Katholieke Universiteit Leuven

Departement Elektrotechniek ESAT-SISTA/TR 05-184

Optimally Regularized Adaptive Filtering Algorithms for Room Acoustic Signal Enhancement[1]

Toon van Waterschoot[2] [3], Geert Rombouts[2], and Marc Moonen[2]

Published in *Signal Processing*, vol. 88, no. 3, Mar. 2008, pp. 594-611.

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/vanwaterschoot/reports/05-184.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD(SISTA), Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, Tel. +32 16 321927, Fax +32 16 321970, WWW: *http://www.esat.kuleuven.be/sista-cosic-docarch.* E-mail: *toon.vanwaterschoot@esat.kuleuven.be.*

[3]Toon van Waterschoot is a Research Assistant with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council: CoE EF/05/006 Optimization in Engineering, the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (“Dynamical systems, control and optimization”, 2007-2011), the Concerted Research Action GOA-AMBioRICS, and IWT project 040803: “SMS4PA-II: Sound Management System for Public Address Systems”. The scientific responsibility is assumed by its authors.

Available online at www.sciencedirect.com
ScienceDirect
Signal Processing 88 (2008) 594–611
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement

Toon van Waterschoot*, Geert Rombouts, Marc Moonen

Katholieke Universiteit Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B3001 Leuven, Belgium

Received 14 July 2006; received in revised form 28 August 2007; accepted 10 September 2007
Available online 19 September 2007

Abstract

In many room acoustic signal processing applications, a room impulse response identification is needed to eliminate undesired effects such as echo, feedback, or reverberation. This is typically done using an adaptive filter driven by a speech or audio input signal. However, such signals exhibit poor excitation properties, which cause standard adaptive filtering algorithms to be very sensitive to disturbing signals, especially in the underdetermined case. A popular remedy is regularization, which is usually implemented with a scaled identity regularization matrix. This type of regularization is governed by a single regularization parameter, the value of which is often chosen in an arbitrary way. We propose to regularize the adaptive filter using a non-identity regularization matrix, in which prior knowledge on the unknown room impulse response may be incorporated. When knowledge of the disturbing signal is also used to add prefiltering and weighting in the adaptation, a new family of regularized adaptive filtering algorithms is obtained, which is shown to be optimal in a mean square error sense. Existing regularized algorithms can then be obtained as special cases, assuming limited or no prior knowledge is available. When combined with a recently proposed method of extracting prior knowledge from the acoustic setup, our algorithms exhibit superior convergence behaviour compared to existing algorithms in different simulation scenarios, while the additional computational cost is small.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Regularization; Adaptive filtering; Room acoustics; Prior knowledge; Signal enhancement

1. Introduction

Several room acoustic signal processing applications require the identification of a room impulse response (RIR). The most well-known application is probably acoustic echo cancellation (AEC) [1], in which an RIR estimate is needed for the prediction of an echo component that is present in a microphone signal. Other applications include adaptive feedback cancellation [2], which is similar to AEC but involves a closed signal loop, and active noise control [3], where both the so-called primary and secondary signal path have to be identified.

We will focus on the AEC scenario for illustrating the general RIR identification problem, see Fig. 1. We assume that the true RIR coefficients of the echo path between the loudspeaker and the

*Corresponding author. Tel.: +32 16 321927; fax: +32 16 321970.
E-mail addresses: toon.vanwaterschoot@esat.kuleuven.be (T. van Waterschoot), geert.rombouts@esat.kuleuven.be (G. Rombouts), marc.moonen@esat.kuleuven.be (M. Moonen).
URL: http://homes.esat.kuleuven.be/~tvanwate (T. van Waterschoot).

0165-1684/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2007.09.001

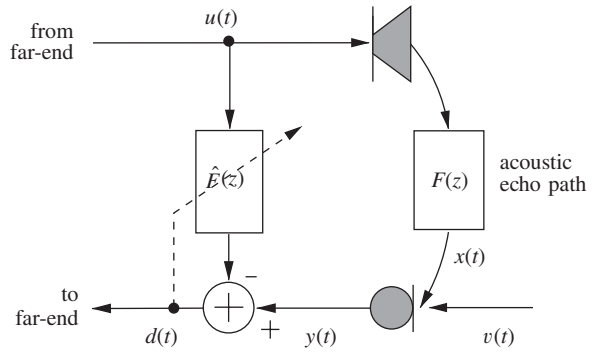


Fig. 1. Room impulse response identification in an AEC scenario. $F(z)$ and $\hat{F}(z)$ represent the transfer functions associated with the filter coefficients in $\mathbf{f}$ and $\hat{\mathbf{f}}(t)$, respectively.

microphone are collected in the parameter vector

$$\mathbf{f} \triangleq [f_0 \;\; f_1 \;\; \ldots \;\; f_{n_F}]^{\mathrm{T}}, \tag{1}$$

of known length $n_F + 1$. The loudspeaker produces a signal $u(t)$ (often referred to as the "far-end" signal), which generates an echo signal $x(t) = \mathbf{u}^{\mathrm{T}}(t)\mathbf{f}$ at the microphone position, with the loudspeaker signal vector defined as

$$\mathbf{u}(t) \triangleq [u(t) \;\; u(t-1) \;\; \ldots \;\; u(t-n_F)]^{\mathrm{T}}. \tag{2}$$

The echo signal $x(t)$ is picked up by the microphone in addition to a local signal $v(t)$, which is defined as the source signal (often referred to as the "near-end" signal), hence the microphone signal can be written as $y(t) = x(t) + v(t)$. It is important to note that the source signal is in most applications the signal of interest; however, from a system identification point of view, $v(t)$ is a disturbance to the estimation of the RIR. If at time t an RIR estimate $\hat{\mathbf{f}}(t)$ is available, then an echo-compensated signal can be calculated as $d(t) = y(t) - \mathbf{u}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t)$, which approximates the source signal $v(t)$. Note that, throughout this paper, we assume that the true RIR order n_F is known, and equal to the adaptive filter order. Methods for dealing with the insufficient-order case can be found in, e.g., [4,5].

Nearly all linear adaptive filtering algorithms are related to the well-known least squares (LS) estimate of the RIR, given by

$$\hat{\mathbf{f}}(t) = (\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}, \tag{3}$$

with the data model for the discrete time range $[1, t]$ defined as

$$\mathbf{y} = \mathbf{U}\mathbf{f} + \mathbf{v}, \tag{4}$$

and

$$\mathbf{U} \triangleq [\mathbf{u}(t) \;\; \mathbf{u}(t-1) \;\; \ldots \;\; \mathbf{u}(1)]^{\mathrm{T}}, \tag{5}$$

$$\mathbf{y} \triangleq [y(t) \;\; y(t-1) \;\; \ldots \;\; y(1)]^{\mathrm{T}}, \tag{6}$$

$$\mathbf{v} \triangleq [v(t) \;\; v(t-1) \;\; \ldots \;\; v(1)]^{\mathrm{T}}. \tag{7}$$

A common problem in room acoustic applications is that the matrix $\mathbf{U}^{\mathrm{T}}\mathbf{U}$, which is inverted in the LS estimate (3), is ill-conditioned or even singular due to poor excitation. Indeed, the identifiability of the RIR $\mathbf{f}$ is only guaranteed if the loudspeaker signal $u(t)$ is persistently exciting of order n_F [6]. However, the dynamics of a typical RIR can often only be captured with several thousands of coefficients. On the other hand, the loudspeaker signal is usually a speech or audio signal, and may exhibit a nearly harmonic spectrum, such that its excitation order is far below n_F. If such an ill-conditioned situation occurs, there will typically be a large variance on the resulting RIR estimate when the source signal $v(t)$ is non-zero, as can be seen from the LS estimate covariance matrix:

$$\mathrm{cov}\{\hat{\mathbf{f}}(t)\} = (\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{R}_{\mathbf{v}}\mathbf{U}(\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}, \tag{8}$$

with the (zero-mean) source signal covariance matrix defined as

$$\mathbf{R}_{\mathbf{v}} = E\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\}, \tag{9}$$

and with $E\{\cdot\}$ and $\mathrm{cov}\{\cdot\}$ the expectation and covariance operators, respectively.

Regularization is a standard technique to turn such an ill-posed problem into a well-posed one, namely by improving the condition of the matrix $\mathbf{U}^{\mathrm{T}}\mathbf{U}$ in the LS estimate (3). The regularization method that is most often adopted in LS estimation is the one proposed by Tikhonov et al. [7] (and reformulated in a linear algebra setting by Neumaier [8]). The Tikhonov regularization method applied to the LS estimate, consists in adding a scaled identity matrix to the matrix $\mathbf{U}^{\mathrm{T}}\mathbf{U}$, i.e.,

$$\hat{\mathbf{f}}(t) = (\mathbf{U}^{\mathrm{T}}\mathbf{U} + \alpha\mathbf{I})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{y}, \tag{10}$$

where α denotes the regularization parameter. This is in fact the minimizing estimate of a modified LS criterion in which the squared Euclidian norm of the RIR estimate is added to the sum of squared errors, and weighted with the regularization parameter α:

$$\min_{\hat{\mathbf{f}}(t)}\{[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}(t)]^{\mathrm{T}}[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}(t)] + \alpha\|\hat{\mathbf{f}}(t)\|_2^2\}. \tag{11}$$

From this criterion, it can be seen that a large RIR estimate is penalized by the regularization term,

leading to an estimate that is biased towards zero. However, allowing this bias in particular results in a reduction of the estimation variance.

The Tikhonov regularized LS estimate in (10) may be calculated recursively by initializing the adaptive filter input correlation matrix as $\mathbf{R}(0)=\alpha\mathbf{I}$, and applying the standard recursive LS (RLS) algorithm:

$$\hat{\mathbf{f}}(t)=\hat{\mathbf{f}}(t-1)+\mathbf{R}^{-1}(t)\mathbf{u}(t)\varepsilon(t), \tag{12}$$

$$\mathbf{R}(t)=\mathbf{R}(t-1)+\mathbf{u}(t)\mathbf{u}^{\mathrm{T}}(t), \tag{13}$$

$$\varepsilon(t)=y(t)-\mathbf{u}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1). \tag{14}$$

Note that the a priori residual $\varepsilon(t)$ in (14) differs from the a posteriori residual $d(t)$, which is sent to the far-end side, in that it depends on the previous RIR estimate $\hat{\mathbf{f}}(t-1)$. If, for tracking purposes, an exponential window is applied in the above RLS algorithm, then the regularization matrix $\alpha\mathbf{I}$ will be exponentially weighted, along with older data (which results in a technique known as soft-constrained initialization [9]). If we wish to maintain the regularization throughout the complete time range $[1,t]$, an additional term should be added to the correlation matrix update in order to compensate for the weighting of the regularization matrix. Moreover, with this modification, the RLS algorithm would not converge to the (exponentially weighted) LS estimate, unless a correction term is also subtracted from the RIR weight update. The *Tikhonov regularized RLS algorithm* with exponential weighting factor λ is then given by:

$$\hat{\mathbf{f}}(t)=\hat{\mathbf{f}}(t-1)+\mathbf{R}^{-1}(t)\times[\mathbf{u}(t)\varepsilon(t)-(1-\lambda)\alpha\hat{\mathbf{f}}(t-1)], \tag{15}$$

$$\mathbf{R}(t)=\lambda\mathbf{R}(t-1)+\mathbf{u}(t)\mathbf{u}^{\mathrm{T}}(t)+(1-\lambda)\alpha\mathbf{I}, \tag{16}$$

$$\varepsilon(t)=y(t)-\mathbf{u}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1). \tag{17}$$

This algorithm was proposed by Horita et al. [10], and was called the Leaky RLS algorithm because of its similarity with the Leaky least mean squares (LMS) algorithm [11] if the RIR weight update in (15) is rewritten as

$$\hat{\mathbf{f}}(t)=[\mathbf{I}-(1-\lambda)\alpha\mathbf{R}^{-1}(t)]\hat{\mathbf{f}}(t-1)+\mathbf{R}^{-1}(t)\mathbf{u}(t)\varepsilon(t). \tag{18}$$

