

# New Insights into the RLS Algorithm

## Jacob Benesty

INRS-EMT, Université du Québec, 800 de la Gauchetière Ouest, Suite 6900, Montréal, Québec, Canada H5A 1K6  
Email: benesty@inrs-emt.ubc.ca

## Tomas Gänsler

Agere Systems Inc., 1110 American Parkway NE, Allentown, PA 18109-3229, USA  
Email: gaensler@agere.com

Received 21 July 2003; Revised 9 October 2003; Recommended for Publication by Hideaki Sakai

The recursive least squares (RLS) algorithm is one of the most popular adaptive algorithms that can be found in the literature, due to the fact that it is easily and exactly derived from the normal equations. In this paper, we give another interpretation of the RLS algorithm and show the importance of linear interpolation error energies in the RLS structure. We also give a very efficient way to recursively estimate the condition number of the input signal covariance matrix thanks to fast versions of the RLS algorithm. Finally, we quantify the misalignment of the RLS algorithm with respect to the condition number.

**Keywords and phrases:** adaptive algorithms, normal equations, RLS, fast RLS, condition number, linear interpolation.

## 1. INTRODUCTION

Adaptive algorithms play a very important role in many diverse applications such as communications, acoustics, speech, radar, sonar, seismology, and biomedical engineering [1, 2, 3, 4]. Among the most well-known adaptive filters are the recursive least squares (RLS) and fast RLS (FRLS) algorithms. The latter is a computationally fast version of the former. Even though the RLS is not as widely used in practice as the least mean square (LMS), it has a very significant theoretical interest since it belongs to the Kalman filters family [5]. Also, many adaptive algorithms (including the LMS) can be seen as approximations of the RLS. Therefore, there is always a need to interpret and understand in new ways the different variables that are built in the RLS algorithm.

The convergence rate, the misalignment, and the numerical stability of adaptive algorithms depend on the condition number of the input signal covariance matrix. The higher this condition number is, the slower the convergence rate is and/or the less stable the algorithm is. For ill-conditioned input signals (like speech), the LMS converges very slowly and the stability and the misalignment of the FRLS are more affected. Thus, there is a need to compute the condition number in order to monitor the behavior of adaptive filters. Unfortunately, there are no simple ways to estimate this condition number.

The objective of this paper is threefold. We first give another interpretation of the RLS algorithm and show the importance of linear interpolation error energies in the RLS

structure. Second, we derive a very simple way to recursively estimate the condition number. The proposed method is very efficient when combined with the FRLS algorithm; it requires only  $L$  more multiplications per iteration, where  $L$  is the length of the adaptive filter. Finally, we show exactly how the misalignment of the RLS algorithm is affected by the condition number, output signal-to-noise ratio (SNR), and parameter choice.

## 2. RLS ALGORITHM

In this section, we briefly derive the classical RLS algorithm in a system identification context. We try to estimate the impulse response of an unknown, linear, and time-invariant system by using the least squares method.

We define the a priori error signal  $e(n)$  at time  $n$  as follows:

$$e(n) = y(n) - \hat{y}(n), \quad (1)$$

where

$$y(n) = \mathbf{h}_t^T \mathbf{x}(n) + w(n) \quad (2)$$

is the system output,

$$\mathbf{h}_t = [h_{t,0} \ h_{t,1} \ \cdots \ h_{t,L-1}]^T \quad (3)$$

is the true (subscript  $t$ ) impulse response of the system, the superscript  $T$  denotes the transpose of a vector or a matrix,

$$\mathbf{x}(n) = [x(n) \ x(n-1) \ \cdots \ x(n-L+1)]^T \quad (4)$$

is a vector containing the last  $L$  samples of the input signal  $x$ , and  $w$  is a white Gaussian noise (uncorrelated with  $x$ ) with variance  $\sigma_w^2$ . In (1),

$$\hat{y}(n) = \mathbf{h}^T(n-1)\mathbf{x}(n) \quad (5)$$

is the model filter output and

$$\mathbf{h}(n-1) = [h_0(n-1) \ h_1(n-1) \ \cdots \ h_{L-1}(n-1)]^T \quad (6)$$

is the model filter of length  $L$ .

We also define the popular RLS error criterion with respect to the modelling filter:

$$J_{LS}(n) = \sum_{m=0}^n \lambda^{n-m} [y(m) - \mathbf{h}^T(n)\mathbf{x}(m)]^2, \quad (7)$$

where  $\lambda$  ( $0 < \lambda < 1$ ) is a forgetting factor. The minimization of (7) leads to the normal equations

$$\mathbf{R}(n)\mathbf{h}(n) = \mathbf{r}(n), \quad (8)$$

where

$$\mathbf{R}(n) = \sum_{m=0}^n \lambda^{n-m} \mathbf{x}(m)\mathbf{x}^T(m) \quad (9)$$

is an estimate of the input signal covariance matrix and

$$\mathbf{r}(n) = \sum_{m=0}^n \lambda^{n-m} \mathbf{x}(m)y(m) \quad (10)$$

is an estimate of the cross-correlation vector between  $x$  and  $y$ .

From the normal equations (8), we easily derive the classical update for the RLS algorithm [1, 3]:

$$\begin{aligned} e(n) &= y(n) - \mathbf{h}^T(n-1)\mathbf{x}(n), \\ \mathbf{h}(n) &= \mathbf{h}(n-1) + \mathbf{R}^{-1}(n)\mathbf{x}(n)e(n). \end{aligned} \quad (11)$$

A fast version of this algorithm can be deduced by computing recursively the a priori Kalman gain vector  $\mathbf{k}'(n) = \mathbf{R}^{-1}(n-1)\mathbf{x}(n)$  [1]. The a posteriori Kalman gain vector  $\mathbf{k}(n) = \mathbf{R}^{-1}(n)\mathbf{x}(n)$  is related to  $\mathbf{k}'(n)$  by [1]:

$$\mathbf{k}(n) = \lambda^{-1}\varphi(n)\mathbf{k}'(n), \quad (12)$$

where

$$\varphi(n) = \frac{\lambda}{\lambda + \mathbf{x}^T(n)\mathbf{R}^{-1}(n-1)\mathbf{x}(n)}. \quad (13)$$

### 3. AN RLS ALGORITHM BASED ON THE INTERPOLATION ERRORS

In this section, we show another way to write the RLS algorithm. This new formulation, based on linear interpolation, gives a better insight of the adaptive algorithm structure.

