SERIES FEEDFOWARD INTERCONNECTED ADAPTIVE DEVICES

John MacLaren Walsh and C. Richard Johnson, Jr.

School of Electrical and Computer Engineering
Cornell University
Ithaca, New York 14853-0001
Email: jmw56@cornell.edu, johnson@ece.cornell.edu

ABSTRACT

We examine performance issues encountered when connecting two adaptive signal processing devices in a series feedfoward fashion. After introducing the notion of an adaptive element, we develop a simple behavior theory for series feedfoward connected adaptive elements given a few reasonable assumptions. We then move on to predict ways such interconnections of adaptive elements may misbehave. A brief example of misbehavior in a pair of interconnected adaptive receiver components is included to highlight the implications of this emerging theory.

1. INTRODUCTION

Interconnected adaptive systems are everywhere. They are a part of us, physiologically (e.g. the human visual system), and in our social interactions (e.g. the national economy). It is not surprising, then, that we engineer them into devices we build (e.g. adaptive control systems, adaptive communications receivers, and adaptive image processors). Oftentimes these systems are so complex that it is very difficult to determine or predict their behavior as a whole, while on a smaller scale (e.g. on the level of a single business or person in the human economy, a single neuron in the visual system, or a single adaptive device in an adaptive receiver) we can model their behavior quite accurately. Just as ancient Greek philosophy posited the existence of the atom to begin a study of complex chemistry, we posit the existence of an adaptive element: the smallest scale upon which a system can be considered adaptive. Because we wish to study systems of interconnected adaptive elements, we first seek to characterize how one or two adaptive elements might behave when connected together. This paper, along with [1], contains a couple of modest, somewhat obvious, and easy to reach beginnings of a theory of interconnected adaptive elements. In particular it characterizes the behavior and misbehavior of a particularly simple binary (that is, two element) adaptive compound in which the elements are connected together in series feedfoward fashion. Such a simple interconnection is rife in communication systems receivers, which is our application of interest here. To begin, we must characterize mathematically what we mean by ”adaptive element.”

2. WHAT IS AN ADAPTIVE ELEMENT?

We use the phrase ”adaptive element” as a shortened form of ”adaptive signal processing element.” The term ”adaptive”, in this instance, indicates that the element changes it processing based on parameters it infers from its input. We begin by providing a mathematical description of an adaptive element. We can discern two tasks which an adaptive element must perform:

1. Process the input to create the output, and
2. Adapt its method of processing the input to create the output based on previous inputs and outputs.

Mathematically speaking, this separation into tasks suggests the idea that there are two subsystems of equations involved with an adaptive element. We call these two subsystems the subelements or sub-atomic particles. The *processing subelement* processes the input signal to create an output signal, and the *adaptation subelement* determines how it should do so. Because we wish to be able to describe the adaptive element in a mathematical manner, we assume that the communication between the adaptive subelement and the processing subelement occurs in the form of a parameter vector, which we call the ***adaptive state***. The adaptive state totally determines the manner in which the adaptive element processes the current input to create the current output. Thus, a mathematical description of the processing subelement is

$$y_k = g(a_k, x_k)$$

where $x_k \in \mathbb{R}^p$ is the input, $a_k \in \mathbb{R}^n$ is the adaptive state, and $y_k \in \mathbb{R}^Q$ is the output of the adaptive signal processing element at time k. On the other hand, a mathematical description of the adaptation element is

$$a_{k+1} = f(k, a_k, x_k)$$

Where, as in Figure 1, $a_k \in \mathbb{R}^n$ is the adaptive state at positive integer time instant k and $x_k \in \mathbb{R}^P$ is the input vector at time k. The right hand pane of Figure 1 emphasizes the separation of the adaptive structure into two substructures, one that controls the adaptation by changing the adaptive states, and one that creates the input from the output based on the adaptive state. Thus, the adaptation subelement lies within the lower of the two boxes, while the processing subelement lies within the upper of the two boxes. The left pane shows a diagram that is equivalent, yet more compact, which we will use from this point on.

3. CASCADED ADAPTIVE ELEMENTS

Now that we have defined an adaptive element, we immediately turn our attention to systems containing two adaptive elements in

J. M. Walsh and C. R. Johnson, Jr. were supported in part by Applied Signal Technology and NSF Grant CCR-0310023.

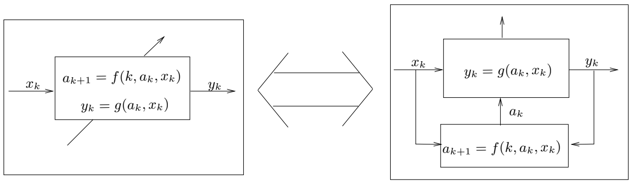

Fig. 1. Two equivalent representations of an adaptive element.

