Robustness and Convergence of Adaptive Schemes in Blind Equalization*

Markus Rupp
Wireless Research Laboratory
Lucent Technologies
791 Holmdel-Keyport Rd.
Holmdel NJ 07733-0400

Ali H. Sayed
Department of Electrical Engineering
School of Engineering and Applied Science
University of California
Los Angeles, CA 90095-1594

Abstract

We pursue a time-domain feedback analysis of adaptive schemes with nonlinear update relations. We consider commonly used algorithms in blind equalization and study their performance in a purely deterministic framework. The derivation employs insights from system theory and feedback analysis, and it clarifies the combined effects of the step-size parameter and the nature of the nonlinear functional on the convergence and robustness performance of adaptive schemes.

1 INTRODUCTION

In recent work [1]-[4], the authors have formulated a time-domain feedback approach for the analysis and design of adaptive schemes with emphasis on robust performance and improved convergence in the presence of measurement noise and modeling uncertainties. In particular, we have addressed the following two issues:

1. We have shown how to select the adaptation gain (step-size) in order to guarantee a robust behaviour in the presence of noise and modeling uncertainties.

2. We have also shown how to select the adaptation gain in order to guarantee faster convergence.

In this paper, we briefly outline extensions of this formulation to adaptive schemes that involve nonlinear update laws, with special emphasis on blind and non-blind equalization schemes in communications.

Notation. We use small boldface letters to denote vectors, "$*$" for Hermitian conjugation, "T" for transposition, and $\|\mathbf{x}\|$ for the Euclidean norm of a vector. All vectors are column vectors except for the input data vector denoted by $\mathbf{u}_i$, which is taken to be a row vector. We also use the shift operator q^{-1}, defined by $q^{-1}s(i) = s(i-1)$, to denote the unit time delay.

*This work was supported in part by the National Science Foundation under Award No. MIP-9409319.

2 NONLINEAR ADAPTIVE SCHEMES

Fig. 1 depicts a nonlinear training structure that arises in channel equalization. The figure shows a sequence $\{s(i)\}$ (usually complex and of constant modulus) being transmitted through an unknown channel $C(q^{-1})$. The receiver is assumed to have an adaptive M-th order FIR structure with weights $\mathbf{w}_{i-1}$, followed by a nonlinear decision device f. The output of the decision device is used to compute an error quantity $e_o(i)$ that is employed in the training algorithm:

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(i)\mathbf{u}_i^* e_o(i)\ , \qquad (1)$$

with $\mathbf{u}_i = [\ u(i),\ \ldots,\ u(i-M+1)\ ]$. The definition of the error quantity $e_o(i)$ depends on whether the equalizer operates in a blind mode or not, which in turn determines the nature of the additional measurement used in Fig. 1. In non-blind operation, or training mode, the measurement is $s(i-D)$ (a delayed version of $s(i)$) and $e_o(i) = e_{o,n}(i) = s(i-D) - f[\mathbf{u}_i\mathbf{w}_{i-1}]$. In blind operation, or data mode, $e_o(i)$ is taken as $e_o(i) = e_{o,b}(i) = f[\mathbf{u}_i\mathbf{w}_{i-1}] - \mathbf{u}_i\mathbf{w}_{i-1}$. We assume for our analysis that there exists an optimal receiver $\mathbf{w}$ with such a structure, FIR followed by the nonlinearity, and which guarantees detection, viz., $f[\mathbf{u}_i\mathbf{w}] = s(i-D)$.

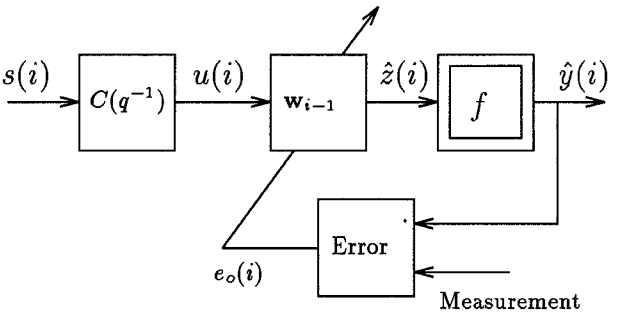

Figure 1: *Structure of the nonlinear adaptive equalizer.*

Table 1 lists several nonlinear functions that have been

used and studied in the context of channel equalization (see, e.g.,[5]-[14]).