We would like to minimize the criterion [6, 7]:

$$\begin{aligned} J_{\text{int},i}(n) &= \sum_{m=0}^n \lambda^{n-m} \left[ -\sum_{l=0}^{L-1} c_{il}(n)x(m-l) \right]^2 \\ &= \sum_{m=0}^n \lambda^{n-m} [ -\mathbf{c}_i^T(n)\mathbf{x}(m) ]^2 \\ &= \mathbf{c}_i^T(n)\mathbf{R}(n)\mathbf{c}_i(n), \end{aligned} \quad (14)$$

with the constraint

$$\mathbf{c}_i^T(n)\mathbf{u}_i = c_{ii} = -1, \quad (15)$$

where

$$\mathbf{c}_i(n) = [c_{i0}(n) \ c_{i1}(n) \ \cdots \ c_{i(L-1)}(n)]^T \quad (16)$$

is the  $i$ th ( $0 \leq i \leq L-1$ ) interpolator of the signal  $\mathbf{x}(n)$  and

$$\mathbf{u}_i = [0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0]^T \quad (17)$$

is a vector of length  $L$ , where its  $i$ th component is equal to one and all others are zero. By using the Lagrange multipliers, it is easy to see that the solution to this optimization problem is

$$\mathbf{R}(n)\mathbf{c}_i(n) = -E_i(n)\mathbf{u}_i, \quad (18)$$

where

$$E_i(n) = \mathbf{c}_i^T(n)\mathbf{R}(n)\mathbf{c}_i(n) = \frac{1}{\mathbf{u}_i^T\mathbf{R}^{-1}(n)\mathbf{u}_i} \quad (19)$$

is the interpolation error energy.

From (18) we find

$$-\frac{\mathbf{c}_i(n)}{E_i(n)} = \mathbf{R}^{-1}(n)\mathbf{u}_i, \quad (20)$$

hence the  $i$ th column of  $\mathbf{R}^{-1}(n)$  is  $-\mathbf{c}_i(n)/E_i(n)$ . We can now deduce that  $\mathbf{R}^{-1}(n)$  can be factorized as follows:

$$\begin{aligned} \mathbf{R}^{-1}(n) &= \begin{bmatrix} 1 & -c_{10}(n) & \cdots & -c_{(L-1)0}(n) \\ -c_{01}(n) & 1 & \cdots & -c_{(L-1)1}(n) \\ \vdots & \vdots & \ddots & \vdots \\ -c_{0(L-1)}(n) & -c_{1(L-1)}(n) & \cdots & 1 \end{bmatrix} \\ &\times \begin{bmatrix} \frac{1}{E_0(n)} & 0 & \cdots & 0 \\ 0 & \frac{1}{E_1(n)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{E_{L-1}(n)} \end{bmatrix} \\ &= \mathbf{C}^T(n)\mathbf{D}_e^{-1}(n). \end{aligned} \quad (21)$$

Furthermore, since  $\mathbf{R}^{-1}(n)$  is a symmetric matrix, (21) can be written as

$$\begin{aligned} \mathbf{R}^{-1}(n) &= \begin{bmatrix} \frac{1}{E_0(n)} & 0 & \cdots & 0 \\ 0 & \frac{1}{E_1(n)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{E_{L-1}(n)} \end{bmatrix} \\ &\times \begin{bmatrix} 1 & -c_{01}(n) & \cdots & -c_{0(L-1)}(n) \\ -c_{10}(n) & 1 & \cdots & -c_{1(L-1)}(n) \\ \vdots & \vdots & \ddots & \vdots \\ -c_{(L-1)0}(n) & -c_{(L-1)1}(n) & \cdots & 1 \end{bmatrix} \\ &= \mathbf{D}_e^{-1}(n)\mathbf{C}(n). \end{aligned} \quad (22)$$

The first and last columns of  $\mathbf{R}^{-1}(n)$  contain, respectively, the normalized forward and backward predictors and all the columns between contain the normalized interpolators.

We define, respectively, the a priori and a posteriori interpolation error signals as

$$e_i(n) = -\mathbf{c}_i^T(n-1)\mathbf{x}(n), \quad \varepsilon_i(n) = -\mathbf{c}_i^T(n)\mathbf{x}(n). \quad (23)$$

Using expression (22), we now have an interesting interpretation of the a priori and a posteriori Kalman gain vectors:

$$\begin{aligned} \mathbf{k}'(n) &= \mathbf{R}^{-1}(n-1)\mathbf{x}(n) \\ &= \begin{bmatrix} \frac{e_0(n)}{E_0(n-1)} & \frac{e_1(n)}{E_1(n-1)} & \cdots & \frac{e_{L-1}(n)}{E_{L-1}(n-1)} \end{bmatrix}^T, \\ \mathbf{k}(n) &= \mathbf{R}^{-1}(n)\mathbf{x}(n) \\ &= \begin{bmatrix} \frac{\varepsilon_0(n)}{E_0(n)} & \frac{\varepsilon_1(n)}{E_1(n)} & \cdots & \frac{\varepsilon_{L-1}(n)}{E_{L-1}(n)} \end{bmatrix}^T. \end{aligned} \quad (24)$$

The  $i$ th component of the a priori (resp., a posteriori) Kalman gain vector is the  $i$ th a priori (resp., a posteriori) interpolation error signal normalized with the  $i$ th interpolation error energy at time  $n-1$  (resp.,  $n$ ).

Writing (18) at time  $n$  and  $n-1$ , we obtain

$$-\frac{\mathbf{R}(n)\mathbf{c}_i(n)}{E_i(n)} = \mathbf{u}_i = -\frac{\lambda\mathbf{R}(n-1)\mathbf{c}_i(n-1)}{\lambda E_i(n-1)}. \quad (25)$$

Replacing  $\lambda\mathbf{R}(n-1)$  in (25) by

$$\lambda\mathbf{R}(n-1) = \mathbf{R}(n) - \mathbf{x}(n)\mathbf{x}^T(n), \quad (26)$$

we get

$$\mathbf{c}_i(n) = \frac{E_i(n)}{\lambda E_i(n-1)} [\mathbf{c}_i(n-1) + \mathbf{k}(n)e_i(n)]. \quad (27)$$

Now, if we premultiply both sides of (27) by  $\mathbf{u}_i^T$ , we can easily find that

$$E_i(n) = \lambda E_i(n-1) + e_i(n)e_i(n). \quad (28)$$

This means that the interpolation error energy can be computed recursively. This relation is well known for the forward ( $i=0$ ) and backward ( $i=L$ ) predictors [1]. It is used to obtain fast versions of the RLS algorithm.

