COEFFICIENT ROUNDING AND TRUNCATION EFFECTS ON RECURSIVE LEAST SQUARES ALGORITHMS

S.H. Ardalan

CENTER FOR COMMUNICATIONS AND SIGNAL PROCESSING

North Carolina State University

Raleigh, NC

CCSP-TR-84/3

March 15, 1984

ABSTRACT

The effects of coefficient truncation and rounding upon the convergence properties of the Recursive Least Squares algorithm are examined. It is shown that in a finite wordlength implementation of RLS algorithms one should avoid truncation and use rounding instead. The statistical wordlength of the Kalman gain vector is derived based on a specified performance measure. The result shows that speech signals require more bits for quantization than white noise. An upper bound on the Kalman gain vector norm is also derived.

Table of Contents

1. Introduction

This technical report analyzes the effects of coefficient rounding and truncation on Recursive Least Squares (RLS) algorithms. By coefficients we mean the state estimate vector elements which are sequentially updated by the RLS algorithm. Errors in the coefficients occur in the updating equation when the result of the multiplication of the Kalman gain and estimation error are rounded or truncated. In this paper we will, as a first step, assume that the Kalman gain is computed with infinite precision prior to rounding.

In section 2, the statistical wordlength of the Kalman gain vector required in order to meet a specified performance measure is derived. The result indicates the dependence of the SWL on filter order and signal statistics. Finally, in section 3 an upper bound for the Kalman gain vector norm is obtained. This norm is required for scaling purposes. The result shows that the norm depends on signal statistics and increases for correlated signals.

2. Effect of Coefficient Rounding and Truncation

In this technical report, we will analyze the effect of using infinite precision in the computation of the Kalman gain while truncation or rounding is used to compute the update of the estimate. One can show that with infinite precision, the Kalman update equations may be written as [1,2]:

$$\theta(t) = \theta(t-1) + K(t)\varepsilon(t) \tag{1.1}$$

with truncation or rounding after the calculation of $K(t)\varepsilon(t)$ we have,

$$\theta'(t) = \theta'(t-1) + \frac{1}{t}\left\{R^{-1}(t)\phi(t)\varepsilon(t)+\psi(t)\right\} \tag{1.2}$$

where

$E\ \{\psi(t)\} = \bar{\psi} \neq 0$ for truncation and

$E\ \{\psi(t)\} = 0$ for rounding.

Using the method of associated differential equation we can write for the mean value of the estimate

$$\frac{d\theta'}{d\tau} = E\ \left\{R^{-1}(\tau)\phi(t)\varepsilon(t)+\psi(t)\right\} \tag{1.4}$$

$$\frac{d\theta'}{d\tau} = R^{-1}\ (\tau)G(\theta)\ (\theta^*-\theta)+\bar{\psi} \tag{1.5}$$

Since infinite precision is used to calculate $R(\tau)$ then $\lim_{\tau\to\infty} R(\tau)\to G(\theta)$ and (1.5) becomes

$$\frac{d\theta'}{d\tau} = (\theta^*-\theta) + \bar{\psi} \tag{1.6}$$

So that the mean of the state estimate after convergence is

$$\theta = \theta^* + \bar{\psi} \tag{1.7}$$

In this analysis we are not considering the effect of truncation or rounding on the variance of Θ. However, the degredation of the mean of the steady state estimate θ is sought. From (1.7), we observe that rounding does not cause a degredation in the mean. However, truncation produces an offset. In order to analyze the effect of this offset on the poles of $H(z)$ in the application of RLS to linear prediction where θ is the vector of predictor coefficients, we introduce the following result due to Wilkinson [3].

If z_r is an isolated zero of the polynomial

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0 \tag{1.8}$$

Then $z_r(\varepsilon)$ is a root of

$$f(z) + \varepsilon g(z) \tag{1.9}$$

and is given by

$$z_r(\varepsilon) = z_r + \sum_{1}^{\infty} P_k \varepsilon^k \tag{1.10}$$

For $\varepsilon \to 0$, Wilkinson derives the very useful results,

$$z_r(\varepsilon) - z_r \sim -\varepsilon \frac{g(z_r)}{f'(z_r)} \quad (\varepsilon \to 0) \tag{1.11}$$

