PROPERTIES OF THE *KURTOSIS* PERFORMANCE SURFACE IN LINEAR ESTIMATION: APPLICATION TO ADAPTIVE FILTERING

Pedro I. Hübscher

INPE - National Institute for Space Research
São José dos Campos - SP - Brazil
pedro@lit.inpe.br

*José C. M. Bermudez**

UFSC - Federal University of Santa Catarina
Department of Electrical Engineering
Florianópolis - SC - Brazil
j.bermudez@ieee.org

ABSTRACT

This paper presents an analysis of the *kurtosis* performance surface as applied to linear estimation. The analysis concentrates on a modified *kurtosis* (MK) function used in implementations of the Least Mean Kurtosis (LMK) adaptive algorithm. The MK function is shown to make the LMK algorithm applicable even for Gaussian inputs. The minimum of the MK function is derived and shown to be unique and to correspond to the Wiener solution of the mean square error (MSE) estimation problem. A quantitative comparison of the MSE and MK functions explains why the LMK adaptive algorithm is faster than MSE-based algorithms during the initial learning phase, becoming slower as it approaches steady-state.

1. INTRODUCTION

Linear estimation is applied to a wide range of adaptive signal processing and adaptive control problems [1, 2]. The design of optimal adaptive systems requires detailed knowledge of both the underlying theoretical problem and the properties of the adaptive algorithm employed. This knowledge is usually acquired from the analysis of the system behavior, including the derivation of analytical models to predict the performance of the adaptive algorithm when applied to the system.

The analysis of the adaptive algorithm includes the prediction of the adaptive weight vector behavior during the learning period and in steady-state. Steady-state efficiency is then studied by comparing the mean converged weight vector with the stationary points of the performance surface. Hence, evaluation of the adaptive algorithm requires the knowledge of the performance surface properties. In addition, adaptive algorithms can be based on different performance surfaces. Thus, addressing the relative performances of two such algorithms requires the understanding of the relationships between the two surfaces.

The performance surface most employed in linear estimation is the mean square error (MSE). Usually, the MSE presents a second order dependence on the adaptive filter weights, has one global minimum and is mathematically manageable. Recent results, however, have shown that cost functions based on higher order moments (larger than 2) of the estimation error, can lead to adaptive algorithms that outperform MSE-based algorithms in important situations. One of these cost functions is based on the *kurtosis* of the estimation error [2]. Recursive minimization of this cost function using a stochastic approximation for its gradient leads to the Least Mean *Kurtosis* (LMK) adaptive algorithm [3, 4].

The LMK algorithm seeks to minimize an approximation for the negative of the error signal *kurtosis*. It belongs to the family of stochastic gradient algorithms [4, 5]. The *kurtosis* is related to the fourth order cumulant of the error signal. Cumulants of order greater than two are equal to zero for zero-mean Gaussian processes. Thus, a true *kurtosis*-based cost function would make estimation ineffective for Gaussian inputs. However, practical implementation issues require approximations on the true *kurtosis* expression. These approximations, which can be recursive [4] or non-recursive [3], introduce changes in the true *kurtosis* that end up improving the performance of the resulting algorithm, when compared to MSE-based algorithms, even for Gaussian signals [3]. In fact, previous results show that the LMK algorithm can outperform the LMS algorithm for Gaussian inputs and several noise distributions [3, 4, 6], including the Gaussian. This superior transient performance when compared to LMS has raised considerable attention to the LMK algorithm. However, little is known about its peformance surface, either on its own right or as compared to the MSE surface.

This paper derives important properties of the *kurtosis* performance surface. The expression for the modified cost function used to implement the LMK algorithm is derived as a function of the input and noise statistics. The minimum of the modified performance surface is determined and shown to be unique. It is also shown that this minimum corresponds to the minimum of the MSE performance surface (the Wiener solution). Analysis of these properties explains why the LMK algorithm converges faster than LMS during the initial learning phase and slower than LMS as it approaches steady-state.