A different regularization technique, which is more often applied to RLS adaptive filtering algorithms, is known as Levenberg–Marquardt regularization [12]. This method is very similar to Tikhonov regularization; however, no correction term is subtracted from the RIR weight update. The *Levenberg–Marquardt regularized RLS algorithm* is given by:

$$\hat{\mathbf{f}}(t)=\hat{\mathbf{f}}(t-1)+\mathbf{R}^{-1}(t)\mathbf{u}(t)\varepsilon(t), \tag{19}$$

$$\mathbf{R}(t)=\lambda\mathbf{R}(t-1)+\mathbf{u}(t)\mathbf{u}^{\mathrm{T}}(t)+(1-\lambda)\alpha\mathbf{I}, \tag{20}$$

$$\varepsilon(t)=y(t)-\mathbf{u}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1). \tag{21}$$

The Levenberg–Marquardt method was initially developed by Levenberg [13] and Marquardt [14] as a hybrid technique for nonlinear optimization. To achieve an optimal trade-off between convergence speed and robustness, the Gauss–Newton method (with update term $\mathbf{u}(t)\mathbf{u}^{\mathrm{T}}(t)$) was combined with the steepest descent method (with update term $(1-\lambda)\mathbf{I}$) in the correlation matrix update (20), by means of a steering factor α. The interpretation of the Levenberg–Marquardt method as a regularization technique was first given by Ljung and Söderström [12], and found its way into acoustic signal processing through the work of Gay [15] and Benesty et al. [1], among others. However, as already mentioned, due to the absence of the correction term in the RIR weight update (compare (19) to (15)), the Levenberg–Marquardt regularized RLS algorithm does not converge to the Tikhonov regularized LS estimate in (10).

In room acoustic applications, the underdetermined RLS (URLS) family [16], of which the normalized least mean squares (NLMS) algorithm and the affine projection algorithm (APA) are the most well-known members, is much more appealing from a computational point of view. However, due to their underdetermined nature, these algorithms are even more susceptible to convergence problems resulting from poor excitation. Therefore, regularization is also included in nearly every algorithm from the URLS family. The *regularized APA* is similar to the Levenberg–Marquardt regularized RLS algorithm, in that a scaled identity matrix is added to the adaptive filter input correlation matrix before inversion:

$$\hat{\mathbf{f}}(t)=\hat{\mathbf{f}}(t-1)+\mu\mathbf{U}_M(t)\times[\mathbf{U}_M^{\mathrm{T}}(t)\mathbf{U}_M(t)+\alpha\mathbf{I}]^{-1}\boldsymbol{\varepsilon}_M(t), \tag{22}$$

$$\boldsymbol{\varepsilon}_M(t)=\mathbf{y}_M(t)-\mathbf{U}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1), \tag{23}$$

where now the identity matrix is of dimension $M\times M$, with M the projection order, μ represents

the step size, and

$$\mathbf{U}_M(t) \triangleq [\mathbf{u}(t) \ \ldots \ \mathbf{u}(t-M+1)], \tag{24}$$

$$\mathbf{y}_M(t) \triangleq [y(t) \ \ldots \ y(t-M+1)]^{\mathrm{T}}. \tag{25}$$

The well-known *NLMS algorithm* can then be obtained from the APA by setting $M = 1$:

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu \frac{\mathbf{u}(t)\varepsilon(t)}{\mathbf{u}^{\mathrm{T}}(t)\mathbf{u}(t) + \alpha}, \tag{26}$$

$$\varepsilon(t) = y(t) - \mathbf{u}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1). \tag{27}$$

The aim of this paper is two-fold. First of all, we will show how the different regularized algorithms fit into a common framework, and point out the relationship between regularization and some other concepts in adaptive filtering such as leakage and proportionate adaptation. Second, we raise the question whether these existing regularization techniques are optimal in some sense, and to which extent they may be improved by using more than just one regularization parameter. Many choices for the regularization parameter α in the above algorithms have been proposed, ranging from any small number [1,16–23], to a parameter incorporating prior knowledge on the near-end signal [8,15,24–26], or on both the near-end signal and the unknown RIR [27,28]. In many acoustic signal enhancement applications, the regularization parameter is an a priori estimate of the long-term source signal power, as suggested in [15]. Also, some algorithms have been proposed in which the regularization parameter can be estimated online [29–31]. The use of multiple regularization parameters has previously been proposed for frequency domain and subband adaptive filtering, by assigning a different regularization parameter to each frequency band, see, e.g., [32]. Also, in neural networks [33, Chapter 5], it is customary to increase the number of regularization parameters by adding a more general penalty term $\alpha\|\mathbf{L}\hat{\mathbf{f}}(t)\|_2^2$ instead of $\alpha\|\hat{\mathbf{f}}(t)\|_2^2$ in the Tikhonov regularized LS criterion (11), as suggested in [34,35], with $\mathbf{L}$ typically chosen as a discrete approximation to some derivative operator for imposing smoothness on the RIR estimate.

In this paper, the main motivation for using more than one regularization parameter is related to the availability of prior knowledge. There exists a well-known equivalence between Tikhonov regularization and Bayesian maximum a posteriori (MAP) or minimum mean square error (MMSE) estimation. In a Bayesian framework, prior knowledge on the unknown RIR parameters can be incorporated in a non-identity regularization matrix. This approach is most often pointed to as "Bayesian regularization" (e.g., in the neural networks literature [36,37]), or also as "generalized Tikhonov regularization" (in the inverse problems literature [38]). In this paper, these concepts are extrapolated to linear adaptive filter theory, and applied to room acoustic problems. Recently, the RIR estimation problem was also tackled using Bayesian regularization in [39], where an iterative optimization algorithm was applied using both non-negativity constraints and $L1$-norm sparsity regularization. As a consequence, this method relies on non-negativity and sparseness of the RIR coefficients, which may be valid assumptions for synthetic RIR's generated using the image source method [40] (as used in the simulations in [39]), but not for realistic RIR's.

In Section 2, we review some results from Bayesian linear MMSE estimation [41], which are essential for deriving the proposed adaptive filtering algorithms. We will formulate a non-recursive RIR estimate that is optimal in the sense of minimizing the mean square deviation of the RIR estimate from the true RIR in a Bayesian framework. This linear MMSE (LMMSE) RIR estimate depends on the statistical assumptions on the source signal $v(t)$, as well as on the true RIR $\mathbf{f}$, and can also be obtained as the minimizing estimate of a weighted and regularized LS criterion. This observation has then motivated us to derive from the latter criterion a set of regularized adaptive filtering algorithms, which contains the existing regularized algorithms as special cases.

The MSE optimally regularized algorithms rely on knowledge of the source signal and true RIR statistics, which are unknown in most applications. In a previous paper, we have proposed adaptive filtering algorithms based on the prediction error method [12], with which the source signal statistics can be estimated concurrently with the unknown RIR (both for AEC [42,43] and for adaptive feedback cancellation [2,44]). Also, we have proposed a method for constructing the true RIR statistics from prior knowledge on the acoustic setup [45]. In Section 3, these techniques are briefly reviewed, and we comment on how they can provide the necessary statistical information for the algorithms proposed in Sections 4 and 5.

In Section 4, we focus on MSE optimal regularization of RLS algorithms. We will propose both a Tikhonov regularized (TR-RLS) algorithm and a Levenberg–Marquardt regularized (LMR-RLS)

algorithm, with a regularization matrix which may contain prior knowledge on the true RIR. It is shown that the TR-RLS and LMR-RLS algorithms may, in fact, be obtained by minimization of the same criterion, and only differ in the statistical assumptions on the true RIR. Furthermore, we point out the increase in computational complexity that is inherent in regularizing the RLS algorithm, as because of the regularization, the matrix inversion lemma (MIL) cannot be applied anymore to the adaptive filter input correlation matrix update (which can also be seen from the update equations (16) and (20)). We discuss two methods for avoiding the explicit inversion of the RLS input correlation matrix, and hence reducing the complexity from $O(n_F^3)$ to $O(n_F^2)$ computations per iteration, in particular by: (1) approximating the regularization matrix by a sum of rank-one matrices [12] (also known as dynamic regularization [18]), and (2) applying a rectangular instead of an exponential data window [46]. Another technique for obtaining $O(n_F^2)$ complexity, namely by indirect updating of the input correlation matrix through its eigenvalue decomposition [10,29], is only feasible if the regularization matrix is a scaled identity matrix.

Regularization of URLS algorithms is tackled in Section 5, where in particular we discuss the regularized NLMS and APA algorithms. Minimization of the weighted and regularized LS criterion that led to the TR-RLS and LMR-RLS algorithms, yields in the underdetermined case an algorithm that exhibits features known as leakage and proportionate adaptation. However, from extensive simulations, we have learned that leakage (even in the simple case of the leaky LMS algorithm [11]) may significantly deteriorate the adaptive filter performance in room acoustic applications. Adopting the same statistical assumptions on the true RIR that led to the LMR-RLS algorithm, now leads to a set of regularized NLMS and APA algorithms, without leakage. Moreover, the proposed LMR-NLMS and LMR-APA algorithms are formally equivalent to the proportionate NLMS (PNLMS) [23,47] and proportionate APA (PAPA) [21] algorithms. This equivalence leads to a new interpretation of the proportionate adaptation concept, and hence to an MSE optimal choice of the proportionate adaptation parameters.

The proposed regularized adaptive filtering algorithms have been tested by means of computer simulations of different scenarios and applications. Simulation results are discussed in Section 6. Finally, Section 7 concludes the paper.

2. MSE optimal weighting and regularization

As discussed in Section 1, most regularized linear adaptive filtering algorithms described earlier, are based on the Tikhonov regularized LS estimate in (10) and are tuned by a single regularization parameter α. A more general approach towards regularization is to replace the scaled Euclidian norm $\alpha\|\hat{\mathbf{f}}(t)\|_2^2$ in the Tikhonov regularized LS criterion in (11) by the weighted Euclidian norm of the deviation $[\hat{\mathbf{f}}(t)-\boldsymbol{\xi}]$ of the RIR estimate $\hat{\mathbf{f}}(t)$ from some reference value $\boldsymbol{\xi}$, with an $(n_F+1)\times(n_F+1)$ weighting matrix $\boldsymbol{\Phi}$. We will also include a $t\times t$ weighting matrix $\mathbf{W}$ in the LS term, which will be useful further on:

$$\min_{\hat{\mathbf{f}}(t)}\{[\mathbf{y}-\mathbf{U}\hat{\mathbf{f}}(t)]^{\mathrm{T}}\mathbf{W}[\mathbf{y}-\mathbf{U}\hat{\mathbf{f}}(t)] + [\hat{\mathbf{f}}(t)-\boldsymbol{\xi}]^{\mathrm{T}}\boldsymbol{\Phi}[\hat{\mathbf{f}}(t)-\boldsymbol{\xi}]\}. \tag{28}$$

Minimizing the above criterion leads to a weighted LS estimate that features a general regularization matrix $\boldsymbol{\Phi}$, not necessarily equal to a scaled identity matrix:

$$\hat{\mathbf{f}}(t)=\boldsymbol{\xi}+(\mathbf{U}^{\mathrm{T}}\mathbf{W}\mathbf{U}+\boldsymbol{\Phi})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{W}(\mathbf{y}-\mathbf{U}\boldsymbol{\xi}). \tag{29}$$