Also, the interpolator vectors can be computed recursively:

$$\mathbf{c}_i(n) = \frac{1}{1 - k_i(n)e_i(n)} [\mathbf{c}_i(n-1) + \mathbf{k}(n)e_i(n)]. \quad (29)$$

If we premultiply both sides of (29) by  $-\mathbf{x}^T(n)$ , we obtain a relation between the a priori and a posteriori interpolation error signals:

$$\frac{\varepsilon_i(n)}{e_i(n)} = \frac{\varphi(n)}{1 - k_i(n)e_i(n)}. \quad (30)$$

We now give another interpretation of the RLS algorithm:

$$\begin{aligned} h_l(n) &= h_l(n-1) + \frac{\varepsilon_l(n)e(n)}{E_l(n)} \\ &= h_l(n-1) + \varphi(n) \frac{e_l(n)e(n)}{\lambda E_l(n-1)}, \quad l = 0, 1, \dots, L-1. \end{aligned} \quad (31)$$

In Sections 4 and 5, we will show how the linear interpolation error energies appear naturally in the condition number formulation.

#### 4. CONDITION NUMBER OF THE INPUT SIGNAL COVARIANCE MATRIX

Usually, the condition number is computed by using the 2-norm matrix. In the context of RLS equations, it is more convenient to use a different norm as explained below.

The covariance matrix  $\mathbf{R}(n)$  is symmetric and positive definite. It can be diagonalized as follows:

$$\mathbf{Q}^T(n)\mathbf{R}(n)\mathbf{Q}(n) = \Lambda(n), \quad (32)$$

where

$$\begin{aligned} \mathbf{Q}^T(n)\mathbf{Q}(n) &= \mathbf{Q}(n)\mathbf{Q}^T(n) = \mathbf{I}, \\ \Lambda(n) &= \text{diag} [\lambda_0(n), \lambda_1(n), \dots, \lambda_{L-1}(n)], \end{aligned} \quad (33)$$

and  $0 < \lambda_0(n) \leq \lambda_1(n) \leq \dots \leq \lambda_{L-1}(n)$ . By definition, the square root of  $\mathbf{R}(n)$  is

$$\mathbf{R}^{1/2}(n) = \mathbf{Q}(n)\Lambda^{1/2}(n)\mathbf{Q}^T(n). \quad (34)$$

The condition number of a matrix  $\mathbf{R}(n)$  is [8]

$$\chi[\mathbf{R}(n)] = \|\mathbf{R}(n)\| \|\mathbf{R}^{-1}(n)\|, \quad (35)$$

where  $\|\cdot\|$  can be any matrix norm. Note that  $\chi[\mathbf{R}(n)]$  depends on the underlying norm and the subscripts will be used to distinguish the different condition numbers. Usually, we take the convention that  $\chi[\mathbf{R}(n)] = \infty$  for a singular matrix  $\mathbf{R}(n)$ .

Consider the following norm:

$$\|\mathbf{R}(n)\|_E = \left\{ \frac{1}{L} \operatorname{tr} [\mathbf{R}^T(n)\mathbf{R}(n)] \right\}^{1/2}. \quad (36)$$

We can easily check that, indeed,  $\|\cdot\|_E$  is a matrix norm since for any real matrices  $\mathbf{A}$  and  $\mathbf{B}$  and a real scalar  $\gamma$ , the following three conditions are satisfied:

- (i)  $\|\mathbf{A}\|_E \geq 0$  and  $\|\mathbf{A}\|_E = 0$  if and only if  $\mathbf{A} = \mathbf{0}_{L \times L}$ ,
- (ii)  $\|\mathbf{A} + \mathbf{B}\|_E \leq \|\mathbf{A}\|_E + \|\mathbf{B}\|_E$ ,
- (iii)  $\|\gamma\mathbf{A}\|_E = |\gamma| \|\mathbf{A}\|_E$ .

Also, the E-norm of the identity matrix is equal to one.

We have

$$\begin{aligned} \|\mathbf{R}^{1/2}(n)\|_E &= \left\{ \frac{1}{L} \operatorname{tr} [\mathbf{R}(n)] \right\}^{1/2} = \left\{ \frac{1}{L} \sum_{l=0}^{L-1} \lambda_l(n) \right\}^{1/2}, \\ \|\mathbf{R}^{-1/2}(n)\|_E &= \left\{ \frac{1}{L} \operatorname{tr} [\mathbf{R}^{-1}(n)] \right\}^{1/2} = \left\{ \frac{1}{L} \sum_{l=0}^{L-1} \frac{1}{\lambda_l(n)} \right\}^{1/2}. \end{aligned} \quad (37)$$

Hence, the condition number of  $\mathbf{R}^{1/2}(n)$  associated with  $\|\cdot\|_E$  is

$$\chi_E[\mathbf{R}^{1/2}(n)] = \|\mathbf{R}^{1/2}(n)\|_E \|\mathbf{R}^{-1/2}(n)\|_E \geq 1. \quad (38)$$

If  $\chi[\mathbf{R}(n)]$  is large, then  $\mathbf{R}(n)$  is said to be an ill-conditioned matrix. Note that this is a norm-dependent property. However, according to [8], any two condition numbers  $\chi_\alpha[\mathbf{R}(n)]$  and  $\chi_\beta[\mathbf{R}(n)]$  are equivalent in that constants  $c_1$  and  $c_2$  can be found for which

$$c_1 \chi_\alpha[\mathbf{R}(n)] \leq \chi_\beta[\mathbf{R}(n)] \leq c_2 \chi_\alpha[\mathbf{R}(n)]. \quad (39)$$

For example, for the 1- and 2-norm matrices, we can show [8] that

$$\frac{1}{L^2} \chi_2[\mathbf{R}(n)] \leq \frac{1}{L} \chi_1[\mathbf{R}(n)] \leq \chi_2[\mathbf{R}(n)]. \quad (40)$$