2. THE LINEAR ESTIMATION PROBLEM

Figure 1 shows the linear estimation problem studied, where:

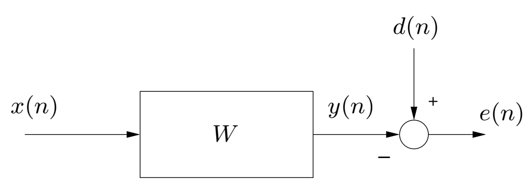

Fig. 1. The linear estimation problem.

*This work has been supported in part by CNPq (Brazilian Ministry of Science and Technology) under Grant No. 352084/92-8

- $x(n)$: input signal, stationary, Gaussian, zero-mean and with variance σ_x^2;
- $d(n)$: desired signal;
- $y(n)$: filter output;
- $e(n)$: estimation error;
- $W = [w_1, w_2, ..., w_N]^T$: filter weigths.

It is assumed that the desired signal $d(n)$ can be modeled as

$$d(n) = W^{o^T} X(n) + z(n) \tag{1}$$

where $W^0 = [w_1^0, w_2^0, ..., w_N^0]^T$ is a constant vector, $X(n) = [x(n), x(n-1), ..., x(n-N+1)]^T$ is the observed data vector, $z(n)$ represents the modeling errors in $d(n) = W^{o^T} X(n)$. $z(n)$ is assumed stationary, white, statistically independent of any other signal and with an even pdf ($f_z(z) = f_z(-z)$).

3. PERFORMANCE SURFACES

The *kurtosis* of the zero-mean estimation error is given by

$$Cum_4[e(n)] = E[e^4(n)] - 3E^2[e^2(n)] \tag{2}$$

where Cum_4 means fourth order cumulant. The cost function J_{LMK} used with the LMK algorithm is defined as the negative of the fourth order cumulant [4], [6]:

$$J_{LMK} = 3E^2[e^2(n)] - E[e^4(n)]. \tag{3}$$

The gradient of J_{LMK} with respect to the weight vector is given by

$$\nabla J_{LMK} = \frac{\partial J_{LMK}}{\partial W} = 6E[e^2(n)]\frac{\partial E[e^2(n)]}{\partial W} - \frac{\partial E[e^4(n)]}{\partial W}. \tag{4}$$

Using a stochastic approximation for the above gradient[1], the estimated gradient is given by

$$\begin{aligned}\widehat{\nabla} J_{LMK}(n) &= \left\{6E[e^2(n)]\frac{\partial e^2(n)}{\partial W} - \frac{\partial e^4(n)}{\partial W}\right\} \\ &= \left\{12E[e^2(n)]e(n) - 4e^3(n)\right\}\frac{\partial e(n)}{\partial W} \\ &= -4\left\{3E[e^2(n)]e(n) - e^3(n)\right\}X(n)\end{aligned} \tag{5}$$

where $\partial e(n)/\partial W = -X(n)$.

For real-time implementation, $E[e^2(n)]$ in (5) has to be estimated. This can be done using the recursion [4]

$$E[e^2(n)] = \beta E[e^2(n-1)] + e^2(n), \ 0 < \beta < 1 \tag{6}$$

with $E[e^2(-1)] = 0$. Using the first three terms of the solution of (6) leads to a non-recursive estimation that is accurate for small β:

$$E[e^2(n)] \approx e^2(n) + \beta e^2(n-1) + \beta^2 e^2(n-2). \tag{7}$$

Note that (7) is a good approximation to $E[e^2(n)]$ only for very small values of β. Nevertheless, it has been used for any $0 < \beta < 1$ [3]–[7]. This and the stochastic gradient approximation lead to an algorithm that operates on a *modified kurtosis* surface, which is quite different from (3). The detailed analysis of the resulting algorithm's performance as a function of β is beyond the scope of this work.