The properties of the above estimate will depend on the choice of the weighting matrices $\mathbf{W}$ and $\boldsymbol{\Phi}$, and of the reference value $\boldsymbol{\xi}$. Note that by choosing $\mathbf{W}=\sigma\mathbf{I}$ and $\boldsymbol{\Phi}=v\mathbf{I}$ such that $\sigma^{-1}v=\alpha$, and $\boldsymbol{\xi}=\mathbf{0}$, we obtain the traditional Tikhonov regularized LS estimate, given in (10). A desirable property of a linear estimate is that it has minimum variance [41]. However, for biased estimates, such as the estimate in (29), both the bias and the variance should be minimized. A straightforward choice is to minimize the mean square error (MSE) between the estimated and true RIR, i.e., with $\mathrm{tr}\{\cdot\}$ the trace operator,

$$\min_{\hat{\mathbf{f}}(t)} E\{[\hat{\mathbf{f}}(t)-\mathbf{f}]^{\mathrm{T}}[\hat{\mathbf{f}}(t)-\mathbf{f}]\} = \min_{\hat{\mathbf{f}}(t)} E\{\mathrm{tr}\{[\hat{\mathbf{f}}(t)-\mathbf{f}][\hat{\mathbf{f}}(t)-\mathbf{f}]^{\mathrm{T}}\}\}, \tag{30}$$

since then the variance and the squared bias are weighted equally [41]. Minimization of the above criterion in a deterministic framework leads to an estimate that depends on the unknown RIR $\mathbf{f}$. It is more useful to minimize (30) in a Bayesian framework, where not only the source signal vector $\mathbf{v}$, but also the true RIR $\mathbf{f}$ is viewed as one particular realization of a stochastic vector process. If we assume a Gaussian prior for both vector processes, with first and second order moments given by

$$E\{\mathbf{v}\}=\mathbf{0}, \tag{31}$$

$$\mathrm{cov}\{\mathbf{v}\}=E\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\}=\mathbf{R}_{\mathbf{v}}, \tag{32}$$

and

$$E\{\mathbf{f}\} = \mathbf{f_0}, \tag{33}$$

$$\mathrm{cov}\{\mathbf{f}\} = E\{(\mathbf{f}-\mathbf{f_0})(\mathbf{f}-\mathbf{f_0})^{\mathrm{T}}\} = \mathbf{R_f}, \tag{34}$$

then the minimizing estimate of the MSE criterion in (30) is given by (see, e.g., [41,48, Chapter 4])

$$\hat{\mathbf{f}}(t) = \mathbf{f_0} + (\mathbf{U}^{\mathrm{T}}\mathbf{R_v}^{-1}\mathbf{U} + \mathbf{R_f}^{-1})^{-1} \times \mathbf{U}^{\mathrm{T}}\mathbf{R_v}^{-1}(\mathbf{y} - \mathbf{U}\mathbf{f_0}). \tag{35}$$

Comparing this estimate with the weighted and regularized LS estimate in (29), we conclude that the MSE optimal choice for the weighting matrices $\mathbf{W}$ and $\mathbf{\Phi}$, and for the reference value $\boldsymbol{\xi}$, is

$$\mathbf{W} = \mathbf{R_v}^{-1}, \tag{36}$$

$$\mathbf{\Phi} = \mathbf{R_f}^{-1}, \tag{37}$$

$$\boldsymbol{\xi} = \mathbf{f_0}. \tag{38}$$

Hence the criterion for deriving MSE optimally weighted and regularized LS-based adaptive filtering algorithms can be formulated as (see, e.g., [49, Chapter 5]

$$\min_{\hat{\mathbf{f}}(t)}\{[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}(t)]^{\mathrm{T}}\mathbf{R_v}^{-1}[\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}(t)] + [\hat{\mathbf{f}}(t) - \mathbf{f_0}]^{\mathrm{T}}\mathbf{R_f}^{-1}[\hat{\mathbf{f}}(t) - \mathbf{f_0}]\}. \tag{39}$$

If no prior knowledge on the true RIR is available (i.e., $\|\mathbf{R_f}\|_F \to \infty$, with $\|\cdot\|_F$ the Frobenius norm), then effectively no regularization is applied. Also, when the source signal power is small, then the data will be more reliable and the impact of the regularization term will decrease. On the other hand, if the source signal power increases, the signal-to-noise ratio (SNR) decreases, and then regularization starts playing a more important role.

3. Source signal and true RIR statistics

The approach of Section 2 is only useful when reliable information on the source signal and on the true RIR is available through the mean $\mathbf{f_0}$ and covariance matrices $\mathbf{R_v}$ and $\mathbf{R_f}$. In Section 4, it will become clear that two particular choices for $\mathbf{f_0}$ are of special interest: choosing $\mathbf{f_0} = \mathbf{0}$ leads to a Tikhonov type of regularization, whereas choosing $\mathbf{f_0} = \hat{\mathbf{f}}(t-1)$ yields a Levenberg–Marquardt type of regularization. The choice of the covariance matrices $\mathbf{R_v}$ and $\mathbf{R_f}$ is discussed below.

3.1. Source signal covariance matrix $\mathbf{R_v}$

In room acoustic applications, the source signal will typically be a speech or audio signal. This is then a non-stationary signal, i.e., having time-varying characteristics, and hence it may be desirable to choose $\mathbf{R_v}$ realization-dependent. To this end, we factorize the inverse covariance matrix $\mathbf{R_v}^{-1}$ as follows [6, Appendix II]:

$$\mathbf{R_v}^{-1} = \mathbf{A}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{A}, \tag{40}$$

with $\mathbf{A}$ a unit upper triangular matrix, i.e.,

$$\mathbf{A} \triangleq \begin{bmatrix} 1 & a_{12} & a_{13} & \dots & a_{1t} \\ 0 & 1 & a_{23} & \dots & a_{2t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \tag{41}$$

and $\mathbf{\Sigma}$ a diagonal matrix. It is now convenient, both from a physical point of view, as well as for computational reasons, to model the source signal $v(t)$ as an autoregressive (AR) process of order n_A, with time-varying AR coefficients $a_i(t)$, and residual signal $r(t)$ [43], i.e.,

$$v(t) = -\sum_{i=1}^{n_A} a_i(t)v(t-i) + r(t). \tag{42}$$

The matrices $\mathbf{A}$ and $\mathbf{\Sigma}$ can then be rewritten as

$$\mathbf{A} = \begin{bmatrix} 1 & a_1(t) & a_2(t) & \dots & a_{n_A}(t) & 0 & \dots & 0 \\ 0 & 1 & a_1(t-1) & \dots & a_{n_A-1}(t-1) & a_{n_A}(t-1) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 \end{bmatrix} \tag{43}$$

and

$$\mathbf{\Sigma} = \mathrm{diag}\{\sigma_r^2(t), \dots, \sigma_r^2(1)\} \triangleq E\left\{\begin{bmatrix} r(t) \\ \vdots \\ r(1) \end{bmatrix}[r(t) \ \dots \ r(1)]\right\}. \tag{44}$$

Now, with (40), (43), and (44), the criterion in (39) can be rewritten by shifting the prefiltering matrix $\mathbf{A}$

into the data term $[\mathbf{y}-\mathbf{U}\hat{\mathbf{f}}(t)]$ as follows:

$$\min_{\hat{\mathbf{f}}(t)}\{[\tilde{\mathbf{y}}-\tilde{\mathbf{U}}\hat{\mathbf{f}}(t)]^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}[\tilde{\mathbf{y}}-\tilde{\mathbf{U}}\hat{\mathbf{f}}(t)] + [\hat{\mathbf{f}}(t)-\mathbf{f_0}]^{\mathrm{T}}\mathbf{R}_{\mathbf{f}}^{-1}[\hat{\mathbf{f}}(t)-\mathbf{f_0}]\}, \quad (45)$$

with the prefiltered loudspeaker and microphone signals defined as

$$\tilde{\mathbf{U}} = [\tilde{\mathbf{u}}(t)\ \ \tilde{\mathbf{u}}(t-1)\ \ \ldots\ \ \tilde{\mathbf{u}}(1)]^{\mathrm{T}} \triangleq \mathbf{A}\mathbf{U}, \quad (46)$$

$$\tilde{\mathbf{y}} = [\tilde{y}(t)\ \ \tilde{y}(t-1)\ \ \ldots\ \ \tilde{y}(1)]^{\mathrm{T}} \triangleq \mathbf{A}\mathbf{y}. \quad (47)$$

The AR coefficients $a_i(t)$ and the residual signal variances $\sigma_r^2(t)$ can be estimated using any of the prediction error identification algorithms described in [43], e.g., the prediction-error-method-based adaptive filtering algorithm with row operations (PEM-AFROW), which is used in the computer simulations described in Section 6. In any case, this estimation procedure, as well as the prefiltering of the loudspeaker and microphone signals as in (46) and (47), can be entirely decoupled from the RIR identification algorithm, and can be appended to the unregularized algorithms described in [43], as well as to the regularized algorithms proposed in this paper. Therefore, in the sequel, we make use of the prefiltered loudspeaker and microphone signals and the residual signal variance, making abstraction of how these quantities are calculated.

3.2. True RIR covariance matrix $\mathbf{R_f}$

Taking into account that, in the sequel, we will focus on two particular choices of the mean [$\mathbf{f_0}=\mathbf{0}$ and $\mathbf{f_0}=\hat{\mathbf{f}}(t-1)$], we should also consider two expressions for the true RIR covariance matrix:

$$\mathbf{R_f} = E\{\mathbf{f}\mathbf{f}^{\mathrm{T}}\} \quad (48)$$

and

$$\mathbf{R_f} = E\{[\mathbf{f}-\hat{\mathbf{f}}(t-1)][\mathbf{f}-\hat{\mathbf{f}}(t-1)]^{\mathrm{T}}\}. \quad (49)$$

Starting from the expression in (48), we will suggest two methods for constructing a diagonal estimate of the true RIR covariance matrix $\mathbf{R_f}$. The first method is based on an initial RIR measurement, and the second method on the 3-parameter RIR model proposed in [45]. The motivation for restricting the estimates to be diagonal matrices is two-fold. First of all, with a diagonal estimate for $\mathbf{R_f}$, the regularization is governed by n_F+1 parameters, which seems to be a fair compromise between the existing regularization techniques (using one regularization parameter) and the general case of using a non-diagonal estimate for $\mathbf{R_f}$ (with $(n_F+1)^2$ regularization parameters). The second reason is that the matrix-vector products involving $\mathbf{R_f}$ and $\mathbf{R_f}^{-1}$, as they appear in the algorithms proposed in Sections 4 and 5, can be computed with $\mathrm{O}(n_F)$ operations if diagonal estimates are used, in contrast to the $\mathrm{O}(n_F^2)$ operations that are needed in the non-diagonal case.

3.2.1. Initial RIR measurement

If the application at hand allows so, probably the most reliable way of constructing prior knowledge on the true RIR, is to perform a batch RIR identification during the initialization of the signal enhancement system. Suppose that in this way an initial RIR estimate

$$\bar{\mathbf{f}} = [\bar{f}_0\ \ \bar{f}_1\ \ \ldots\ \ \bar{f}_{n_F}]^{\mathrm{T}} \quad (50)$$

is obtained, then a diagonal estimate of the true RIR covariance matrix may be constructed as follows:

$$\hat{\mathbf{R}}_{\mathbf{f},\mathrm{init}} = \mathrm{diag}\{\bar{f}_0^2,\bar{f}_1^2,\ldots,\bar{f}_{n_F}^2\}. \quad (51)$$

The main disadvantage of this method is that it is non-robust to RIR changes, since the estimate of $\mathbf{R_f}$ is based on one particular measurement. Alternatively, several initial measurements under different conditions (e.g., different microphone or loudspeaker positions) could be averaged before constructing $\hat{\mathbf{R}}_{\mathbf{f},\mathrm{init}}$, or, if permitted by the application, the initial measurement could be updated once in a while.

3.2.2. 3-parameter RIR model

In [45], we have proposed a very simple RIR model, based on the observation that RIRs have a typical form, which may be characterized by three parameters, as illustrated in Fig. 2:

- the *initial delay* d, which corresponds to the time needed for the loudspeaker sound wave to reach the microphone through a direct path (i.e., without reflections),
- the *direct path attenuation* A, which determines the peak response in the RIR, and
- the *exponential decay time constant* τ, which models the envelope of the reverberant tail of the RIR.

