TOWARDS OPTIMAL REGULARIZATION BY INCORPORATING PRIOR KNOWLEDGE IN AN ACOUSTIC ECHO CANCELLER

[1,2]*Toon van Waterschoot,* [2]*Geert Rombouts and* [2]*Marc Moonen*

[1]toon.vanwaterschoot@esat.kuleuven.be
[2]Katholieke Universiteit Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

ABSTRACT

The problem of poor excitation is often encountered in acoustic echo cancellation, due to the high coloration of audio signals and the large dimension of the room impulse response parameter vector. Poor excitation leads to a large variance of the impulse response estimate, resulting in a slowly converging adaptive algorithm. The standard solution is to add a scaled identity matrix to the ill-conditioned input correlation matrix, where scaling is performed with an estimate of the near-end background noise power. We illustrate how this type of regularization fits in a linear minimum mean square error framework and how regularization may be improved by incorporating prior knowledge on the room impulse response. Prior knowledge can be constructed based on some physical parameters of the acoustic setup. Off-line simulation results indicate that the proposed regularization technique may yield a low-variance room impulse response estimate.

1. INTRODUCTION

Acoustic echo cancellation (AEC) has been a popular research topic in acoustic signal processing, motivated mainly by the increasing demand for hands-free speech communication. A classical AEC scenario is shown in **Figure 1**. A speech signal $u(t)$ from the far-end side is broadcasted in an acoustic enclosure (the 'room') by means of a loudspeaker. A microphone is present in the room for recording a local signal $v(t)$ (the 'near-end signal') which is to be transmitted back to the far-end side. An acoustic echo path exists between the loudspeaker and the microphone such that the recorded microphone signal $y(t) = x(t) + v(t)$ contains an undesired echo component $x(t)$ in addition to the near-end signal component $v(t)$. If the echo path transfer function is modelled as a finite impulse response (FIR) filter $F(q,t) \triangleq f_0(t) + f_1(t)q^{-1} + \ldots + f_{n_F}(t)q^{-n_F}$, then the echo component can be considered as a filtered version of the loudspeaker signal: $x(t) = F(q,t)u(t)$. Here q denotes the time shift operator, e.g. $q^{-k}u(t) = u(t-k)$. The main objective in AEC is to identify the unknown room impulse response (RIR) $F(q,t)$ and hence to subtract an estimate of the echo component from the microphone signal. In this way an echo-compensated signal $d(t) = y(t) - \hat{F}(q,t)u(t)$ is sent to the far-end side, with $\hat{F}(q,t)$ an estimate of $F(q,t)$.

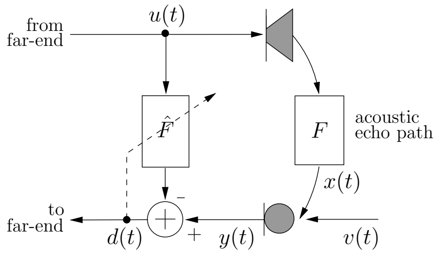

Figure 1: A typical acoustic echo cancellation (AEC) scenario.

It is well-known that audio signals exhibit a high degree of tonality which may result in an ill-conditioned correlation matrix in least squares parameter estimation. The standard solution to this problem is to add a scaled identity matrix to the input correlation matrix and hence reduce its eigenvalue spread. This technique is known as Tikhonov [1] regularization. Using this modified correlation matrix can be interpreted as solving a regularized least squares problem (also known as ridge regression) in which the squared norm of the unknown parameter vector (containing the RIR coefficients) is added to the least squares cost function. With exponential weighting this leads to the leaky RLS algorithm [2], which is the Gauss-Newton counterpart to the more well-known leaky LMS algorithm [3]. Instead, we propose to regularize the cost function by adding the weighted squared norm of the unknown parameter vector. The optimal weighting matrix in a mean square error (MSE) sense is given by linear estimation theory [4]. If the unknown parameter vector is regarded as a realization of a stochastic vector process, the optimal weighting matrix for regularization is given by the covariance matrix of the vector process. This covariance matrix can be either constructed from prior knowledge on the RIR, or estimated concurrently with the RIR.

In **Section 2** we will introduce the concept of regularization in least squares parameter estimation. The linear minimum mean square error estimator is derived in **Section 3** and its relationship to the regularized least squares estimator is pointed out. **Section 4** describes a novel method to gather prior knowledge on a room impulse response for constructing a diagonal non-identity regularization matrix. In **Section 5** we show some off-line simulation results and finally **Section 6** concludes the paper.

