LMS TRACKING BEHAVIOR UNDER PERIODICALLY CHANGING SYSTEMS

Markus Rupp
Bell-Labs, Lucent Technologies, Wireless Research Lab.
791 Holmdel-Keyport Rd., Holmdel, NJ 07733-0400, USA
Fax: (732) 888 7074, Tel: (732) 888 7104
`e-mail:rupp@lucent.com`

ABSTRACT

The tracking behavior of Least Mean Squares (LMS) and Recursive Least Squares (RLS) algorithms have been under considerations for several years. For system identification problems it is usually assumed that the system under consideration is of FIR type with coefficients that are statistically independent random processes with identical distribution (IID), a somewhat artificial assumption that can hardly be found in existing systems. In this article it is assumed that the system is periodically changing over time. This is an important case that occurs in every wireless system where the carrier frequency provided by the transmitter suffers an offset compared to the local oscillator of the receiver. The article presents an analysis of the algorithm tracking behavior as well as algorithms to estimate the frequency offset so that this particular corruption can be removed.

1 Introduction

Tracking analysis of adaptive filters has been of interest for more than a decade. Papers by Eleftheriou and Falconer [1], Marcos and Macchi [2], and Hajivandi and Gardner[3] explain the behavior for LMS and RLS algorithms under certain assumptions (see also Haykin [4]). Hereby an FIR system for identification is assumed that is modeled by a filter weights as random processes with independent behavior but identical distribution, thus i.i.d. This assumption is very heuristic and not justified by any measurements. Therefore, the question remains how closely this assumption resembles "real-world systems".

The investigation in this paper is driven by an application as it typically occurs in cordless or wireless portable phones. Due to the up-mixing of the signals to higher frequencies and finally the down-conversion, a frequency offset is introduced to the received sequence. This offset is due to the crystal inaccuracy and can be up to several kHz for a 900MHz carrier (IS136, see [5]). The received sequence $d(k)$ can now be described as the convolution of the transmitted sequence $u(k)$ and the channel $\mathbf{w}_o$ plus additive noise $v(k)$:

$$d(k) = \mathbf{u}_k \mathbf{w}_o e^{j\Omega k} + v(k)\,. \tag{1}$$

The transmitted symbols are combined in a row vector

$$\mathbf{u}_k = [u(k), u(k-1), .., u(k-M+1)]$$

, M being the order of the channel filter. The frequency offset component $e^{j\Omega k}$ appears as multiplicative component and can be combined with the unknown channel to a periodically changing system. The quest for the LMS algorithm, given by

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu(d(k) - \mathbf{u}_k\mathbf{w}_k)\mathbf{u}_k^* \tag{2}$$

is thus to track a periodically changing channel, a movement that can be much faster than that caused by fading.

Often filter banks are used to roughly estimate this offset and a de-rotation technique is used to lessen it to a certain extent. However, if channel tracking is required even small frequency offsets are very harmful when running the LMS algorithm. Thus, even with filter banks relatively small frequencies in the range [-450Hz,450Hz] based on 900MHz carrier frequency might occur. The paper will mainly deal with relatively small frequencies $\Omega \ll 1$ compared to the sampling rate, since only this is motivated by the application.

2 First Order Analysis

Since the paper deals with periodically changing systems of the form $\mathbf{w}_o e^{j\Omega k}$, the corresponding parameter error vector is given by

$$\tilde{\mathbf{w}}_k \stackrel{\Delta}{=} \mathbf{w}_o e^{j\Omega k} - \mathbf{w}_k. \tag{3}$$

The update equation (2) can be rewritten into state-space form for which

$$\tilde{\mathbf{w}}_{k+1} = (\mathbf{I} - \mu\mathbf{u}_k^*\mathbf{u}_k)\tilde{\mathbf{w}}_k - \mu v(k)\mathbf{u}_k^* - \mathbf{w}_o e^{j\Omega k}(1 - e^{j\Omega}) \tag{4}$$

is obtained. Assuming the widely used independence assumption, the expectation of the parameter error vector can be computed

$$E[\tilde{\mathbf{w}}_{k+1}] = (\mathbf{I} - \mu\mathbf{R})E[\tilde{\mathbf{w}}_k] - \mathbf{w}_o e^{j\Omega k}(1 - e^{j\Omega}) \tag{5}$$

