CONDITION NUMBER OF DATA MATRIX AND PERSISTENT EXCITATION CONDITIONS IN RLS ADAPTIVE FILTERING*

Max Gerken and Phillip M.S. Burt[1]

PTC – Escola Politécnica – USP
Av. Prof. Luciano Gualberto,
travessa 3, n. 158
Cidade Universitária, Butantã
05508-900 - São Paulo - SP
mgk@lcs.poli.usp.br

Maria D. Miranda[2]

Programa de Pós-Graduação em
Engenharia Elétrica
Universidade Presbiteriana Mackenzie
Rua da Consolação, 896 - 5° andar
01302-907 - São Paulo - SP
mdm@mackenzie.com.br

Abstract - Persistent excitation and the condition of the data matrix are considered in the framework of RLS adaptive filtering with exponential weighting. Two persistent excitation conditions that are used in convergence and numerical stability analysis of RLS algorithms are shown to be equivalent. The boundedness of the data matrix condition number is also shown to be equivalent to the considered conditions as long as the input signal energy is lower and upper bounded. Some related inequalities are presented that give insight into the numerical stability behavior of RLS algorithms. The relations of excitation persistency with concepts like predictability, spectral content of the excitation signal, identifiability, exponential convergence and numerical stability of RLS algorithms are briefly addressed in order to give an overview of the subject.

Resumo - O conceito de excitação persistente e o condicionamento da matriz de dados são considerados no âmbito dos algoritmos adaptativos RLS com ponderação exponencial. É demonstrada a eqüivalência de duas condições de excitação persistente utilizadas na análise de convergência e estabilidade numérica de algoritmos da família RLS. É mostrado também que a existência de um limite superior para a dispersão dos valores singulares da matriz de dados é equivalente a estas condições de persistência de excitação, desde que a energia do sinal de entrada seja limitada inferiormente e superiormente. Além disso, são apresentadas algumas desigualdades que são úteis para o entendimento do comportamento numérico de algoritmos adaptativos RLS. As relações entre persistência de excitação e sinais perfeitamente predizíveis, conteúdo espectral de sinais, identificação de sistemas, convergência exponencial e estabilidade numérica de algoritmos RLS são abordadas de modo a permitir uma visão mais ampla do assunto.

Keywords: Adaptive filtering, RLS algorithms, persistent excitation.

1. INTRODUCTION

Identification of an unknown system is a central issue in various applications of the communications area such as channel equalization, echo cancelation in communication channels, or acoustic echo cancelation for teleconferencing or for handsfree telefone terminals, and many others [1, 2, 3]. It is the procedure of determining a model for the unknown system in terms of a set of measurements of the input and output response signals and in terms of an appropriate chosen cost function that is optimized with respect to the parameters of the unknown model. Very often, such a procedure is implemented adaptively, that is, the model is improved as new measurements are received and this knowledge is incorporated. There are three fundamental aspects that must be considered for a successful implementation an identification procedure: the signals must provide sufficient information in order to assure identifiability, the cost function and the optimization method must be appropriately chosen and an efficient well behaved adaptive procedure should be used. By considering the Least Squares (LS) optimization method this paper focus on the first point and also its relations with the convergence behaviour of adaptive LS algorithms, particularly with their numerical stability.

The standard LS problem with exponential weighting is to choose a coefficient vector $\mathbf{w}_M(n)$ to minimize the norm of an estimation error vector given by

$$\mathbf{e}_M(n) = \mathbf{d}(n) - \mathbf{A}_M(n)\mathbf{w}_M(n),$$

where $\mathbf{d}(n)$ is the desired response vector and $\mathbf{A}_M(n)$ is the data matrix, defined as $\mathbf{A}_M(n) = [A(i,j)] = [\lambda^{(n-i+1)/2}u(i-j)]$, with λ being the weighting factor and $u(n)$ the pre-windowed input sequence, i. e. $u(n) = 0$ for $n < 0$. In a LS problem of order M, which is the dimension of the coefficient vector $\mathbf{w}_M(n)$, the data matrix $\mathbf{A}_M(n)$ has dimensions $(n+1) \times M$ and the quadratic norm $\mathbf{e}_M^T(n)\mathbf{e}_M(n)$ is minimized by the optimal coefficient vector

$$\mathbf{w}_M^o(n) = \Phi_M^{-1}(n)\Omega_M(n), \tag{1}$$

where $\Phi_M(n) = \mathbf{A}_M^T(n)\mathbf{A}_M(n)$ is the autocorrelation matrix of the pre-windowed exponentially weighted input sequence, and $\Omega_M(n) = \mathbf{A}_M^T(n)\mathbf{d}(n)$ is the cross-correlation vector between the input data and the desired response.

