Short-Time Instabilities in the LMS Algorithm

MARKUS RUPP*

Department of Electrical and Computer Engineering,
University of California, Santa Barbara, CA 93106
e-mail: markus@agni.ece.ucsb.edu

Abstract– The least mean square (LMS) algorithm is known to converge in the mean and in the mean square. This does not imply that the algorithm converges strictly at every time step k. In short-time periods the algorithm's convergence can burst up and cause severe disturbances in speech applications. As long as Gaussian processes are used to drive the filter input and the order of the filter is relatively large, the occurrence of these instabilities is very rare. However, for other statistics this does not need to be true. This contribution closes this gap in the literature by discussing potential short-time unstable behavior of the LMS algorithm. For spherically invariant random processes (SIRP), like Gaussian, Laplacian, and K_0, the probabilities for the occurrence of instability at a single time instant k are investigated.

I. INTRODUCTION

For the past few decades the convergence of the least mean square (LMS) algorithm has been analyzed, depending on the statistics, either in the mean or in the mean square. However, this does not necessarily state that the algorithm is converging at every time step for a specific step-size μ. Depending on the statistics of the input sequence there is still a nonzero probability that the algorithm will be unstable for a short-time, even if the step-size μ is very small. In speech applications, like echo or noise control, these short-time instabilities lead to bursting, severely disturbing a conversation. If the input sequence is bounded, as in the case of a uniform distributed process, then there exists a step-size that is small enough to ensure the convergence of the algorithm at every time step. However, for unbounded processes, like Gaussian or K_0, the probability of failure will be greater than zero, even if the step-size is very small.

If the algorithm is written in state space form using the weight-error vector $\epsilon(k)$, the difference of the estimated taps at time k and the optimal solution, the update equations are

$$\underline{\epsilon}(k+1) = \left(\underline{I} - \mu\underline{u}(k)\underline{u}^T(k)\right)\underline{\epsilon}(k) - \mu\underline{u}(k)r(k) \qquad (1)$$

where $\underline{I}$ is an M dimensional identity matrix, $\underline{u}(k)$ is a $(1 \times M)$ vector with the M past samples of the input sequence and $r(k)$ is additive output noise. Considering only the homogeneous equation and not taken the noise into account, the contraction mapping property of the algorithm can be applied. Although the paper does not deal with the normalized LMS algorithm, the normalized step-size

$$\alpha_k = \frac{\mu}{\underline{u}^T(k)\underline{u}(k)}$$

is used here, but only for the purpose of clarity and will be substituted back later. The squared L_2–norm of the weight-error vector $\underline{\epsilon}(k)$ is investigated next

$$\|\underline{\epsilon}(k+1)\|_2^2 = \|\underline{\epsilon}(k)\|_2^2 \left(1 + \frac{\left(\underline{\epsilon}^T(k)\underline{u}(k)\right)^2}{\|\underline{\epsilon}(k)\|_2^2\|\underline{u}(k)\|_2^2}\left(\alpha_k^2 - 2\alpha_k\right)\right). \qquad (2)$$

The term $k_{\epsilon u}(k) = \frac{\left(\underline{\epsilon}^T(k)\underline{u}(k)\right)^2}{\|\underline{\epsilon}(k)\|_2^2\|\underline{u}(k)\|_2^2}$ lies between zero and one. Thus, for $0 < \alpha_k < 2$ the update equation is a contraction mapping of the weight-error vector. Consequently, the algorithm fails at time k if $\alpha_k > 2$ depending on the term $k_{\epsilon u}(k)$. Since there is nothing known about $k_{\epsilon u}(k)$ the worst case is assumed and $k_{\epsilon u}(k)$ is set equal to one. Thus, the algorithm is unstable at time instant k if

$$\|\underline{u}(k)\|_2^2 > \frac{2}{\mu}. \qquad (3)$$

A common limit for the mean square approach is $\mu < 2/(\sigma_u^2 M)$ (see [2]). In order to obtain more convenient expressions, a normalized step-size

$$\mu' = \mu M$$

will be used throughout the paper. Thus, the probability of failure of the algorithm at time instant k will be defined as

$$P_f = P\left(\|\underline{u}(k)\|_2^2 \geq \frac{2M}{\mu'}\right). \qquad (4)$$

Since the squared L_2–norm is never negative, a very rough approximation can be given using the Tchebycheff inequality (see [3])

$$P\left(\|\underline{u}(k)\|_2^2 \geq \frac{2M}{\mu'}\right) \leq \frac{\mu'}{2}. \qquad (5)$$