We now show the same principle for the E- and 2-norm matrices. We recall that

$$\chi_2[\mathbf{R}(n)] = \frac{\lambda_{L-1}(n)}{\lambda_0(n)}. \quad (41)$$

Since  $\operatorname{tr}[\mathbf{R}^{-1}(n)] \geq 1/\lambda_0(n)$  and  $\operatorname{tr}[\mathbf{R}(n)] \geq \lambda_{L-1}(n)$ , we have

$$\operatorname{tr}[\mathbf{R}(n)] \operatorname{tr}[\mathbf{R}^{-1}(n)] \geq \frac{\operatorname{tr}[\mathbf{R}(n)]}{\lambda_0(n)} \geq \frac{\lambda_{L-1}(n)}{\lambda_0(n)}. \quad (42)$$

Also, since  $\operatorname{tr}[\mathbf{R}(n)] \leq L\lambda_{L-1}(n)$  and  $\operatorname{tr}[\mathbf{R}^{-1}(n)] \leq L/\lambda_0(n)$ , we obtain

$$\operatorname{tr}[\mathbf{R}(n)] \operatorname{tr}[\mathbf{R}^{-1}(n)] \leq L \frac{\operatorname{tr}[\mathbf{R}(n)]}{\lambda_0(n)} \leq L^2 \frac{\lambda_{L-1}(n)}{\lambda_0(n)}. \quad (43)$$

Therefore, we deduce that

$$\frac{1}{L^2} \chi_2[\mathbf{R}(n)] \leq \chi_E^2[\mathbf{R}^{1/2}(n)] \leq \chi_2[\mathbf{R}(n)]. \quad (44)$$

According to the previous expression,  $\chi_E^2[\mathbf{R}^{1/2}(n)]$  is then a measure of the condition number of the matrix  $\mathbf{R}(n)$ . In Section 5, we will show how to recursively compute  $\chi_E^2[\mathbf{R}^{1/2}(n)]$ .

## 5. RECURSIVE COMPUTATION OF THE CONDITION NUMBER

The positive number  $\|\mathbf{R}^{1/2}(n)\|_E^2$  can be easily calculated recursively. Indeed, taking the trace of

$$\mathbf{R}(n) = \lambda \mathbf{R}(n-1) + \mathbf{x}(n)\mathbf{x}^T(n), \quad (45)$$

we get

$$\operatorname{tr}[\mathbf{R}(n)] = \lambda \operatorname{tr}[\mathbf{R}(n-1)] + \mathbf{x}^T(n)\mathbf{x}(n). \quad (46)$$

Therefore,

$$\|\mathbf{R}^{1/2}(n)\|_E^2 = \lambda \|\mathbf{R}^{1/2}(n-1)\|_E^2 + \frac{\mathbf{x}^T(n)\mathbf{x}(n)}{L}. \quad (47)$$

Note that the inner product  $\mathbf{x}^T(n)\mathbf{x}(n)$  can also be computed in a recursive way with two multiplications only at each iteration.

Now we need to determine  $\|\mathbf{R}^{-1/2}(n)\|_E^2$ . Thanks to (22), we find that

$$\operatorname{tr}[\mathbf{R}^{-1}(n)] = \sum_{l=0}^{L-1} \frac{1}{E_l(n)}. \quad (48)$$

Using (24), we have

$$\mathbf{k}^T(n)\mathbf{k}'(n) = \sum_{l=0}^{L-1} \frac{e_l(n)\varepsilon_l(n)}{E_l(n)E_l(n-1)}, \quad (49)$$

and replacing in the previous expression:

$$E_l(n) - \lambda E_l(n-1) = e_l(n)\varepsilon_l(n), \quad (50)$$

we obtain

$$\mathbf{k}^T(n)\mathbf{k}'(n) = \sum_{l=0}^{L-1} \frac{1}{E_l(n-1)} - \lambda \sum_{l=0}^{L-1} \frac{1}{E_l(n)}. \quad (51)$$

Thus,

$$\begin{aligned} \text{tr}[\mathbf{R}^{-1}(n)] &= \sum_{l=0}^{L-1} \frac{1}{E_l(n)} \\ &= \lambda^{-1} \left[ \sum_{l=0}^{L-1} \frac{1}{E_l(n-1)} - \mathbf{k}^T(n)\mathbf{k}'(n) \right]. \end{aligned} \quad (52)$$

Finally,

$$\begin{aligned} \|\mathbf{R}^{-1/2}(n)\|_E^2 &= \lambda^{-1} \left[ \|\mathbf{R}^{-1/2}(n-1)\|_E^2 - \frac{\lambda^{-1}\varphi(n)\mathbf{k}'^T(n)\mathbf{k}'(n)}{L} \right] \\ &= \lambda^{-1} \left[ \|\mathbf{R}^{-1/2}(n-1)\|_E^2 - \frac{\lambda^{-1}\varphi(n)}{L} \sum_{l=0}^{L-1} \frac{e_l^2(n)}{E_l(n-1)} \right]. \end{aligned} \quad (53)$$

By using (47) and (53), we see that we easily compute  $\chi_E^2[\mathbf{R}^{1/2}(n)]$  recursively with only an order of  $L$  multiplications per iteration given that  $\mathbf{k}'(n)$  is known.

Note that we could have used the inverse of  $\mathbf{R}(n)$ ,

$$\mathbf{R}^{-1}(n) = \lambda^{-1}\mathbf{R}^{-1}(n-1) - \lambda^{-2}\varphi(n)\mathbf{k}'(n)\mathbf{k}'^T(n), \quad (54)$$

to estimate  $\|\mathbf{R}^{-1/2}(n)\|_E^2$ , but we have chosen here to use the interpolation formulation to better understand the link among all variables in the RLS algorithm, and especially to emphasize the role of the interpolation error energies since  $\text{tr}[\mathbf{R}^{-1}(n)] = \sum_{l=0}^{L-1} 1/E_l(n)$ , even though there are indirect ways to compute this value. Clearly, everything can be written in terms of  $E_l(n)$  and this formulation is more natural for the condition number estimation. For example, in the extreme cases of an input signal close to a white noise or to a predictable process, the value  $\max_l[E_l(n)]/\min_l[E_l(n)]$  gives a good idea of the condition number of the corresponding signal covariance matrix.

