On the Stability of the Least-Mean Fourth (LMF) Algorithm

Vítor H. Nascimento* and José Carlos M. Bermudez+

***Abstract*— We show that the least-mean fourth (LMF) adaptive algorithm is not mean-square stable when the regressor input is not strictly bounded (as happens, for example, if the input has a Gaussian distribution). This happens no matter how small the step-size is made. We prove this result for a slight modification of the Gaussian distribution in a length $M = 1$ filter (in order to simplify our arguments), and provide several examples of divergence when the regressor is Gaussian.**

Our results provide tools for filter designers to better understand what can happen when the LMF algorithm is used, and in which situations it might not be a good idea to use this algorithm.

I. INTRODUCTION

The least-mean fourth (LMF) algorithm was proposed almost 20 years ago [1] as an alternative to the least-mean square (LMS) algorithm. The goal was to achieve a lower steady-state misadjustment for a given speed of convergence using a different cost-function. It is not difficult to intuitively understand how this is accomplished if we compare the update laws of both algorithms:

LMS:

$$\begin{aligned} \boldsymbol{W}_2(n+1) &= \boldsymbol{W}_2(n) + \mu e_2(n)\boldsymbol{X}(n), \\ e_2(n) &= d(n) - \boldsymbol{W}_2(n)^T\boldsymbol{X}(n), \end{aligned} \tag{1}$$

LMF:

$$\begin{aligned} \boldsymbol{W}(n+1) &= \boldsymbol{W}(n) + \mu e(n)^3\boldsymbol{X}(n), \\ e(n) &= d(n) - \boldsymbol{W}(n)^T\boldsymbol{X}(n), \end{aligned} \tag{2}$$

where $\boldsymbol{W}_2(n)$ and $\boldsymbol{W}(n) \in \mathbb{R}^M$ are current estimates of a parameter (column) vector $\boldsymbol{W}_o \in \mathbb{R}^M$. $\boldsymbol{X}(n) \in \mathbb{R}^M$ is a known regressor vector, and $d(n)$ is a known scalar sequence, usually called *desired* sequence.

It is well-known [2], [3] that, if $\{d(n), \boldsymbol{X}(n)\}$ are zero-mean, jointly wide-sense stationary sequences, one can always model the relationship between $d(n)$ and $\boldsymbol{X}(n)$ as

$$d(n) = \boldsymbol{W}_o^T\boldsymbol{X}(n) + e_0(n), \tag{3}$$

*Depto de Eng. de Sistemas Eletrônicos, Escola Politécnica, Universidade de São Paulo, e-mail: vitor@lps.usp.br

+Depto de Engenharia Elétrica, Universidade Federal de Santa Catarina, e-mail: j.bermudez@ieee.org

where $e_0(n)$ is a zero-mean scalar sequence, uncorrelated with $\boldsymbol{X}(n)$ and with variance $\mathrm{E}\, e_0(n)^2 = \sigma_0^2$ ($\mathrm{E}(\cdot)$ is the statistical expectation operator). In this context, $\boldsymbol{W}_o$ is called the *Wiener solution*. The LMS estimate $\boldsymbol{W}_2(n)$ converges in the mean to $\boldsymbol{W}_o$ with a finite covariance matrix, as long as the step-size μ is small enough. It is also known that, for small μ, the LMS steady-state mean-square estimation error (MSE) is approximately given by

$$\lim_{n\to\infty} \mathrm{E}\, e_2(n)^2 \approx \sigma_0^2 + \mu\, \sigma_0^2 \frac{\mathrm{Tr}(\boldsymbol{R}_x)}{2}, \tag{4}$$

where $\boldsymbol{R}_x = \mathrm{E}\, \boldsymbol{X}(n)\boldsymbol{X}(n)^T$ is the autocorrelation matrix of $\boldsymbol{X}(n)$, and $\mathrm{Tr}(\boldsymbol{R}_x)$ is its trace.