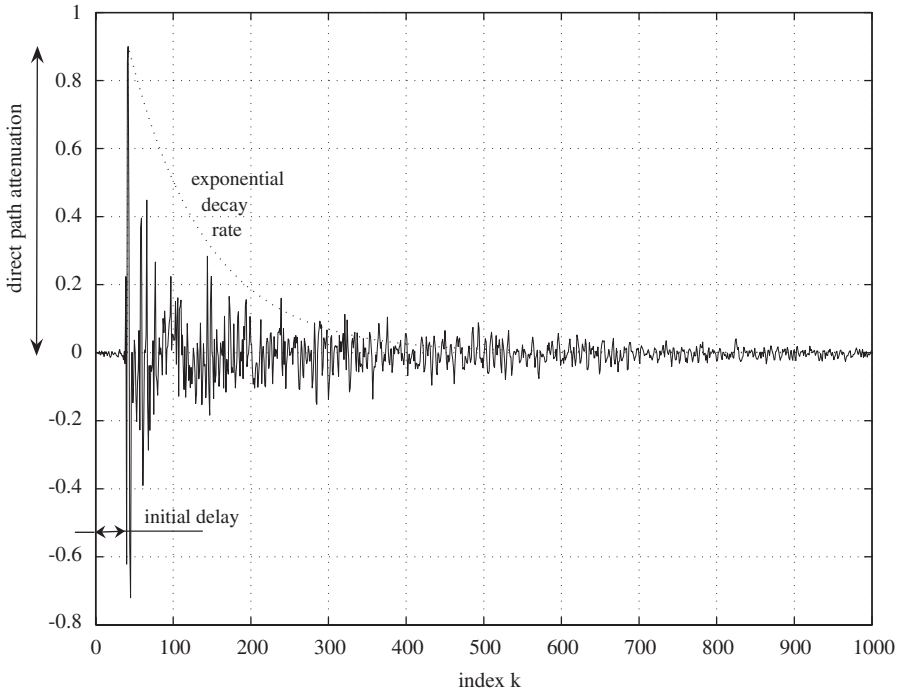


Fig. 2. Room impulse response characterized by three parameters.

These three parameters may be estimated from the acoustic setup (distance between loudspeaker and microphone, acoustic absorption of the walls, room volume, etc.), e.g., using Sabine's reverberation formulas [50]. Hence they can be considered as prior knowledge. If these three parameters are taken into account, a diagonal estimate of the true RIR covariance matrix may be constructed as

$$\hat{\mathbf{R}}_{\mathbf{f},\mathbf{3}} = A\,\mathrm{diag}\left\{\underbrace{\beta,\ldots,\beta}_{d},\ 1,\underbrace{\mathrm{e}^{-\frac{2}{\tau}},\ldots,\mathrm{e}^{-2\frac{n_F-d}{\tau}}}_{n_F-d}\right\}, \quad (52)$$

where β is a small number. As an example, the RIR in Fig. 2 may be approximated by choosing $d = 38$, $A = 0.9$, $\tau = 100$ and $\beta = 10^{-6}$. The advantage of this method is that the parameter τ is invariant to RIR changes due to an arbitrary microphone or loudspeaker movement. Moreover, the other two parameters, d and A, are invariant to microphone or loudspeaker movements as long as the distance between the loudspeaker and the microphone remains constant. Hence the 3-parameter model is found to be more robust to RIR changes than the model based on an initial RIR measurement.

4. Regularized RLS algorithms

An MSE optimally weighted and regularized RLS algorithm with an exponential weighting factor λ can be derived from criterion (45), namely,

$$\min_{\hat{\mathbf{f}}(t)}\{[\tilde{\mathbf{y}} - \tilde{\mathbf{U}}\hat{\mathbf{f}}(t)]^{\mathrm{T}}\boldsymbol{\Sigma}_{\lambda}^{-1}[\tilde{\mathbf{y}} - \tilde{\mathbf{U}}\hat{\mathbf{f}}(t)] + [\hat{\mathbf{f}}(t) - \mathbf{f_0}]^{\mathrm{T}}\mathbf{R}_{\mathbf{f}}^{-1}[\hat{\mathbf{f}}(t) - \mathbf{f_0}]\}, \quad (53)$$

where the exponential data weighting is combined with the MSE optimal weighting in the diagonal matrix $\boldsymbol{\Sigma}_{\lambda}^{-1}$:

$$\boldsymbol{\Sigma}_{\lambda}^{-1} \triangleq \mathrm{diag}\left\{\frac{1}{\sigma_r^2(t)}, \frac{\lambda}{\sigma_r^2(t-1)}, \ldots, \frac{\lambda^{t-1}}{\sigma_r^2(1)}\right\}. \quad (54)$$

The derivation of the RLS algorithm corresponding to (53) and (54) is omitted here, since it is quite similar to deriving the standard RLS algorithm (see, e.g., [9]) or the leaky RLS algorithm [10]. The resulting algorithm is given by

$$\hat{\mathbf{f}}(t) = \mathbf{f_0} + [\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}(t)\mathbf{R}_{\mathbf{f}}^{-1}][\hat{\mathbf{f}}(t-1) - \mathbf{f_0}] + \frac{1}{\sigma_r^2(t)}\mathbf{R}^{-1}(t)\tilde{\mathbf{u}}(t)\tilde{\varepsilon}(t), \quad (55)$$

$$\mathbf{R}(t) = \lambda\mathbf{R}(t-1) + \frac{1}{\sigma_r^2(t)}\tilde{\mathbf{u}}(t)\tilde{\mathbf{u}}^{\mathrm{T}}(t) + (1-\lambda)\mathbf{R}_{\mathbf{f}}^{-1}, \tag{56}$$

$$\tilde{\varepsilon}(t) = \tilde{y}(t) - \tilde{\mathbf{u}}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1), \tag{57}$$

where $\tilde{\mathbf{u}}(t)$, $\tilde{y}(t)$, and $\tilde{\varepsilon}(t)$ represent the loudspeaker signal vector, microphone signal, and a priori residual signal, respectively, after prefiltering with the prediction error filter in the top row of $\mathbf{A}$ in (43), and $\sigma_r^2(t)$ represents the residual signal variance from the source signal linear prediction (see Section 3.1 and [43]).

To point out the relationship of the above algorithm with the existing regularized RLS algorithms, described in Section 1, we will focus on two particular choices for the mean $\mathbf{f_0}$ of the true RIR distribution. Choosing $\mathbf{f_0} = \mathbf{0}$, leads to the MSE optimal *Tikhonov regularized RLS* (*TR-RLS*) *algorithm*, shown in the upper part of Table 1, which is a generalization of the traditional Tikhonov regularized RLS algorithm (or leaky RLS algorithm) in (15)–(17). In Section 1, the traditional Levenberg–Marquardt regularized RLS algorithm was considered as an approximation to the traditional Tikhonov regularized RLS algorithm by omitting the correction term in the RIR weight update. However, we now observe that an MSE optimal *Levenberg–Marquardt regularized RLS* (*LMR-RLS*) *algorithm* can also be obtained from the algorithm in (55)–(57) by setting $\mathbf{f_0} = \hat{\mathbf{f}}(t-1)$.

Table 1
MSE optimally regularized RLS algorithms

TR-RLS—Tikhonov regularized recursive least squares

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mathbf{R}^{-1}(t)\left[\frac{1}{\sigma_r^2(t)}\tilde{\mathbf{u}}(t)\tilde{\varepsilon}(t) - (1-\lambda)\mathbf{R}_{\mathbf{f}}^{-1}\hat{\mathbf{f}}(t-1)\right],$$

$$\mathbf{R}(t) = \lambda\mathbf{R}(t-1) + \frac{1}{\sigma_r^2(t)}\tilde{\mathbf{u}}(t)\tilde{\mathbf{u}}^{\mathrm{T}}(t) + (1-\lambda)\mathbf{R}_{\mathbf{f}}^{-1},$$

$$\tilde{\varepsilon}(t) = \tilde{y}(t) - \tilde{\mathbf{u}}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1).$$

LMR-RLS—Levenberg–Marquardt regularized recursive least squares

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \frac{1}{\sigma_r^2(t)}\mathbf{R}^{-1}(t)\tilde{\mathbf{u}}(t)\tilde{\varepsilon}(t),$$

$$\mathbf{R}(t) = \lambda\mathbf{R}(t-1) + \frac{1}{\sigma_r^2(t)}\tilde{\mathbf{u}}(t)\tilde{\mathbf{u}}^{\mathrm{T}}(t) + (1-\lambda)\mathbf{R}_{\mathbf{f}}^{-1},$$

$$\tilde{\varepsilon}(t) = \tilde{y}(t) - \tilde{\mathbf{u}}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1).$$

The resulting LMR-RLS algorithm is shown in the lower part of Table 1.

A severe drawback of all the regularized RLS algorithms, either with the traditional scaled identity matrix, or with a regularization matrix incorporating prior knowledge, is that the inversion of the regularized correlation matrix $\mathbf{R}(t)$ cannot be performed cheaply, i.e., in $O(n_F^2)$ operations, based on the MIL. While in the standard RLS algorithm the correlation matrix is updated with a rank-one matrix in each iteration, in the regularized algorithms a full-rank regularization matrix is added in each iteration. Nevertheless, some procedures have been proposed to reduce the complexity of RLS algorithms regularized with a scaled identity matrix, to $O(n_F^2)$ operations per iteration. We illustrate how two such procedures can also be applied in the case of an MSE optimal regularization matrix $\mathbf{R}_{\mathbf{f}}^{-1}$ that is not necessarily a scaled identity matrix. In fact, these procedures can be applied to arbitrary, non-diagonal regularization matrices, and obviously also to the diagonal matrices $\hat{\mathbf{R}}_{\mathbf{f},\mathrm{init}}$ and $\hat{\mathbf{R}}_{\mathbf{f},3}$ described in Section 3.

- A first procedure starts from the decomposition of the $(n_F+1)\times(n_F+1)$ regularization matrix in a sum of n_F+1 rank-one matrices, or more specifically, a sum of n_F+1 outer vector products. This decomposition is easily constructed for a diagonal regularization matrix, e.g., for the inverse of the matrix $\hat{\mathbf{R}}_{\mathbf{f},\mathrm{init}}$ proposed in (51):

$$\hat{\mathbf{R}}_{\mathbf{f},\mathrm{init}}^{-1} = \sum_{k=0}^{n_F} \boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}}, \tag{58}$$

with

$$\boldsymbol{\phi}_k^{\mathrm{T}} \triangleq [0 \ \ldots \ 0 \ \bar{f}_k^{-1} \ 0 \ \ldots \ 0]. \tag{59}$$

The regularized RLS correlation matrix update can then be approximated by adding each outer product $\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}}$ with appropriate scaling only once in n_F+1 iteration steps, instead of adding the full rank matrix $(1-\lambda)\mathbf{R}_{\mathbf{f}}^{-1}$ in each iteration step:

$$\begin{aligned}\mathbf{R}(t) = \lambda\mathbf{R}(t-1) &+ \frac{1}{\sigma_r^2(t)}\tilde{\mathbf{u}}(t)\tilde{\mathbf{u}}^{\mathrm{T}}(t) \\ &+ (1-\lambda)(n_F+1)\boldsymbol{\phi}_k\boldsymbol{\phi}_k^{\mathrm{T}}, \\ k &= \mathrm{mod}\,(t, n_F+1).\end{aligned} \tag{60}$$

Since this approximate update involves two rank-one matrix updates, applying the MIL to (60) yields an inverse correlation matrix update which requires the inversion of a 2×2 matrix, and can hence be computed in $O(n_F^2)$ operations. This

method was first proposed in [12] and adapted later on to the so-called dynamic regularization method for AEC [18].