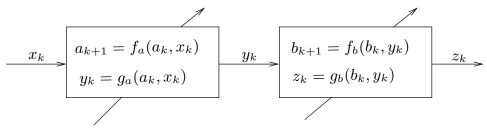

Fig. 2. Two adaptive devices connected in series feed-forward form.

cascade form, as shown in Figure 2. We refer to this configuration as a Series Feed-Forward Binary Adaptive Compound (SFFBAC). We now develop a theorem that predicts the way in which these cascaded adaptive elements will behave. Due to the success of SFFBACs in practice [2], we would expect that, if the first adaptive element is contractive[1] to desirable parameters a_k^*, and the second adaptive element is contractive to desirable parameters b_k^* whenever the first adaptive element is operating perfectly (ie, the parameters are at a_k^*), then the two connected elements should eventually reach a_k^* and b_k^*. Such an idea is common engineering intuition. We determined a set of engineering intuitive qualitative conditions like these, the Distributed Solution Conditions, which lead to proper operation of the SFFBAC, in [1]. We now set about proving that these qualitative/intuitive conditions actually guarantee proper operation of the SFFBACs in a mathematically rigorous way using the following theorem. We also indicate that the conditions are robust to small modelling errors and time variation in the desired points.

Theorem 1 (DS SFFBACs Time Variation and Noise) *Consider the SFFBAC shown in Figure 2 suffering from disturbances (noise n_k^a, n_k^b) and time variation.*

$$a_{k+1} = f_a(k, a_k, x_k) + n_k^a \tag{1}$$

$$y_k = g_a(a_k, x_k)$$

$$b_{k+1} = f_b(k, b_k, y_k) + n_k^b = f_b(k, b_k, g_a(a_k, x_k)) + n_k^b \tag{2}$$

The desired equilibrium trajectories are denoted a_k^* and b_k^*:

$$a_k^* = f_a(k, a_k^*, x_k) \quad \forall k$$

$$b_k^* = f_b(k, b_k^*, g_a(a_k^*, x_k)) \quad \forall k$$

and we assume that the first adaptive element is uniformly contractive to its desired parameters

$$\begin{array}{c} \|f_a(\xi, x_k) - f_a(a_k^*, x_k)\| < \alpha\|\xi - a_k^*\| \\ \alpha < 1 \quad \forall \xi \in \mathcal{B}_{a_k^*} \quad \forall k \end{array} \tag{3}$$

and that, given an ideal first adaptive element exactly at its desired parameter settings, the second adaptive element is uniformly contractive to its desired parameters

$$\begin{array}{c} \|f_b(\nu, g_a(a_k^*, x_k)) - f_b(b_k^*, g_a(a_k^*, x_k))\| < \beta\|\nu - b_k^*\| \\ \beta < 1 \quad \forall \nu \in \mathcal{B}_{b_k^*} \quad \forall k \end{array} \tag{4}$$

where $\mathcal{B}_{a_k^*} = \{\xi | \|\xi - a_k^*\| < r_a\}$ and $\mathcal{B}_{b_k^*} = \{\nu | \|\nu - b_k^*\| < r_b\}$ and we assume $\alpha \neq \beta$[2]. Assume Lipschitz continuity in the coupling of the first element's input-output equation and the second element's state equation.

$$\begin{array}{c} \|f_b(\nu, g_a(\xi_1, x_k)) - f_b(b, g_a(\xi_2, x_k))\| < \chi\|\xi_1 - \xi_2\| \\ \forall \xi_1, \xi_2 \in \mathcal{B}_{a_k^*} \ \forall \nu \in \mathcal{B}_{b_k^*} \end{array} \tag{5}$$

Finally, assume small enough noise and slow enough time variation.