Since the tracking behavior is of interest, initial conditions are neglected and the steady-state of (5) is solved. Assume a solution of the form

$$E[\tilde{\mathbf{w}}_k] = \mathbf{a} e^{j\Omega k} \tag{6}$$

and the following steady-state solution for $\mathbf{a}$ can be obtained:

$$\mathbf{a} = \left(\mathbf{I} - \frac{\mu}{1-e^{j\Omega}}\mathbf{R}\right)^{-1} \mathbf{w}_o \,. \tag{7}$$

Thus, the expected value of the parameter error vector becomes time varying as well, or, equivalently, the expected value of the estimated channel parameter

$$\lim_{k\to\infty} E[\mathbf{w}_k] = \left(\mathbf{I} - \left[\mathbf{I} - \frac{\mu}{1-e^{j\Omega}}\mathbf{R}\right]^{-1}\right) \mathbf{w}_o e^{j\Omega\, k} \,. \tag{8}$$

In brief, the estimate resulting from the LMS algorithm is lagging behind the frequency by a certain angle which is proportional to the frequency offset. The larger the frequency offset, or the smaller the step-size μ, the larger the angle.

2.1 Experiment

The following experiment shows the accuracy of the result (8). The experiment runs the algorithm for $N = 8000$ samples and 30dB SNR with step-sizes $\mu = [0.03, 0.1, 0.2]$ and frequencies $\Omega = [0.0001, 0.001, 0.01]$, driven by a white random sequence of constant modulus ($\mathbf{R} = \mathbf{I}$). The quality measure is given by

$$\begin{aligned} E[|\mathbf{w}_N - \mathbf{w}_o e^{j\Omega N}|^2] &\approx |E[\mathbf{w}_N - \mathbf{w}_o e^{j\Omega}]|^2 \\ &= \frac{|1-e^{j\Omega}|^2}{|1-\mu-e^{j\Omega}|^2}|\mathbf{w}_o|^2 \end{aligned} \tag{9}$$

which can be further accurately approximated by

$$E[|\mathbf{w}_N - \mathbf{w}_o e^{j\Omega N}|^2] \approx \frac{\Omega^2}{\mu^2}|\mathbf{w}_o|^2 \tag{10}$$

as long as the frequency is small and the step-size not too small ($\cos(\Omega) \gg 1-\mu$). The system was chosen to $\mathbf{w}_o = [1, 10, 1]$. The following Table 1 shows the results of the nine experiments (three step-sizes $\times$ three frequencies). The values given in the table are relative errors when compared the results of the algorithm with the predicted values from (9). Only very small errors occur validating the accuracy of the result.

μ / Ω	0.0001	0.001	0.01
0.03	0	0.028	0.02
0.1	0.11	0.11	0.01
0.2	0.13	0.08	0.017

Table 1: Relative errors of experimental results versus (9).

3 Frequency Offset Estimation

After understanding the basic properties of LMS under periodic systems, it is now possible to introduce frequency offset estimates derived from observable terms. Two methods are proposed in the following.

3.1 Method 1

Since the estimate of the system is lagging behind the true solution but essentially following the movement with the correct frequency, two successive estimates can be taken to get rid of the time varying part. For the noiseless case when driven with a white random process one obtains

$$\begin{aligned} E[\mathbf{w}_k^H \mathbf{w}_{k+1}] &\approx E[\mathbf{w}_k^H]E[\mathbf{w}_{k+1}] \\ &= |\mathbf{w}_o|^2 \left|1 - \frac{1-e^{j\Omega}}{1-\mu-e^{j\Omega}}\right|^2 e^{j\Omega} \,. \end{aligned}$$

Only the last part is a complex number containing the desired information. The first estimate of the frequency offset is thus given by

$$\hat{\Omega}_1(k) = \angle(\mathbf{w}_k^H \mathbf{w}_{k+1}) \,. \tag{11}$$

A similar method has been reported in [6], however by applying a least-squares method only on the sync word of each slot. The considerations here show that even LMS is capable of estimating the frequency offset in the same way without loosing any accuracy.