Of course, this implies that $P_f \leq 1$ for $\mu' \geq 2$. If, however, exact equalities are desired the statistics of the input process have to be known. Since Gaussian statistics are very

*This work was supported by a scholarship from DAAD (German Academic Exchange Service) and the scientific division of NATO (Wissenschaftsausschuß der NATO).

frequently used in analyzing the LMS algorithm, P_f is discussed for these statistics in the next section. The approach can be extended to the large class of spherically invariant random processes (SIRP) that closely resemble speech sequences. This is presented in the third section. Examples for K_0 and Laplace densities are given.

II. Gaussian input sequence

If it is assumed that the input sequence is uncorrelated Gaussian with unit variance, the probability P_f can be written in terms of the squared radius $s = r^2$

$$P_f(x) = P(s > x)\,. \tag{6}$$

The radial density of an M dimensional Gaussian process can be given by

$$p_s(s) = \frac{s^{\frac{M}{2}-1}}{2^{\frac{M}{2}}\Gamma(\frac{M}{2})} e^{-s/2} U(s) \tag{7}$$

where $U(s)$ is the unit step function. Thus, the desired probability can be calculated:

$$\begin{aligned} P_f(x) &= \frac{1}{2^{\frac{M}{2}}\Gamma(\frac{M}{2})}\int_x^\infty s^{\frac{M}{2}-1}e^{-s/2}ds \\ &= 1 - \frac{1}{2^{\frac{M}{2}}\Gamma(\frac{M}{2})}\int_0^x s^{\frac{M}{2}-1}e^{-s/2}ds \\ &= 1 - \Gamma_{\text{inc}}\left(\frac{x}{2}, \frac{M}{2}\right)\,. \end{aligned} \tag{8}$$

The function $\Gamma_{\text{inc}}(x, a)$ is known as the incomplete gamma function [1] and can be computed using Matlab. Figure 1 depicts the failure probability for various choices of the filter order M over the step-size μ'. It can be seen that for $\mu' \le 1$ P_f is very small, whereas it increases drastically for $\mu' > 2$. The semilog plot on the right side emphasizes the behavior for small μ'. This can be calculated explicitly. For $\frac{M}{2}$ being an integer value the incomplete gamma function can be given as a series (see [1])

$$1 - \Gamma_{\text{inc}}\left(\frac{M}{\mu'}, \frac{M}{2}\right) = \sum_{k=0}^{\frac{M}{2}-1} \frac{\left(\frac{M}{\mu'}\right)^k}{k!} e^{-\frac{M}{\mu'}}\,. \tag{9}$$

As long as $\frac{M}{\mu'} \gg \frac{M}{2}$ the sum can be approximated by its largest value

$$1 - \Gamma_{\text{inc}}\left(\frac{M}{\mu'}, \frac{M}{2}\right) \approx \frac{\left(\frac{M}{\mu'}\right)^{\frac{M}{2}-1}}{(\frac{M}{2}-1)!} e^{-\frac{M}{\mu'}}\,. \tag{10}$$

Thus, for small step-sizes the probability of failure can be given explicitly. If the filter order M is fixed, then the failure probability behaves essentially like $e^{-\frac{M}{\mu'}}$, or in the semilog plot like $-\frac{M}{\mu'}$, i.e. if the normalized step-size μ' is doubled the failure rate is increased to half of the exponent.

For large step-sizes μ' the behavior can easily be detected from (9):

$$P_f \approx e^{-\frac{M}{\mu'}} \qquad ;\text{for } \mu' > M. \tag{11}$$

Although this is not of practical importance, it gives an indication of the behavior.

Since the limit is usually given for $\mu' = 2$, P_f for this step-size is of special interest and needs further calculation. As can be observed in Figure 1 the value for $\mu' = 2$ tends to 0.5 if M increases. It is proven next that indeed P_f tends to 0.5 for $\mu' = 2$ if M tends to infinity.

$$\lim_{M\to\infty} 1 - \Gamma_{\text{inc}}\left(\frac{M}{2}, \frac{M}{2}\right) = \frac{1}{2}\,. \tag{12}$$

The filter order M is assumed to be even. Thus, the incomplete gamma function can again be written as a series

$$1 - \Gamma_{\text{inc}}\left(\frac{M}{2}, \frac{M}{2}\right) = \sum_{k=0}^{\frac{M}{2}-1} \frac{\left(\frac{M}{2}\right)^k}{k!} e^{-\frac{M}{2}}\,. \tag{13}$$