$$\|a_{k+1}^* - a_k^*\| < d_1 \quad \|b_{k+1}^* - b_k^*\| < d_2 \tag{6}$$

$$\|n_k^a\| < c_a \quad \|n_k^b\| < c_b \tag{7}$$

Then, if the initializations are accurate enough, the disturbance is small enough, and the time variation is slow enough, such that

$$r_a > \frac{d_1 + c_a}{1 - \alpha} + \|a_1 - a_1^*\| \tag{8}$$

$$\begin{array}{rl} r_b > & \beta^{k_{max}}\|b_1 - b_1^*\| + \frac{\alpha\chi(\alpha^{k_{max}} - \beta^{k_{max}})}{\alpha - \beta}\|a_1 - a_1^*\| \\ & + \frac{d_1 + c_a}{(1-\alpha)(1-\beta)} + \frac{d_2 + c_b}{1 - \beta} \end{array} \tag{9}$$

where

$$k_{max} = \frac{\ln\left(\frac{\ln(\beta)}{\ln(\alpha)}\left(1 - \frac{(\alpha-\beta)\|b_1 - b_1^*\|}{\chi\|a_1 - a_1^*\|}\right)\right)}{\ln(\frac{\alpha}{\beta})} \tag{10}$$

then the whole system is exponentially stable to a ball of size $\frac{d_1 + c_a}{1-\alpha}$ around a_k^* and a ball of size $\frac{d_2 + c_b}{1-\beta} + \frac{\chi(d_1 + c_a)}{(1-\alpha)(1-\beta)}$ around b_k^*:

$$\|a_{k+1} - a_{k+1}^*\| \le \alpha^k\|a_1 - a_1^*\| + \frac{d_1 + c_a}{1 - \alpha} \tag{11}$$

$$\begin{array}{rl} \|b_{k+1} - b_{k+1}^*\| \le & \beta^k\|b_1 - b_1^*\| + \frac{\chi(\alpha^k - \beta^k)}{\alpha - \beta}\|a_1 - a_1^*\| \\ & + \frac{d_2 + c_b}{1-\beta} + \frac{\chi(d_1 + c_a)}{(1-\alpha)(1-\beta)} \end{array} \tag{12}$$

♦ Subtracting a_{k+1}^* from both sides, adding $a_k^* - a_k^*$ to the right hand side of (1), and using the triangle inequality gives

$$\|a_{k+1} - a_{k+1}^*\| \le \|f_a(k, a_k, x_k) - a_k^*\| + \|a_{k+1}^* - a_k^*\| + \|n_k^a\|$$

assuming that $a_i \in \mathcal{B}_{a_i^*} \forall i$ and using (6), (7), and (3) we have

$$\|a_{k+1} - a_{k+1}^*\| \le \alpha\|a_k - a_k^*\| + c_a + d_1$$

which, after running the recursion and using the sum of an infinite geometric series gives (11). Thus, to guarantee that $a_i \in \mathcal{B}_{a_i^*} \forall i$, we require (8). Proceeding in a similar manner with the second adaptive element, we have

$$\|b_{k+1} - b_{k+1}^*\| \le \|f_b(k, b_k, g_a(a_k, x_k)) - b_k^*\| + c_b + d_2$$

Adding $f_b(k, b_k, g_a(a_k^*, x_k)) - f_b(k, b_k, g_a(a_k^*, x_k))$ inside the norm on the right hand side, using the triangle inequality again, assuming that $b_i \in \mathcal{B}_{b_i^*} \ \forall i$, and using (4) and (5) gives

$$\|b_{k+1} - b_{k+1}^*\| \le \beta\|b_k - b_k^*\| + \chi\|a_k - a_k^*\| + c_b + d_2$$

[1]After each iteration the parameter is closer to the desired parameters that it was previously.

[2]For the special case when $\alpha = \beta$, see [1].

Substituting in (11) and running the recursion gives

$$\|b_{k+1} - b^*_{k+1}\| \leq \beta^k \|b_1 - b^*_1\| + \sum_{i=0}^{k-1} \beta^i \chi \left(\alpha^{k-i-1} \|a_1 - a^*_1\| + \frac{d_1+c_a}{1-\alpha} \right) + \frac{d_2+c_b}{1-\beta}$$

Pulling out an α^{k-1} and using the sum of a finite/infinite geometric series gives (12). To finish we must check our assumption that $b_i \in \mathcal{B}_{b^*_i}\ \forall i$. If the b_k with the largest $\|b_k - b^*_k\|$ is in $\mathcal{B}_{b^*_k}$, then all of the other b_is will be in $\mathcal{B}_{b^*_k}$. To find the largest, we take the derivative of (12) with respect to k and setting it equal to zero gives

$$\|b_1 - b^*_1\| - \frac{\chi\|a_1 - a^*_1\|}{\alpha - \beta} = -\frac{\ln(\alpha)}{\ln(\beta)} \frac{\chi\|a_1 - a^*_1\|}{\alpha - \beta} e^{\ln(\frac{\alpha}{\beta})k_{max}}$$

which indicates that the time instant at which the right hand side of (12) is maximized is k_{max} as given in (10). This indicates that we should require (9) to guarantee that we always stay within the balls in which our assumptions are valid. ⧫