- A second procedure was proposed in the context of regularization of fast RLS algorithms [46], and consists in replacing the exponential data window with a rectangular window. If the adaptive filter input correlation matrix is initialized to the regularization matrix, i.e., $\mathbf{R}(0)=\mathbf{R}_{\mathbf{f}}^{-1}$, then no restoration of the regularization matrix needs to be performed during the operation of the algorithm, if a sliding window approach is followed. Hence the correlation matrix update consists of two rank-one matrix updates:

$$\begin{aligned}\mathbf{R}(t)=\mathbf{R}(t-1)&+\frac{1}{\sigma_r^2(t)}\tilde{\mathbf{u}}(t)\tilde{\mathbf{u}}^{\mathrm{T}}(t)\\&-\frac{1}{\sigma_r^2(t-n_F-1)}\tilde{\mathbf{u}}(t-n_F-1)\\&\times\tilde{\mathbf{u}}^{\mathrm{T}}(t-n_F-1),\end{aligned}\tag{61}$$

and applying the MIL leads to an $O(n_F^2)$ update of the inverse correlation matrix, which involves the inversion of a 2×2 matrix. However, this method is not very appealing from a numerical point of view, since a sliding rectangular data window approach suffers from linear round-off error build-up.

5. Regularized APA and NLMS algorithms

Considering the APA and the NLMS algorithm as URLS algorithms [16], it seems straightforward to derive regularized APA and NLMS algorithms from the same criterion that led to the MSE optimally weighted and regularized RLS algorithms in Section 4, and to exploit the underdetermined nature of the algorithms to reduce the dimension of the input correlation matrix that has to be inverted. Rewriting the criterion in (45) for the underdetermined case, leads to

$$\begin{aligned}\min_{\hat{\mathbf{f}}(t)}\{&[\tilde{\mathbf{y}}_M(t)-\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t)]^{\mathrm{T}}\boldsymbol{\Sigma}_M^{-1}(t)[\tilde{\mathbf{y}}_M(t)\\&-\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t)]+[\hat{\mathbf{f}}(t)-\mathbf{f_0}]^{\mathrm{T}}\mathbf{R}_{\mathbf{f}}^{-1}[\hat{\mathbf{f}}(t)-\mathbf{f_0}]\},\end{aligned}\tag{62}$$

with M the projection order, as defined earlier. Consistent with the notation in (22) and (23), we define the prefiltered data matrices $\tilde{\mathbf{U}}_M(t)$ and $\tilde{\mathbf{y}}_M(t)$ as

$$\begin{aligned}\tilde{\mathbf{U}}_M(t)&\triangleq[\tilde{\mathbf{u}}(t)\ \ \tilde{\mathbf{u}}(t-1)\ \ \ldots\ \ \tilde{\mathbf{u}}(t-M+1)]&(63)\\&=[\mathbf{u}(t)\ \ \mathbf{u}(t-1)\ \ \ldots\ \ \mathbf{u}(t-M-n_A+1)]\\&\quad\times\mathbf{A}_M^{\mathrm{T}}(t),&(64)\end{aligned}$$

$$\begin{aligned}\tilde{\mathbf{y}}_M(t)&\triangleq[\tilde{y}(t)\ \ \tilde{y}(t-1)\ \ \ldots\ \ \tilde{y}(t-M+1)]^{\mathrm{T}}&(65)\\&=\mathbf{A}_M(t)[y(t)\ \ y(t-1)\\&\quad\ldots\ \ y(t-M-n_A+1)]^{\mathrm{T}},&(66)\end{aligned}$$

where $\mathbf{A}_M(t)$ is defined as the $M\times(M+n_A)$ top left submatrix of $\mathbf{A}$ in (43). The diagonal weighting matrix $\boldsymbol{\Sigma}_M(t)$ appearing in (62) is defined as

$$\boldsymbol{\Sigma}_M(t)\triangleq\mathrm{diag}\{\sigma_r^2(t),\ldots,\sigma_r^2(t-M+1)\}.\tag{67}$$

Minimizing the criterion in (62), and subsequently applying the MIL to reduce the dimension of the $(n_F+1)\times(n_F+1)$ inverse correlation matrix to $M\times M$, leads to the following underdetermined estimate:

$$\begin{aligned}\hat{\mathbf{f}}(t)=\mathbf{f_0}&+\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)[\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)+\boldsymbol{\Sigma}_M(t)]^{-1}\\&\times[\tilde{\mathbf{y}}_M(t)-\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{f_0}].\end{aligned}\tag{68}$$

Note that this estimate is non-recursive, i.e., it does not depend on the previous estimate $\hat{\mathbf{f}}(t-1)$. A recursive algorithm can be obtained by explicitly bringing in the dependency on $\hat{\mathbf{f}}(t-1)$, as follows:

$$\begin{aligned}\hat{\mathbf{f}}(t)&=\mathbf{f_0}+\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)[\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)+\boldsymbol{\Sigma}_M(t)]^{-1}\\&\quad\times[\tilde{\mathbf{y}}_M(t)-\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1)\\&\quad+\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1)-\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{f_0})],&(69)\\&=\mathbf{f_0}+\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)[\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)+\boldsymbol{\Sigma}_M(t)]^{-1}\\&\quad\times\mathbf{U}_M^{\mathrm{T}}(t)[\hat{\mathbf{f}}(t-1)-\mathbf{f_0}]\\&\quad+\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)[\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)+\boldsymbol{\Sigma}_M(t)]^{-1}\\&\quad\times[\tilde{\mathbf{y}}_M(t)-\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1)].&(70)\end{aligned}$$

Finally, by again applying the MIL to the second term on the right-hand side, we rewrite the estimate in (70) into an expression which is computationally less interesting, but more convenient to understand the estimator's properties:

$$\begin{aligned}\hat{\mathbf{f}}(t)=\mathbf{f_0}&+\{\mathbf{I}-[\tilde{\mathbf{U}}_M(t)\boldsymbol{\Sigma}_M^{-1}(t)\mathbf{U}_M^{\mathrm{T}}(t)\\&+\mathbf{R}_{\mathbf{f}}^{-1}]^{-1}\mathbf{R}_{\mathbf{f}}^{-1}\}[\hat{\mathbf{f}}(t-1)-\mathbf{f_0}]\\&+\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)[\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{R}_{\mathbf{f}}\tilde{\mathbf{U}}_M(t)\\&+\boldsymbol{\Sigma}_M(t)]^{-1}[\tilde{\mathbf{y}}_M(t)-\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1)].\end{aligned}\tag{71}$$

From (71), it can be seen that the minimizing estimate of the underdetermined criterion in (62), consists of three terms: the mean value $\mathbf{f_0}$ of the true RIR distribution, a leakage term depending on the deviation $[\hat{\mathbf{f}}(t-1)-\mathbf{f_0}]$ of the previous estimate from the mean value, and a proportionate adaptation term. However, from extensive simulations, we have concluded that leakage (even in the simple case of the leaky LMS algorithm [11]) does not have a beneficial effect on the

convergence behaviour of URLS algorithms in room acoustic applications. We will therefore not consider Tikhonov regularized URLS algorithms here (even though their update formulae are readily derived with $\mathbf{f_0} = \mathbf{0}$). The leakage term in (71) disappears by choosing $\mathbf{f_0} = \hat{\mathbf{f}}(t-1)$, and hence we end up with a Levenberg–Marquardt type of regularization. If we finally introduce a relaxation factor μ, then the resulting algorithms are the *Levenberg–Marquardt regularized APA* (*LMR-APA*) and, when the projection order is set to $M = 1$, the *Levenberg–Marquardt regularized NLMS* (*LMR-NLMS*) *algorithm*, which are shown in Table 2. Note that the traditionally regularized APA and NLMS algorithms, described in Section 1, can be obtained as special cases of these algorithms, by omitting the prefiltering step, and by choosing $\mathbf{\Sigma}_M(t) = \sigma\mathbf{I}$ and $\mathbf{R_f} = v\mathbf{I}$ such that $\sigma v^{-1} = \alpha$.

The LMR-APA and LMR-NLMS algorithms are closely related to the proportionate APA (PAPA) [21] and proportionate NLMS (PNLMS) [47] algorithms. In the PNLMS RIR weight update,

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu\frac{\mathbf{G}(t)\mathbf{u}(t)\varepsilon(t)}{\mathbf{u}^{\mathrm{T}}(t)\mathbf{G}(t)\mathbf{u}(t) + \alpha}, \tag{72}$$

the diagonal matrix $\mathbf{G}(t)$ is constructed as follows:

$$\mathbf{G}(t) = \frac{1}{\bar{g}(t)}\mathrm{diag}\{g_0(t), \ldots, g_{n_F}(t)\}, \tag{73}$$

with, for $k = 0, \ldots, n_F$,

$$g_k(t) \triangleq \max\{\rho \cdot \max\{\delta, |\hat{f}_0(t-1)|, \ldots, |\hat{f}_{n_F}(t-1)|\}, |\hat{f}_k(t-1)|\}, \tag{74}$$

$$\bar{g}(t) \triangleq \frac{1}{n_F+1}\sum_{k=0}^{n_F} g_k(t), \tag{75}$$

where ρ and δ are small positive parameters. This choice of $\mathbf{G}(t)$ was made somewhat intuitively by Duttweiler [47], with the aim of allocating a larger portion of the available adaptation energy to larger adaptive filter weights, and hence to speed up the convergence. Later on, Chen et al. [23] provided an interpretation of the above choice of $\mathbf{G}(t)$ in terms of Bayesian priors. Suppose that some prior knowledge is available on the true RIR through its probability density function (PDF) $p(\mathbf{f})$, then it is advantageous to choose the PNLMS parameters $g_k(t)$ proportional to the negative logarithm of the prior distribution [23], i.e.,

$$g_k(t) \propto -\ln[p(f_k)], \quad k = 0, \ldots, n_F. \tag{76}$$

It can be shown that, slightly deviating from the treatment in [23],[1]

$$-\ln[p(f_k)] \sim |f_k - f_{0,k}| \quad \text{for a Laplacian prior,} \tag{77}$$

$$\sim (f_k - f_{0,k})^2 \quad \text{for a Gaussian prior,} \tag{78}$$

where f_k and $f_{0,k}$ represent the $(k+1)$th elements of the vectors $\mathbf{f}$ and $\mathbf{f_0}$, respectively. In the PNLMS algorithm, the negative logarithm of a Laplacian prior on the true RIR is approximated using a smoothed estimate of $|f_k|$, given in (74), which is obtained by taking the L_∞-norm of $|\hat{f}_k(t-1)|$, $\rho|\hat{\mathbf{f}}(t-1)|$ and $\rho\delta$, hence implicitly assuming that the mean $\mathbf{f_0} = \mathbf{0}$. In other words, the prior knowledge on the true RIR is extracted from the previous RIR estimate $\hat{\mathbf{f}}(t-1)$. Obviously, the prior knowledge could also be constructed using a Gaussian prior, in which case the matrix $\mathbf{G}(t)$ is in fact a (normalized) estimate of the true RIR covariance matrix $\mathbf{R_f}$ in (49). Also, the smoothed estimate could be calculated using an L_2-norm instead of an L_∞-norm, which appears to have a beneficial effect on the PNLMS convergence.