It is easy to combine the estimation of the condition number with an FRLS algorithm. There exist several methods to compute the a priori Kalman gain vector  $\mathbf{k}'(n)$  in a very efficient way. Once this gain vector is determined, the estimation of  $\chi_E^2[\mathbf{R}^{1/2}(n)]$  at each iteration follows immediately with roughly  $L$  more multiplications. Algorithm 1 shows the combination of an FRLS algorithm with the condition number estimation of the input signal covariance matrix.

## 6. MISALIGNMENT AND CONDITION NUMBER

We define the normalized misalignment in dB as follows:

$$m_0(n) = 10 \log_{10} E \left[ \frac{\|\mathbf{h}_t - \mathbf{h}(n)\|_2^2}{\|\mathbf{h}_t\|_2^2} \right], \quad (55)$$

where  $\|\cdot\|_2$  denotes the 2-norm vector. Equation (55) measures the mismatch between the true impulse response and the modelling filter.

Initialization.

$$\mathbf{h}(0) = \mathbf{k}'(0) = \mathbf{a}(0) = \mathbf{b}(0) = \mathbf{0},$$

$$\alpha(0) = \lambda,$$

$$E_a(0) = E_0, \text{ (positive constant),}$$

$$\|\mathbf{R}^{1/2}(0)\|_E^2 = \frac{E_0}{L} \sum_{l=0}^{L-1} \lambda^{-l},$$

$$\|\mathbf{R}^{-1/2}(0)\|_E^2 = \frac{1}{LE_0} \sum_{l=0}^{L-1} \lambda^l.$$

Prediction.

$$e_a(n) = x(n) - \mathbf{a}^T(n-1)\mathbf{x}(n-1),$$

$$\alpha_1(n) = \alpha(n-1) + e_a^2(n)/E_a(n-1),$$

$$\begin{bmatrix} \mathbf{t}(n) \\ \mathbf{m}(n) \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{k}'(n-1) \end{bmatrix} + \begin{bmatrix} 1 \\ -\mathbf{a}(n-1) \end{bmatrix} e_a(n)/E_a(n-1),$$

$$E_a(n) = \lambda[E_a(n-1) + e_a^2(n)/\alpha(n-1)],$$

$$\mathbf{a}(n) = \mathbf{a}(n-1) + \mathbf{k}'(n-1)e_a(n)/\alpha(n-1),$$

$$e_b(n) = x(n-L) - \mathbf{b}^T(n-1)\mathbf{x}(n),$$

$$\mathbf{k}'(n) = \mathbf{t}(n) + \mathbf{b}(n-1)m(n),$$

$$\alpha(n) = \alpha_1(n) - e_b(n)m(n),$$

$$\mathbf{b}(n) = \mathbf{b}(n-1) + \mathbf{k}'(n)e_b(n)/\alpha(n).$$

Filtering.

$$e(n) = y(n) - \mathbf{h}^T(n-1)\mathbf{x}(n),$$

$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{k}'(n)e(n)/\alpha(n).$$

Condition Number.

$$\|\mathbf{R}^{1/2}(n)\|_E^2 = \lambda \|\mathbf{R}^{1/2}(n-1)\|_E^2 + \frac{\mathbf{x}^T(n)\mathbf{x}(n)}{L},$$

$$\|\mathbf{R}^{-1/2}(n)\|_E^2 = \lambda^{-1} \left[ \|\mathbf{R}^{-1/2}(n-1)\|_E^2 - \frac{\mathbf{k}'^T(n)\mathbf{k}'(n)}{L\alpha(n)} \right],$$

$$\chi_E^2[\mathbf{R}^{1/2}(n)] = \|\mathbf{R}^{1/2}(n)\|_E^2 \|\mathbf{R}^{-1/2}(n)\|_E^2.$$

ALGORITHM 1: The FRLS algorithm and estimation of the condition number.

It can easily be shown, under certain conditions, that [9]

$$E \left[ \|\mathbf{h}_t - \mathbf{h}(n)\|_2^2 \right] \approx \frac{1}{2} \sigma_w^2 \text{tr}[\mathbf{R}^{-1}(n)]. \quad (56)$$

Hence, we can write (56) in terms of the interpolation error energies:

$$E \left[ \|\mathbf{h}_t - \mathbf{h}(n)\|_2^2 \right] \approx \frac{1}{2} \sigma_w^2 \sum_{l=0}^{L-1} \frac{1}{E_l(n)}. \quad (57)$$

However, we are more interested here to write (56) in terms

of the condition number. Indeed, we have

$$\begin{aligned}\|\mathbf{R}^{1/2}(n)\|_{\text{E}}^2 &= \frac{1}{L} \text{tr}[\mathbf{R}(n)], \\ \|\mathbf{R}^{-1/2}(n)\|_{\text{E}}^2 &= \frac{1}{L} \sum_{l=0}^{L-1} \frac{1}{E_l(n)}.\end{aligned}\quad (58)$$

But

$$\begin{aligned}\text{tr}[\mathbf{R}(n)] &= \text{tr}\left[\sum_{m=0}^n \lambda^{n-m} \mathbf{x}(m) \mathbf{x}^T(m)\right] \\ &= \sum_{m=0}^n \lambda^{n-m} \mathbf{x}^T(m) \mathbf{x}(m) \\ &\approx \frac{L}{1-\lambda} \sigma_x^2,\end{aligned}\quad (59)$$

for  $n$  large and for a stationary signal  $x$  with power  $\sigma_x^2$ . The condition number is then

$$\chi_{\text{E}}^2[\mathbf{R}^{1/2}(n)] \approx \frac{\sigma_x^2}{(1-\lambda)L} \sum_{l=0}^{L-1} \frac{1}{E_l(n)}, \quad (60)$$

and expression (57) becomes

$$E\left[\|\mathbf{h}_t - \mathbf{h}(n)\|_2^2\right] \approx \frac{(1-\lambda)L}{2} \frac{\sigma_w^2}{\sigma_x^2} \chi_{\text{E}}^2[\mathbf{R}^{1/2}(n)]. \quad (61)$$

If we divide both sides of (61) by  $\|\mathbf{h}_t\|_2^2$ , we get

$$E\left[\frac{\|\mathbf{h}_t - \mathbf{h}(n)\|_2^2}{\|\mathbf{h}_t\|_2^2}\right] \approx \frac{(1-\lambda)L}{2} \frac{\sigma_w^2}{\|\mathbf{h}_t\|_2^2 \sigma_x^2} \chi_{\text{E}}^2[\mathbf{R}^{1/2}(n)]. \quad (62)$$