Since $\varepsilon g(z)$ describes a pertubation of the coefficients of $f(z)$, the above result is useful in the analysis of the effects of quantization of the coefficients on the location of the zeroes. For linear prediction we have

$$H(z) = \frac{1}{1 + \sum_{i=1}^{n} \theta_i z^{-1}} \tag{1.12}$$

let

$$f(z) = z^n + \theta_1 z^{n-1} + \ldots + \theta_n \tag{1.13}$$

Then the polynomial due to truncation is

$$q(z) = z^n + (\theta_1 + \bar{\psi}) z^{n-1} + \ldots \theta_n + \bar{\psi} \tag{1.14}$$

Or

$$q(z) = f(z) + \bar{\psi}\{z^{n-1} + z^{n-2} + \ldots + 1\} \tag{1.15}$$

If $\psi \to 0$ then we can use the result (1.10) to obtain,

$$z_r(\varepsilon)-z_r \sim -\bar{\psi} \frac{\sum_{k=0}^{n-1} z_r^k}{f'(z_r)} = -\bar{\psi} \frac{\sum_{k-0}^{n-1} z_r^k}{\prod_{i=1\ i\neq r}^{n} (z_r-z_i)} \tag{1.16}$$

The result can be interpreted as follows: If the zero z_r is close to the unit circle then the numerator in (1.16) will be large, increasing the pole dislocation due to truncation. This might lead the pole to locate outside the unit circle causing instability. However, if $|z_r|$ is smaller than 1 then the numerator will decrease.

The term $f'(z_r)$ in the denominator is interesting. For example, if the poles are clustered then $f'(z_r)$ will be small. This will cause a large error in the pole location causing in turn a large degredation or even divergence. This can be seen for a second order process with clustered poles. In this case the poles are

$$z_r,\ z'_r \text{ with } z'_r = z_r+\delta \text{ where } \delta \text{ is small.} \tag{1.17}$$

and

$$f'(z) = 2(z-zr)-\delta \tag{1.18}$$

Substituting for $f'(z_r)$ in (1.16) we get

$$z_r(\varepsilon)-z_r \sim -\bar{\psi} \frac{\sum_{k=0}^{n-1} z_r^k}{-\delta} = \bar{\psi} \frac{\sum_{k=0}^{n-1} z_r^k}{\delta} \tag{1.19}$$

Pole clustering (narrow band spectrum) implies δ small. From (4.19) this results in a large pole dislocation causing severe degredation and possible divergence.

The above analysis suggests that in a finite wordlength implementation of the RLS algorithms truncation of the estimates after updating must be avoided and rounding used instead. This is especially important when correlated signals are used since they represent narrow band spectrums with clustered poles.

3. Statistical Word Length for the Quantization of the Kalman Gain

This section deals with the derivation of the statistical word length (SWL) required to quantize the Kalman gain vector elements in order to meet certain performance measures. The method used is based on the statistical wordlength measure first employed by Knowles and Okayto [4]. This method was expanded upon by Moroney [5] where he analyzed coefficient quantization effects in the digital implementation of digital feedback compensators. In this section, Moroney's results are applied to the RLS system identification problem in order to derive the SWL of the quantized Kalman gain vector.

Consider a general scalar measure of peformance f. Let f_∞ denote the ideal infinite precision measure;

$$f_\infty = f(c_1, C_2, \ldots c_N) \tag{2.1}$$

where c_i, $i=1,\ldots,N$ are the coefficients that are to be quantized. The degredation in performance, df, due to quantization can be expanded in a Taylor's series about the ideal f_∞. Keeping first order terms,

$$df(c_1, c_2, \ldots, c_N) \approx \sum_{i=1}^{N} \left(\left.\frac{\partial f}{\partial c_i}\right|_\infty dc_i \right) , \tag{2.2}$$

where dc_i is the error due to quantization and $\left.\left(\frac{\partial f}{\partial c_i}\right)\right|_\infty$ is the first partial derivative of f evaluated at the infinite precision coefficient value.

If $\Delta = 2^{-k}$ is the quantization step size then if rounding is used

$$\frac{-\Delta}{2} < dc_i < \Delta/2$$

Therefore,

$$df < \frac{\Delta}{2} \sum_{i=1}^{N} \left| \left(\frac{\partial f}{\partial c_i}\bigg|_{\infty} \right) \right| \tag{2.3}$$

which is a very pessimistic wordlength estimate. The basic idea behind statistical wordlength is to produce a less pessimistic estimate by treating an ensemble of structures. Thus the coefficient errors dc_i (assuming rounding) can be thought of as uniformly distributed zero-mean uncorrellated random variables, each of variance $\frac{\Delta^2}{12}$. Thus, we treat the error df as a random value. Therefore,

$$(\sigma_{df})^2 = \frac{\Delta^2}{12} \sum_{i=1}^{N} \left(\frac{\partial fi}{\partial c_i}\bigg|_{\infty} \right)^2 \qquad \text{df zero mean} \tag{2.4}$$