Equalization type/algorithm	$f[z]$
Direct-decision 2-PSK	$\text{sign}[z]$
Direct-decision equalizer	$\text{dec}[z]$
CMA (Godard 2-2)	$z\lvert z\rvert^2$
Norm. CMA (Godard 1-2)	$\frac{z}{\lvert z\rvert}$
Sato's algorithm	$\gamma\text{sign}[z]$

Table 1: *Nonlinear devices for equalization.*

In channel equalization, we are interested in the limiting behaviour of the adaptive scheme (1) as time progresses to infinity. In particular, our objective is to exhibit conditions on $\{f, \mu(i)\}$ under which $e_a(i) = \mathbf{u}_i[\mathbf{w}-\mathbf{w}_{i-1}] = \mathbf{u}_i\tilde{\mathbf{w}}_i \rightarrow 0$ as $i \rightarrow \infty$ and, consequently, $\hat{z}(i) \rightarrow z(i)$ and $\hat{y}(i) \rightarrow y(i)$.

We shall assume, without loss of generality, that the update equation (1) is only employed when $e_o(i) \neq 0$ (i.e., we ignore the non-active steps and focus only on updates that involve nonzero error terms $e_o(i)$). In this case, our objective becomes the following: given a sequence of updates with nonzero errors $e_o(i)$, do the resulting weight vector estimates $\mathbf{w}_i$ tend to a value that guarantees $e_a(i) \rightarrow 0$ (and, consequently, $e_o(i) \rightarrow 0$)?

2.1 The Reference Model

The above questions are based on the assumption that there exists an optimal receiver that is capable of mapping the received symbols to the correct transmitted ones, viz., an optimal receiver that consists of a linear combiner $\mathbf{w}$ in cascade with the same nonlinear function $f[\cdot]$. We refer to this optimal configuration as the reference model. Fig. 2 depicts the optimal receiver (upper branch) together with the structure for adaptive equalizer (lower branch).

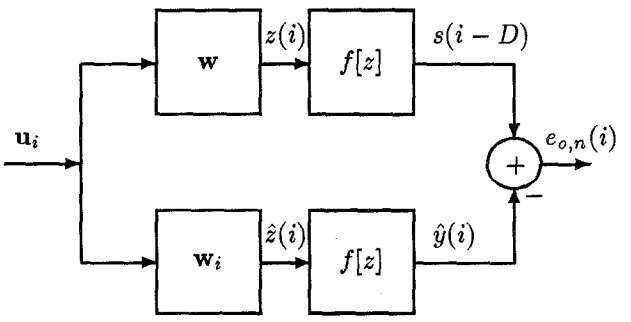

Figure 2: *Model reference structure for nonlinear equalization.*

2.2 Non-Blind Mode of Operation

Fig. 3 shows a feedback mapping of a-posteriori and a-priori errors following the methods described in [1]-[3], where h is the function that relates e_a and $e_{o,n}$, $e_{o,n}(i) = h[z(i), \hat{z}(i)]e_a(i)$, and is given by

$$h[z(i), \hat{z}(i)] = \frac{f[z(i)] - f[\hat{z}(i)]}{z(i) - \hat{z}(i)} . \qquad (2)$$

