Steady-State and Tracking Analyses of the Sign Algorithm Without the Explicit Use of the Independence Assumption

Nabil R. Yousef, *Student Member, IEEE,* and Ali H. Sayed, *Senior Member, IEEE*

***Abstract*—In this letter, we present a simple approach to the steady-state and tracking analyses of the sign algorithm that avoids the explicit use of the independence assumption.**

***Index Terms*—Adaptive filter, least mean square (LMS), independence assumption, sign-ups, steady-state, tracking.**

I. Introduction

CONSIDER noisy measurements $\{d(i)\}$ that arise from the linear model

$$d(i) = \mathbf{u}_i \mathbf{w}^o + v(i) \tag{1}$$

where

- $\mathbf{w}^o$ unknown *column* vector that we wish to estimate;
- $v(i)$ accounts for measurement noise and modeling errors;
- $\mathbf{u}_i$ row input (or regressor) vector.

The most popular adaptive algorithm for the estimation of $\mathbf{w}^o$ is the least mean squares (LMS) algorithm, which is given by

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

where

- $\mathbf{w}_i$ estimate for $\mathbf{w}^o$ at iteration i;
- μ step size;
- $e(i)$ so-called output estimation error;
- $e(i)$ $d(i) - \mathbf{u}_i \mathbf{w}_i$;
- $*$ complex conjugate transposition.

In high speed data communications, the symbol interval may not be long enough to execute an iteration of the LMS algorithm. This makes multiplication-free variants of the LMS algorithm very convenient for these applications. Among these variants is the sign algorithm (SA) [1], [2], which is given by[1]

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu\, \mathbf{u}_i^* \mathrm{sign}(e(i)). \tag{2}$$

In addition to its simplicity, the SA possesses many other advantages such as high robustness and convenient tracking properties [3].

An important performance measure for an adaptive filter is its steady-state mean square error (MSE), which is defined as

$$\mathtt{MSE} = \lim_{i\to\infty} \mathrm{E}(|e(i)|^2) = \lim_{i\to\infty} \mathrm{E}(|v(i) + \mathbf{u}_i \tilde{\mathbf{w}}_i|^2)$$

where $\tilde{\mathbf{w}}_i = \mathbf{w}^o - \mathbf{w}_i$ denotes the weight error vector. Under the often realistic assumption that

A1) The noise sequence $\{v(i)\}$ is iid and statistically independent of the regressor sequence $\{\mathbf{u}_i\}$.

We find that the MSE is equivalently given by

$$\mathtt{MSE} = \sigma_v^2 + \lim_{i\to\infty} \mathrm{E}(|\mathbf{u}_i \tilde{\mathbf{w}}_i|^2). \tag{3}$$

Another assumption that is widely used in the literature for evaluating (3) is to assume, in addition to A1, that the regression vector $\mathbf{u}_i$ is independent of $\tilde{\mathbf{w}}_i$ (see, e.g., [2]–[7]). With this in mind, one proceeds to determine a recursion for the variance of the weight error vector $\mathrm{E}(\tilde{\mathbf{w}}_i \tilde{\mathbf{w}}_i^*)$. Then the MSE is computed by using the steady-state value of this variance $\mathrm{E}(\tilde{\mathbf{w}}_\infty \tilde{\mathbf{w}}_\infty^*)$. In this letter, we present a more direct approach, which avoids the explicit use of the independence assumption and the need for evaluating $\mathrm{E}(\tilde{\mathbf{w}}_\infty \tilde{\mathbf{w}}_\infty^*)$.

II. Fundamental Energy Relation

We start by defining the following so-called *a priori* and *a posteriori* estimation errors

$$e_a(i) = \mathbf{u}_i \tilde{\mathbf{w}}_i, \quad e_p(i) = \mathbf{u}_i \tilde{\mathbf{w}}_{i+1}.$$

Using the data model (1), it is easy to see that the errors $\{e(i), e_a(i)\}$ are related via $e(i) = e_a(i) + v(i)$. If we further subtract $\mathbf{w}^o$ from both sides of (2) and multiply by $\mathbf{u}_i$ from the left, we also find that the errors $\{e_p(i), e_a(i), e(i)\}$ are related for nonzero $\mathbf{u}_i$ via

$$e_p(i) = e_a(i) - \mu \|\mathbf{u}_i\|^2 \mathrm{sign}(e(i)). \tag{4}$$

Substituting (4) into (2), we obtain the update relation

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - \frac{\mathbf{u}_i^*}{\|\mathbf{u}_i\|^2}[e_a(i) - e_p(i)].$$

Manuscript received October 19, 1999. This material was based on work supported in part by the National Science Foundation Awards CCR-9732376 and ECS-9820765. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. G. Ramponi.