There is a great number of Recursive Least Squares (RLS) algorithms that may be used to solve the problem sketched above [4]. Many of them are fast algorithms, whose computational complexity is proportional to M. Examples

---

of implementation of those algorithms in hardware and software and their application to channel equalization and echo cancelation can be found in [1, 2, 3, 5, 6]. Several of those algorithms deal directly with the coefficient vector $w_M(n)$ while others use it only indirectly, for example through the coefficients of a lattice structure. As a matter of fact, regardless of the particular RLS algorithm, the literature reports that the conditioning of the data matrix or, equivalently, of the autocorrelation matrix, bears influence on the numerical behavior of RLS algorithms [4, 7, 8, 9, 10].

More specifically, how near or far $A_M(n)$ is from 'rank deficiency' may be measured using the spectral condition number, or the singular value dispersion, defined as

$$\chi(n) = \frac{s_{\max}(n)}{s_{\min}(n)},$$

with $s_{\max}(n)$ and $s_{\min}(n)$ being respectively the greatest and the smallest of the M singular values of $A_M(n)$. For high singular value dispersions the data matrix $A_M(n)$ is nearly rank deficient. Therefore, a large condition number $\chi(n)$ shows that the data and correlation matrices are ill conditioned and that numerical errors, introduced by manipulations of those matrices, may be largely amplified. Generally, larger condition numbers will result in larger numerical error levels [4, 8, 10].

In the literature this behavior is also associated to the lack of persistency of excitation, a concept used in connection with the analysis of stability and exponential convergence of adaptive algorithms [11, 12, 13]. Loosely speaking, a persistently exciting input signal is one that is sufficiently varied so that the resulting data provides as much information about the input/output characteristics of stable linear systems on which it acts as would be provided by all possible inputs. This means, as will be presented more precisely later, that the concept of persistent excitation is strongly connected to the ability of identifying an unknown system. Moreover, it has also been found that keeping the input signal persistently exciting is of paramount importance in practical implementations of RLS algorithms, especially when exponential weighting is used. The absence of persistent excitation may result in instability of several adaptive algorithms [7, 14].

The appropriate conditions to assure excitation persistency are not unique, and are algorithm dependent [8, 14]. In the case of RLS algorithms an excitation persistency condition used to guarantee exponential convergence [11, 14] is that the vector $u_M^T(k) = [u(k)\ u(k-1)\ \ldots\ u(k-M+1)]$ must satisfy

$$0 < a_1 I \leq \frac{1}{N} \sum_{k=n}^{n+N-1} u_M(k) u_M^T(k) \leq a_2 I < \infty, \quad (2)$$

for all $n \geq 0$, some positive integer N, some $a_1 > 0$ and some $a_2 > 0$. Here I stands for the identity matrix and, given two matrices B and C, $B < C$ means that $C - B$ is positive definite and $B \leq C$ means that $C - B$ is positive semi-definite. Another persistency condition useful for numerical stability analysis of RLS algorithms [7], particularly of QR-

LSL algorithms [8, 9], is

$$0 < b_1 I \leq A_M^T(n) A_M(n) \leq b_2 I < \infty, \quad (3)$$

for all $n \geq M - 1$, some $b_1 > 0$ and some $b_2 > 0$.

Sometimes excitation persistency is directly associated to the condition number of the autocorrelation or data matrices. This fact and the different persistency conditions cause some confusion among beginners in the subject of adaptive filtering.

The goal of this note is to show that the excitation persistency conditions (2) and (3) are equivalent and to clearly relate them to a bounded condition number of the data matrix. It is shown that a weighted input signal energy uniformly bounded from above and below is an adequate restriction to guarantee that a bounded condition number of the data matrix is equivalent to the mentioned persistency conditions. Some related inequalities are also presented that are useful for numerical stability analysis of RLS algorithms and give some insight into this subject.

The paper is organized as follows. First, for the sake of clarity and precision the equivalence of the mentioned persistent conditions is presented in the form of a lemma. Then, some remarks and interpretations are presented. The relations of excitation persistency with concepts like predictability, spectral content of the excitation signal, identifiability, exponential convergence and numerical stability of RLS algorithms are briefly addressed in order to give an overview of the subject. At this point some references are given, where the interested reader can find further details. The paper is closed by some comments about the importance and application of the presented results. The demonstration of the Lemma is presented in the Appendix, so that it may be skipped in a first reading.