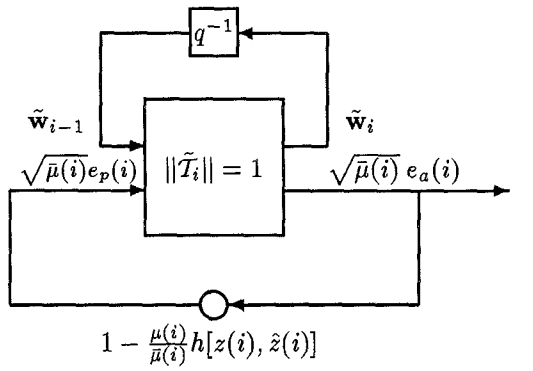

Figure 3: *Feedback structure implied by non-blind equalization.*

If we define

$$\Delta(N) = \max_{0\leq i\leq N} \left|1 - \frac{\mu(i)}{\bar{\mu}(i)}h[z(i), \hat{z}(i)]\right| ,$$

$$\gamma(N) = \max_{0\leq i\leq N} \frac{\mu(i)}{\bar{\mu}(i)} ,$$

and the normalized step-size $\bar{\mu}(i) = 1/\|\mathbf{u}_i\|^2$, then it can be checked (along the lines of [1]-[3]) that if $\Delta(N) < 1$ then the following bounds on the weighted energies of the a-priori estimation errors hold:

$$\sqrt{\sum_{i=0}^{N} \bar{\mu}(i)\,|e_a(i)|^2} \leq \frac{1}{1-\Delta(N)}\,\|\tilde{\mathbf{w}}_{-1}\| . \qquad (3)$$

$$\sqrt{\sum_{i=0}^{N} \mu(i)\,|e_a(i)|^2} \leq \frac{\gamma^{1/2}(N)}{1-\Delta(N)}\,\|\tilde{\mathbf{w}}_{-1}\| . \qquad (4)$$

Relations (3) and (4) are desirable because they imply, when they hold, that in the limit (as $N \rightarrow \infty$) the weighted energy of the a-priori estimation errors remains bounded and, hence, that $\{\sqrt{\mu(i)}\, e_a(i)\}$ and $\{\sqrt{\bar{\mu}(i)}\, e_a(i)\}$ tend to zero.

The condition $\Delta(N) < 1$ requires (in terms of the real and imaginary parts of h) that

$$\left[1 - \frac{\mu(i)}{\bar{\mu}(i)}h_R(i)\right]^2 + \frac{\mu^2(i)}{\bar{\mu}^2(i)}\,h_I^2(i) < 1 , \qquad (5)$$

which shows that h should necessarily be positive-real. These conditions can be verified for many of the algorithms listed in Table 1.

2.2.1 2-PSK and CMA

For example, for 2-PSK it can be verified that if $|\mathbf{u}_i\mathbf{w}|$ and $|\mathbf{u}_i\mathbf{w}_{i-1}|$ are uniformly bounded from above, and if $\mu(i)$ is chosen such that $0 < \mu(i) < |\mathbf{u}_i\mathbf{w}_{i-1}|/\|\mathbf{u}_i\|^2$, then $\sqrt{\mu(i)}\, e_a(i) \to 0$ as $i \to \infty$.

Likewise, for the CM algorithm, if $|\mathbf{u}_i\mathbf{w}|$ and $|\mathbf{u}_i\mathbf{w}_{i-1}|$ are uniformly bounded from below and if $\mu(i)$ is chosen such that (5) is satisfied, where $h[z(i), \hat{z}(i)]$ is evaluated as

$$h[z(i), \hat{z}(i)] = \frac{s(i-D) - \mathbf{u}_i\mathbf{w}_{i-1}|\mathbf{u}_i\mathbf{w}_{i-1}|^2}{s(i-D)|s(i-D)|^{-\frac{2}{3}} - \mathbf{u}_i\mathbf{w}_{i-1}} ,$$

then we also obtain $\sqrt{\mu(i)}\, e_a(i) \to 0$. For a projection step-size $\mu(i) = \alpha\bar{\mu}(i)$, we also see that α should guarantee

$$0 < |1 - \alpha h[\mathbf{u}_i\mathbf{w}, \mathbf{u}_i\mathbf{w}_{i-1}]| < 1 , \qquad (6)$$

which, in view of the positive-realness of the function h, can in general be guaranteed for small enough α.