The second term on the right-hand side of (4) is the steady-state excess MSE, which is caused by the fluctuations of $\boldsymbol{W}_2(n)$ around $\boldsymbol{W}_o$ after convergence. This term is proportional to μ.

It can also be shown that the rate of convergence of $\mathrm{E}\, e_2(n)^2$ is $1 - 2\mu\lambda_{\min}$ for small μ, where $\lambda_{\min}$ is the smallest eigenvalue of $\boldsymbol{R}_x$.

One can see that μ controls the behavior of the algorithm, and that two important goals are competing: for fast convergence, one would use a large step-size μ, but to achieve low steady-state MSE, a smaller step-size would be better. One intuitive way to understand the LMF algorithm is to consider it as a variant to LMS with a variable step-size $\bar{\mu}(n) = e(n)^2\mu$. When the error is large, adaptation is faster, when the error is small, adaptation is slower, resulting in a fast convergence with small steady-state error.

Regarding the LMF algorithm in this way also highlights its main drawback: if the error gets too large, the "equivalent step-size" $\bar{\mu}(n)$ will get large, and one would expect the algorithm to diverge. This is exactly what happens.

Recent works [4], [5], [6], [7] studied the behavior of the LMF algorithm for Gaussian noise and regressors, finding approximate mean-square stability conditions. Several other works also studied the stability of LMF. For example, [8] proves that $\boldsymbol{W}(n)$ converges to a ball around $\boldsymbol{W}_o$ when the regressor vector sequence is bounded, i.e., when there is a $B < \infty$ such that

$||\boldsymbol{X}(n)|| < B$ for all n ($||\cdot||$ is the Euclidean norm).

In this work we argue that there is always a nonzero probability of divergence in any given realization of the LMF algorithm when the entries of $\boldsymbol{X}(n)$ have a probability density function (pdf) with infinite support, i.e., there is a small (but nonzero) probability that an entry is larger than any $C > 0$. This is what happens, for example, with the Gaussian distribution.

We prove this property for a simple case, when $M = 1$ (scalar filter) and the distribution of $X(n)$ is a slight modification of the normal, with pdf given by

$$p_X(x) = \begin{cases} 0, & \text{if } |x| < \epsilon, \\ \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(|x|-\epsilon)^2}{2\sigma_x^2}}, & \text{if } |x| \geq \epsilon, \end{cases} \tag{5}$$

for $\epsilon > 0$.

This result means that the LMF algorithm is *not* mean-square stable with these near-Gaussian inputs. In other words, the steady-state mean-square error (MSE) is unbounded.

Notice that this result does not imply that *every* realization of the LMF algorithm will result in divergence. In fact, for finite time intervals, the probability of divergence on a single realization of the algorithm tends to zero as the step-size is decreased to zero, as we show in a few examples further on. We are currently working on an approximation for the probability of divergence. The purpose of this paper is only to show the important property that this probability of divergence is nonzero even for very small step-sizes.

In light of this result, we can better understand the approximations given in the literature for the MSE of the LMF algorithm. For small step-sizes, the probability of divergence is very small and the approximations in the literature are in fact computing $\mathrm{E}\{e(n)^2|$ the filter coefficients did not diverge$\}$. Thus, there is not a step-size boundary $\mu = \mu_{\max}$ above which the algorithm starts diverging. What happens is that the probability of divergence increases with μ. This property has a similarity with what happens with LMS, as explained in [9] — LMS has a range of step-sizes for which the algorithm converges with probability one, but diverges in the mean-square (MS) sense; a range for which the algorithm diverges almost always (and in the MS sense); and a range for which it converges in the MS sense. One of the main differences between LMS and LMF is that the last case only happens for the LMF algorithm when both the regressor and the noise are bounded.

In the next sections we prove our assertions for a simple situation, and provide several simulations corroborating our theoretical results.