Unfortunately, in most cases, (3) and (4) do not hold for the instantaneous update of an adaptive algorithm, especially in the large family of stochastic gradient algorithms. However, in this case, most often they are true for the expected system, which is the probabilistic mean of the adaptive algorithm. Given some regularity conditions, one can guarantee an error between the mean system and the true system which shrinks at least linearly with the step size[3] μ using stochastic hovering theorems [3]. Thus henceforth we will analyze the behavior of the averaged system with the realization that the instantaneous system will behave arbitrarily close to it if we shrink μ small enough.

Now that we have given conditions under which we can expect the SFFBAC to behave, that is for the trajectories a_k and b_k to approach their desired values, we wish to predict ways in which such an interconnected system may fail. Intuitively, if the first adaptive element has a step size (μ_a) which is chosen too small to be able to track the desired parameter settings a^{opt}_k, the second adaptive element may be kept from moving to its desired parameter settings. The next theorem makes this intuitively simple idea mathematically rigorous and proves it.

Theorem 2 (Zero Manifold Movement in SFFBACs) *Consider the SFFBAC shown in Figure 2, and suppose its elements take the form*

$$\hat{a}_{k+1} = \hat{a}_k + \mu f_a(\hat{a}_k - a^{opt}_k) \qquad (13)$$
$$b_{k+1} = b_k + \mu_b f_b(k, b_k, a_k)$$

Let f_a be a continuously differentiable function such that $f_a(0) = 0$ and $\alpha \equiv \|I + \mu_a df_a(0)\| < 1$, where $df_a(0)$ is the derivative of f_a evaluated at 0. Suppose a^{opt}_k is periodic, with an average component of a^*, such that:

$$a^* = \frac{1}{N} \sum_{i=k}^{N+k} a^{opt}_k \quad \forall k$$

Define a function $b^*(k, a)$ such that[4]

$$0 = f_b(k, b^*(k, a), a) \qquad \forall k, a \in \mathcal{D} \qquad (14)$$

[3] hidden within $f_a(k, a_k, x_k)$ and $f_b(k, b_k, y_k)$ in our previous notation, but included explicitly in the next theorem.

[4] Note that such a function is guaranteed to exist by the Implicit Function Theorem as long as $\frac{df_b}{db}|_{(a,b)=(a^*,b^{opt}_k)} \neq 0$ and we consider r_a and r_b small enough.

We call $b^*(k, a)$ the zero-manifold of the adaptive algorithm.

$$g(k, b, a) = \begin{cases} \frac{\|b+\mu_b f_b(k,b,a) - b^*(k,a)\|}{\|b - b^*(k,a)\|} & b \neq b^*(k, a) \\ \lim\limits_{b \to b^*(k,a)} \frac{\|b+\mu_b f_b(k,b,a) - b^*(k,a)\|}{\|b - b^*(k,a)\|} & b = b^*(k, a) \end{cases}$$

Define

$$\beta_{r_a, r_b} = \sup_{k \in \mathcal{N}, a \in \mathcal{B}_{a^*}(r_a)} \ \sup_{b \in \mathcal{B}_{b^*(k,a)}(r_b)} g(k, b, a) \qquad (15)$$

with $\mathcal{B}_{b^*(k,a)}(r_b) = \{b | \|b - b^*(k,a)\| \leq r_b\}$ and $\mathcal{B}_{a^*}(r_a) = \{a | \|a - a^*\| \leq r_a\}$ so that we have

$$\|b + \mu_b f_b(k, b, a) - b^*(k, a)\| \leq \beta_{r_a, r_b} \|b - b^*(k, a)\|$$

$$\forall a \in \mathcal{B}_{a^*}(r_a),\ b \in \mathcal{B}_{b^*(k,a)}(r_b)$$

We assume that the algorithms were chosen such that when the first element is at its optimal setting a^{opt}_k the second adaptive element is uniformly contractive to b^{opt}_k. Thus $\exists r_a,\ h_b$ such that $\beta_{r_a, r_b} < 1\ \forall r_b \leq h_b$ and $b^{opt}_k = b^*(k, a^{opt}_k)$, as is typically done when using distributed solution tactics. Also, assume