For large N, the central limit theorem can be applied to justify a gaussian distribution for df. Thus,

$$\Pr\left[\, |df| < 2\sigma_{df} \right] = 0.954$$

If

$$f_{\infty} - \frac{E_0}{2} < f < f_{\infty} + \frac{E_0}{2}$$

Then

$$\sigma_{df} = \frac{E_0}{2}$$

Hence

$$\Delta = \frac{\sqrt{3}\quad E_0}{\left\{ \sum_{i=1}^{N} \left(\frac{\partial fi}{\partial c_i}\bigg|_{\infty} \right)^2 \right\}^{1/2}} \tag{2.5}$$

The statistical wordlength can be defined as

$$SWL = \ell + \log_2 \frac{1}{\Delta} \tag{2.6}$$

where ℓ represents the integer portion of the coefficient. The above result is quite useful in digital filtering when f is the magnitude of the spectrum for example. However, in optimal filter design and the RLS algorithms usually

$$\left.\frac{\partial J}{\partial c_i}\right|_\infty = 0 \ , \ i=1,\ldots,N \tag{2.7}$$

when J is the performance measure. For example in RLS estimation it is the expected value of the square of the estimate error. Note that the Kalman gain coefficients are derived by requiring that (2.7) holds. Therefore, for optimal filters we must use second order terms in (2.2). Thus,

$$dJ \approx \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\left.\frac{\partial^2 J}{\partial c_i \partial c_j}\right|_\infty dc_i dc_j \right) \quad . \tag{2.8}$$

$$E\,\{dJ\} = \frac{1}{2} \sum_{i=1}^{N} \left(\left.\frac{\partial^2 J}{\partial c_i^2}\right|_\infty \right) E[(dc_i)^2] \quad . \tag{2.9}$$

As seen from (2.9) the mean value of $d_J \neq 0$ unlike the case for df.

Before deriving an expression for the variance of dJ, let us define the random variable ε to be the square of dc_i. Its mean and variance can be shown to be

$$E(\varepsilon) = \bar{\varepsilon} = (\Delta^2)/12 \ ; \ E(\varepsilon^2) = \overline{\varepsilon^2} = (\Delta^4)/180 \tag{2.10}$$

Thus,

$$E[(dJ)^2] = \frac{(\overline{\varepsilon^2})}{4} \sum_{i=1}^{N} \left(\left.\frac{\partial^2 J}{\partial c_i^2}\right|_\infty \right)^2 + \frac{(\bar{\varepsilon})^2}{4} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i\neq j}}^{N} \left(\left.\frac{\partial^2 J}{\partial c_i^2}\right|_\infty \right) \left(\left.\frac{\partial^2 J}{\partial c_j^2}\right|_\infty \right) \tag{2.11}$$

$$+ \frac{(\bar{\varepsilon})^2}{2} \sum_{\substack{i=1 \\ i\neq j}}^{N} \sum_{j=1}^{N} \left(\left.\frac{\partial^2 J}{\partial c_i \partial c_j}\right|_\infty \right)^2 \quad ,$$

$$(\sigma_{dJ})^2 = \frac{\sigma_\varepsilon^2}{4} \sum_{i=1}^{N} \left(\frac{\partial^2 J}{\partial c_i^2}\Big|_\infty\right)^2 + (\bar{\varepsilon})^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\partial^2 J}{\partial c_i \partial c_j}\Big|_\infty\right)^2 \tag{2.12}$$

Since the degredation in J can only increase, we have

$$J_\infty + E_0 = J_\infty + \overline{dJ} + 2\sigma_{dJ} \tag{2.13}$$

This choice of σ_{dj} gives a 97.5% confidence level in terms of remaining below the allowed deviation E_0. Writing,

$$E_0 = \overline{dJ} + 2\sigma_{dJ} \tag{2.14}$$

we obtain

$$\frac{1}{\Delta^2} = \frac{1}{24E_0} \sum_{i=1}^{N} \left(\frac{\partial^2 J}{\partial c_i^2}\Big|_\infty\right) + \frac{1}{6E_0} \left[\sum_{\substack{i=1 \\ i>j}}^{N} \sum_{j=1}^{N} \left(\frac{\partial^2 J}{\partial c_i \partial c_j}\Big|_\infty\right)^2 + \frac{1}{5} \sum_{i=0}^{N} \left(\frac{\partial^2 J}{\partial c_i^2}\Big|_\infty\right)^2 \right]^{1/2} \tag{2.15}$$

From which,

$$SWL = \ell + \frac{1}{2} \log_2 \left(\frac{1}{\Delta^2}\right) \quad .$$

The result (2.15) was derived by Moroney [7]. We will now apply this result to the RLS algoirthm.