II. A SIMPLE EXAMPLE OF INSTABILITY

Our goal here is to give a simple example showing that LMF will have a nonzero probability of divergence for a rather nice distribution of the regressor input, no matter how small we choose the (nonzero) step-size. We believe that this scalar example explains clearly what is the mechanism of divergence, so there is no need to expand the example for longer filters.

A. *Proof of instability for scalar filters*

Consider the LMF algorithm (2) applied with filter length $M = 1$ to identify a constant W_o, given a scalar independent and identically distributed (iid) sequence $X(n)$ with pdf given by (5). Assume also that $d(n)$ is given by

$$d(n) = W_o X(n),$$

i.e., there is no noise. Defining the weight estimation error

$$V(n) = W_o - W(n),$$

the LMF weight-error update equation is written

$$V(n+1) = \big(1 - \mu X(n)^4 V(n)^2\big) V(n). \tag{6}$$

We will show first that there is a value $0 < K < \infty$ such that, if for any n it holds that $|V(n)| > K$, then $\lim_{n\to\infty} |V(n)| = \infty$. Later we will show that the probability of $|V(n+1)| > K$ given $|V(n)| = \alpha$ is nonzero for all $\alpha > 0$.

Let K be such that, say,

$$\mu\epsilon^4 K^2 - 1 > 3 \Leftrightarrow K > \frac{2}{\sqrt{\mu}\epsilon^2}. \tag{7}$$

Given inequality (7) and since $|X(n)| \geq \epsilon$ by (5), it necessarily holds that

$$|1 - \mu X(n)^4 K^2| > |1 - \mu\epsilon^4 K^2| > 3. \tag{8}$$

If we now assume that $|V(n)| > K$, (6) yields

$$\begin{aligned} |V(n+1)| &= \big|1 - \mu X(n)^4 V(n)^2\big| |V(n)| > \\ &> |1 - \mu\epsilon^4 K^2| |V(n)| > 3|V(n)|, \end{aligned} \tag{9}$$

and we conclude that $|V(n)| \to \infty$ if at any time instant it happens that $|V(n)| > K$.

We complete our argument by showing that the probability of $|V(n+1)| > K$ given $|V(n)| = \alpha$ is nonzero for any $\alpha > 0$. Define, for a given $\alpha > 0$,

$$\beta(\alpha) \triangleq \frac{K}{\mu\alpha^3} + \frac{1}{\mu\alpha^2}. \tag{10}$$

Then, for $|V(n)| = \alpha$, an input $X(n)$ such that $X(n)^4 > \beta(\alpha)$ leads to

$$\mu X(n)^4 V(n)^2 - 1 > \frac{K}{|V(n)|}, \tag{11}$$

and thus

$$|V(n+1)| = |1 - \mu X(n)^4 V(n)^2||V(n)| > K. \tag{12}$$

Expressions (9) and (12) show that $|V(n)| \to \infty$ if $X(n)^4 > \beta(\alpha)$ for any given $|V(n)| = \alpha$. Thus, to prove that there is a nonzero probability of divergence, it remains to show that there is a nonzero probability that $X(n)^4 > \beta(\alpha)$, given that $|V(n)| = \alpha$.

Using (5), is follows that

$$\begin{aligned}&\Pr\{|V(n+1)| > K \mid \ |V(n)| = \alpha\}\\ &> \Pr\left\{ X(n)^4 > \beta(\alpha) \ \middle| \ \ |V(n)| = \alpha\right\}\\ &= 2\int_{\beta(\alpha)^{1/4}}^{\infty} p_X(x)dx > 0,\end{aligned}$$

where $\Pr\{A|B\}$ is the probability of ocurrence of A given B. This concludes the proof.

B. Gaussian regressors

When $X(n)$ is normal ($\epsilon = 0$ in (5)), the simple proof above does not apply. However, we now present simulations showing that the result still holds.