2.2.2 An Alternative Nonlinearity

Note that the nonlinear decision functions used in 2-PSK, CM, and normalized CM lead to positive-real functions $h[z, \hat{z}]$, but their real parts are not necessarily bounded. On the other hand, assume we employ as a nonlinear device the function

$$f[z] = \frac{4|z|}{(1+|z|)^2}\, z . \qquad (7)$$

Then we always get $f[z] = z$ for $|z| = 1$. The maximum magnitude of the corresponding function $h[z, \hat{z}]$,

$$\max_{z,\hat{z}} |h[z, \hat{z}]| = \max_{z,\hat{z}} \left| \frac{f[z] - f[\hat{z}]}{z - \hat{z}} \right| ,$$

can be calculated by considering the limit as $z \to \hat{z}$ (a large value can only occur if the values of z and $\hat{z}$ are very close). Assume they both have the same phase but different amplitudes, say r and $\hat{r}$. We then compute the maximum magnitude of $h[z, \hat{z}]$ by applying L'Hospital's rule and obtain (as $r \to \hat{r}$) $\max_{z,\hat{z}} |h[z, \hat{z}]| = 8r/(1+r)^3$. The maximum value of this ratio occurs for $r = 0.5$. Therefore, the magnitude of the real part of $h[z, \hat{z}]$ is bounded from above by $4/(1.5)^3 = 1.1852$:

$$0 \le \text{Real}\,\{h[z, \hat{z}]\} \le \frac{4}{(1.5)^3} .$$

The nonlinear function (7) therefore allows, in contrast to the earlier cases (2-PSK, CMA, and normalized CMA), a step-size parameter that will not in general be as small as before. This is because the real part of $h[z, \hat{z}]$ is bounded. But since the larger the value of the step-size the faster the convergence speed, this function is therefore expected to lead to a higher convergence rate in the training phase. Simulation results are included at the end of this paper to support this statement.

2.3 Blind Mode of Operation

In the blind mode of operation, the feedback path is modified as shown in Fig. 4, with $(1-h)$ replacing h and where $v_z(i) = f[\mathbf{u}_i\mathbf{w}] - \mathbf{u}_i\mathbf{w}$ denotes the distortion introduced by the channel and by the optimal receiver $\mathbf{w}$.

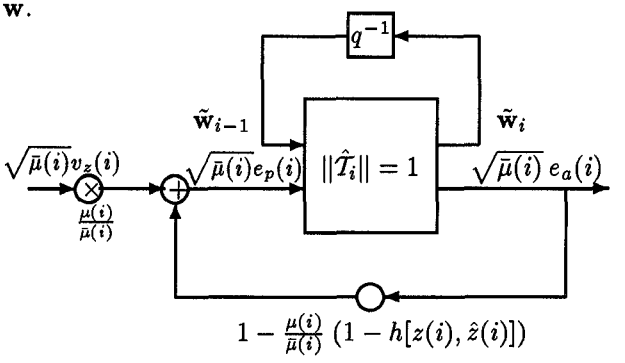

Figure 4: *Structure for blind operation.*

A contractive map will now require

$$\left| 1 - \frac{\mu(i)}{\bar{\mu}(i)}(1 - h[z(i), \hat{z}(i)]) \right| < 1 \qquad (8)$$

for all possible combinations of $z(i)$ and $\hat{z}(i)$ over the desired interval of time. A necessary condition for this to hold is to require the function $1-h$ to be positive real. This is in contrast to the non-blind training mode, which requires h itself to be positive real.

One can verify the following for 2-PSK operation. Assume the optimal receiver guarantees $|\mathbf{u}_i\mathbf{w}| = 1$ and its distortion $v_z(\cdot)$ is negligible or has finite energy. If the adaptive weights are only updated whenever $|\mathbf{u}_i\mathbf{w}_{i-1}| > 1$ and if $\mu(i)$ is chosen according to

$$\mu(i) < 2\bar{\mu}(i)\, \frac{|\hat{z}(i)| + 1}{|\hat{z}(i)| - 1}$$

then $\sqrt{\mu(i)}\, e_a(i) \to 0$.