3.2 Method 2

An even simpler method in terms of complexity is given by the following estimate

$$\hat{\Omega}_2(k) = \angle(d(k)^*(\mathbf{u}_k \mathbf{w}_k)) \,, \tag{12}$$

i.e., two values that need to be computed anyway: $\mathbf{u}_k\mathbf{w}_k$ is the a-priori estimate for the received sequence $d(k)$. Therefore, the method only requires an additional complex multiplication each iteration. The expectation of $d(k)^*(\mathbf{u}_k\mathbf{w}_k)$ given by

$$E[d(k)^*(\mathbf{u}_k\mathbf{w}_k)] = |\mathbf{u}_k\mathbf{w}_o|^2 \left(1 - \frac{1-e^{j\Omega}}{1-\mu-e^{j\Omega}}\right) \,. \tag{13}$$

The evaluation of the angle of the second term leads to the desired information, however, not as immediately as in the first method. For small Ω and not too small μ, an approximation results in

$$E[\hat{\Omega}_2(k)] \approx \frac{\Omega}{\mu} \,. \tag{14}$$

Figure 1 shows the phase of (13) as well as the approximation (14). The frequency is based on 24300Hz sampling rate and a step-size of $\mu = 0.25$, an example that is typical for IS-136 based TDMA cellular phones. For the first 200Hz, the approximation is very precise and even up to 600Hz the accuracy is reasonable.

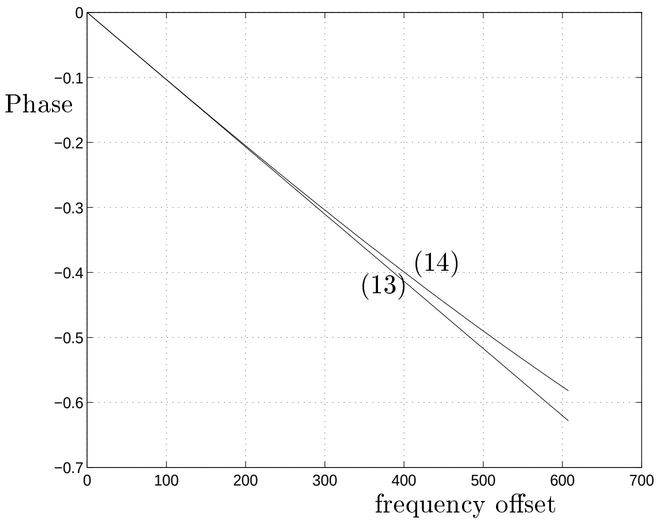

Figure 1: *Estimated angle (13) over frequency offset and corresponding approximation (14).*

3.3 Practical Issues for Implementation

Both methods described above can be implemented in various ways. Both have in common that a complex value is estimated first and in a second step its angle. Accuracy is usually improved by time-averaging the observations. The question occurs if the averaging should be done on the original data and then the angle computation, a nonlinear mapping, should follow or vice-versa. This is a very general problem in estimation and will be treated in the following. It will be assumed that a set of observed values $x(k)$ is given by

$$x(k) = x + v(k) \quad \text{for } k = 1..N\,. \tag{15}$$

It is well-known that averaging over N observations improves the quality of the estimate. If afterwards a nonlinear mapping $f(\cdot)$ is applied, the relations might be different. Two methods will therefore be investigated:

- **Method 1:** The observations are averaged first and then the nonlinear function is applied:

$$y_1(k) = f\left(\frac{1}{N}\sum_{i=1}^{N} x(k)\right) \tag{16}$$

- **Method 2:** The nonlinear mapping is applied at every observation and the averaging is done afterwards:

$$y_2(k) = \frac{1}{N}\sum_{i=1}^{N} f\left(x(k)\right)\,. \tag{17}$$

It is now assumed that the noise component $v(k)$ is small compared to the value x of interest. Then the function f can be given as

$$f(x+v(k)) = f(x)+f'(x)v(k)+\frac{f''(x)}{2}v^2(k)+O(v^3(k))\,. \tag{18}$$

Since we assume the noise to be small, the last term can be omitted. The question now is if $y_1(k)$ is a better estimate than $y_2(k)$. For the first method we obtain

$$\begin{aligned} y_1(k) &= f\left[\frac{1}{N}\sum_{i=1}^{N} x(k)\right] = f\left[x+\frac{1}{N}\sum_{i=1}^{N} v(k)\right] \\ &= f(x)+f'(x)\frac{1}{N}\sum_{i=1}^{N} v(k) \\ &+ \frac{f''(x)}{2}\left(\frac{1}{N}\sum_{i=1}^{N} v(k)\right)^2\,. \end{aligned} \tag{19}$$

whereas the second method delivers

$$\begin{aligned} y_2(k) &= \frac{1}{N}\sum_{i=1}^{N} f(x(k)) \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[f(x)+f'(x)v(k)+\frac{f''(x)}{2}v^2(k)\right] \\ &= f(x)+f'(x)\frac{1}{N}\sum_{i=1}^{N} v(k) \\ &+ \frac{f''(x)}{2}\frac{1}{N}\sum_{i=1}^{N} v^2(k)\,. \end{aligned} \tag{20}$$

Comparing (19) and (20) shows that both expressions only defer in the third, quadratic term. Because of the Cauchy-Schwarz inequality it can be shown that

$$\left(\frac{1}{N}\sum_{i=1}^{N} v(k)\right)^2 \le \frac{1}{N}\sum_{i=1}^{N} v^2(k)\,,$$

in other words, method 1 leads to smaller errors than method 2. On the other hand, the first method requires only one nonlinear function after N observations, whereas method 2 requires N of them. From a point of DSP implementation method 1 is not only of higher accuracy but also of much lower complexity.

3.4 Experimental Results

Both methods have the advantage that they are easy to implement in particular under fixed point conditions. For low frequencies $\Omega \ll 1$, it is not necessary to compute the inverse tangent function but using the argument instead leads to accurate results. The following experiment is run under IS-136 conditions, i.e., for cellular TDMA phones at 900MHz carrier frequency. We assume a severe channel corruption $\mathbf{w}_o = [0.1, 1.0, 1.0]$

and the noise variance $\sigma_v^2 = 0.01$, which leads to about 23dB SNR. The information is transmitted in slots of 162 symbols which add up to 176 if the adjacent sync word is augmented. A sample frequency of 24300Hz maps the normalized frequencies $\Omega = 0.001, 0.01, 0.1$ in approximately 4Hz,40Hz,400Hz, respectively. Further flat fading with a Doppler speed of 200Km/h is introduced during the transmission. 100 slots are transmitted and both methods applied to each slot of data beginning with the eleventh symbol so that the initial channel estimate is of less importance. The step-size is chosen to $\mu = 0.24$. The estimates are obtained by simply averaging the complex values $\mathbf{w}_k^H \mathbf{w}_{k+1}$ and $d(k)^* \mathbf{u}_k \mathbf{w}_k$, respectively. For the three frequencies the 100 run-estimates were: [0.0013,0.0102,0.0971] for method 1 and [0.0041,0.0331,0.334] for method 2. The relative accuracy, given in terms of standard deviation relative to the mean is given by

Frequency Ω	0.001	0.01	0.1
Method 1	2.27	0.35	0.044
Method 2	2.37	0.35	0.078

Table 2: Relative error for both frequency offset methods.