Substituting (7) in (5) yields an expression for the estimated gradient of the now modified performance surface:

$$\begin{aligned}\widehat{\nabla}\hat{J}_{LMK}(n) = &-8e^3(n)X(n) - 12\beta e^2(n-1)e(n)X(n) \\ &- 12\beta^2 e^2(n-2)e(n)X(n).\end{aligned} \tag{8}$$

Using (8), the weight update equation for the LMK adaptive algorithm becomes

$$\begin{aligned}W(n+1) &= W(n) - \mu_{lmk}\widehat{\nabla}\hat{J}_{LMK}(n) \\ &= W(n) + \mu_{lmk}\{2e^2(n) + 3\beta e^2(n-1) + 3\beta^2 e^2(n-2)\} \\ &\quad \times e(n)X(n)\end{aligned} \tag{9}$$

where μ_{lmk} absorbed a multiplicative factor of 4 from (8). Note that (8) and (9) can now be easily implemented using the available instantaneous signals. However, (8) is the gradient of a modified cost function $\hat{J}_{LMK}$ whose properties are unknown.

3.1. The Modified Performance Surface

From Fig. 1, $e(n) = d(n) - X^T(n)W$ and (8) can be written as

$$\begin{aligned}\widehat{\nabla}\hat{J}_{LMK}(n) = \Big\{&8e^3(n) + 12\beta e^2(n-1)e(n) \\ &+ 12\beta^2 e^2(n-2)e(n)\Big\}\frac{\partial e(n)}{\partial W}.\end{aligned} \tag{10}$$

Integrating (10) and replacing $e^2(n)$ and $e^4(n)$ by their expected values, yields

$$\begin{aligned}\hat{J}_{LMK} = &2E[e^4(n)] + 6\beta E[e^2(n-1)]E[e^2(n)] \\ &+ 6\beta^2 E[e^2(n-2)]E[e^2(n)].\end{aligned} \tag{11}$$

If $x(n)$ and $z(n)$ are stationary, $e(n)$ is also stationary[2] and thus $E[e^2(n-2)] = E[e^2(n-1)] = E[e^2(n)]$. Therefore, defining the weight error vector $V = W - W^o$ and using $e(n) = z(n) - X^T(n)V$, (11) can be written as

$$\begin{aligned}\hat{J}_{LMK} &= 2E[e^4(n)] + 6(\beta+\beta^2)E^2[e(n)] \\ &= 6(1+\beta+\beta^2)\left\{2\sigma_z^2 V^T RV + V^T RVV^T RV\right\} \\ &\quad + 2E[z^4(n)] + 6(\beta+\beta^2)\sigma_z^4\end{aligned} \tag{12}$$

where $\sigma_z^2 = E[z^2(n)]$ and $R = E[X(n)X^T(n)]$.

Eq. (12) is a good approximation of the actual performance surface whose gradient is given by (8). The results in [3] show that the LMK adaptive algorithm derived from (8) converges faster than LMS during the learning phase and slower than LMS as it approaches steady-state for Gaussian input. For $\beta = 0$, the Least Mean Fourth (LMF) adaptive algorithm is obtained [7]. In the next sections we derive the properties of the performance surfaces given by (3) and (12).

[1] Substituting $\frac{\partial e^2(n)}{\partial W}$ and $\frac{\partial e^4(n)}{\partial W}$ for $\frac{\partial E[e^2(n)]}{\partial W}$ and $\frac{\partial E[e^4(n)]}{\partial W}$, respectively.

[2] W and W^o are constants.