2. PERSISTENT EXCITATION AND CONDITION NUMBER OF DATA MATRIX

Lemma: Consider $0 < \lambda < 1$ and a pre-windowed signal $u(n)$ satisfying $u(n) = 0$ for $n < 0$ and $u(0) \neq 0$. Then, the following statements are equivalent:

(i) The weighted energy $\xi_u(n) = \sum_{\ell=0}^{n} \lambda^{n-\ell} u^2(\ell)$ of the input signal $u(n)$ is uniformly lower and upper bounded, and the spectral condition number of the data matrix $A_M(n)$ is uniformly upper bounded, that is, there exist positive constants β_1, β_2 and r_o such that

$$0 < \beta_1 \leq \xi_u(n) \leq \beta_2 < \infty, \text{ for } n \geq 0, \quad (4)$$

and

$$\chi(n) = \frac{s_{max}(n)}{s_{min}(n)} \leq r_o < \infty, \text{ for } n \geq M - 1, \quad (5)$$

where $s_{\min}(n)$ and $s_{\max}(n)$ are respectively the smallest and greatest singular values of the data matrix $A_M(n)$.

(ii) Positive constants b_1 e b_2 can be found such that for $n \geq M - 1$

$$0 < b_1 I \leq A_M^T(n) A_M(n) \leq b_2 I < \infty, \quad (6)$$

with $\mathbf{A}_M(n) = [A(i,j)] = [\lambda^{(n-i+1)/2} u(i-j)]$.

(iii) There are some positive integer N and some positive constants a_1 and a_2 such that for all $n \geq 0$

$$0 < a_1 \mathbf{I} \leq \frac{1}{N} \sum_{k=n}^{n+N-1} \mathbf{u}_M(k)\mathbf{u}_M^T(k) \leq a_2 \mathbf{I} < \infty, \quad (7)$$

where $\mathbf{u}_M^T(k) = [u(k)\ u(k-1)\ \ldots\ u(k-M+1)]$.

Moreover:

(i.a) If positive bounds b_1 and b_2 as in (ii) are known and $n \geq M - 1$, then the inequalities of (i) can be written as:

$$0 < b_1 \leq s_{\min}^2(n) \leq \xi_u(n) \leq s_{\max}^2(n) \leq b_2 < \infty \quad (8)$$

and

$$\chi(n) \leq \sqrt{\frac{b_2}{b_1}}. \quad (9)$$

(ii.a) If bounds β_1, β_2 and r_o of (i) are known and $n \geq M - 1$, the inequalities of (ii) can be written as follows. Lower bound:

$$0 < \frac{\beta_1}{r_o^2}\mathbf{I} \leq \frac{\xi_u(n)}{r_o^2}\mathbf{I} \leq s_{\min}^2(n)\mathbf{I} \leq \mathbf{A}_M^T(n)\mathbf{A}_M(n), \quad (10)$$

Upper bound:

$$\mathbf{A}_M^T(n)\mathbf{A}_M(n) \leq s_{\max}^2(n)\mathbf{I} \leq r_o^2\xi_u(n)\mathbf{I} \leq r_o^2\beta_2\mathbf{I} < \infty, \quad (11)$$

or

$$\begin{aligned}
\mathbf{A}_M^T(n)\mathbf{A}_M(n) &\leq s_{\max}^2(n)\mathbf{I} \leq \sum_{\ell=0}^{M-1}\xi_u(n-\ell)\mathbf{I} \\
&\leq M\beta_2\mathbf{I} < \infty, \quad (12)
\end{aligned}$$

or

$$\begin{aligned}
\mathbf{A}_M^T(n)\mathbf{A}_M(n) &\leq s_{\max}^2(n)\mathbf{I} \leq \frac{\xi_u(n)}{\lambda^{M-1}}\frac{1-\lambda^M}{1-\lambda}\mathbf{I} \\
&\leq \frac{\xi_u(n)}{\lambda^{M-1}}M\mathbf{I} \leq \frac{\beta_2 M}{\lambda^{M-1}}\mathbf{I} < \infty. \quad (13)
\end{aligned}$$

Moreover, if a positive integer N and positive bounds a_1 and a_2 as in (iii) are known, then for $n \geq N$ the inequalities of (ii) can be written as:

$$0 < N\lambda^{2N}a_1\mathbf{I} \leq \mathbf{A}_M^T(n)\mathbf{A}_M(n) \leq \frac{N}{1-\lambda}a_2\mathbf{I} < \infty. \quad (14)$$

(iii.a) If some positive bounds b_1 and b_2 satisfying (ii) are known, then the inequalities of (iii) can be written as:

$$\begin{aligned}
0 &< \frac{1-\lambda}{N}b_1\mathbf{I} \leq \frac{1}{N}\sum_{k=n}^{n+N-1}\mathbf{u}_M(k)\mathbf{u}_M^T(k) \\
&\leq \frac{1}{N\lambda^{N-1}}b_2\mathbf{I} < \infty, \quad (15)
\end{aligned}$$

with N being the smallest integer satisfying

$$N \geq \max\left\{\frac{\log(\lambda b_1/b_2)}{\log(\lambda)}, M\right\}.$$

Proof: The proof is presented in the Appendix.