Finally, we have a formula for the normalized misalignment in dB (which is valid only after convergence of the RLS algorithm):

$$\begin{aligned}m_0(n) &\approx 10 \log_{10} \frac{(1-\lambda)L}{2} + 10 \log_{10} \frac{\sigma_w^2}{\|\mathbf{h}_t\|_2^2 \sigma_x^2} \\ &\quad + 10 \log_{10} \chi_{\text{E}}^2[\mathbf{R}^{1/2}(n)].\end{aligned}\quad (63)$$

Expression (63) depends on three terms or three factors: the exponential window, the level of noise at the system output, and the condition number. The closer the exponential window is to one, the better the misalignment is, but the tracking abilities of the RLS algorithm will suffer a lot. A high level of noise as well as an input signal with a large condition number will obviously degrade the misalignment. With a fixed exponential window and noise, it is interesting to see how the misalignment will degrade by increasing the condition number of the input signal. For example, by increasing the condition number from 1 to 10, the misalignment will degrade by 10 dB; the simulations confirm this.

Usually, we take for the exponential window

$$\lambda = 1 - \frac{1}{K_0 L}, \quad (64)$$

where  $K_0 \geq 3$ . Also, the second term in (63) represents roughly the inverse output SNR in dB. We can then rewrite (63) as follows:

$$m_0(n) \approx -10 \log_{10}(2K_0) - \text{oSNR} + 10 \log_{10} \chi_{\text{E}}^2[\mathbf{R}^{1/2}(n)]. \quad (65)$$

For example, if we take  $K_0 = 5$  and an output SNR (oSNR) of 39 dB, we obtain

$$m_0(n) \approx -49 + 10 \log_{10} \chi_{\text{E}}^2[\mathbf{R}^{1/2}(n)]. \quad (66)$$

If the input signal is a white noise,  $\chi_{\text{E}}^2[\mathbf{R}^{1/2}(n)] = 1$ , then  $m_0(n) \approx -49$  dB. This will be confirmed in the following section.

## 7. SIMULATIONS

In this section, we present some results on the condition number estimation and how this number affects the misalignment in a system identification context. We try to estimate an impulse response  $\mathbf{h}_t$  of length  $L = 512$ . The same length is used for the adaptive filter  $\mathbf{h}(n)$ . We run the FRLS algorithm with a forgetting factor  $\lambda = 1 - 1/(5L)$ . Performance of the estimation is measured by means of the normalized misalignment (55). The input signal  $x(n)$  is a speech signal sampled at 8 kHz. The output signal  $y(n)$  is obtained by convolving  $\mathbf{h}_t$  with  $x(n)$  and adding a white Gaussian noise signal with an SNR of 39 dB. In order to evaluate the condition number in different situations, a white Gaussian signal is added to the input  $x(n)$  with different SNRs. The range of the input SNR is  $-10$  dB to 50 dB. Therefore, with an input SNR equal to  $-10$  dB (the white noise dominates the speech), we can expect the condition number of the input signal covariance matrix to be close to 1, while with an input SNR of 50 dB (the speech largely dominates the white noise), the condition number will be high. Figures 1, 2, 3, 4, 5, 6, and 7 show the evolution in time of the input signal, the normalized misalignment (we approximate the normalized misalignment with its instantaneous value), and the condition number of the input signal covariance matrix with different input SNRs (from  $-10$  dB to 50 dB). We can see that as the input SNR increases, the condition number degrades as expected since the speech signal is ill-conditioned. As a result, the normalized misalignment is greatly affected by a large value of the condition number. As expected, the value of the misalignment after convergence in Figure 1 is equal to  $-49$  dB and the condition number is almost one. Now compare this to Figure 3. In Figure 3, the misalignment is equal to  $-40$  dB and the average condition number is 8.2. The higher condition number in this case degrades the misalignment by 9 dB, which is exactly the degradation predicted by formula (63). We can verify the same trend with the other simulations.

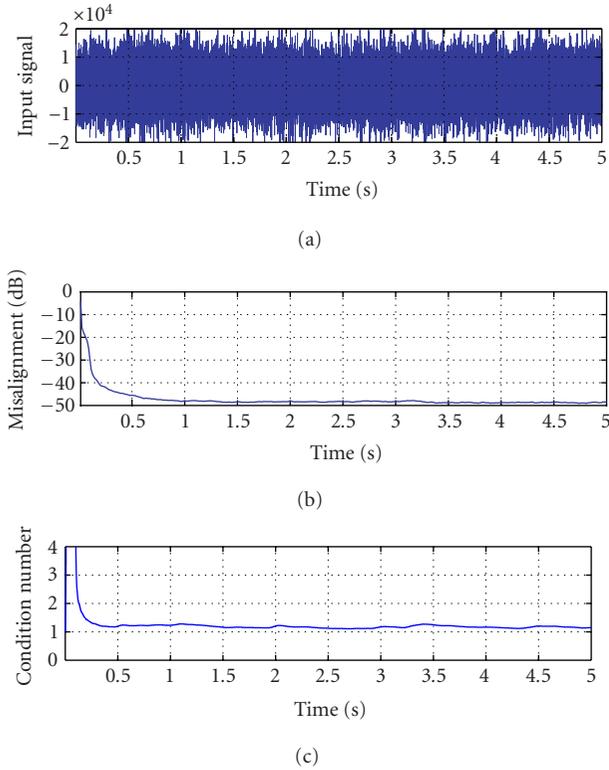


FIGURE 1: Evolution in time of the (a) input signal, (b) normalized misalignment, and (c) condition number of the input signal covariance matrix. The input SNR is  $-10$  dB.