4. PROPERTIES OF THE TRUE KURTOSIS PERFORMANCE SURFACE (J_{LMK})

For comparison purposes, we first study the properties of the true *kurtosis* performance surface (3) for a Gaussian input signal $x(n)$. With $V = W - W^o$ and using $e(n) = z(n) - X^T(n)V$ in (3), the expression for the true *kurtosis* performance surface becomes

$$J_{LMK} = 3E^2[(z(n) - X^T(n)V)^2] - E[(z(n) - X^T(n)V)^4]. \tag{13}$$

Using the same methodology used in [3, 7], it can be shown that:

$$\begin{aligned} E^2[(z(n) - X^T(n)V)^2] = \sigma_z^4 + 2\sigma_z^2 V^T RV \\ + V^T RVV^T RV \end{aligned} \tag{14}$$

and

$$\begin{aligned} E[(z(n) - X^T(n)V)^4] = 3V^T RVV^T RV \\ + 6\sigma_z^2 V^T RV + E[z^4(n)]. \end{aligned} \tag{15}$$

Substituting (14) and (15) in (13), yields

$$J_{LMK} = 3\sigma_z^4 - E[z^4(n)]. \tag{16}$$

The r.h.s. of (16) is constant and equal to the negative of $Cum_4[z(n)]$. Its gradient with respect to V is equal to zero, which shows that the *kurtosis* performance surface is useless for estimation problems with Gaussian input signals.

5. PROPERTIES OF THE MODIFIED PERFORMANCE SURFACE ($\hat{J}_{LMK}$)

5.1. Stationary Points of $\hat{J}_{LMK}$

The gradient of (12) with respect to V is easily determined to be

$$\frac{\partial \hat{J}_{LMK}}{\partial V} = 24(1 + \beta + \beta^2)\Big\{\sigma_z^2 + V^T RV\Big\}RV \tag{17}$$

and the stationary points are the solutions of $\partial \hat{J}_{LMK}/\partial V = 0$.

Since R is positive semidefinite [1], $V^T RV \geq 0$. Thus, the solution of $\partial \hat{J}_{LMK}/\partial V = 0$ is $V = 0$ and the stationary point of $\hat{J}_{LMK}$ corresponds to $W = W^o$, which is exactly the Wiener solution of the linear minimum MSE problem.

To show that $V = 0$ is a minimum, we must study the properties of the Hessian matrix. The Hessian matrix of $\hat{J}_{LMK}$ at $V = 0$ is given by

$$\begin{aligned} H\Big|_{V=0} &= \frac{\partial^2 \hat{J}_{LMK}}{\partial V^2}\Big|_{V=0} = \frac{\partial}{\partial V}\left\{\frac{\partial \hat{J}_{LMK}}{\partial V}\right\}\Big|_{V=0} \\ &= \frac{\partial}{\partial V}\left\{24(1 + \beta + \beta^2)\sigma_z^2 RV\right\}\Big|_{V=0} \\ &+ \frac{\partial}{\partial V}\left\{24(1 + \beta + \beta^2)V^T RVRV\right\}\Big|_{V=0} \\ &= 24(1 + \beta + \beta^2)\sigma_z^2 R \end{aligned} \tag{18}$$

which is positive definite if R is assumed positive definite, a reasonable assumption in most practical applications [1]. In this case, $W = W^o$ is the point of minimum of $\hat{J}_{LMK}$. Substituting $V = 0$ in (12) yields

$$\hat{J}_{LMK_{min}} = 2E[z^4(n)] + 6(\beta + \beta^2)\sigma_z^4 \tag{19}$$

which differs from (16) by $3E[z^4(n)] + 3(2\beta^2 + 2\beta - 1)\sigma_z^4$. This is a consequence of the modifications introduced in the true *kurtosis* performance surface J_{LMK} to allow a practical implementation of the LMK algorithm.

5.2. MSE Obtained from $\hat{J}_{LMK}$

Performance comparisons among different adaptive algorithms require the use of a common figure of merit. The most used figure of merit is the MSE. It provides a physical interpretation of the results as error power and is used as cost function in deriving several adaptive filtering algorithms.