The authors are with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@eu.ucla.edu).

Publisher Item Identifier S 1070-9908(00)09405-0.

[1]For real data, $\mathrm{sign}(x) = 1$ if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. For complex data, the sign function is defined as $\mathrm{sign}[a + jb] = (1/\sqrt{2})(\mathrm{sign}[a] + j\,\mathrm{sign}\,[b])$

By evaluating the energies of both sides of this equation, we find that (see [8] and [9])

$$\|\tilde{\mathbf{w}}_{i+1}\|^2 + \frac{1}{\|\mathbf{u}_i\|^2}|e_a(i)|^2 = \|\tilde{\mathbf{w}}_i\|^2 + \frac{1}{\|\mathbf{u}_i\|^2}|e_p(i)|^2. \quad (5)$$

When $\mathbf{u}_i = 0$, it is obviously true that

$$\|\tilde{\mathbf{w}}_{i+1}\|^2 = \|\tilde{\mathbf{w}}_i\|^2. \quad (6)$$

Both results (5) and (6) can be grouped together into a single equation by defining

$$\overline{\mu}(i) = \left(\|\mathbf{u}_i\|^2\right)^\dagger$$

in terms of the pseudo-inverse of a scalar, so that we obtain

$$\|\tilde{\mathbf{w}}_{i+1}\|^2 + \overline{\mu}(i)|e_a(i)|^2 = \|\tilde{\mathbf{w}}_i\|^2 + \overline{\mu}(i)|e_p(i)|^2. \quad (7)$$

No approximations or assumptions are needed to establish the energy conservation relation (7). It is an exact relation that shows how the energies of the weight error vectors at two successive time instants are related to the energies of the *a priori* and *a posteriori* estimation errors.

III. Steady-State Analysis

In this letter, we are interested in knowing what performance we can expect from the filter if it reaches steady state. Thus, by taking expectations of both sides of (4) and noting that $\mathrm{E}(\|\tilde{\mathbf{w}}_{i+1}\|^2) = \mathrm{E}(\|\tilde{\mathbf{w}}_i\|^2)$ in steady state, we get

$$\mathrm{E}(\overline{\mu}(i)|e_a(i)|^2) = \mathrm{E}(\overline{\mu}(i)|e_p(i)|^2).$$

Using (7), the above collapses to the following relation in terms of $\{e_a(i), v(i)\}$ only (recall that $e(i) = e_a(i) + v(i)$):

$$\mathrm{E}(\overline{\mu}(i)|e_a(i)|^2) = \mathrm{E}\left(\overline{\mu}(i)\left|e_a(i) - \frac{\mu}{\overline{\mu}(i)}\mathrm{sign}(e(i))\right|^2\right). \quad (8)$$

Expanding (8), terms involving $\overline{\mu}(i)$ cancel out, and we obtain the equality:

$$2\mu\mathrm{E}\left(e_a(i)\mathrm{sign}\left(e_a(i) + v(i)\right)\right) = \mu^2\mathrm{Tr}(\mathbf{R}) \quad (9)$$

where $\mathbf{R} = \mathrm{E}(\mathbf{u}_i^*\mathbf{u}_i)$. This equation can now be solved for the desired MSE, which is given from (3) by $\mathtt{MSE} = \sigma_v^2 + \mathrm{E}(|e_a(i)|^2)$. We stress that (9) is an *exact* relation that holds *without* any approximations or assumptions (except for the assumption that the filter reaches steady-state). The procedure of finding the MSE through (9) avoids the need for evaluating $\mathrm{E}(\|\tilde{\mathbf{w}}_\infty\|^2)$. This is because in steady state and in view of the energy-preserving relation (7), the effect of the weight error variance is canceled out.