Likewise, assume the weight updates in the normalized CM algorithm are performed only whenever $|\mathbf{u}_i\mathbf{w}_{i-1}| > 1 + \epsilon$ for some given positive $\epsilon \ll 1$. Assume also that the optimal receiver guarantees $|\mathbf{u}_i\mathbf{w}| < 1 + \epsilon$. If $\mu(i) < 2\bar{\mu}(i)$, and if the optimal channel-receiver distortion is negligible or has finite energy, then we also obtain $\sqrt{\mu(i)}\, e_a(i) \to 0$.

3 SIMULATION RESULTS

We start with the non-blind mode of operation.

3.1 Non-Blind Mode of Operation

3.1.1 CMA and New Nonlinear Device

The channel is chosen as $C(q^{-1}) = 1 + 1.2q^{-1}$, which leads to a purely non-causal inverse. The receiver length is $M = 3$. Our first simulation compares the performance of the normalized CM algorithm and the nonlinear device suggested in (7). The transmitted signals are general CM signals with randomly chosen phases. The nonlinear device is $f[z] = z/|z|$. A small value for α is chosen, viz., $\alpha = 0.1$ in (6). The initial weight vector was selected as $\mathbf{w}_{-1}^T = [0, 0, 1]$.

Fig. 5 compares the learning curves (averaged over 50 runs) of the normalized CM algorithm and the non-blind algorithm that corresponds to the proposed nonlinear function $f[z] = 4z|z|/(1 + |z|)^2$.

The error that is depicted in Fig. 5, and which has been used for a fair comparison of both algorithms, is $e(i) = s(i - D) - \hat{z}(i)/|\hat{z}(i)|$.

As the figure demonstrates, the new nonlinear device leads to considerably smaller errors. But since both filter lengths are small, both equalizers are not capable of completely reconstructing the transmitted signal.

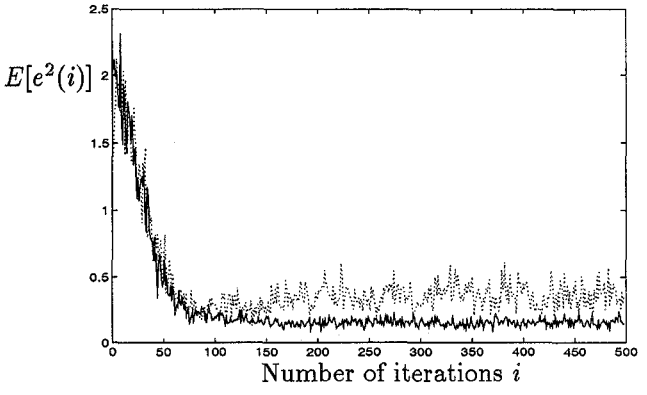

Figure 5: *Learning curves for general CM signals with equalizer filter length $M = 3$ for normalized CM algorithm (dotted line) and for new nonlinear device (continuous line).*

3.1.2 2-PSK Case

In this experiment the channel is chosen as $C(q^{-1}) = 1 + 0.9q^{-1}$, and the receiver length is taken as $M = 3$. The initial weight vector is $\mathbf{w}_{-1} = [0, 1, 0]^T$.

The step-size parameter was chosen in two ways: a non-normalized mode where $\mu(i) < \bar{\mu}(i)$ (as is the case with standard gradient algorithms [2]) and a normalized mode where $\mu(i) < \bar{\mu}(i)|\hat{z}(i)|$ as suggested by the discussion at Sec. 2.2.1. Fig. 6 depicts the results for the two modes, where α denotes $\mu(i)/\bar{\mu}(i)$. The figure shows the Bit-Error-Rates (BER); the ratio of falsely detected bits to the overall transmitted bits. The algorithms were run for $N = 200$ steps and the results averaged over 20 Monte Carlo runs. The new normalization shows superior behaviour compared to the non-normalized algorithm for every step-size chosen. The improvement is by about a factor of four.