For the RLS algorithm we seek to minimize the trace of the state error autocovariance matrix P(k).

$$J(k) = \text{Trace } [P(k)] \tag{2.16}$$

we have

$$\frac{\partial J(k)}{\partial \underline{K}(k)} = 0 \tag{2.17}$$

where $\underline{K}(k)$ is the Kalman gain vector. Therefore, we must use second order terms to evaluate degradations in J due to the quantization of $\underline{K}(k)$. From the derivation of the Kalman gain [6] we have

$$\frac{\partial J(k)}{\partial \underline{k}(k)} = -2\left[I - K(k)x^T(k)\right] P(k-1)x(k) + 2\underline{K}(k)r(k)$$

$$\frac{\partial^2 J(k)}{\partial \underline{k}^2(k)} = B(k), \text{ where } b_{ij} = \frac{\partial^2 J}{\partial k_i \partial k_j} \tag{2.18}$$

Now,

$$\frac{\partial J}{\partial k_i} = 2\underline{K}_i[\underline{x}^T(k)P(k-1)\underline{x}(k)+r(k)]-2c_i \qquad (2.19)$$

where c_i is an element of the vector $P(k-1)x(k)$.

Proceeding,

$$\frac{\partial^2 J}{\partial k_i \partial k_j} = 0 \qquad (2.20)$$

which is a simplifying result.

$$\frac{\partial^2 J}{\partial k_i^2} = 2\ [\underline{x}^T(k)P(k-1)\underline{x}(k)+r(k)] \qquad (2.21)$$

This is an important result which relates the terms effecting quantization degredation to the scalar denominator term used in the update of the matrix $P(k)$. The term in the brackets is also relevant in the fast RLS algorithms. Using (2.20) we obtain for $\frac{1}{\Delta^2}$ from (2.15)

$$\frac{1}{\Delta^2} = \frac{1}{24E_0} \sum_{i=1}^{N} \left(\frac{\partial^2 J}{\partial c_i^2}\Big|_\infty\right) + \frac{1}{6E_0}\left[\frac{1}{5}\sum_{i=1}^{N}\left(\frac{\partial^2 J}{\partial c_i^2}\Big|_\infty\right)^2\right]^{1/2} \qquad (2.22)$$

Let

$$\theta(k) = \underline{x}^T(k)P(k-1)\underline{x}(k) + r(k) \qquad (2.23)$$

Then

$$\frac{\partial^2 J}{\partial k_i^2} = 2\theta(k) \text{ for } i=1,\ldots,N \qquad (2.24)$$

Substituting into (2.22)

$$\frac{1}{\Delta^2} = \frac{1}{24E_0} \times 2N\theta(k) + \frac{1}{6E_0}\left(\frac{4}{5}\ N\theta^2(k)\right)^{1/2} = \frac{1}{12E_0}\ N\theta(k) + \frac{\sqrt{N}}{6\sqrt{5E}}\ \theta(k) \qquad (2.25)$$

$$\frac{1}{\Delta^2} = \frac{\theta(k)}{6E_0}\left(\frac{N}{2} + \frac{\sqrt{N}}{\sqrt{5}}\right) \qquad (2.26)$$

Finally,

$$SWL = 1 + {}^{1}/{}_{2}\log_2\left\{\frac{\theta(k)}{6E_0}\left(\frac{N}{2} + \frac{\sqrt{N}}{\sqrt{5}}\right)\right\} \qquad (2.27)$$