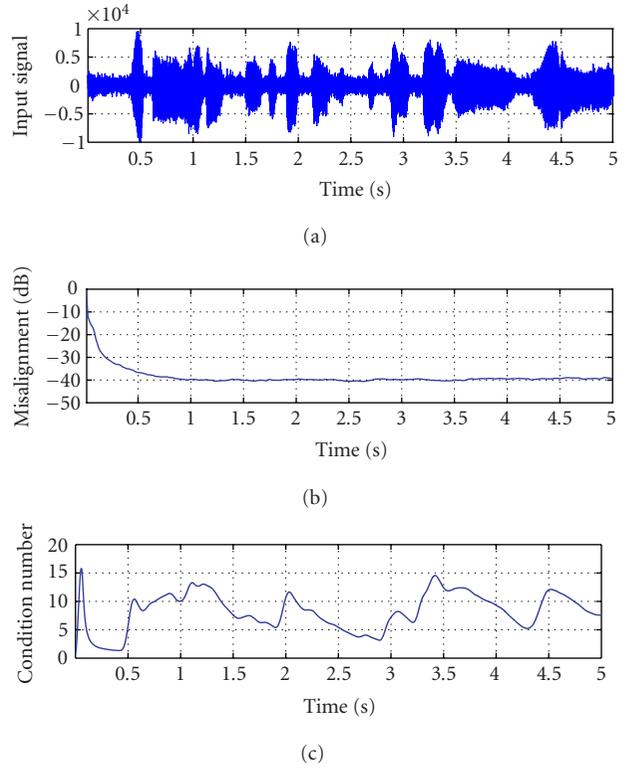


FIGURE 3: The presentation is the same as in Figure 1. The input SNR is  $10$  dB.

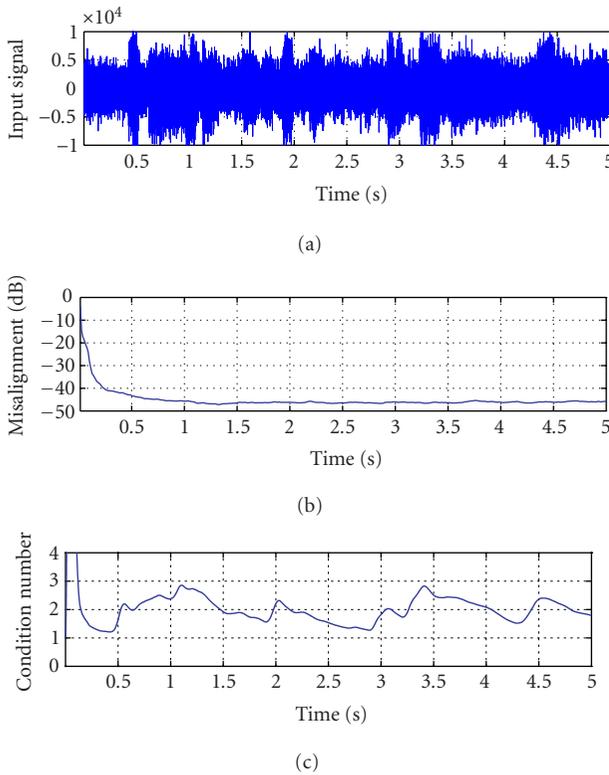


FIGURE 2: The presentation is the same as in Figure 1. The input SNR is  $0$  dB.

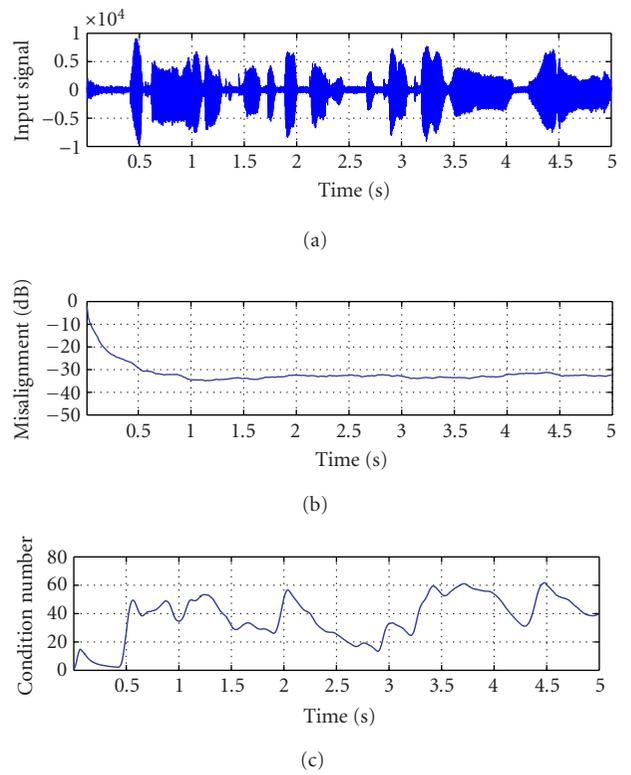


FIGURE 4: The presentation is the same as in Figure 1. The input SNR is  $20$  dB.

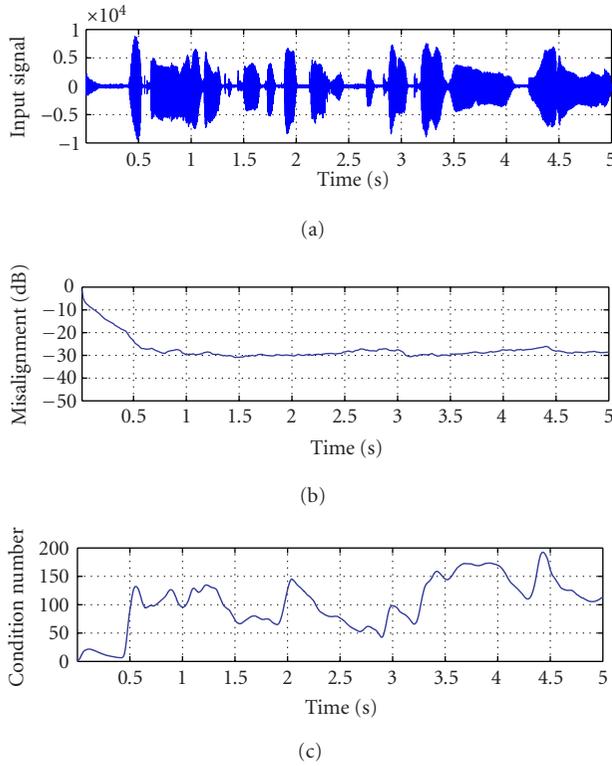


FIGURE 5: The presentation is the same as in Figure 1. The input SNR is 30 dB.