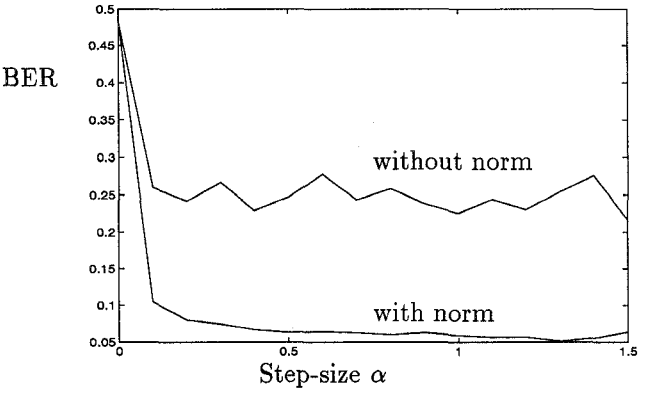

Figure 6: *BER for 2-PSK with various step-sizes α for normalized and non-normalized mode.*

3.2 Blind Mode of Operation

As argued in Sec. 2.3, a convergent (and robust) performance in the blind mode of operation can be guaranteed as long as the operation of the adaptive equalizer is restricted to "large" enough values of $\hat{z}(i)$.

3.2.1 2-PSK Case

In the following experiment, we continue to employ the channel $C(q^{-1}) = 1 + 0.9q^{-1}$ and a receiver length $M = 3$. The initial coefficients were $\mathbf{w}_{-1} = [0.8, -0.2, 0]^T$.

We ran several simulations with $\alpha = 0.1 = \mu(i)/\bar{\mu}(i)$ and for different values of β. That is, we compared the results when the update was done only for values $|\hat{z}| > \beta$ for various β.

Fig. 7 depicts BER values as a function of β. For $\beta = 0$, the standard 2-PSK algorithm is obtained. The experiment was run for $N = 200$ steps and the results averaged over 20 runs. For values of $\beta > 0$, the algorithm shows a considerably improved behaviour. However, the larger the β, the smaller becomes the improvement, since then the updates become less frequent.

3.2.2 Normalized CM Algorithm

The next experiment considers the normalized CM algorithm. As indicated in Sec.2.3, the convergence (and

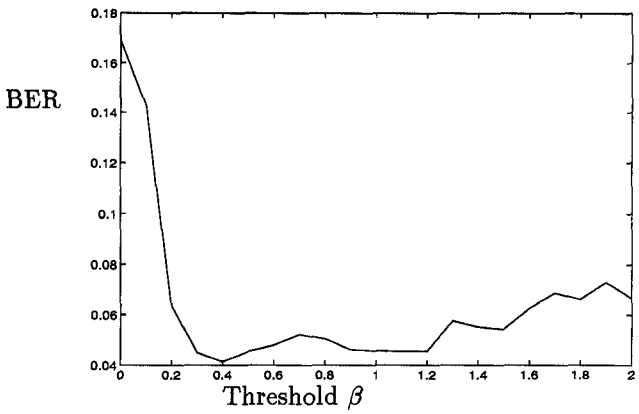

Figure 7: *BER for 2-PSK with different thresholds* β *(*$\alpha = 0.1$*).*

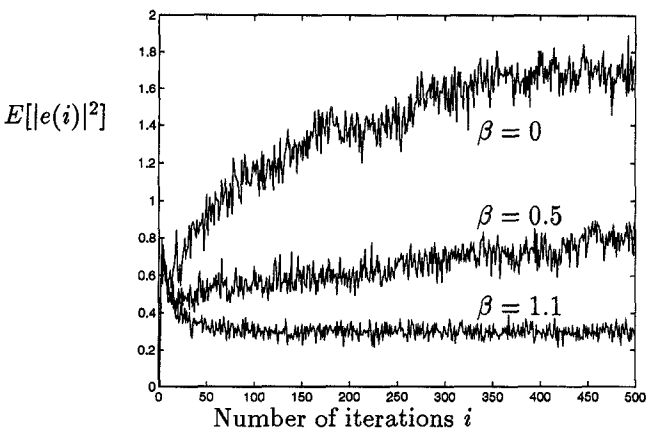

Figure 8: *Instantaneous error energy for normalized CM algorithm with step-size* $\alpha = 1$ *and threshold* $\beta = 0, 0.5, 1.1$.

robustness) performance can be improved by updating only for $|\hat{z}(i)| > 1 + \epsilon$.