In order to discuss this result more fully, we derive an approximate steady state upper bound on $\theta(k)$. We note that

$$P(k) = \left(\sum_{i=1}^{k} x(i)x^T(i) \right)^{-1} r(k) \tag{2.28}$$

But we can write,

$$A = E\{x(k)x^T(k)\} \approx \frac{1}{k} \left(\sum_{i=1}^{k} x(i)x^T(i) \right) \tag{2.29}$$

where A is the autocorrelation matrix of the signal x(k). From (2.28) and (2.29)

$$P(k) \sim \frac{1}{k} A^{-1} r(k) = r(k) \frac{1}{k} U^{-1} \Lambda^{-1} U \tag{5.30}$$

where we have written

$$A = U \Lambda U^{-1} \tag{2.31}$$

where Λ is diagonal matrix of the eigenvalues of A_j λ_i , $i=1,\ldots,N$. From (2.23), substituting for P(k) from (2.30),

$$\theta(k) \approx \frac{1}{k} r(k) x^T(k) A^{-1} x(k) + r(k) \tag{2.32}$$

Taking the norm of boths sides of (2.32),

$$\theta(k) < \frac{1}{k} r(k) \|x(k)\| \|A^{-1}\| \|x(k)\| + r(k) \tag{2.33}$$

Let

$$s(k) = \|x(k)\|^2 \tag{2.34}$$

$$\|A^{-1}\| = 1/\lambda_{min} \tag{2.35}$$

we have

$$\theta(k) < r(k) \frac{1}{k} \frac{s(k)}{\lambda min} + r(k) \tag{2.36}$$

$r(k) = \xi_0$ is the noise power of the estimate. Substituting (2.36) into (2.27)

$$\text{SWL} < \ell + \frac{1}{2}\log_2\left\{\frac{\xi_0}{6E_0}\left(\frac{\sqrt{N}}{\sqrt{5}}\right)\left(\frac{1}{k}\frac{s(k)}{\lambda \min} + 1\right)\right\} \tag{2.37}$$

Discussion: Suppose that the input signal x(k) is white gaussian. Then $\lambda_i = \sigma_x^2$ i=1,...,N and (2.37) becomes,

$$\text{SWL} < \ell + \frac{1}{2}\log_2\left\{\frac{\xi_0}{6E_0}\left(\frac{\sqrt{N}}{\sqrt{5}}\right)\left(\frac{N}{k} + 1\right)\right\}, \quad s(k) \sim N\sigma_x^2 \tag{2.38}$$

Since we have assumed N large, then

$$\text{SWL} < \ell + \frac{1}{2}\log_2\frac{\xi_0}{6E_0} + \frac{1}{2}\log_2\left(\frac{N^2}{2} + \frac{N\sqrt{N}}{\sqrt{5}}\right) + \frac{1}{2}\log_2\frac{1}{k} \tag{2.39}$$

The result (2.39) contains a significant observation. Even if E_0 is large (that is a large degredation can be tolerated) the second term in (2.39) implies that the number of bits required will still increase with the filter length, N. The table below evaluates the expression for various values of N.

N	$\frac{1}{2}\log_2\left(\frac{N^2}{E} + \frac{N\sqrt{N}}{\sqrt{5}}\right)$ in bits	$\frac{1}{2}\log_2\frac{1}{k}$ k=N
10	3	-1.66
50	5	-2.82
100	6	-3.3
150	7	-3.6
200	7	-3.8
250	7.5	-3.98
512	8.5	-4.5

The term $\log_2\frac{\xi_0}{6E_0}$ is zero only if $\xi_0 = 6E_0$ or $E_0 = \frac{\xi_0}{6}$. Degredations above this value will result in less bits required as implied in the table above since $\log_2\left(\frac{\xi_0}{6E_0}\right)$ will be negative. Since for ideal modeling of the system

$J(k) \to \xi_0$ as $k \to \infty$ the condition

$$E_0 = \frac{\xi_0}{6} \tag{2.40}$$

implies a degredation of $J_\infty + E_0 = \xi_0 + \frac{\xi_0}{6} = \xi_0\,(\frac{7}{6})$ or

$$10 \log \frac{\xi_0}{\xi_0\,(\frac{7}{6})} = 0.67 \text{ dB} \tag{2.41}$$

However, these results are valid if the filter does not diverge.