The performance surface of the MSE is given by [1]

$$\xi = \sigma_z^2 + V^T RV. \tag{20}$$

From (12), $V^T RV$ can be written as

$$\begin{aligned} &\Big\{V^T RV\Big\}^2 + 2\sigma_z^2 \Big\{V^T RV\Big\} \\ &+ \frac{2E[z^4(n)] + 6(\beta + \beta^2)\sigma_z^4 - \hat{J}_{LMK}}{6(1 + \beta + \beta^2)} = 0. \end{aligned} \tag{21}$$

Considering only the positive solution of (21) (R positive definite) and substituting it in (20), leads to a direct relationship between the MSE and $\hat{J}_{LMK}$:

$$\begin{aligned} \xi = f(\hat{J}_{LMK}) &= \sigma_z^2 + V^T RV \\ &= \frac{1}{2}\sqrt{\frac{4\{3\sigma_z^4 - E[z^4(n)]\} + 2\hat{J}_{LMK}}{3(1 + \beta + \beta^2)}}. \end{aligned} \tag{22}$$

Substituting (19) in (22) yields $\xi(\hat{J}_{LMK_{min}}) = \sigma_z^2$, which confirms the result that the minima of both surfaces correspond to the Wiener solution $W = W^o$. This expression also provides a way to compare the convergence rates of adaptive algorithms derived from these two surfaces.

5.3. Comparison of Convergence Rates

To compare the convergence ratios of adaptive algorithms based on the MSE and on the $\hat{J}_{LMK}$ surfaces, we write the square of (22) as

$$4\xi^2 = \frac{12\sigma_z^4 - 4E[z^4(n)] + 2\hat{J}_{LMK}}{C} \tag{23}$$

where $C = 3(1 + \beta + \beta^2) \geq 3$ for $\beta \geq 0$.

Taking the derivatives of both sides of (23) with respect to V and considering their magnitudes,

$$\frac{\left|\frac{\partial \hat{J}_{LMK}}{\partial V}\right|}{\left|\frac{\partial \xi}{\partial V}\right|} = 4C\ \xi. \tag{24}$$

Analysis of (24) shows that the relationship between $\partial \hat{J}_{LMK}/\partial V$ and $\partial \xi/\partial V$ can be classified into two distinct regions. These regions are related to the convergence speed of the LMK and LMS adaptive algorithms, respectively:

a) During the transient phase, when ξ is large ($\xi \geq \frac{1}{4C}$),

$$\left|\frac{\partial \hat{J}_{LMK}}{\partial V}\right| \geq \left|\frac{\partial \xi}{\partial V}\right| \tag{25}$$

and thus the LMK algorithm converges faster than the LMS algorithm.

b) Close to steady-state, when ξ is small ($\xi < \frac{1}{4C}$),

$$\left|\frac{\partial \hat{J}_{LMK}}{\partial V}\right| < \left|\frac{\partial \xi}{\partial V}\right| \tag{26}$$

and thus LMK converges slower than LMS.

The quantitative measure of the difference in convergence rates just derived agrees with a qualitative analysis of the weight updating equations of the two algorithms. The weight updating equations for the LMS [1] and LMK [3] algorithms are given by

$$V_{lms}(n+1) = V_{lms}(n) + \mu_{lms} e(n) X(n) \tag{27}$$

and

$$\begin{aligned} &V_{lmk}(n+1) = V_{lmk}(n) \\ &+ \mu_{lmk}\{2e^2(n) + 3\beta e^2(n-1) + 3\beta^2 e^2(n-2)\} e(n) X(n). \end{aligned} \tag{28}$$

Thus, the LMK algorithm can be interpreted as an LMS algorithm with variable step size given by,

$$\mu_{lms}(n) = \mu_{lmk}\{2e^2(n) + 3\beta e^2(n-1) + 3\beta^2 e^2(n-2)\}. \tag{29}$$