Fig. 8 depicts three averaged learning curves obtained from averaging $|e(i)|^2 = |s(i-D) - \hat{z}(i)/|\hat{z}(i)||^2$ over 200 runs. The weight vector update employed a normalized step-size $\alpha = 1.0$ and was used only for $|\hat{z}(i)| > \beta$. The step-size was set to zero otherwise. The algorithm was run for $M = 3$ with $\mathbf{w}_{-1} = [2, 0, 0]^T$, for the channel $C(q^{-1}) = 1 + 1.2q^{-1}$. Note that a large initial value for $\mathbf{w}_{-1}$ is necessary if $\beta > 1$ since otherwise the algorithm never updates.

As the figure demonstrates, the normalized CMA has superior performance if it is not updated at every time instant. For $\beta = 0.5$ we have improvements over $\beta = 0$. For $\beta = 1.1$ the learning curves show the best performance.

We found $\epsilon = \max_i |v_z(i)| \approx 0.28$ in this case. For $\beta = 1 + \epsilon = 1.28$, the behaviour of the algorithm is close to the one from $\beta = 1.1$. Since $v_z(i)$ works as an additive noise, the output error of the normalized CM algorithm in steady-state is expected to be larger than $2|v_z(i)|^2$ which is in very good agreement with the simulation. Further experiments showed that the error energy can be reduced by decreasing the step-size. The tightness of the bound $\mu(i) < 2\bar{\mu}(i)$ has also been investigated. Further experiments showed that with $\beta = 1.5$ the algorithm converged for step-sizes $\alpha < 2$ as expected. Other situations in which $\epsilon = \max_i |v_z(i)|$ is very small did not show much improvement. In fact when the estimate $\hat{z}$ is close to z, $\text{Real}\{1 - h[z, \hat{z}]\}$ becomes positive even if $|\hat{z}(i)| < 1$.

References

[1] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive gradient algorithms via the Small Gain Theorem," *Proc. SPIE*, vol. 2563, pp. 458–469, San Diego, CA, July 1995.

[2] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1428-1440, June 1996.

[3] M. Rupp and A. H. Sayed, "Robustness of Gauss-Newton recursive methods: A deterministic feedback analysis," *Signal Processing*, vol 50, no. 3, pp. 165–188, June 1996.

[4] A. H. Sayed and M. Rupp, "Error energy bounds for adaptive gradient algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1982-1989, Aug. 1996.

[5] S. Haykin, *Blind Deconvolution*, NJ: Prentice Hall, 1994.

[6] Y. Sato, "A method of self-recovering equalization for multi level amplitude modulation, *IEEE Trans. Commun.*, vol. COM-23, pp. 679-682, June 1975.

[7] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867-1875, Nov. 1980.

[8] J.E. Mazo,"Analysis of decision-directed equalizer convergence," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1857-1877, Dec. 1980.

[9] M. G. Larimore and J. R. Treichler, "Convergence behaviour of the constant modulus algorithm," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 13–16, Boston, MA, 1983.

[10] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 161–176, 1984.

[11] A. Benveniste, M. Goursat, "Blind Equalizers," *IEEE Transactions on Communications*, vol. 32, no. 8, pp. 871-883, Aug. 1984.

[12] G.J. Foschini, "Equalization without altering or detect data," *AT&T Technical Journal*, pp. 1885-1991, Oct. 1985.

[13] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 292-297, Jan. 1990.

[14] Y. Li and Z. Ding, "Convergence analysis of finite length blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. SP-43, pp. 2120-2129, Sep. 1995.