Assume that LMF is applied to the same situation as before, but with $\epsilon = 0$. In our simulations, we evaluated:

- The probability of divergence of LMF, measured as follows: we ran $L = 10^6$ realizations of the algorithm, starting from the same initial condition $V(0) = 1$ and with zero noise. We counted a "divergence" everytime the absolute error $|V(n)|$ became larger than 10^{100} (choosing this value in a very large range does not affect the results),
- The probability $P_>$ of $|V(1)| > |V(0)|$,
- The value $V_{1/2}$ for which the probability $\Pr\big\{|V(n+1)| > |V(n)| \mid |V(n)| = V_{1/2}\big\} = 0.5$,
- The probability $P_{>V_{1/2}}$ that $|V(1)| > V_{1/2}$, given the initial condition.

The probability of divergence was obtained experimentally. All other values can be computed as follows.

We start by computing the pdf of $V(n+1)$ given $V(n)$. From (6), it is clear that

$$\begin{aligned}&\Pr\{V(n+1) < z|V(n) = Z > 0\}\\ &\quad = \Pr\{(1 - \mu X(n)^4 Z^2)Z < z\}\end{aligned}$$

or, equivalently,

$$\begin{aligned}&\Pr\{V(n+1) < z|V(n) = Z > 0\}\\ &\quad = \Pr\left\{X(n)^4 > \frac{1 - z/Z}{\mu Z^2}\right\}\end{aligned}$$

The pdf of $X(n)^4$ is given by

$$\begin{aligned}p_{X^4}(y) &= \frac{d\Pr\{X^4 < y\}}{dy}\\ &= \frac{d\left(2\int_0^{y^{1/4}} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} dx\right)}{dy}\\ &= \frac{1}{\sqrt{8\pi}\sigma_x y^{3/4}} e^{-\frac{\sqrt{y}}{2\sigma_x^2}}, \ y \geq 0.\end{aligned} \tag{13}$$

Thus

$$\begin{aligned}&\Pr\{V(n+1) < z \mid V(n) = Z > 0\}\\ &\quad = \Pr\left\{X(n)^4 > \frac{1 - z/Z}{\mu Z^2}\right\}\\ &\quad = \int_{\frac{1-z/Z}{\mu Z^2}}^{\infty} \frac{1}{\sqrt{8\pi}\sigma_x y^{3/4}} e^{-\frac{\sqrt{y}}{2\sigma_x^2}} dy,\end{aligned} \tag{14}$$

Finally, the desired pdf is obtained by differentiating (14) with respect to z:

$$\begin{aligned}&p_{V(n+1) \ | \ V(n)}(z \mid V(n) = Z)\\ &= \frac{d\Pr\{V(n+1) < z|V(n) = Z > 0\}}{dz}\\ &= \frac{1}{\sqrt{8\pi}\sigma_x \mu^{1/4} Z^{3/4}(Z - z)^{3/4}} e^{-\frac{\sqrt{Z-z}}{2\sigma_x^2\sqrt{\mu Z^3}}}.\end{aligned} \tag{15}$$

Assuming $\sigma_x^2 = 1$, we can use (15) with $V(0) = 1$ (deterministic) to determine the probabilities

$$P_> = \Pr\{|V(1)| > |V(0)|\}$$

and

$$P_{>V_{1/2}} = \Pr\{|V(1)| > |V_{1/2}| \mid V(0) = 1\},$$

and the point $V_{1/2} > 0$ for which

$$\Pr\{|V(1)| > V(0) \mid V(0) = V_{1/2}\} = 0.5.$$

These values are given, for several choices of μ, in Table I. Column $P_>$ gives the probability of $|V(1)| > |V(0)|$ (for $|V(0)| = 1$), column $V_{1/2}$ gives the value of $|V(0)|$ for which the corresponding $P_>$ would be 0.5. Column $P_{>V_{1/2}}$ gives the probability that $|V(1)| > V_{1/2}$, given that $|V(0)| = 1$, and the last column gives the observed probability of divergence. The table also shows N_{div}, the observed number of realizations of the LMF algorithm for which $|V(n)| > 10^{100}$, as explained above.