Remark 1: With $\mathbf{h}^T = [h_0\ h_1\ \ldots\ h_{M-1}]$ and

$$\mathbf{A}_M^T(n)\mathbf{A}_M(n) = \sum_{k=0}^{n}\lambda^{n-k}\mathbf{u}_M(k)\mathbf{u}_M^T(k) \quad (16)$$

it follows that

$$\begin{aligned}
\mathbf{h}^T\mathbf{A}_M^T(n)\mathbf{A}_M(n)\mathbf{h} &= \sum_{k=0}^{n}\lambda^{n-k}\mathbf{h}^T\mathbf{u}_M(k)\mathbf{u}_M^T(k)\mathbf{h} \\
&= \sum_{k=0}^{n}\lambda^{n-k}y^2(k) \triangleq \xi_y(n), \quad (17)
\end{aligned}$$

where

$$y(k) = \sum_{\ell=0}^{M-1}h_\ell u(k-\ell). \quad (18)$$

Now, without loss of generality, considering only unit norm vectors $\mathbf{h}$, from (6) it follows that

$$0 < b_1 \leq \xi_y(n) \leq b_2 < \infty, \quad \forall \mathbf{h} \neq \mathbf{0}.$$

This implies that $y(n)$ cannot converge uniformly towards zero as n increases, unless $\mathbf{h} = \mathbf{0}$ is chosen (see (17)). On the other hand, if $\mathbf{h} \neq \mathbf{0}$ could be found such that $\sum_{\ell=0}^{M-1}h_\ell u(n-\ell) = 0$, it would imply that $u(n) = -\sum_{\ell=1}^{M-1}(h_\ell/h_0)u(n-\ell)$, for $h_0 \neq 0$[1]. This would mean that $u(n)$ is perfectly predictable from at least $M-1$ samples from its past. Since (18) can not vanish identically as k increases, $u(n)$ can not converge to such a situation with increasing n. In this sense, persistence condition (ii) guarantees that $u(n)$ is not a predictable sequence given $M-1$ samples from its past. A similar reasoning was used in the analysis of numerical stability of QR-LSL algorithms [8, 9].

Remark 2: A useful frequency domain interpretation follows quite straightforward from Remark 1 [15]. Considering initially a deterministic signal $u(n)$, the frequency domain version of Equation (18) is

$$Y(e^{j\omega}) = H(e^{j\omega})U(e^{j\omega}),$$

with $Y(e^{j\omega})$, $H(e^{j\omega})$ and $U(e^{j\omega})$ being respectively the Discrete Time Fourier Transforms of $y(n)$, h_ℓ, $0 \leq \ell \leq M-1$, and $u(n)$. Since any possible $Y(e^{j\omega})$ can not vanish identically, it follows that at least $U(e^{j\omega})$ is nonzero at M distinct frequencies. Note that if $U(e^{j\omega}) \neq 0$ only at $M-1$, or less, distinct frequencies, then h_ℓ could be chosen such

---

[1]If $h_0 = 0$ there would exist $0 < \ell_0 \leq M-1$ such that $h_{\ell_0} \neq 0$ and $u(n-\ell_0) = -\sum_{\ell=\ell_0+1}^{M-1}(h_\ell/h_{\ell_0})u(n-\ell)$. Thus, for any n, $u(n-\ell_0)$ would be predictable from less than $M-1$ samples from its past and so would be $u(n)$.

that the $M - 1$ zeros of its z-Transform $H(z)$ lie on the unit circumference where $U(e^{j\omega})$ is nonzero.

If $u(n)$ is a stationary stochastic sequence it results from statement (iii) of the Lemma that $E\{u_M(k)u_M^T(k)\}$ is positive definite, with $E\{.\}$ being the statistical expectation operator. Thus, considering $\mathbf{h} \neq \mathbf{0}$,

$$\mathbf{h}^T E\{\mathbf{u}_M(k)\mathbf{u}_M^T(k)\}\mathbf{h} = E\{\mathbf{h}^T\mathbf{u}_M(k)\mathbf{u}_M^T(k)\mathbf{h}\}$$
$$= E\{y^2(k)\} > 0.$$

Therefore, it is not possible to obtain $\mathbf{h} \neq \mathbf{0}$ such that $y(k)$ of Equation (18) vanish identically. Introducing the power spectral density function of $u(n)$,

$$\Psi_u(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} E\{\mathbf{u}_M(k)\mathbf{u}_M^T(k)\}e^{-j\omega k},$$

it follows that

$$E\{\mathbf{h}^T\mathbf{u}_M(k)\mathbf{u}_M^T(k)\mathbf{h}\} =$$
$$E\{y^2(k)\} = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \Psi_u(e^{j\omega})\left|H(e^{j\omega})\right|^2 d\omega. \qquad (19)$$

Thus, if $E\{\mathbf{u}_M(k)\mathbf{u}_M^T(k)\}$ is positive definite it results that $\Psi_u(e^{j\omega})$, the power spectral density function of $u(n)$, must be at least nonzero at M distinct frequencies. Again, note that if $\Psi_u(e^{j\omega}) \neq 0$ only at $M - 1$, or less, distinct frequencies, then h_ℓ could be chosen such that the $M - 1$ zeros of its z-Transform $H(z)$ lie on the unit circumference exactly where $\Psi_u(e^{j\omega})$ is nonzero. In this case the integral in Equation (??) would be zero.