Thus, method 1 shows generally better behavior than method 2 although requires higher computational complexity.

Further experiments (in the context of IS-136 standards) in an existing Viterbi-MLSE algorithm with LMS for channel tracking showed very robust behavior even for method 2, as long the frequency deviation was less than 200Hz at 24300Hz sampling frequency. Even at 15dB SNR the relative estimation errors for the frequency were smaller than 1%. Since the frequencies were so small compared to the sampling frequency that the nonlinear function for mapping the complex values to their angles could completely be omitted and replace by the argument. This further reduced complexity to $N = 162$ complex MAC operations and one division.

4 The Multi-Frequency Case

Until now only the case of one frequency in the movement of the unknown system has been considered since its application is the undesired frequency shift in wireless systems. Now, it will be assumed that the system to identify is given by

$$\mathbf{w}(k) = \sum_{i=0}^{N-1} \mathbf{w}_i \, e^{j\Omega_i k} \tag{21}$$

Following the same method as before, the following equation is obtained

$$\begin{aligned} \tilde{\mathbf{w}}_{k+1} &= \left[\mathbf{I} - \mu \mathbf{u}_k \mathbf{u}_k^H\right] \tilde{\mathbf{w}}_{k-1} \qquad (22) \\ &- \sum_{i=0}^{N-1} \mathbf{w}_i \, e^{j\Omega_i k} (e^{j\Omega_i} - 1) - \mu \mathbf{u}_k v(k) \, . \end{aligned}$$

Computing the expectation with respect to the driving process $\mathbf{u}_k$ and the noise $v(k)$, a solution for tracking (steady-state solution) can be found as

$$\mathbf{a}(k) = \sum_{i=0}^{N-1} \mathbf{a}_i \, e^{j\Omega_i k} \, . \tag{23}$$

Comparing the terms with each frequency Ω_i the components $\mathbf{a}_i$ become

$$\mathbf{a}_i = \left[\mathbf{I} - \frac{\mu}{1 - e^{j\Omega_i}} \mathbf{R}\right]^{-1} \mathbf{w}_i \, . \tag{24}$$

In order to obtain the minimum error for every component the matrix term has to be minimized. Minimizing the matrix term can be done by diagonalizing it first and minimize for every eigenvalue. Thus, minimizing the tracking error $\|\mathbf{a}_i - \mathbf{w}_i\|^2$ is equivalent to the requirement

$$\min_{\mu} \frac{\mu^2 \lambda_l^2}{|e^{j\Omega_i} - 1 + \mu \lambda_l|^2} \, . \tag{25}$$

The optimal step-size is thus obtained for

$$\mu_{opt} = \frac{2}{\lambda_l} \tag{26}$$

the same result as for standard LMS without frequency offset. The various frequency offset components show no influence on the optimal step-size.

References

[1] E. Eleftheriou, D.D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoustics, Speech and Signal Proc.*, vol. ASSP-34, no. 5, pp. 1097-1110, Oct. 1986.

[2] S. Marcos, O. Macchi, "Tracking capability of the least mean square algorithm: application to an asynchronous echo canceller," *IEEE Trans. Acoustics, Speech and Signal Proc.*, vol. ASSP-35, no. 11, pp. 1570-1578, Nov. 1987.

[3] M. Hajivandi, W.A. Gardner, "Measures of tracking performance for the LMS algorithm," *IEEE Trans. Acoustics, Speech and Signal Proc.*, vol. ASSP-38, no. 11, pp. 1953-1958, Nov. 1990.

[4] Simon Haykin, *Adaptive Filter Theory*, 3. edition, Prentice Hall 1996.

[5] *TIA/EIA/Interim Standard IS 136,* Telecommunication Association, Dec. 1994.

[6] A. Bahai, M. Saraf, "A frequency offset estimation technique for nonstationary channels," *Proc. of ICASSP 97*, pp. 3897-3900, April 1997.