It is recalled that Bernoulli trials can be written as Poisson trials if the number of trials n tends to infinity, the probability p tends to zero, but the product np tends to a fixed number a (Poisson theorem [3])

$$\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow{n\to\infty} \frac{a^k}{k!} e^{-a} \tag{14}$$

and for $np \gg 1, n \gg 1$ [3]

$$\sum_{k=0}^{k_2} \binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{1}{2} + \text{erf}\left(\frac{k_2 - np}{\sqrt{np(1-p)}}\right)\,. \tag{15}$$

The following is identified: $a = \frac{M}{2}$ and $k_2 = \frac{M}{2} - 1$. Hence

$$\begin{aligned} 1 - \lim_{M\to\infty} \Gamma_{\text{inc}}\left(\frac{M}{2}, \frac{M}{2}\right) &= \lim_{M\to\infty} \sum_{k=0}^{\frac{M}{2}-1} \frac{\left(\frac{M}{2}\right)^k}{k!} e^{-\frac{M}{2}} \\ &= \lim_{M\to\infty} \frac{1}{2} + \text{erf}\left(\frac{-1}{\sqrt{\frac{M}{2}}}\right) \\ &= 0.5 \qquad \text{q.e.d.} \end{aligned} \tag{16}$$

Although $\mu' = 2$ is typically given as stability limit [2], it is shown here that the probability for short-time instability at this limit equals at least 0.5 even for large filter orders. The value for maximal convergence speed $\mu' = 1$ has a very low probability of failure and this probability vanishes if M tends to infinity. However, if other than Gaussian statistics are applied, the situation can become much more drastic.

III. SIRP input sequence

If it is assumed that all high order joint density functions are only dependent on the radius and not on any angles, spherically invariant processes are obtained. For a detailed description of their numerous properties the reader is referred to [4, 5]. A very suitable description of these pdfs can be given in terms of the Meijer's G-functions [1]. They are defined by a Mellin-Barnes integral

$$G_{p\,q}^{m\,n}\left(z \,\middle|\, {}^{a_p}_{b_q}\right) = \frac{1}{2\pi j} \int_C z^s \frac{\prod_{l=1}^m \Gamma(b_l - s)\prod_{l=1}^n \Gamma(1 - a_l + s)}{\prod_{l=m+1}^q \Gamma(1 - b_l + s)\prod_{l=n+1}^p \Gamma(a_l - s)} ds \tag{17}$$

where the two parameter sets a_p, b_q are divided into two subsets. In the case where only simple poles appear, the integral can be given as a series with hypergeometric functions and computed numerically. Applications of G-functions to LMS, NLMS and Delayed Update LMS (DLMS) algorithms can be found in [6, 7]. Here, only two types of G-functions are under consideration: G_{01}^{10} and G_{02}^{20} with one and two parameters, respectively, which are known very well. Both of them can be described explicitly

$$\begin{aligned} G_{01}^{10}\left(x \,\middle|\, _{b_1}\right) &= x^{b_1} e^{-x} \\ G_{02}^{20}\left(x \,\middle|\, _{b_1\, b_2}\right) &= 2x^{\frac{b_1+b_2}{2}} K_{b_1-b_2}\left(2\sqrt{x}\right) \end{aligned}$$

where $K_n(x)$ is the modified Bessel- or McDonald function of the second kind for order n (see [1]). Since investigations with these two expressions become very difficult, G-functions will be used. They have the advantage that almost every linear operation is simply a change in the parameter sets. As shown in [4, 5] pdfs of bandlimited speech signals can be described by chosing the two parameters b_1 and b_2. A suitable form for pdfs is

$$p_x(x) = A G_{pq}^{mn}\left(\lambda x^2 \,\middle|\, {}^{a_p}_{b_q}\right) \tag{18}$$

where A and λ are used to normalize the function in order to be a valid pdf. A and λ can be calculated by knowing the parameters a_p, b_q

$$\begin{aligned} \lambda &= (-1)^{n+m-q} \frac{\prod_{l=1}^{q}\left(b_l + \frac{1}{2}\right)}{\prod_{l=1}^{p}\left(a_l + \frac{1}{2}\right)}, \\ A &= \lambda^{\frac{1}{2}} \frac{\prod_{l=m+1}^{q}\Gamma\left(\frac{1}{2} - b_l\right)\prod_{l=n+1}^{p}\Gamma\left(\frac{1}{2} + a_l\right)}{\prod_{l=1}^{m}\Gamma\left(\frac{1}{2} + b_l\right)\prod_{l=1}^{n}\Gamma\left(\frac{1}{2} - a_l\right)}. \end{aligned}$$