Remark 3: The singular value decomposition of $\mathbf{A}_M(n)$ yields

$$\mathbf{A}_M(n) = \mathbf{U}_M(n)\mathbf{S}_M(n)\mathbf{V}_M(n), \qquad (20)$$

where $\mathbf{S}_M(n) = \text{diag}\{s_i(n), 1 \leq i \leq M\}$ and $\mathbf{U}_M(n)$ and $\mathbf{V}_M(n)$ are orthogonal matrices, respectively with dimensions $(n+1) \times M$ and $M \times M$. Thus, the corresponding autocorrelation matrix and its inverse are given by

$$\Phi_M(n) = \mathbf{A}_M^T(n)\mathbf{A}_M(n) = \mathbf{V}_M^T(n)\mathbf{S}_M^2(n)\mathbf{V}_M(n) \qquad (21)$$

and

$$\Phi_M^{-1}(n) = \mathbf{V}_M^T(n)\mathbf{S}_M^{-2}(n)\mathbf{V}_M(n), \qquad (22)$$

where $\mathbf{S}_M^{-2}(n) = \text{diag}\{s_i^{-2}(n), 1 \leq i \leq M\}$. Therefore, excitation persistency condition (ii) (see (i.a)) guarantees the largest eigenvalue of the inverse autocorrelation matrix to be upper bounded by b_1^{-1} and its smallest eigenvalue to be positive and bounded away from zero by b_2^{-1} for all $n \geq M - 1$. Roughly speaking, this assures that the recursive update of $\Phi_M^{-1}(k)$, characteristic for several RLS algorithms, behaves well. This behavior is of paramount importance to assure exponential convergence [11, 14] and numerical stability of RLS algorithms that deal directly with $\Phi_M^{-1}(k)$ [7].

Remark 4: The upper bound given by (12) is not only tighter then the one of (13), which may be of interest when the explicit dependence of the energy $\xi_u(n)$ is desired, but also, for sufficiently high condition numbers, tighter than the upper bound of (11). Moreover, (12) shows that for an ill-conditioned signal with finite energy the upper bound of (6) does not depend of the condition number, as it is determined only by the order M and the input signal energy bound. As shown by (10), it is the lower bound of (6) that depends on the condition number, being inversely proportional to it. Therefore, for a data matrix where $\chi(n)$ increases with n, at least one eigenvalue of $\Phi_M^{-1}(n)$ will explode. As observed in [4, 14] such a behavior causes severe numerical problems.

Remark 5: Considering the identification problem of a system excited by $u(n)$ that produces an output $d(n)$, conditions (i) to (iii) guarantee that the solution of (1) is well defined and unique for all $n \geq M - 1$. In other words, they guarantee the identification of an unique FIR model of M coefficients for the system generating $d(n)$. Thus, the input signal $u(n)$ is sufficiently varied such that both $u(n)$ and $d(n)$ provide enough information about the input/output characteristics of the system to be identified. This enables a signal $u(n)$ satisfying (i) to (iii) to be named a persistently exciting signal of degree M.

3. FINAL REMARKS

As observed, persistency of excitation is strongly connected to identifiability conditions. This is of fundamental importance for many applications in the telecommunications area. For example, the equalization of a linear communication channel is the attempt to identify a delayed version of the channel's inverse transfer function. Following Remark 5, if the received signal (at the input of the equalizer) is persistent of degree M then a well defined FIR solution of order $M - 1$ can be found by using adaptive algorithms from the RLS family. Moreover, following Remarks 3 and 4, the persistency of the received signal is important to guarantee a stable numerical behavior of the adaptive algorithm. Of course, it can not be expected that practical signals always satisfy persistency conditions, so that an important issue for practical applications is to investigate how to design algorithms that present stable behaviour even for non-persistent signals (for more details on this subject the interested reader is referred to [8, 9, 10]). To handle this problem, a clear comprehension and a precise definition of what excitation persistency means are essential.