To proceed, we introduce the assumption that the signals $e(i)$ and $v(i)$ are jointly Gaussian in steady state (so that $e_a(i)$ and $v(i)$ are jointly Gaussian). This assumption is reasonable for sufficiently small step sizes and in steady-state operation (which is the state we are interested in). To see this, assume the original data $\mathbf{u}_i$ and $v(i)$ are Gaussian. This does not imply that $e(i)$ is Gaussian, because $e(i)$ depends on the data $\{\mathbf{u}_j, v(j)\}$ in a nonlinear fashion through $\tilde{\mathbf{w}}_i$. However, if we assume that $\tilde{\mathbf{w}}_i$ is a constant (and not a random variable), then $e(i)$ will be Gaussian. This suggests that in steady-state operation and when the step size is small enough that the weight error vector varies slowly with time, we can expect the Gaussianity assumption on $\{e(i), v(i)\}$ to be reasonable.

With this Gaussianity assumption, we can now proceed to determine the MSE of the sign algorithm. We may remark that earlier works in the literature have already studied this same problem and arrived at an expression for the MSE [4], [6]. This was achieved in [4] as follows. First, a transient analysis is performed in order to derive a recursion for the variance of the weight error vector $\mathrm{E}(\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^*)$. The derivation of this recursion requires a certain independence assumption in addition to the Gaussianity of $e(i)$. Second, the steady-state value of the weight error variance $\mathrm{E}(\tilde{\mathbf{w}}_\infty\tilde{\mathbf{w}}_\infty^*)$ is used to evaluate the MSE. In this letter, by focusing on the steady-state operation of the filter right away, and by relying on the energy relation (7), we avoid the transient analysis and the independence condition. The derivation in the sequel relies solely on the Gaussianity assumption of $e(i)$ and $v(i)$ in steady state.

Using A1 and Price's theorem for both real and complex-valued data [10],[2] we obtain

$$\mathrm{E}(e_a(i\,\mathrm{sign}(e_a(i) + v(i))) = \sqrt{\frac{2}{\pi}}\frac{\mathrm{E}(|e_a(i)|^2)}{\sqrt{\sigma_v^2 + \mathrm{E}|e_a(i)|^2}}. \quad (10)$$

Substituting into (9) and solving for $\mathtt{MSE} = \sigma_v^2 + \mathrm{E}(|e_a(i)|^2)$, we find that

$$\mathtt{MSE} = \sigma_v^2 + \frac{\alpha}{2}\cdot\left(\alpha + \sqrt{\alpha^2 + 4\sigma_v^2}\right) \quad (11)$$

where $\alpha = \sqrt{\pi/8}\mu\mathrm{Tr}(\mathrm{R})$. This is the same result that was obtained in [4] by taking the limit of $\mathrm{E}(\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^*)$. Here, we have obtained it more directly and without explicitly using the independence assumption.

IV. Tracking Analysis

In a nonstationary environment, the data $\{d(i)\}$ is assumed to arise from a linear model of the form $d(i) = \mathbf{u}_i\mathbf{w}_i^o + v(i)$, where the unknown system $\mathbf{w}_i^o$ is now time variant. It is often assumed that the variation in $\mathbf{w}_i^o$ is according to the model, $\mathbf{w}_{i+1}^o = \mathbf{w}_i^o + \mathbf{q}_i$, where $\mathbf{q}_i$ denotes the random perturbation [3], [6]. The purpose of the tracking analysis of an adaptive filter is to study its ability to track such time variations. To evaluate the tracking performance of the SA, we first redefine the weight error vector as $\tilde{\mathbf{w}}_i = \mathbf{w}_i^o - \mathbf{w}_i$, and the *a posteriori* estimation error as $e_p(i) = \mathbf{u}_i(\tilde{\mathbf{w}}_{i+1} - \mathbf{q}_i)$. Then $\tilde{\mathbf{w}}_i$ satisfies

$$\tilde{\mathbf{w}}_{i+1} = \tilde{\mathbf{w}}_i - \mu(i)\mathbf{u}_i^*\mathrm{sign}(e(i)) + \mathbf{q}_i. \quad (12)$$

If we further multiply (12) by $\mathbf{u}_i$ from the left, we obtain that (4) and (5) still hold for the nonstationary case, while (7) becomes:

$$\|\tilde{\mathbf{w}}_{i+1} - \mathbf{q}_i\|^2 + \overline{\mu}(i)|e_a(i)|^2 = \|\tilde{\mathbf{w}}_i\|^2 + \overline{\mu}(i)|e_p(i)|^2. \quad (13)$$

For mathematical tractability of the tracking analysis, we impose the following assumption, which is typical in the context

[2]For two jointly Gaussian random variables x and y, we have $\mathrm{E}(x\mathrm{sign}(y)) = \sqrt{2/\pi}\cdot(1/\sigma_y)\mathrm{E}(xy)$.

of tracking analysis of adaptive filters (see, e.g., [3], [6], and [7]).