$$r_a > \|a_1 - a^*\| + \frac{d}{1 - \alpha} + \epsilon(\mu_a) \qquad (16)$$

and

$$r_b > \|b_1 - b^*(a^*, 1)\| + \frac{c}{1 - \beta_{r_a, r_b}} \qquad (17)$$

where we used the following constants:

$$d = \sup_{a \in \mathcal{B}_0(h_a)} \mu_a \|f_a(a_k + a^* - a^{opt}_k) - df_a(0) \left(a_k + a^* - a^{opt}_k\right)\|$$

$$c = \sup_k \|b^*(k, a^*) - b^*(k+1, a^*)\|$$

and the function, $\epsilon(\mu)$ is $O(\mu_a)$. If these assumptions hold, the first adaptive element's state converges to a ball around a^*, whose size can be made arbitrarily small by shrinking μ_a, and the second adaptive element's state converges to a ball around $b^*(k, a^*)$.

$$\|a_{k+1} - a^*\| \leq \alpha^k \|a_1 - a^*\| + \frac{d}{1 - \alpha} + O(\mu_a) \qquad (18)$$

$$\|b_{k+1} - b^*(k+1, a^*)\| \leq \beta^k_{r_a, r_b} \|b_1 - b^*(1, a^*)\| + \frac{c}{1 - \beta_{r_a, r_b}} \qquad (19)$$

Thus, as $\mu_a \to 0$, it is possible that neither algorithm converges to its optimal trajectory! Instead, the first element converges to the average, a^*, of its optimal periodic trajectory, and the second algorithm moves around in a zero manifold, $b^*(k, a^*)$. Since it is possible that neither adaptive element is at their optimum point, there could be performance degradation in the system.

⧫ Define $a_k = \hat{a}_k - a^*$, and substitute it in (13) to get:

$$a_{k+1} = a_k + \mu_a f_a(a_k + a^* - a^{opt}_k) \qquad (20)$$

We define the "linearization error"

$$h(a, k) = f_a(a_k + a^* - a^{opt}_k) - df_a(0) \left(a_k + a^* - a^{opt}_k\right)$$

This enables us to write

$$a_{k+1} = a_k + \mu_a df_a(0) \left(a_k + a^* - a^{opt}_k\right) + \mu_a h(a_k, k)$$

We define another system

$$\bar{a}_{k+1} = \bar{a}_k + \mu_a df_a(0)\bar{a}_k + \mu_a h(\bar{a}_k, k)$$

We can use the deterministic averaging hovering theorem (DHT) (see [3] and [1]) to relate these two systems. We check the applicability of this theory by looking at the total perturbation:

$$\begin{aligned} p(k,a) &= \textstyle\sum_{i=1}^{k} \left[df_a(0)\left(a + a^* - a_i^{opt}\right) + h(a,i)\right. \\ &\quad \left. - \left(df_a(0) + h(a,i)\right)\right] \\ &= \textstyle\sum_{i=1}^{k} df_a(0)\left(a^* - a_i^{opt}\right) \end{aligned}$$

which, considering (20), must be bounded, since it is periodic and drawn from the real numbers. The Lipschitz continuity, which is also required by the DHT, can be shown with

$$p(k,a) - p(k,a') = 0 \equiv L_p$$

For the DHT, [1], we also need $\|a_{k+1}^* - a_k^*\| \leq c \; \forall k$ and a condition on the initializations. The first holds with $c = 0$, due to the fixed point we show stability to, a^*. Assuming the DHT applies, we have the bound

$$\|\bar{a}_{k+1} - a_{k+1}\| \leq \frac{2 - \alpha^{T/\mu}}{1 - \alpha^{T/\mu}} \mu (B_p + L_p h + L_p B_f T) e^{\lambda_f T} \tag{21}$$

An important step in applying the deterministic hovering theorem to get (19), comes the form we chose for $\alpha \equiv \|I + \mu df_a(0)\|$, which, for small μ and stable averaged systems is $\approx 1 - \varrho\mu$ for some constant ϱ. Given this information, we can guarantee the right hand side in (21) is $O(\mu)$ by the deterministic hovering theorem. Details for this technique can be found in the hovering theorem section of [3] or [1]. Since we know from the triangle inequality that

$$\|a_{k+1} - a^*\| \leq \|a_{k+1} - \bar{a}_{k+1}\| + \|\bar{a}_{k+1} - a^*\| \tag{22}$$

we now focus on verifying that the second term in the sum is bounded. To do so, we can use Theorem 1, considering $h(a_k, k)$ as a disturbance, bounded using

$$\mu_a \|h(a_k, k)\| \leq d \equiv \sup_{a \in \mathcal{B}_0(h_a)} \mu_a \|h(a,k)\|$$