APPENDIX

Proof of the Lemma:

((i) ⇐ (ii)) From (20) and (21) it readily follows

$$b_2\mathbf{I} - \mathbf{A}_M^T(n)\mathbf{A}_M(n) \geq 0 \Leftrightarrow b_2\mathbf{I} - \mathbf{S}_M^2(n) \geq 0$$
$$\Leftrightarrow b_2 \geq s_{\text{max}}^2(n) \qquad (23)$$

and

$$\mathbf{A}_M^T(n)\mathbf{A}_M(n) - b_1\mathbf{I} \geq 0 \Leftrightarrow \mathbf{S}_M^2(n) - b_1\mathbf{I} \geq 0$$
$$\Leftrightarrow b_1 \leq s_{\text{min}}^2(n). \qquad (24)$$

Consequently, if constants $0 < b_1 < b_2 < \infty$ exist satisfying (6) then for any[2] $n \geq M - 1$ the excitation persistency condition (6) can be written as

$$0 < b_1\mathbf{I} \leq s_{min}^2(n)\mathbf{I} \leq \mathbf{A}_M^T(n)\mathbf{A}_M(n) \leq s_{max}^2(n)\mathbf{I}$$
$$\leq b_2\mathbf{I} < \infty \tag{25}$$

and the condition number $\chi(n)$ is upper bounded:

$$\chi(n) \leq \sqrt{\frac{b_2}{b_1}}. \tag{26}$$

Moreover, considering unit norm vectors $\mathbf{h}$ in (17), it results from (25) that

$$0 < b_1 \leq s_{min}^2(n) \leq \xi_y(n) \leq s_{max}^2(n) \leq b_2 < \infty$$

for $n \geq M - 1$. In particular, $y(k) = u(k)$ if $\mathbf{h}^T = [1\ 0\ \ldots\ 0]$. As a consequence, for $n \geq M - 1$ one obtains

$$0 < b_1 \leq s_{min}^2(n) \leq \xi_u(n) \leq s_{max}^2(n) \leq b_2 < \infty.$$

If $n \geq 0$ is to be taken instead of $n \geq M - 1$, the lower bound β_1 will be at least $min\{b_1,\ u(0)\lambda^{M-1}\}$ and the upper bound β_2 will be $max\{b_2,\ \xi_u(k),\ k = 0,\ 1,\ \ldots,\ M - 1\}$.

((i) $\Rightarrow$ (ii)) On the other hand, the condition number $\chi(n)$ may be bounded even in situations where $s_{max}(n)$ and $s_{min}(n)$ are both arbitrarily low or arbitrarily high. For that reason it is necessary to introduce a further condition (as in (i)) to make the existence of an upper bound for $\chi(n)$ equivalent to (6). One possibility, as used in (i), is to require the input signal energy to be lower and upper bounded as in (4).

Considering now (4) and (5), a lower bound for $s_{min}^2(n)$ and an upper bound for $s_{max}^2(n)$ can be obtained as follows.

The autocorrelation matrix $\Phi_M(n) = \mathbf{A}_M^T(n)\mathbf{A}_M(n)$ is symmetric and at least positive semi-definite. Therefore, it may be decomposed as in (21) with $\mathbf{S}_M^2(n) = diag\{s_i^2(n)\}\big|_{i=1}^M$ being the eigenvalue matrix and $\mathbf{V}_M(n)$ being the eigenvector matrix with M orthonormal columns $\mathbf{v}_i(n)$. An unit norm column vector $\mathbf{h}$ of length M can be expanded using the eigenvector basis:

$$\mathbf{h} = \sum_{i=1}^M \alpha_i\mathbf{v}_i(n),\quad\text{with}\quad \sum_{i=1}^M \alpha_i^2 = 1.$$

In this case

$$\xi_y(n) = \mathbf{h}^T\mathbf{A}_M^T(n)\mathbf{A}_M(n)\mathbf{h} = \sum_{i=1}^M \alpha_i^2 s_i^2(n)$$

follows, and consequently

$$s_{min}^2(n) \leq \xi_y(n) \leq s_{max}^2(n)$$

holds for any unit norm vector $\mathbf{h}$. Therefore, with $\mathbf{h}^T = [1, 0, \ldots, 0]$ and (5), it results

$$\frac{s_{max}^2(n)}{r_o^2} \leq \xi_u(n) \leq r_o^2 s_{min}^2(n).$$

Now, considering (4) and (21), the following bounds follow:

$$0 < \frac{\beta_1}{r_o^2}\mathbf{I} \leq \frac{\xi_u(n)}{r_o^2}\mathbf{I} \leq s_{min}^2(n)\mathbf{I} \leq \mathbf{A}_M^T(n)\mathbf{A}_M(n) \tag{27}$$

$$\mathbf{A}_M^T(n)\mathbf{A}_M(n) \leq s_{max}^2(n)\mathbf{I} \leq r_o^2\xi_u(n)\mathbf{I} \leq r_o^2\beta_2\mathbf{I} < \infty. \tag{28}$$

Alternatively, other upper bounds can be obtained. For this purpose one should note that the diagonal elements of $\Phi_M(n) = [\phi_M^{(i,j)}(n)]$ are delayed energies of the input signal, i.e. $\phi_M^{(i,i)}(n) = \xi_u(n - i + 1)$, for $1 \leq i \leq M$. This result may be found in [4] or may be deduced directly from $\Phi_M(n) = \mathbf{A}_M^T(n)\mathbf{A}_M(n) = \sum_{k=0}^n \lambda^{n-k}\mathbf{u}(k)\mathbf{u}^T(k)$.