The function $p_{r_M}(r)$ describes the density of the L_2-norm of the $(1 \times M)$ vector $\underline{u}(k)$ and can be given explicitly

$$p_{r_M}(r) = 2A\frac{\sqrt{\pi}}{\Gamma\left(\frac{M}{2}\right)} G_{p+1\,q+1}^{m+1\,n}\left(\lambda r^2 \,\middle|\, {}^{a_p,0}_{\frac{M-1}{2},b_q}\right). \tag{19}$$

Hence, the probability of failure can be written as

$$\begin{aligned} P_f(x) &= 1 - \int_0^{\sqrt{x}} p_{r_M}(r)dr \\ &= 1 - 2A\frac{\sqrt{\pi}}{\Gamma\left(\frac{M}{2}\right)} \int_0^{\sqrt{x}} G_{p+1\,q+1}^{m+1\,n}\left(\lambda r^2 \,\middle|\, {}^{a_p,0}_{\frac{M-1}{2},b_q}\right) dr \\ &= 1 - A\frac{\sqrt{\pi\lambda}}{\Gamma\left(\frac{M}{2}\right)} x G_{p+2\,q+2}^{m+1\,n+1}\left(\lambda x \,\middle|\, {}^{0,a_p-\frac{1}{2},-\frac{1}{2}}_{\frac{M}{2}-1,b_q-\frac{1}{2},-1}\right) \end{aligned} \tag{20}$$

If $x = \frac{2M}{\mu'}$, the failure probability can be given for the G_{02}^{20} function

$$P_f = 1 - A\frac{\sqrt{\pi\lambda}}{\Gamma\left(\frac{M}{2}\right)}\frac{2M}{\mu'} G_{2\,4}^{3\,1}\left(\frac{2M\lambda}{\mu'} \,\middle|\, {}^{0,-\frac{1}{2}}_{\frac{M}{2}-1,b_1-\frac{1}{2},b_2-\frac{1}{2},-1}\right). \tag{21}$$

Figure 2 presents the failure behavior for a Laplace density. This can be described by a G_{02}^{20} function with parameters $b_1 = 0, b_2 = 0.5$. The function does not become singular at zero. A more extreme behavior can be expected from the K_0-density that is obtained for $b_1 = b_2 = 0$. For both densities the variation of the curves is not very large if different filter orders M are applied. The step-sizes for fastest convergence have been calculated in [6]. They are $\mu' \approx 1/2$ and $1/3$ for the Laplacian and the K_0-density, respectively. For these values, however, the failure rate is around one percent. The higher the filter order M, the smaller the failure probability. For larger values of the normalized step-size μ' the behavior changes. Even for large filter orders M the failure probabilities for $\mu' = 2$ are lower than 0.5, and the higher the filter order is, the higher the failure rate, P_f. In comparison to a Gaussian signal the probability of a failure has increased drastically and even for small step-sizes the probability of failure is relatively high. Thus, using speech signals, either very small step-sizes should be used resulting in slower convergence, or the usage of the LMS algorithm should be avoided. The projection LMS algorithm (NLMS) is then to prefer, since because of its contraction mapping property the algorithm can guarantee convergence at every time step k, and thus a much better behavior results.

IV. CONCLUSIONS

In this contribution it has been shown that although the LMS algorithm converges in mean and mean-square, short-time periods exist in which the algorithm does not converge, even if the step-size is very small. The failure probability of the algorithm was calculated for several spherically invariant processes. For Gaussian processes the failure rates are only high if the normalized step-size exceeds the limit two. For step-sizes lower or equal to the optimal value the effect vanishes for relatively large filter orders M. However, if other spherically invariant processes, like Laplace and K_0 processes, are applied, the failure probabilities are still relatively large, even if small step-sizes are used. Thus, speech signals can cause severe problems to the LMS algorithm.