A2) The sequence $\{\mathbf{q}_i\}$ is a stationary sequence of independent zero-mean vectors whose autocorrelation matrix $\mathbf{Q} = \mathrm{E}\mathbf{q}_i\mathbf{q}_i^*$ is positive definite. Furthermore, $\{\mathbf{q}_i\}$ is mutually statistically independent of the sequences $\{\mathbf{u}_i\}$ and $\{v(i)\}$.

Using (4) and A.2, it is straightforward to verify that the variance relation (8) should now be replaced by

$$\mathrm{E}\left(\overline{\mu}(i)|e_a(i)|^2\right) = \mathrm{Tr}(\mathbf{Q}) + \mathrm{E}\left(\overline{\mu}(i)\left|e_a(i) - \frac{\mu}{\overline{\mu}(i)}\mathrm{sign}(e(i))\right|^2\right).$$

Comparing the above with (8), we see that both expressions differ by the single term $\mathrm{Tr}(\mathbf{Q})$. This indicates that the MSE in the nonstationary case can be obtained almost by inspection from the result for the stationary case. Indeed, expanding the above equality we obtain [compare with (9)]

$$2\mu\mathrm{E}(e_a(i)\mathrm{sign}\,(e_a(i) + v(i))) = \mathrm{Tr}(\mathbf{Q}) + \mu^2\mathrm{Tr}(\mathbf{R}). \quad (14)$$

Using (10), we find that the MSE is still given by (11), where α is now given by

$$\alpha = \sqrt{\frac{\pi}{8}}\left(\mu^{-1}\mathrm{Tr}(\mathbf{Q}) + \mu\mathrm{Tr}(\mathbf{R})\right).$$

This is the same result obtained in [6] by repeating the analysis of [4]. Here, we have obtained the result in a more direct way by simply comparing the energy relations (9) and (14) for the stationary and nonstationary cases.

V. Concluding Remarks

We may remark that the approach of this letter can also be applied to more general adaptive schemes (see, e.g., [11]–[13]).

References

[1] N. A. M. Verhoeckx, H. C. Van Den Elzen, F. A. M. Snijders, and P. J. Van Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 768–781, Dec. 1979.

[2] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 510–518, June 1981.

[3] E. Eweda, "Comparison of RLS, LMS, and sign algorithms for tracking randomly time-varying channels," *IEEE Trans. Signal Processing*, vol. 42, pp. 2937–2944, Nov. 1994.

[4] V. Mathews and S. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 450–454, Apr. 1987.

[5] E. Masry and F. Bullo, "Convergence analysis of the sign algorithm for adaptive filtering," *IEEE Trans. Inform. Theory*, vol. 41, pp. 489–495, Mar. 1995.

[6] S. Cho and V. Mathews, "Tracking analysis of sign algorithm in nonstationary environments," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2046–2057, Dec. 1990.

[7] E. Eweda, "Tracking analysis of the sign algorithm without the Gaussian constraint," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 115–122, Jan. 1998.

[8] A. H. Sayed and M. Rupp, "A time-domain feedback analysis of adaptive algorithms via the small gain theorem," *Proc. SPIE*, vol. 2563, pp. 458–469, July 1995.

[9] M. Rupp and A. H. Sayed, "A time-domain feedback analysis of filtered-error adaptive gradient algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1428–1439, Jun. 1996.

[10] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 69–72, June 1958.

[11] J. Mai and A. H. Sayed, "A feedback approach to the steady-state performance of fractionally-spaced blind adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 48, pp. 80–91, Jan. 2000.

[12] N. R. Yousef and A. H. Sayed, "A feedback analysis of the tracking performance of blind adaptive equalization algorithms," in *Proc. IEEE Conf. Decision and Control*, vol. 1, Phoenix, AZ, Dec. 1999, pp. 174–179.

[13] ——, "A unified approach to the steady-state and tracking analyzes of adaptive filters," *IEEE Trans. Signal Processing*, to be published.