Thus, from

$$\sum_{\ell=0}^{M-1} \xi_u(n - \ell) = \text{trace}\left\{\mathbf{A}_M^T(n)\mathbf{A}_M(n)\right\} = \sum_{i=1}^M s_i^2(n)$$

one obtains

$$s_{max}^2(n) \leq \sum_{i=1}^M s_i^2(n) = \sum_{\ell=0}^{M-1} \xi_u(n - \ell) \leq M\beta_2. \tag{29}$$

Moreover, since $\xi_u(n) = \lambda^i\xi_u(n - i) + \sum_{\ell=0}^{i-1}\lambda^\ell u^2(n - \ell)$, which implies $\xi_u(n)/\lambda^i \geq \xi_u(n - i)$, another upper bound may be obtained:

$$s_{max}^2(n) \leq \sum_{i=0}^{M-1}\xi_u(n-i) \leq \frac{\xi_u(n)}{\lambda^{M-1}}\sum_{\ell=0}^{M-1}\lambda^\ell =$$
$$\frac{\xi_u(n)}{\lambda^{M-1}}\frac{1-\lambda^M}{1-\lambda} \leq \frac{\xi_u(n)}{\lambda^{M-1}}M \leq \frac{\beta_2 M}{\lambda^{M-1}}. \tag{30}$$

((ii) $\Leftarrow$ (iii)) Before going on with the proof of the equivalence of (ii) and (iii) it should be noted that given two $M \times M$ matrices $\mathbf{B}$ and $\mathbf{C}$ such that

$$\mathbf{h}^T\mathbf{C}\mathbf{h} \leq \mathbf{h}^T\mathbf{B}\mathbf{h}$$

for all $\mathbf{h} \neq 0$, then $\mathbf{C} \leq \mathbf{B}$. Thus, for proving the equivalence of (ii) and (iii), inequalities (6) and (7) could be taken multiplied from the left by $\mathbf{h}^T$ and from the rigth by $\mathbf{h}$, for any $\mathbf{h}$ satisfying $\mathbf{h}^T\mathbf{h} = 1$.

Having this in mind, from (7) one obtains for any $\mathbf{h} \neq 0$

$$\frac{1}{N}\sum_{k=n}^{n+N-1} y^2(k) \leq a_2, \forall n \geq 0 \quad \Rightarrow \quad y^2(k) \leq Na_2, \forall k \geq 0. \tag{31}$$

Thus, the upper bound of (6) follows immediately:

$$\sum_{k=0}^n \lambda^{n-k}y^2(k) \leq Na_2\sum_{k=0}^n \lambda^{n-k} = Na_2\frac{1-\lambda^n}{1-\lambda} \leq \frac{Na_2}{1-\lambda}. \tag{32}$$

Considering $n \geq N$, using

$$\frac{1}{N}\sum_{k=n}^{n+N-1} y^2(k) \geq a_1, \tag{33}$$

---

[2]Considering $u(n) = 0$ for $n < 0$ and $u(0) \neq 0$, $n \geq M - 1 > 0$ should be taken, because for $n < M - 1$ the data matrix $\mathbf{A}_M(n)$ has at least one null column and consequently $s_{min}(n) = 0$.

90

obtained from (7) and taking L as the integer part of n/N, the lower bound of (6) is obtained as follows:

$$
\sum_{k=0}^{n} \lambda^{n-k} y^2(k) =
$$
$$
\sum_{\ell=0}^{L-1} \sum_{k=\ell N}^{(\ell+1)N-1} \lambda^{n-k} y^2(k) + \sum_{k=LN}^{n} \lambda^{n-k} y^2(k) =
$$
$$
\sum_{\ell=0}^{L-1} \lambda^{n-\ell N} \sum_{k=\ell N}^{(\ell+1)N-1} \lambda^{\ell N-k} y^2(k) + \sum_{k=LN}^{n} \lambda^{n-k} y^2(k) \geq
$$
$$
\sum_{\ell=0}^{L-1} \lambda^{n-\ell N} N a_1 \geq
$$
$$
N a_1 \lambda^{2N} \sum_{\ell=0}^{L-1} \lambda^{n-(\ell+2)N} \geq
$$
$$
N a_1 \lambda^{2N}
$$

(34)

For $n < N$ it readily follows that

$$
\sum_{k=0}^{n} \lambda^{n-k} y^2(k) \geq \lambda^N y^2(0) ,
$$

(35)

setting a lower bound for this case.

