon = 0$. Hence $f'(0^+) = 0$. By symmetry a similar statement will be true for $\epsilon \leq 0$ and thus the desired derivative is zero. The conclusion is:

$$\frac{\partial}{\partial \hat{\theta}_0(k)} E\{\text{sgn}(z_k) x_{k-1}\} = 0$$

which is the lower left term in (5.9).

Now we study the upper left term in (5.9),

$$\frac{\partial}{\partial \hat{\theta}_0(k)} E\{\text{sgn}(z_k) x_k\}.$$

Since $x_k + \beta x_{k-1} = a_k$, it is enough to study the derivative of

$$m(\epsilon) \triangleq E\{a_k \text{sgn}(cx_k + \bar{\theta}_1 x_{k-1})\} \\ = E\{a_k \text{sgn}((\bar{\theta}_1 - \beta\epsilon)x_{k-1} + ca_k)\}.$$

With $k(\epsilon) \geq 0$ as before ($\epsilon \geq 0$ sufficiently small) define

$$n(\epsilon, \eta) = E\{a_k \text{sgn}(x_{k-1} + k(\epsilon)a_k) \mid a_k = \eta\}$$

meaning we can express (much as before)

$$m(\epsilon) = \frac{1}{2d} \int_0^d (n(\epsilon, \eta) + n(\epsilon, -\eta)) d\eta$$

where

$$q(\epsilon, \eta) \triangleq n(\epsilon, \eta) + n(\epsilon, -\eta) \\ = \eta \left\{ \int_{-k(\epsilon)\eta}^{+\infty} - \int_{-\infty}^{-k(\epsilon)\eta} - \int_{+k(\epsilon)\eta}^{+\infty} + \int_{-\infty}^{+k(\epsilon)\eta} \right\} p(x) dx \\ = 4\eta \int_0^{+k(\epsilon)\eta} p(x_{k-1}) dx_{k-1}.$$

Then we can evaluate

$$\frac{\partial q(\epsilon, \eta)}{\partial \epsilon} = 4\eta^2 p(k(\epsilon)\eta) k'(\epsilon)$$

which at $\epsilon = 0^+$ gives

$$\frac{\partial q(\epsilon, \eta)}{\partial \epsilon} \Big|_{\epsilon=0^+} = 4\eta^2 \frac{p(0)}{\bar{\theta}_1} > 0.$$

Therefore

$$\frac{\partial m(\epsilon)}{\partial \epsilon} \Big|_{\epsilon=0^+} = \frac{1}{2d} \int_0^d \frac{\partial q(\epsilon, \eta)}{\partial \epsilon} \Big|_{\epsilon=0^+} d\eta = \frac{2p(0)}{d\bar{\theta}_1} \int_0^d \eta^2 d\eta = \frac{2}{3} \frac{d^2 p(0)}{\bar{\theta}_1}$$

which supplies the missing term of (5.9). (Working with $\epsilon < 0$, meaning also $h(\epsilon, \eta) \leq 0$ and taking limits yields, not surprisingly, the same result.)

The remainder of this appendix is devoted to proving inequality (5.10). Let $p(x)$ denote the density of x_k (the system output), then $\beta^{-1} p(x\beta^{-1})$ is the density for βx_{k-1} . Noting that βx_{k-1} and a_k are independent random variables implies from (5.1) that

$$p(x) = \int_{-\infty}^{+\infty} r(x-z) \frac{1}{\beta} p\left(\frac{z}{\beta}\right) dz = \frac{1}{2d} \int_{x-d}^{x+d} \frac{1}{\beta} p\left(\frac{z}{\beta}\right) dz$$

where $r(\cdot)$ is the (uniform $[-d, +d]$) density of a_k . So

$$p(0) = \frac{1}{2d\beta} \int_{-d}^{+d} p\left(\frac{z}{\beta}\right) dz = \frac{1}{2d} \int_{-d\beta^{-1}}^{+d\beta^{-1}} p(y) dy \leq \frac{1}{2d}.$$