[1]Toon van Waterschoot is a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'), 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.

2. REGULARIZATION IN LEAST SQUARES PARAMETER ESTIMATION

The off-line identification of the RIR $F(q,t)$ can be considered as a linear estimation problem with the RIR coefficients collected in the parameter vector $\mathbf{f}$:

$$\begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} u(1) & \dots & u(1-n_F) \\ \vdots & \ddots & \vdots \\ u(N) & \dots & u(N-n_F) \end{bmatrix} \cdot \begin{bmatrix} f_0 \\ \vdots \\ f_{n_F} \end{bmatrix} + \begin{bmatrix} v(1) \\ \vdots \\ v(N) \end{bmatrix},$$

or in matrix notation

$$\mathbf{y} = \mathbf{U}\mathbf{f} + \mathbf{v}. \tag{1}$$

An estimator for $\mathbf{f}$ may be obtained by minimizing the least squares (LS) criterion

$$\min_{\hat{\mathbf{f}}} V_{LS}(\hat{\mathbf{f}}) = \min_{\hat{\mathbf{f}}} (\mathbf{y} - \mathbf{U}\hat{\mathbf{f}})^T (\mathbf{y} - \mathbf{U}\hat{\mathbf{f}}) \tag{2}$$

which results in the well-known LS estimator

$$\boxed{\hat{\mathbf{f}}_{\mathbf{LS}} = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y}.} \tag{3}$$

When the input signal $u(t)$ is not (or hardly) persistently exciting, as is often the case when using audio signals, the matrix $\mathbf{U}^T\mathbf{U}$ may be ill-conditioned or even singular. A common solution is to apply Tikhonov regularization [1] by adding a scaled identity matrix to $\mathbf{U}^T\mathbf{U}$:

$$\boxed{\hat{\mathbf{f}}_{\mathbf{RgLS}} = (\mathbf{U}^T\mathbf{U} + \alpha\mathbf{I})^{-1}\mathbf{U}^T\mathbf{y}.} \tag{4}$$

In [5] it was shown that choosing the regularization parameter α in the neighbourhood of the near-end background noise power σ_v^2 has a beneficial influence when $\mathbf{f}$ is identified recursively. We will show below that:

- the estimator in (4) with $\alpha = \sigma_v^2$ is optimal in a minimum mean square error (MMSE) sense if the near-end background noise is drawn from a zero-mean stationary white noise process with variance σ_v^2 and if the unknown parameter vector $\mathbf{f}$ can be considered as a realization of a zero-mean, unit-variance vector white noise process,
- it may be convenient to replace the scaled unity matrix $\alpha\mathbf{I}$ in (4) by a non-identity regularization matrix $\mathbf{P}$:

$$\hat{\mathbf{f}}_{\mathbf{RgLS}} = (\mathbf{U}^T\mathbf{U} + \mathbf{P})^{-1}\mathbf{U}^T\mathbf{y}. \tag{5}$$

3. LINEAR MIMIMUM MEAN SQUARE ERROR ESTIMATION

Let us derive an expression for the minimum mean square error (MMSE) estimator of $\mathbf{f}$:

$$\min_{\hat{\mathbf{f}}} V_{MMSE}(\hat{\mathbf{f}}) = \min_{\hat{\mathbf{f}}} E(\hat{\mathbf{f}} - \mathbf{f})^T(\hat{\mathbf{f}} - \mathbf{f}) \tag{6}$$

under the following assumptions:

- The estimator is a linear function of the data in $\mathbf{y}$:

$$\hat{\mathbf{f}}_{\mathbf{MMSE}} = \mathbf{Z}^T\mathbf{y}. \tag{7}$$

- The measurement noise in $\mathbf{v}$ is drawn from a stationary white noise process with zero mean and variance σ_v^2:

$$\boldsymbol{\mu}_{\mathbf{v}} \triangleq E\mathbf{v} = \mathbf{0}, \tag{8}$$

$$\mathbf{R}_{\mathbf{v}} \triangleq \mathrm{cov}(\mathbf{v}) = E\mathbf{v}\mathbf{v}^T = \sigma_v^2\mathbf{I}. \tag{9}$$

- The true parameter vector $\mathbf{f}$ is considered as a random variable on which some prior knowledge may be available. More specifically, let the prior probability density function (PDF) $p(\mathbf{f})$ be characterized by its first and second order moments:

$$\boldsymbol{\mu}_{\mathbf{f}} \triangleq E\mathbf{f}, \tag{10}$$

$$\mathbf{R}_{\mathbf{f}} \triangleq \mathrm{cov}(\mathbf{f}) = E(\mathbf{f} - E\mathbf{f})(\mathbf{f} - E\mathbf{f})^T. \tag{11}$$