((ii) $\Rightarrow$ (iii)) With $N \geq M$ and $n \geq 0$, the upper bound of (7) follows immediately from (6), (16) and (17):

$$
\frac{1}{N} \sum_{k=n}^{n+N-1} y^2(k) \leq \frac{1}{N\lambda^{N-1}} \sum_{k=n}^{n+N-1} \lambda^{n+N-1-k} y^2(k)
$$
$$
\leq \frac{1}{N\lambda^{N-1}} \sum_{k=0}^{n+N-1} \lambda^{n+N-1-k} y^2(k)
$$
$$
\leq \frac{1}{N\lambda^{N-1}} b_2
$$

(36)

To obtain the lower bound of (7) one should again take $N \geq M$ such that from (6), (16) and (17)

$$
b_1 \leq \sum_{k=0}^{n+N-1} \lambda^{n+N-1-k} y^2(k)
$$
$$
\leq \lambda^N \sum_{k=0}^{n-1} \lambda^{n-1-k} y^2(k) + \sum_{k=n}^{n+N-1} \lambda^{n+N-1-k} y^2(k)
$$
$$
\leq \lambda^N b_2 + \sum_{k=n}^{n+N-1} y^2(k)
$$

(37)

follows. Taking now

$$
N \geq \max \left\{ \frac{\log(\lambda b_1/b_2)}{\log(\lambda)}, M \right\}
$$

such that

$$
\lambda^N b_2 \leq \lambda b_1 ,
$$

inequality (37) leads to

$$
\frac{1}{N} \sum_{k=n}^{n+N-1} y^2(k) \geq \frac{1-\lambda}{N} b_1.
$$

(38)

REFERENCES

[1] G.G. Glentis, K. Berberidis, S. Theodorakis: Efficient Least Squares Adaptive Algorithms for FIR Transversal Filtering. *IEEE Signal Processing Magazine*, Vol. 16, n. 4, July 1999, pp.13-41.

[2] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsh, H. Puder, T. Schertler, G. Schmidt, J. Tilp: Acoustic Echo Cancelation Control: An Application of Very-High-Order Adaptive Filters, *IEEE Signal Processing Magazine*, Vol. 16, n. 4, July 1999, pp. 42-69.

[3] M.D. Miranda, M. Gerken, Magno T. M. da Silva: Efficient Implementation of the a priori error-feedback LSL algorithm, *IEE Electronics Letters*, Vol. 35, No. 16, August 1999, pp. 1308-1309.

[4] S. Haykin: *Adaptive Filter Theory*, Prentice Hall Int., 2nd. Ed., 1991.

[5] C. Virouland, M. Terré: Implantation a Base D'Operatours CORDIC d'un Algorithme de Moindres Carrés Rapides par Decomposition QR, Annales du XV Colloque Gretsi, Juan Les Pins, Septembre 1995, pp.973-976.

[6] Y. Wang, Y. Terada, M. Matsui, K. Iida, K. Nakayama: Development of High Quality Acoustic Subband Echo Canceller Using Dual-Filter Structure and Fast Recursive Least Squares Algorithm, *Proc. IEEE ICASSP-2000*, Istanbul, Turkey, June 2000, Vol. VI, pp.3674-3677.

[7] S. Ljung, L. Ljung: Error propagation properties of recursive least-squares adaptation algorithms, *Automatica*, vol.21, no.2, 1985, pp.157-167.

[8] P.A. Regalia: Numerical stability properties of a QR-based fast least squares algorithm, *IEEE Trans. on Signal Processing*, vol.41, no.6, June 1993, pp.2096-2109.

[9] M. Miranda, M. Gerken: A Hybrid Least Squares QR-Lattice Algorithm Using A Priori Errors, *IEEE Trans. on Signal Processing*, December 1997, pp.2900-2911.

[10] D.T.M. Slock: Backward consistency concept and round-off error propagation dynamics in recursive least-squares algorithms, *Optical Engineering*, vol. 31, no. 6, June 1992, pp.1153-1169.

[11] R.M. Johnstone, C.R. Johnson Jr., R.R. Bitmead, B.D.O. Anderson: Exponential convergence of recursive least squares with exponential forgetting factor, *System and Control Letters*, Vol.2, August 1982, pp.77-82.

[12] T. Söderström, P. Stoica: *Instrumental variable Methods for System Identification*, Springer-Verlag, Berlin, 1983.

[13] B.D.O. Anderson, R.R. Bitmead, C.R. Johnson Jr., P.V. Kokotovic, R.L. Kosiet, I.M.Y. Mareels, L. Praly, B.D. Riedle: *Stability of Adaptive Systems: Passivity and Averaging Analysis*, MIT Press, Cambridge, Massachusetts, 1986.

[14] C.R. Johnson, Jr.: *Lectures on Adaptive Parameter Estimation*, Prentice Hall, Englewood Cliffs, New Jersey, 1988.

[15] P.A. Regalia: *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, 1995.