Suppose now that the signal x(k) is highly correlated such that $(\frac{\lambda_{max}}{\lambda_{min}})$ is large. We can therefore write approximately that

$$s(k) \sim \frac{N}{2}\,\lambda_{max} \tag{2.42}$$

Thus, (2.37) becomes

$$SWL < \ell + \frac{1}{2}\log_2 \frac{\xi_0}{6E_0} + \frac{1}{2}\log_2\left(\frac{N}{2} + \frac{\sqrt{N}}{\sqrt{5}}\right)N + \frac{1}{2}\log\left(\frac{1}{2}\,\frac{\lambda_{max}}{\lambda_{min}}\right) \tag{2.43}$$

Comparing to (2.39) for the case of white gaussian noise the term

$$\frac{1}{2}\log\left(\frac{1}{2}\,\frac{\lambda_{max}}{\lambda_{min}}\right) \tag{2.44}$$

has been added. Since as N increases $\frac{\lambda_{max}}{\lambda_{min}}$ increases for correlated signals, it is implicitly buried in (2.44). For speech it is very natural to have

$$\frac{\lambda_{max}}{\lambda_{min}} > 100$$

which implies that more than 3 bits are required for correllated speech, in this case, compared to white gaussian.

4. <u>An Upper Bound on the Steady State Kalman Gain Vector Norm.</u>

In the finite wordlength implementation of the RLS algorithms, a bound on the Kalman gain vector norm is necessary for scaling purposes. Furthermore, the dependence of this upper bound on signal correlation is sought. The

Kalman gain can be expressed as [6].

$$\underline{K}(k) = P(k)\underline{x}(k) \tag{3.1}$$

where

$$P(k) = \Big(\sum_{i=1}^{N} \underline{x}(k)\underline{x}^T(k) \Big)^{-1} \tag{3.2}$$

After $k > N$ iterations, where N is the vector length, we can approximate P(k) by

$$P(k) \approx \frac{1}{k} A^{-1} \tag{3.3}$$

where A is the autocorrelation matrix of the signal. Taking the norms of both sides of (3.1),

$$\| K(k) \| < \| P(k) \| \ \| \underline{x}(k) \| \tag{3.4}$$

Substituting (3.3) into (3.4)

$$\| K(k) \| < \frac{1}{k} \| A^{-1} \| \ \| \underline{x}(k) \| \tag{3.5}$$

But

$$\| A^{-1} \| = 1/\lambda_{min} \tag{3.6}$$

So that

$$\| K(k) \| < \frac{1}{k} \ \frac{\| x(k) \|}{\lambda_{min}} \tag{3.7}$$

The term $\| x(k) \|$ represents the square root of the short term power of the signal x(k). If N is large, then the rough approximation

$$\| x(k) \| \sim \sqrt{\text{Trace } A} = \sqrt{\textstyle\sum \lambda_i} \tag{3.8}$$

can be made. Therefore,

$$\| K(k) \| < \frac{1}{k} \ \frac{\sqrt{\sum \lambda_i}}{\lambda_{min}} \tag{3.9}$$

<u>Discussion</u>: It is observed from (3.9) that the norm (L_2) of the Kalman gain vector decreases as the number of iterations k increases. Since the norm upper bound is inversely related to λ_{min}, the norm is large for highly correlated signals. In this case the integer length will be large even after many iterations. For white noise, the norm upper bound will be smaller than for highly correlated signals and (3.9) implies that the integer portion of the fixed point implementation will be small. In this case, fraction quantization plays an important role. This is demonstrated in simulations where a decrease in fraction quantization bits leads to larger degredation for white noise than speech. This is due in part by the large integer part in speech compared to white noise and the lesser dependence on the fractional part because of the larger norm.

For white noise, $\lambda_i = \sigma_x^2$ i=1,...,N and (3.9) becomes

$$\|K(k)\| < \frac{\sqrt{N\sigma_x^2}}{\sigma_x^2} = \frac{1}{k}\ \frac{1}{\sigma_x}\sqrt{N} \tag{3.10}$$

This result shows that the norm upper bound roughly speaking, depends on the square root of the filter length N. It also depends inversely on the rms value of the signal σ_x.

5. References

1. L. Ljung, "Analysis of Recursive Stochastic Algorithms," IEEE Trans. on Automatic Control, Vol. AC-22, No. 4, Aug. 1977.

2. Soderstrom, Ljung, and Gustavsson, "A Theoretical Analysis of Recursive Identification Methods", Automatica, Vol. 14, 1978, pp. 231-244.

3. J.H. Wilkinson, Rounding Errors in Algebraic Processes, Prentice-Hall, NJ. 1963.

4. J.B. Knowles and E.M. Olcayto, "Coefficient Accuracy and Digital Filter Response," IEEE Trans. Circuits and Systems, Vol. CAS-23, No. 9, September 1976, pp. 551-562.

5. Moroney, Issues in the Implementation of Digital Feedback Compensators, MIT Press, Cambridge, MA,1983.