Theorem 1 tells us that the second term in (22) is bounded by

$$\|\bar{a}_{k+1} - a^*\| \leq \alpha^k \|\bar{a}_1 - a^*\| + \frac{d}{1-\alpha} \tag{23}$$

Since we are considering the same initial conditions for both the averaged system and the un-averaged system, we have $\bar{a}_1 = a_1$. Using this fact, and combining our bounds from (21) and (23) using (22), we have (18)[5]. Thus, apart from some linearization error, the first adaptive element converges to within an ball of a^* which can be made arbitrarily small by shrinking the step size μ_a. Moving on to the second element, using (15) we have

$$\begin{aligned} \|b_{k+1} - b^*(k+1, a^*)\| \leq \quad & \beta_{r_a, r_b} \|b_k - b^*(k, a^*)\| \\ & + \|b^*(k, a^*) - b^*(k+1, a^*)\| \end{aligned}$$

Defining $c = \sup_k \|b^*(k, a^*) - b^*(k+1, a^*)\|$, and using the sum of an infinite geometric series gives (19), which proves that we converge to a ball around the zero-manifold, which is possibly different from the desired trajectory, $b^{opt} = b^*(k, a_k^{opt})$. To finish we verify we remained within the balls in which our constants were valid. We do so using (18) and (19), to get (16) and (17) from the theorem.⧫

[5] So $\epsilon(\mu_a) =$ the sum of the right hand sides of (21) and (23).

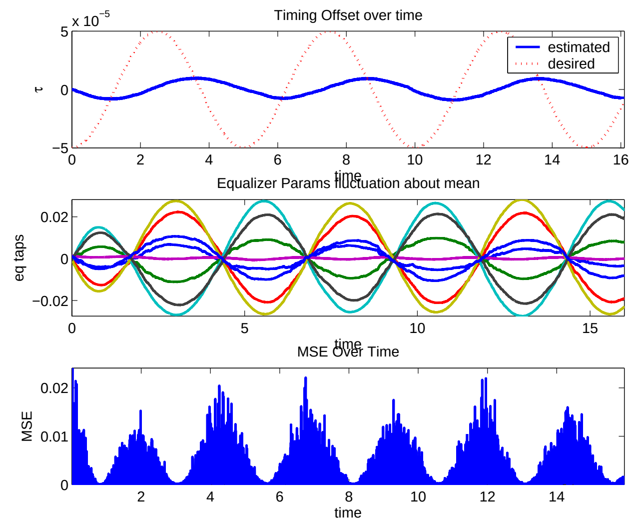

Fig. 3. Misbehavior in a SFFBAC composed of a power maximization timing recovery unit followed by a direct adaptive LMS equalizer.

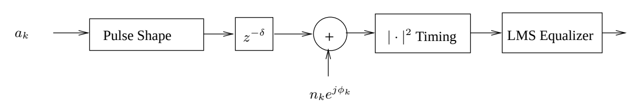

Fig. 4. The communications system featured in the misbehavior example.

4. EXAMPLES AND APPLICATIONS

We explored applications of both theorems to adaptive communications receivers in [1]. We merely reproduce the results here. Figure 3 shows misbehavior in a digital communications receiver containing a baud timing recovery unit using a power maximization algorithm followed by a least mean squares direct adaptive equalizer [2]. The top pane shows that the timing recovery's step size is too small for it to accurately estimate the fluctuating timing offset. The middle pane shows that the equalizer is fluctuating from its desired values which are constant over time because the synchronized channel is not changing. This fluctuation is due to the inaccurate timing offset estimate, and the bottom pane shows that the system is suffering from a periodic increase in mean squared error (decrease in performance) due to this misbehavior.

5. REFERENCES

[1] J. M. Walsh, *Distributed Adaptive Signal Processing: A Study of Series Feedforward Binary Adaptive Compounds.*, Master's Thesis, Cornell University, 2003.

[2] C. R. Johnson, Jr. and W. Sethares, *Telecommunications Breakdown: Concepts of Communication Transmitted via Software-Defined Radio,* Prentice Hall, 2003.

[3] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms: Stability and Performance,* Prentice Hall, Englewood Cliffs, N.J., 1995.