References

1. I.S. Gradsteyn, I.M. Ryzhik, *Table of Integrals*, Series and Products. New York: Academic, 1980.
2. S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1986.
3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw Hill, 1987.
4. H. Brehm, "Description of spherically invariant random processes by means of G-functions," *Springer lecture Notes*, vol. 969, pp. 39–73, 1982.
5. H. Brehm, W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing*, vol. 9, pp. 119–141, 1987.
6. M. Rupp, "The behavior of LMS and NLMS algorithms in the presence of spherically invariant processes," *IEEE Transactions on Signal Processing*, vol. SP-41, no. 3, pp. 1149-1160, Mar. 1993.
7. M. Rupp, R. Frenzel "The behavior of LMS and NLMS algorithms with delayed coefficient update in the presence of spherically invariant processes," *IEEE Trans. on Sig. Proc.*, vol. SP-42, no. 3, pp. 668-672, Mar. 1994.

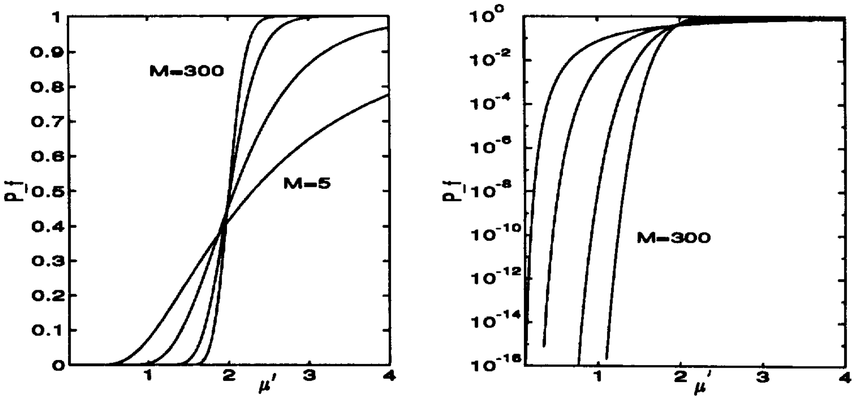

Figure 1: Failure probability P_f over μ' for a Gaussian process with filter orders $M = 5, 20, 100, 300$

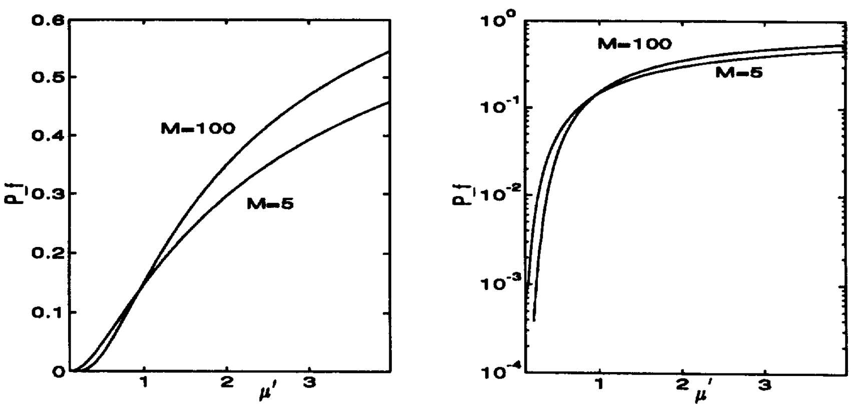

Figure 2: Failure probability P_f over μ' for a Laplace process with filter orders $M = 5, 100$

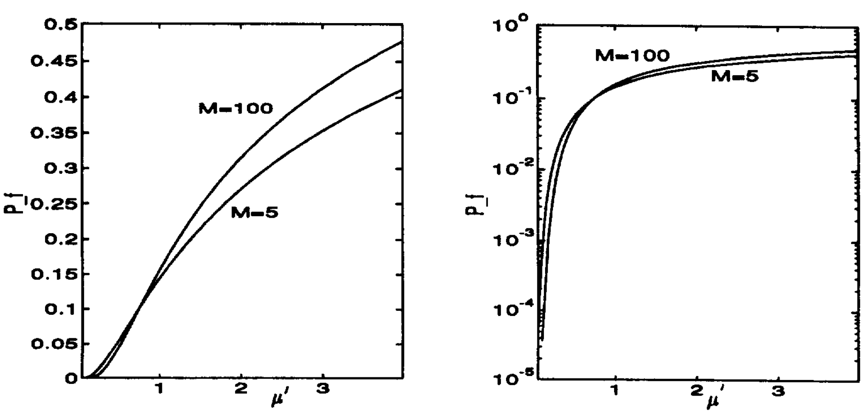

Figure 3: Failure probability P_f over μ' for a K_0 process with filter orders $M = 5, 100$