The last column in Table I shows that the probability of divergence grows with the step-size. Even for the largest step-size in the table, the filter coefficients behave rather nicely in most realizations. Fig. 1 shows three realizations for a scalar filter ($M = 1$ coefficient), with Gaussian iid input $X(n)$ with unit variance, step-size

TABLE I

OBSERVED PROBABILITY OF DIVERGENCE, FOR $M = 1$ AND $|V(0)| = 1$. THE LAST COLUMN GIVES THE OBSERVED PROBABILITY OF DIVERGENCE, USING $L = 10^6$ REALIZATIONS OF THE FILTER. EXPLANATIONS FOR THE OTHER COLUMNS ARE GIVEN IN THE TEXT.

μ	$P_>$	$V_{1/2}$	$P_{>V_{1/2}}$	N_{div}/L
0.01	1.7×10^{-4}	31.1	5.2×10^{-14}	7×10^{-6}
0.02	1.6×10^{-3}	22.0	5.8×10^{-9}	3.0×10^{-4}
0.03	4.3×10^{-3}	17.9	5.4×10^{-7}	1.6×10^{-3}
0.04	7.8×10^{-3}	15.5	6.5×10^{-6}	4.4×10^{-3}
0.05	1.2×10^{-2}	13.9	3.3×10^{-5}	8.5×10^{-3}
0.06	1.6×10^{-2}	12.7	1.0×10^{-4}	1.4×10^{-2}
0.07	2.1×10^{-2}	11.7	2.4×10^{-4}	2.0×10^{-2}
0.08	2.5×10^{-2}	11.0	4.7×10^{-4}	2.8×10^{-2}
0.09	3.0×10^{-2}	10.4	8.0×10^{-4}	3.5×10^{-2}
0.10	3.4×10^{-2}	9.8	1.3×10^{-3}	4.3×10^{-2}
0.20	7.5×10^{-2}	7.0	1.2×10^{-2}	1.3×10^{-1}

$\mu = 0.03$ (probability of divergence of 0.16% according to Table I), and initial condition $V(0) = 1$. The figure shows two realizations where the algorithm converged, and one for which the algorithm diverged. Note that divergence does not take long to become clear. This has been verified to be a typical algorithm behavior. In addition, we note that the probability of divergence depends on the initial condition: the larger the initial error $V(0)$, the larger the probability of divergence.

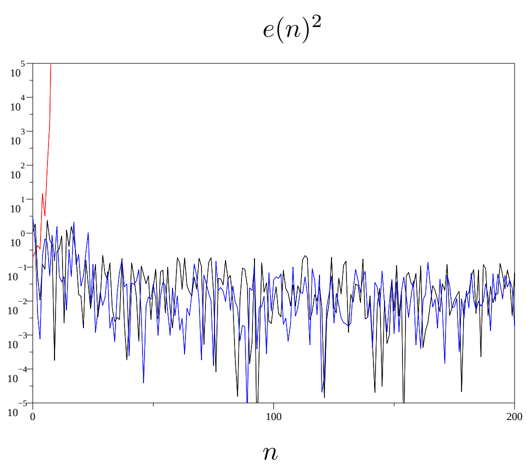

Fig. 1. Three runs of LMF with scalar regressors, true $W_o = 0$, $V(0) = W(0) = 1$, $\mu = 0.03$, $X(n) \sim N(0, 1)$.

We also provide an example of a filter with $M = 10$ coefficients, in order to show that this behavior is not restricted to scalar filters. Fig. 2 shows simulations similar to those of Fig. 1 (again we show three simulations, two converging and one diverging). The parameters used were: $\boldsymbol{X}(n)$ was an iid vector sequence with zero mean and covariance matrix equal to the identity (I), the step-size chosen was $\mu = 0.02$, and the initial condition was $\boldsymbol{V}(0) = [\,1\ 0\ \dots\ 0\,]^T$.