From this point of view, the LMR-NLMS algorithm provides a new choice for the matrix $\mathbf{G}(t)$ in the PNLMS algorithm, which is optimal in the sense of minimizing the underdetermined MSE optimally weighted and regularized criterion in (62). The above comparison of LMR-NLMS with PNLMS can be done similarly for the LMR-APA

Table 2
MSE optimally regularized underdetermined RLS algorithms

LMR-APA—Levenberg–Marquardt regularized APA

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu\mathbf{R_f}\tilde{\mathbf{U}}_M(t)[\tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\mathbf{R_f}\tilde{\mathbf{U}}_M(t) + \mathbf{\Sigma}_M(t)]^{-1}\tilde{\boldsymbol{\varepsilon}}_M(t),$$

$$\tilde{\boldsymbol{\varepsilon}}_M(t) = \tilde{\mathbf{y}}_M(t) - \tilde{\mathbf{U}}_M^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1).$$

LMR-NLMS—Levenberg–Marquardt regularized NLMS

$$\hat{\mathbf{f}}(t) = \hat{\mathbf{f}}(t-1) + \mu\frac{\mathbf{R_f}\tilde{\mathbf{u}}(t)\tilde{\varepsilon}(t)}{\tilde{\mathbf{u}}^{\mathrm{T}}(t)\mathbf{R_f}\tilde{\mathbf{u}}(t) + \sigma_r^2(t)},$$

$$\tilde{\varepsilon}(t) = \tilde{y}(t) - \tilde{\mathbf{u}}^{\mathrm{T}}(t)\hat{\mathbf{f}}(t-1).$$

[1] Similarity between two scalar functions $\zeta(f_k)$ and $\eta(f_k)$ is denoted with the $\sim$ symbol, and defined as follows: $\zeta(f_k) \sim \eta(f_k) \Leftrightarrow d(\zeta(f_k^1), \zeta(f_k^2)) = rd(\eta(f_k^1), \eta(f_k^2))$, where f_k^1 and f_k^2 represent different realizations of the stochastic variable f_k, r is a scaling factor, and the distance operator $d(x, y)$ is defined as the Minkowski or p-norm distance, which reduces to $|x - y|$ in the scalar case, for any $p \in \mathbb{N}$.

and PAPA algorithms. An important difference with the existing proportionate adaptation algorithms is that in the proposed LMR-APA and LMR-NLMS algorithms the regularization matrix $\mathbf{R_f}$ is fixed. As a consequence, the proposed algorithms may respond somewhat slower to an RIR change, but on the other hand they are much less computationally demanding than the PAPA and PNLMS algorithms, in which $\mathbf{G}(t)$ is recalculated with (74) and (75) in each iteration of the adaptive filter.

6. Simulation results

We present the results of five computer simulations, in which a comparison of the proposed and existing adaptive filtering algorithms is made in a variety of RIR identification scenarios. The performance measure for comparing the different algorithms is the misadjustment, which is defined as the normalized Euclidian distance between the estimated and true RIR on a logarithmic scale:

$$\text{misadjustment (dB)} = 20\log_{10}\frac{\|\hat{\mathbf{f}}(t)-\mathbf{f}\|}{\|\mathbf{f}\|}. \tag{79}$$

In all simulations, the sampling frequency $f_s = 8\,\text{kHz}$.

6.1. Regularized RLS and PEM-AFROW algorithms for AEC in a stationary environment

In these two simulations, AEC is performed using regularized RLS and PEM-AFROW algorithms in a stationary echo path situation. The PEM-AFROW algorithm is a recursive prediction error identification algorithm, in which an AR model of the near-end signal is estimated for prefiltering the loudspeaker and microphone signals as in (46) and (47), see [2,42,43]. The echo path is equal to the RIR plotted in Fig. 2, which is of known length $n_F + 1 = 1000$ and is taken from the measured RIR database in [51]. An exponential data window is applied with an effective length of three times the RIR length, i.e., $\lambda = 0.9997$. The far-end signal is a 1.23 s male speech signal (equivalent to $N = 9831$ samples), whereas the near-end signal is either a stationary Gaussian white noise (GWN) signal or a male speech signal, both with variance $\sigma_v^2 = 10^{-3}$. This results in an SNR of 18 dB at the microphone, i.e.,

$$\text{SNR}_{\text{mic}} \triangleq 10\log_{10}\frac{\sum_{t=1}^{N}|x(t)|^2}{\sum_{t=1}^{N}|v(t)|^2} = 18\,\text{dB}. \tag{80}$$

As reference algorithms, we consider the standard RLS algorithm and the traditional Tikhonov and Levenberg–Marquardt regularized RLS algorithms with a scaled identity regularization matrix and $\alpha = \sigma_v^2$, which we will further call the TR-RLS $\alpha\mathbf{I}$ and LMR-RLS $\alpha\mathbf{I}$ algorithms.

In a first simulation, with a near-end GWN signal, these reference algorithms are compared to the proposed TR-RLS $\hat{\mathbf{R}}_{\mathbf{f},3}$ and LMR-RLS $\hat{\mathbf{R}}_{\mathbf{f},3}$ algorithms, in which the 3-parameter model given in (52) is used for constructing a non-identity regularization matrix with $d = 38$, $A = 0.9$, $\tau = 100$ and $\beta = 10^{-6}$. Finally, also the TR-RLS $\hat{\mathbf{R}}_{\mathbf{f},\text{true}}$ and LMR-RLS $\hat{\mathbf{R}}_{\mathbf{f},\text{true}}$ algorithms are simulated, in which regularization is performed using a diagonal matrix as in (51) with $\bar{\mathbf{f}}$ equal to the true RIR $\mathbf{f}$. The latter algorithms illustrate the "best-case" performance that can be obtained with the proposed regularized RLS algorithms in a stationary environment, which could be realized only if an exact initial measurement of the unknown RIR were available. In all regularized algorithms, we assume exact knowledge of the near-end GWN signal variance σ_v^2. Since the near-end signal is a white noise signal, no prefiltering is applied to the loudspeaker and microphone signals.

The convergence curves of the different RLS algorithms are shown in Fig. 3. The main observation is that, as expected, using more prior knowledge on the unknown RIR leads to faster convergence and to a smaller steady-state misadjustment. The initial divergence of the algorithms is due to the lack of excitation in the first 1000 recursion steps (the loudspeaker signal vector $\mathbf{u}(t)$ contains $n_F + 1 - t$ zeros if $t < n_F + 1$). This effect is less pronounced for the regularized algorithms, which moreover recover faster when more prior knowledge is available. After the initial divergence effect has disappeared, it can be seen that the $\alpha\mathbf{I}$ regularized algorithms do not perform better than the standard RLS algorithm. Finally, we note that the difference between Tikhonov and Levenberg–Marquardt regularization appears to be negligible.

In a second simulation, a near-end speech signal is applied, such that prefiltering of the loudspeaker and microphone signals with the inverse near-end signal model is desirable. The regularized PEM-AFROW algorithms are implemented with a Gauss–Newton update, to allow for fair comparison with the regularized RLS algorithms, and the AR model order is set to 12. Since the difference

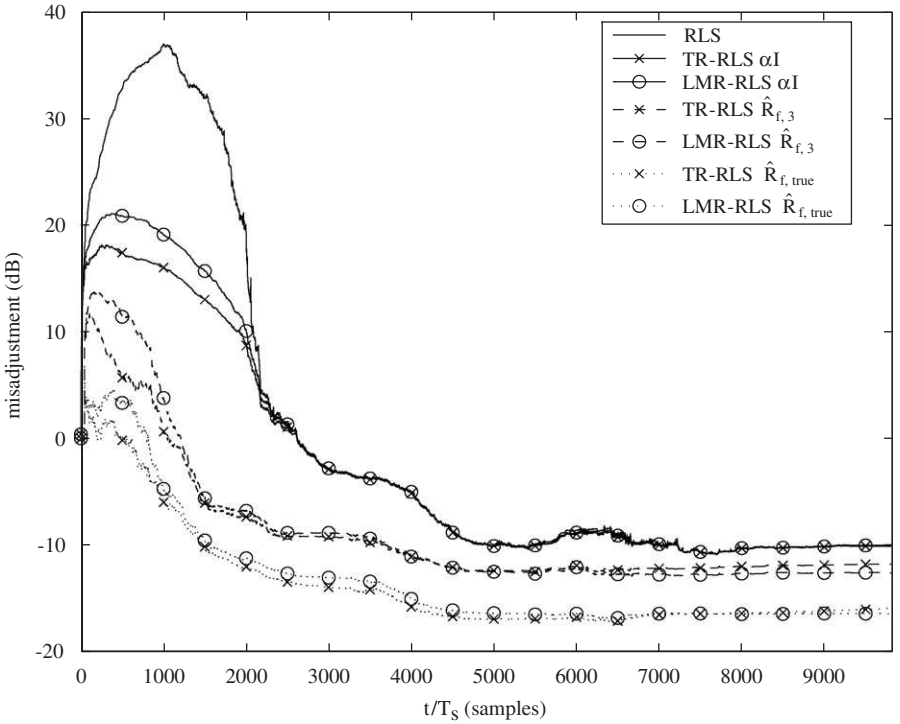


Fig. 3. Convergence curves of regularized RLS algorithms for an AEC application with a near-end GWN signal in a stationary environment.

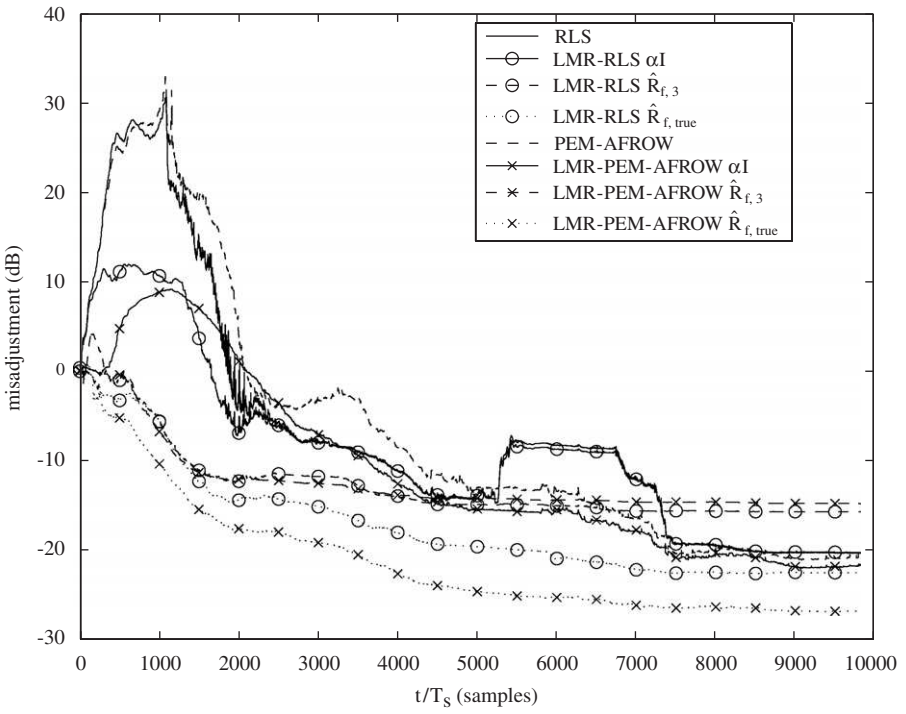


Fig. 4. Convergence curves of regularized RLS and PEM-AFROW algorithms for an AEC application with a near-end speech signal in a stationary environment.

between Tikhonov and Levenberg–Marquardt regularization is small, we have plotted in Fig. 4 only the convergence curves of the Levenberg–Marquardt regularized algorithms. A first observation is that the unregularized RLS algorithm shows a slightly better convergence than the unregularized PEM-AFROW algorithm, except in the time interval $t/T_s \in [5250, 7250]$, where a high-frequency artefact in the near-end signal causes divergence of the RLS algorithm, without affecting the double-talk-robust PEM-AFROW algorithm. Another issue that can be remarked from Fig. 4 is that the $\hat{\mathbf{R}}_{\mathbf{f},3}$ regularized algorithms show better initial convergence than the unregularized and $\alpha\mathbf{I}$ regularized ones, but afterwards tend to converge to a biased solution.