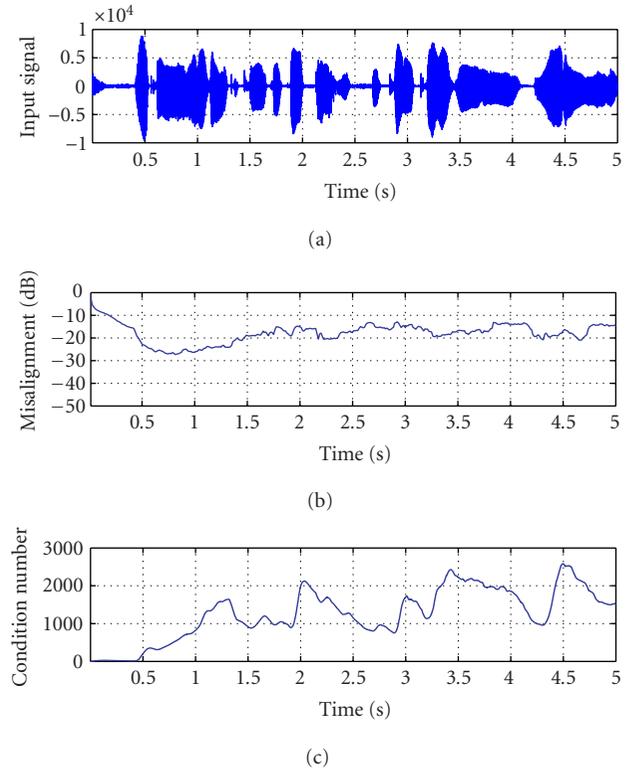


FIGURE 7: The presentation is the same as in Figure 1. The input SNR is 50 dB.

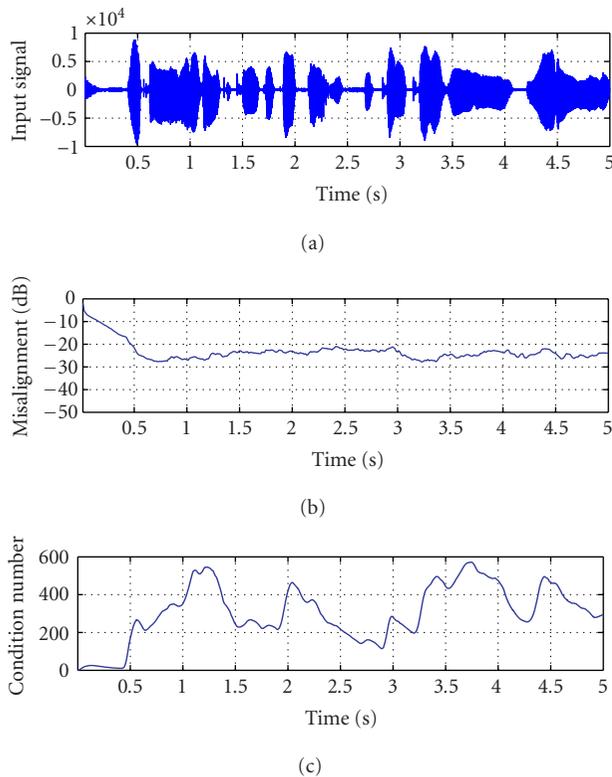


FIGURE 6: The presentation is the same as in Figure 1. The input SNR is 40 dB.

## 8. CONCLUSIONS

The RLS algorithm plays a major role in adaptive signal processing. A very good understanding of its different variables may lead to new concepts and new algorithms. In this paper, we have shown that the update equation of the RLS can be written in terms of the a priori or a posteriori interpolation error signals normalized with their respective interpolation error energies. Hence, the interpolation error energy formulation can be further exploited. This formulation has motivated us to propose a simple and an efficient way to estimate the condition number of the input signal covariance matrix. We have shown that this condition number can be easily integrated in the FRLS structure at a very low cost from an arithmetic complexity point of view. Finally, we have shown how the misalignment of the RLS depends on the condition number. A formula was derived, predicting how the misalignment degrades when the condition number increases. The accuracy of this formula was exemplified by simulations.

## REFERENCES

- [1] M. G. Bellanger, *Adaptive Digital Filters and Signal Analysis*, Marcel Dekker, New York, NY, USA, 1987.
- [2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.
- [3] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 4th edition, 2002.

- [4] J. Benesty and Y. Huang, Eds., *Adaptive Signal Processing: Applications to Real-World Problems*, Springer-Verlag, Berlin, 2003.
- [5] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Processing Magazine*, vol. 11, no. 3, pp. 18–60, 1994.
- [6] S. Kay, "Some results in linear interpolation theory," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 31, no. 3, pp. 746–749, 1983.
- [7] B. Picinbono and J.-M. Kerilis, "Some properties of prediction and interpolation errors," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 525–531, 1988.
- [8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, USA, 1996.
- [9] J. Benesty, T. Gänslér, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer-Verlag, Berlin, 2001.

**Jacob Benesty** was born in 1963. He received M.S. degree in microwaves from Pierre & Marie Curie University, France, in 1987, and his Ph.D. degree in control and signal processing from Orsay University, France, in 1991. During his Ph.D. (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris, France. From January 1994 to July 1995, he worked at Telecom Paris University. From October 1995 to May 2003, he was with Bell Laboratories, Murray Hill, NJ, USA. In May 2003, he joined INRS-EMT, University of Quebec, Montreal, Quebec, Canada, as an Associate Professor. His research interests are in acoustic signal processing and multimedia communications. He is the recipient of the IEEE Signal Processing Society 2001 Best Paper Award. He coauthored the book *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, Berlin, 2001) and coedited/coauthored three more books.



**Tomas Gänslér** was born in Sweden in 1966. He received his M.S. degree in electrical engineering and his Ph.D. degree in signal processing from Lund University, Lund, Sweden, in 1990 and 1996. From 1997 to September 1999, he held a position as an Assistant Professor at Lund University. During 1998, he was employed by Bell Labs, Lucent Technologies, as a Consultant and from October 1999, he joined the technical staff as a member. From 2001, he is with Agere Systems Inc., a spin-off from Lucent Technologies' Microelectronics group. His research interests include robust estimation, adaptive filtering, mono/multichannel echo cancellation, and subband signal processing. He coauthored the books *Advances in Network and Acoustic Echo Cancellation* and *Acoustic Signal Processing for Telecommunication*.