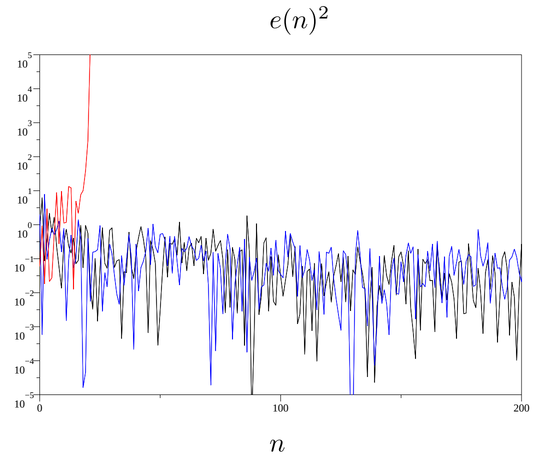

Fig. 2. Three runs of LMF with $M = 10$, true $\boldsymbol{W}_o = [\,0\ 0\ \dots\ 0\,]^T$, $\boldsymbol{V}(0) = \boldsymbol{W}(0) = [\,1\ 0\ \dots\ 0\,]^T$, $\mu = 0.02$, $\boldsymbol{X}(n)$ Gaussian with covariance equal to I.

III. CONCLUSIONS

In this paper we argued that the least-mean fourth algorithm cannot be mean-square stable when the regressor sequence is not strictly bounded. In practice (since all actual regressor sequences are bounded), this means that the algorithm is very sensitive to bursts of large noise (noise of impulsive type).

The behavior of the LMF algorithm in this respect is very different from that of LMS. If the weight error vector $\boldsymbol{V}(n)$ is taken by chance to a large value in a particular realization of the LMS algorithm, it tends to return quickly to reasonable behavior [9]. The LMF algorithm, on the other hand, may become completely unstable if the weight error vector becomes too large. This behavior is due to its cubic nonlinearity.

We are currently working on an approximation for the probability of divergence of LMF, given the initial condition, the step-size, the filter length, the noise distribution, and the covariance matrix of the regressor vector $\boldsymbol{X}(n)$. The goal is to provide designers with tools to decide whether using the LMF algorithm in a particular situation is a sensible choice, and later on, to look for the best ways to increase algorithm robustness.

REFERENCES

[1] E. Walach and B. Widrow, "The least mean fourth (LMF) adaptive algorithm and its family," *IEEE Transactions on Information Theory*, vol. IT-30, no. 2, pp. 275–283, Mar. 1984.

[2] Paulo Sérgio Ramirez Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Kluwer Academic Publishers, 2nd edition, 2002.
[3] Ali H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley-Interscience, 2003.
[4] Shin'ichi Koike, "Stability conditions for adaptive algorithms with non-quadratic error criteria," in *EUSIPCO2000*, Tampere, Finland, Sept. 2000.
[5] Shin'ichi Koike, "Analysis of the least mean fourth algorithm based on gaussian distributed tap weights," in *IEEE International Symposium on Intelligent Signal Processing and Communication Systems, ISPACS 2001*, Nashville, USA, Nov. 2001, IEEE.
[6] Pedro Inácio Hübscher and José Carlos M. Bermudez, "An improved statistical analysis of the least mean fourth (LMF) adaptive algorithm," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 664–671, Mar. 2003.
[7] Pedro I. Hübscher, Vítor H. Nascimento, and José Carlos M. Bermudez, "New results on the stability analysis of the LMF (least mean fourth) adaptive algorithm," in *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2003, vol. VI, pp. VI 369–VI 372.
[8] William A. Sethares, "Adaptive algorithms with nonlinear data and error functions," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2199–2206, Sept. 1992.
[9] Vítor H. Nascimento and Ali H. Sayed, "On the learning mechanism of adaptive filters," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1609–1625, June 2000.