6.2. Regularized NLMS and PEM-AFROW algorithms for AEC in a non-stationary environment

In the next two simulations, the performance of regularized NLMS and PEM-AFROW algorithms is compared for AEC in a non-stationary environment. We switch between three different room impulse responses of known length $n_F + 1 = 1000$, that were measured in our acoustic lab. The first RIR change occurs at $t/T_s = 1.2 \times 10^5$ samples, and consists in a change of 75 cm in the microphone position, in such a way that the distance between the loudspeaker and the microphone remains constant. After the second RIR change, occurring at $t/T_s = 2.5 \times 10^5$ samples, the loudspeaker and microphone positions remain unchanged, but the room is made more reverberant by decreasing the absorption coefficients of the walls and ceiling (the $T60$ reverberation times of the first two RIRs and of the third RIR were roughly estimated as $T60_{1,2} = 0,225\,\text{s}$ and $T60_3 = 0,3\,\text{s}$, respectively). In this simulation, the far-end signal is a 46 s male speech signal (equivalent to $N = 368\,320$ samples), and the near-end signal is either a stationary GWN signal, or a male speech signal, both with variance $\sigma_v^2 = 3 \times 10^{-5}$, resulting in $\text{SNR}_{\text{mic}} = 16\,\text{dB}$.

The reference algorithms are now the (unregularized) standard NLMS algorithm, the proportionate NLMS algorithm with $\rho = 5/(n_F + 1)$ and $\delta = 0.01$ as suggested in [47], and the traditionally regularized NLMS algorithm as given in (26) and (27) with $\alpha = \sigma_v^2$, which is further called LMR-NLMS $\alpha\mathbf{I}$. In the first simulation, with a near-end GWN signal, these algorithms are compared to the LMR-NLMS $\hat{\mathbf{R}}_{\mathbf{f},3}$ and LMR-NLMS $\hat{\mathbf{R}}_{\mathbf{f},\text{true}}$ algorithms in which the regularization matrix is based on the 3-parameter model (with $d = 75$, $A = 0.1022$, $\tau = 70$ and $\beta = 10^{-6}$), and on exact knowledge of the first RIR, respectively. The regularization matrix is not altered after the two RIR changes, such that the robustness of the different regularized algorithms w.r.t. RIR changes can be evaluated. The step size μ is individually tuned for each of the algorithms such that the excess MSE in a stationary environment would approximately be the same for all algorithms. Again, in this case, no prefiltering is applied since the near-end signal is white noise.

The convergence curves and step sizes are shown in Fig. 5. First of all, we observe that the improvement in convergence speed of the existing PNLMS and LMR-NLMS $\alpha\mathbf{I}$ algorithms, as compared to the standard NLMS algorithm, is small. A significantly better convergence behaviour is obtained with the proposed LMR-NLMS $\hat{\mathbf{R}}_{\mathbf{f},3}$ and LMR-NLMS $\hat{\mathbf{R}}_{\mathbf{f},\text{true}}$ algorithms in the time interval $t/T_s = [0, 1.2 \times 10^5]$ samples, where the regularization matrix of the proposed algorithms is based on the "correct" RIR. After the first RIR change, the LMR-NLMS $\hat{\mathbf{R}}_{\mathbf{f},\text{true}}$ algorithm's performance decreases dramatically, whereas the LMR-NLMS $\hat{\mathbf{R}}_{\mathbf{f},3}$ algorithm converges equally fast as in the first time interval. This is not much of a surprise, since the regularization matrix $\hat{\mathbf{R}}_{\mathbf{f},\text{true}}$ based on the true first RIR will be a bad model for the second RIR covariance matrix, whereas the 3-parameter model regularization matrix $\hat{\mathbf{R}}_{\mathbf{f},3}$ of the first RIR will still be valid for the second RIR. Indeed, the microphone repositioning does not alter any of the three parameters on which $\hat{\mathbf{R}}_{\mathbf{f},3}$ is based, since the distance between the loudspeaker and the microphone remains constant. However, after the second RIR change the parameter τ in $\hat{\mathbf{R}}_{\mathbf{f},3}$ will have an inaccurate value, since the room reverberation has increased. This clearly affects the LMR-NLMS $\hat{\mathbf{R}}_{\mathbf{f},3}$ convergence speed, but still this algorithm outperforms all other algorithms.

In the case of a near-end speech signal, the above algorithms are in addition compared to the LMR-PEM-AFROW $\alpha\mathbf{I}$, LMR-PEM-AFROW $\hat{\mathbf{R}}_{\mathbf{f},3}$, and LMR-PEM-AFROW $\hat{\mathbf{R}}_{\mathbf{f},\text{true}}$ algorithms. The PEM-AFROW algorithm again employs an AR(12) model for the near-end signal linear prediction, but now a stochastic gradient update is used for adapting the RIR estimate. The convergence curves shown in Fig. 6 reveal more or less the same issues as in the white noise case. In addition, we observe that the LMR-PEM-AFROW algo-

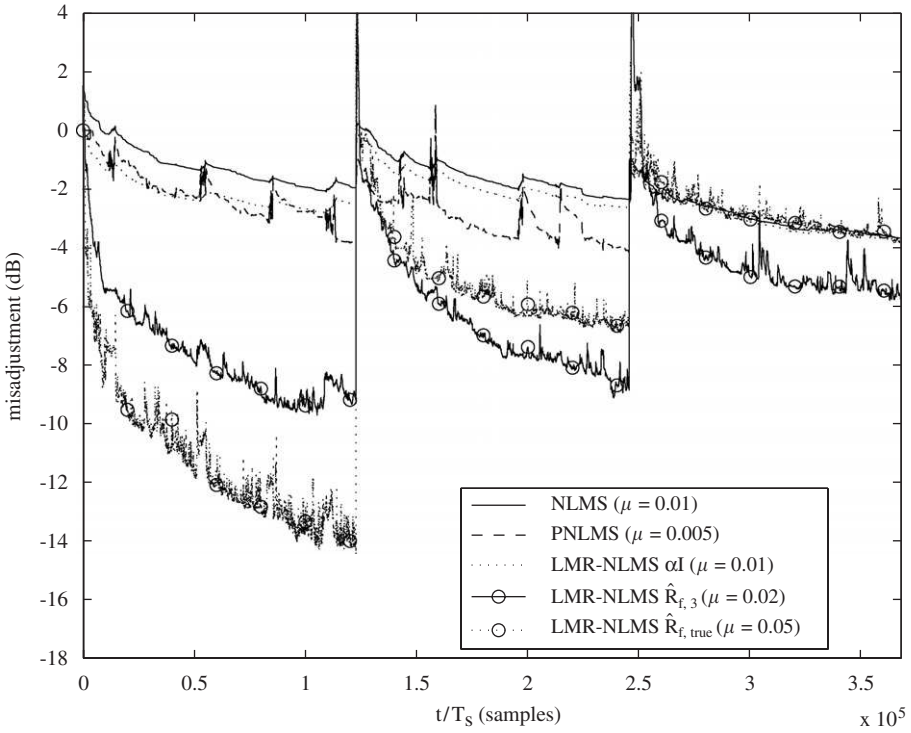


Fig. 5. Convergence curves of regularized NLMS algorithms for an AEC application with a near-end GWN signal and two echo path changes.

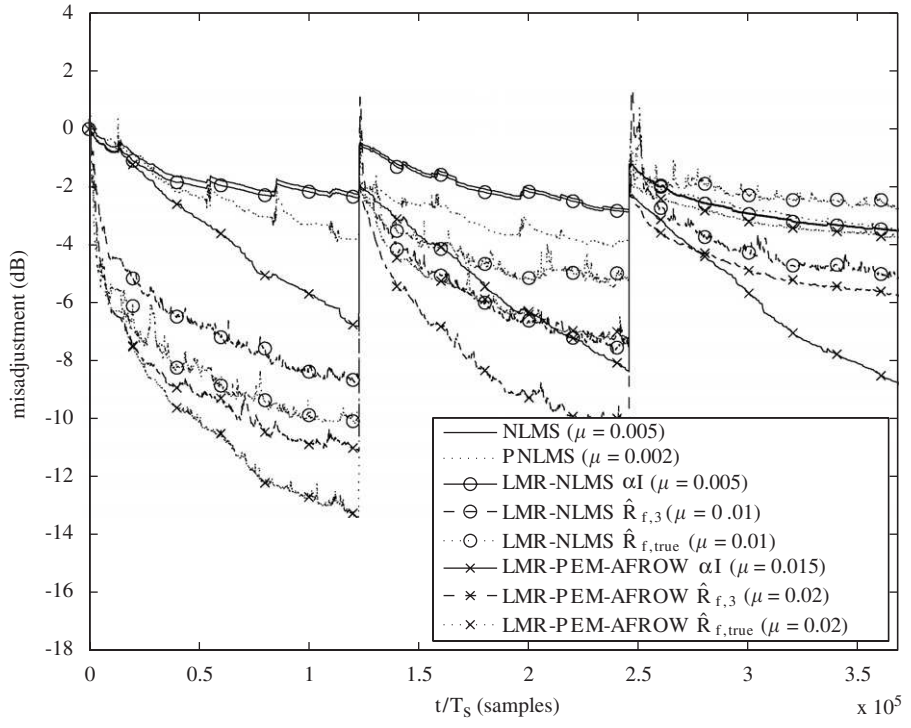


Fig. 6. Convergence curves of regularized NLMS and PEM-AFROW algorithms for an AEC application with a near-end speech signal and two echo path changes.

rithms perform significantly better than their LMR-NLMS counterparts, for the same choice of the regularization matrix.

6.3. Regularized PEM-AFROW algorithms for adaptive feedback cancellation in a stationary environment

In this final simulation, the proposed LMR-PEM-AFROW algorithm is applied in a closed-loop scenario for performing adaptive feedback cancellation. The source signal $v(t)$, equal to the same male speech signal as in the previous simulation, is added to the feedback signal before being amplified and delayed in the forward path. The resulting signal is sent to the loudspeaker, after which it is filtered in the feedback path to yield the feedback signal. As in the AEC simulations with near-end speech, we should also take into account the source signal characteristics in this scenario. We again apply the LMR-PEM-AFROW algorithm with stochastic gradient update, featuring an AR(12) linear prediction of the source signal. In this simulation, the feedback path is equal to the RIR used in the first simulation. The forward path delay is set equal to the PEM-AFROW linear prediction window length $L = 160$ samples, as suggested in [2], and the forward path gain is set to $K = -19\,\mathrm{dB}$, resulting in a closed-loop gain margin of 3 dB. These settings lead to $\mathrm{SNR}_{\mathrm{mic}} = -11\,\mathrm{dB}$.

The LMR-PEM-AFROW algorithm is compared for different choices of the regularization matrix with the (unregularized) standard NLMS algorithm in Fig. 7. It is clear that the LMR-PEM-AFROW algorithm is better suited to the closed-loop estimation problem than the NLMS algorithm. We further note that initially, the LMR-PEM-AFROW $\hat{\mathbf{R}}_{\mathbf{f},3}$ and LMR-PEM-AFROW $\hat{\mathbf{R}}_{\mathbf{f},\mathbf{true}}$ algorithms converge considerably faster than the LMR-PEM-AFROW $\alpha\mathbf{I}$ algorithm, yet after some time the misadjustment of these three algorithms evolves to approximately the same level.