The convergence rate and the misadjustment of the LMS algorithm are proportional to the step size [1, 2]. Thus, (29) shows a large step size when $e(n)$ is large. As $e(n)$ converges towards its steady-state condition, the step size is reduced. The relative behavior of both algorithms in a system identification problem is illustrated in Fig. 2 (average of 50 realizations), which was obtained for $W^o = [0.1085\ 0.2169\ 0.3254\ 0.4339\ 0.5423\ 0.4339\ 0.3254\ 0.2169\ 0.1085]$, ${W^o}^T W^o = 1$, $W(0) = 0$, $\beta = 0.5$. The input signal applied to the system to estimate W^o was an AR(1) process generated as $x(n) = 0.5x(n-1) + g(n)$ where $g(n)$ is white Gaussian with unit variance. Two different additive noise distributions were considered: Gaussian and sinusoidal, the latter expressed as $z(n) = \sqrt{2\sigma_z^2}\sin(377n + \phi)$ with ϕ uniformly distributed in $[-\pi, \pi]$. In both cases $\sigma_z^2 = 0.1$. The step sizes were adjusted for equal steady-state MSE in all cases. $\mu_{LMS} = 0.000215$, $\mu_{LMK} = 0.000180$ for Gaussian noise and $\mu_{LMK} = 0.000586$ for sinusoidal noise. Since the LMS MSE is not affected by the noise distribution, only one curve is shown for the LMS algorithm.

6. CONCLUSION

This paper studied the properties of the performance surface of the linear estimation problem with cost function based on the *kurtosis* of error. The analysis is concentrated on a modified performance surface used in actual implementations of the LMK adaptive algorithm. It was shown that this modified *kurtosis* performance surface has a global minimum that corresponds to the Wiener solution of the MSE surface. A comparative analysis between the MSE and the modified *kurtosis* cost functions has shown why the LMK algorithm outperforms the LMS algorithm during the transient phase, but has slower convergence as it approaches steady-state.

7. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 4^a Ed., Upper Saddle River, NJ: Prentice-Hall, 2002.

[2] D. G. Manolakis, V. K. Ingle e S. M. Kogon, *Statistical and Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, Boston, Mass: McGraw Hill, 2000.

[3] P. I. Hübscher and J. C. M. Bermudez, "A model for the behavior of the least mean kurtosis (LMK) adaptive algorithm with gaussian inputs", *International Telecommunications Symposium - ITS2002*, Natal/RN, Brasil, 2002.

[4] O. Tanrikulu and A. G. Constantinides, "Least Mean Kurtosis: a novel higher-order statistics based adaptive filtering algorithm", *Electronic Letters*, 30(3): 189-190, Feb., 1994.

[5] A. G. Constantinides and D. I. Pazaitis, "A novel kurtosis driven variable step-Case adaptive algorithm", *IEEE Transactions on Signal Processing*, v. 47, n.3, Mar., 1999.

[6] O. Tanrikulu, *Adaptive Signaling Processing Algorithms with Accelerated Convergence and Noise Immunity*, PhD Thesis, Department of Electrical Engineering, Imperial College, London, 1995.

[7] P. I. Hübscher and J. C. M. Bermudez, "An improved statistical analysis of the least mean fourth (LMF) adaptive algorithm", *IEEE Transactions on Signal Processing*, Vol. 51, No. 3, March 2003.

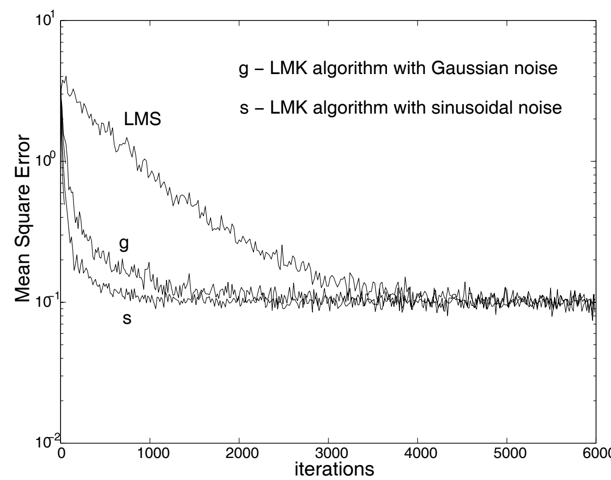

Fig. 2. Relative behavior of the LMK and LMS algorithms.