7. Conclusion

In this paper, we have developed a framework for understanding and optimizing the technique of regularization in adaptive filtering algorithms, with application to room acoustic signal enhancement. The traditional Tikhonov regularized and Levenberg–Marquardt regularized RLS algorithms, as well as the regularized APA and NLMS algorithms

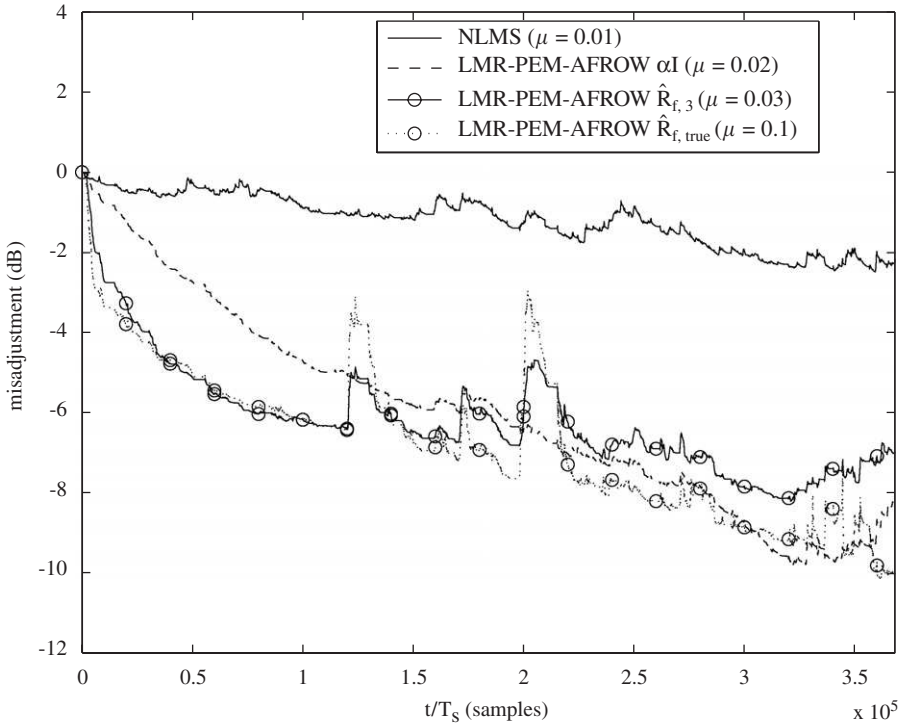


Fig. 7. Convergence curves of regularized PEM-AFROW algorithms for an AFC application in a stationary environment.

fit into this framework as special cases, but are found to be suboptimal in many scenarios. An MSE optimal approach towards regularization has been applied to RIR identification, in which prior knowledge on the unknown RIR and on the source signal may be incorporated in a more general regularization matrix, and in an appropriate pre-filtering and weighting procedure. This approach results in a new family of optimally regularized adaptive filtering algorithms, which are not only linked to the existing regularized algorithms, but also to other adaptive filtering concepts such as leakage and proportionate adaptation. From simulation results, it is clear that exploiting prior knowledge may considerably increase the adaptive filter convergence speed. In addition, using a method for constructing a regularization matrix based on a 3-parameter RIR model, a significant increase in robustness with respect to RIR changes can be achieved.

Acknowledgements

Toon van Waterschoot is a Research Assistant with the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of K.U. Leuven Research Council: CoE EF/05/006 Optimization in Engineering, the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P6/04 (''Dynamical systems, control and optimization'', 2007–2011), the Concerted Research Action GOA-AMBioRICS, and IWT project 040803: ''SMS4PA-II: Sound Management System for Public Address Systems''. The scientific responsibility is assumed by its authors.

References

[1] J. Benesty, T. Gänsler, D.R. Morgan, M.M. Sondhi, S.L. Gay, Advances in Network and Acoustic Echo Cancellation, Springer, Berlin, 2001.

[2] G. Rombouts, T. van Waterschoot, K. Struyve, M. Moonen, Acoustic feedback suppression for long acoustic paths using a nonstationary source model, IEEE Trans. Signal Process. 54 (9) (2006) 3426–3434.

[3] S.J. Elliot, P.A. Nelson, Active noise control, IEEE Signal Process. Mag. 10 (4) (1993) 12–35.

[4] R.D. Poltmann, Stochastic gradient algorithm for system identification using adaptive FIR-filters with too low number of coefficients, IEEE Trans. Circuits Syst. 35 (2) (1988) 247–250.

[5] G. Rombouts, T. van Waterschoot, K. Struyve, P. Verhoeve, M. Moonen, Identification of undermodelled room impulse responses, in: Proceedings of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC '05), Eindhoven, The Netherlands, 2005, pp. 153–156.

[6] L. Ljung, System Identification: Theory for the User, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[7] A. Tikhonov, V. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.

[8] A. Neumaier, Solving ill-conditioned and singular linear systems: a tutorial on regularization, SIAM Rev. 40 (1998) 636–666.

[9] S. Haykin, Adaptive Filter Theory, Prentice-Hall, Englewood Cliffs, NJ, 1996.

[10] E. Horita, K. Sumiya, H. Urakami, S. Mitsuishi, A leaky RLS algorithm: its optimality and implementation, IEEE Trans. Signal Process. 52 (10) (2004) 2924–2932.

[11] K. Mayyas, T. Aboulnasr, Leaky LMS algorithm: MSE analysis for Gaussian data, IEEE Trans. Signal Process. 45 (4) (1997) 927–934.

[12] L. Ljung, T. Söderström, Theory and Practice of Recursive Identification, MIT Press, Cambridge, MA, 1986.

[13] K. Levenberg, A method for the solution of certain nonlinear problems in least squares, Quart. Appl. Math. 2 (1944) 164–168.

[14] D.W. Marquardt, An algorithm for least-squares estimation of non-linear parameters, SIAM J. Appl. Math. 11 (1963) 431–441.

[15] S.L. Gay, Fast projection algorithms with application to voice echo cancellation, Ph.D. Thesis, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, USA, October 1994.

[16] B. Baykal, A.G. Constantinides, Underdetermined-order recursive least-squares adaptive filtering: the concept and algorithms, IEEE Trans. Signal Process. 45 (2) (1997) 346–362.

[17] M. Bodson, An adaptive algorithm with information-dependent data forgetting, in: Proceedings of the 1995 American Control Conference (ACC '95), vol. 5, Seattle, Washington, USA, 1995, pp. 3485–3489.

[18] S.L. Gay, Dynamically regularized fast RLS with application to echo cancellation, in: Proceedings of the 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '96), vol. 2, Atlanta, Georgia, USA, 1996, pp. 957–960.

[19] D.R. Morgan, S.G. Kratzer, On a class of computationally efficient, rapidly converging, generalized NLMS algorithms, IEEE Signal Process. Lett. 3 (8) (1996) 245–247.

[20] T. Gänsler, S.L. Gay, M.M. Sondhi, J. Benesty, Double-talk robust fast converging algorithms for network echo cancellation, in: Proceedings of the 1999 IEEE Workshop on Applications on Signal Processing to Audio and Acoustics (WASPAA '99), New Paltz, NY, USA, 1999, pp. 1–4.

[21] T. Gänsler, J. Benesty, S.L. Gay, M.M. Sondhi, A robust proportionate affine projection algorithm for network echo cancellation, in: Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00), vol. 2, Istanbul, Turkey, 2000, pp. 793–796.

[22] T.Y. Al-Naffouri, A.H. Sayed, An adaptive filter robust to data uncertainties, in: Proceedings of the Allerton Conference on Communication, Control and Computing, vol. 2, Allerton, IL, USA, 2000, pp. 1175–1183.

[23] Z. Chen, S. Haykin, S.L. Gay, Proportionate adaptation: new paradigms in adaptive filtering, in: S. Haykin, B. Widrow (Eds.), Advances in LMS filters, Wiley, New York, 2003 (Chapter 8).
[24] S.L. Gay, S. Tavathia, The fast affine projection algorithm, in: Proceedings of the 1995 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '95), vol. 5, Detroit, Michigan, USA, 1995, pp. 3023–3026.
[25] V. Myllylä, G. Schmidt, Pseudo-optimal regularization for affine projection algorithms, in: Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02), vol. 2, Orlando, Florida, USA, 2002, pp. 1917–1920.
[26] C. Kasparis, R.J. Piechocki, P.N. Fletcher, A.R. Nix, A bootstrap multi-user detector for CDMA based on Tikhonov regularization, in: Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03), vol. 4, 2003, pp. 69–72.
[27] Z.-F. Wang, X.-W. Yu, The minimum mean square error linear estimator and ridge regression, in: Proceedings of the 2002 International Conference on Machine Learning and Cybernetics (ICMLC '02), Beijing, China, 2002, pp. 199–201.
[28] S. Jo, S.W. Kim, Regularized affine projection algorithm using constrained least squares of residuals, IEEE Trans. Signal Process, 2004, submitted for publication.
[29] A.H. Sayed, A. Garulli, S. Chandrasekaran, A fast iterative solution for worst-case parameter estimation with bounded model uncertainties, in: Proceedings of the 1997 American Control Conference (ACC '97), vol. 3, Albuquerque, New Mexico, USA, 1997, pp. 1499–1503.
[30] D.P. Mandic, A generalized normalized gradient descent algorithm, IEEE Signal Process. Lett. 11 (2) (2004) 115–118.
[31] Y.-S. Choi, H.-C. Shin, W.-J. Song, Robust regularization for normalized LMS algorithms, IEEE Trans. Circuits Syst. II Exp. Briefs 53 (8) (2006) 627–631.
[32] E. Chau, H. Sheikhzadeh, R.L. Brennan, Complexity reduction and regularization of a fast affine projection algorithm for oversampled subband adaptive filters, in: Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04), vol. 5, Montreal, QC, Canada, 2004, pp. 109–112.
[33] A. Cichocki, R. Unbehauen, Neural Networks for Optimization and Signal Processing, Wiley, New York, 1993.
[34] J.M. Varah, A practical examination of some numerical methods for linear discrete ill-posed problems, SIAM Rev. 21 (1) (1979) 100–111.
[35] P.C. Hansen, Truncated singular value decomposition solutions to discrete ill-posed problems with ill-determined numerical rank, SIAM J. Sci. Stat. Comput. 11 (3) (1990) 503–518.
[36] D. MacKay, Bayesian interpolation, Neural Comput. 4 (3) (1992) 415–447.
[37] D. Foresee, M. Hagan, Gauss–Newton approximation to Bayesian learning, in: Proceedings of the 1997 International Joint Conference on Neural Networks (IJCNN '97), Houston, Texas, USA, 1997, pp. 1930–1935.
[38] J.P. Kaipio, V. Kolehmainen, M. Vauhkonen, E. Somersalo, Inverse problems with structural prior information, Inverse Probl. 15 (3) (1999) 713–729.
[39] Y. Lin, D.D. Lee, Bayesian regularization and nonnegative deconvolution for room impulse response estimation, IEEE Trans. Signal Process. 54 (3) (2006) 839–847.
[40] J.B. Allen, D.A. Berkley, Image method for efficiently simulating small-room acoustics, J. Acoust. Soc. Am. 65 (4) (1979) 943–950.
[41] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, Upper Saddle River, NJ, 1993.
[42] T. van Waterschoot, M. Moonen, Double-talk robust acoustic echo cancellation with continuous near-end activity, in: Proceedings of the 13th European Signal Processing Conference (EUSIPCO '05), Antalya, Turkey, 2005.
[43] T. van Waterschoot, G. Rombouts, P. Verhoeve, M. Moonen, Double-talk-robust prediction error identification algorithms for acoustic echo cancellation, IEEE Trans. Signal Process. 55 (3) (2007) 846–858.
[44] G. Rombouts, T. van Waterschoot, K. Struyve, M. Moonen, Acoustic feedback suppression for long acoustic paths using a nonstationary source model, in: Proceedings of the 13th European Signal Processing Conference (EUSIPCO '05), Antalya, Turkey, 2005.
[45] T. van Waterschoot, G. Rombouts, M. Moonen, Towards optimal regularization by incorporating prior knowledge in an acoustic echo canceller, in: Proceedings of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC '05), Eindhoven, The Netherlands, 2005, pp. 157–160.
[46] A. Houacine, Regularized fast recursive least squares algorithms, in: Proceedings of the 1990 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '90), Albuquerque, New Mexico, USA, 1990, pp. 1587–1590.
[47] D.L. Duttweiler, Proportionate normalized least-mean-squares adaptation in echo cancelers, IEEE Trans. Speech Audio Process. 9 (5) (2000) 508–518.
[48] D.G. Luenberger, Optimization by Vector Space Methods, Wiley, New York, 1969.
[49] P.C. Hansen, Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion, SIAM, Philadelphia, PA, USA, 1998.
[50] W.C. Sabine, Collected Papers on Acoustics, Peninsula, Los Altos, California, 1992.
[51] J.R. Hopgood, Acoustic impulse response database, ©CEDAR Audio, [Online] Available from: ⟨http://www-sigproc.eng.cam.ac.uk/oldhomes/jrh1008/public_html/Resources/ImpulseResponses/index.html⟩.