Then the linear MMSE estimator can be obtained as the mean of the posterior PDF $p(\mathbf{f}|\mathbf{y})$ after the data have been recorded [4]:

$$\begin{aligned} \hat{\mathbf{f}}_{\mathbf{MMSE}} &= E(\mathbf{f}|\mathbf{y}) \qquad (12) \\ &= \boldsymbol{\mu} + (\mathbf{U}^T\mathbf{R}_{\mathbf{v}}{}^{-1}\mathbf{U} + \mathbf{R}_{\mathbf{f}}{}^{-1})^{-1}\mathbf{U}^T\mathbf{R}_{\mathbf{v}}{}^{-1}(\mathbf{y} - \mathbf{U}\boldsymbol{\mu}). \end{aligned}$$

We will construct the prior knowledge on $\mathbf{f}$ in such a way that $\boldsymbol{\mu}_{\mathbf{f}} = \mathbf{0}$. Then, also using the white noise assumption in (9), the expression for $\hat{\mathbf{f}}_{\mathbf{MMSE}}$ can be rewritten as

$$\boxed{\hat{\mathbf{f}}_{\mathbf{MMSE}} = (\mathbf{U}^T\mathbf{U} + \sigma_v^2\mathbf{R}_{\mathbf{f}}{}^{-1})^{-1}\mathbf{U}^T\mathbf{y}.} \tag{13}$$

From this point of view, applying Tikhonov regularization as in (4) with $\alpha = \sigma_v^2$ is equivalent to assuming that the true parameter vector $\mathbf{f}$ is drawn from a stationary vector white noise process. In room acoustics applications however, more information on the true parameter may be available and an appropriate non-identity covariance matrix $\mathbf{R}_{\mathbf{f}}$ can be constructed.

4. GATHERING PRIOR KNOWLEDGE ON ROOM ACOUSTICS

If an LS estimator with Tikhonov regularization is applied in the AEC problem, the regularization matrix $\mathbf{P} = \alpha\mathbf{I}$ (with the regularization parameter chosen as $\alpha = \sigma_v^2$) looks as in **Figure 2**.

A room impulse response has a very typical form, which may be characterized by three parameters, as illustrated in **Figure 3**:

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

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 [6]. Hence they can be considered as prior knowledge. If these three parameters are taken into account, a diagonal regularization matrix may be constructed that looks as in **Figure 4**.

An idealized case, which is interesting as a reference method, occurs when the true RIR is known. In this case a diagonal regularization matrix may be constructed with the diagonal elements equal to the inverse square values of the true parameter vector coefficients, as illustrated in **Figure 5**.

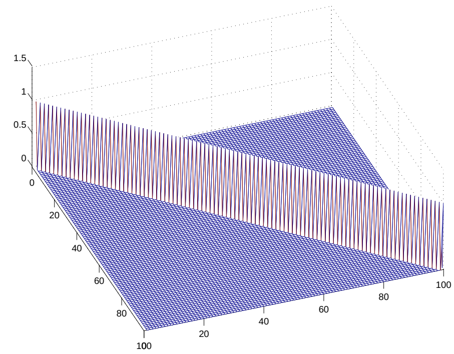

Figure 2: Visualization of traditional regularization matrix (horizontal: matrix indices, vertical: matrix element value)

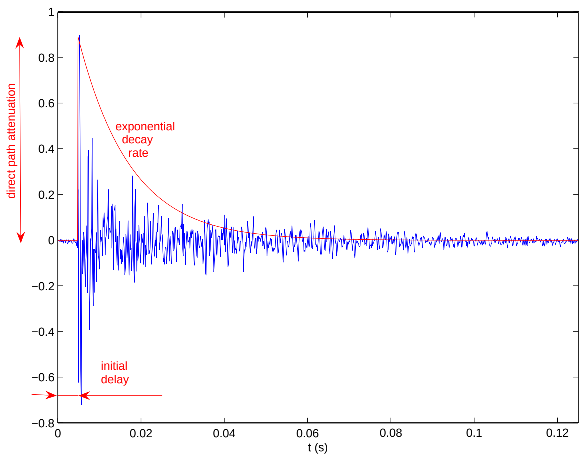

Figure 3: Room impulse response characterized by three parameters.

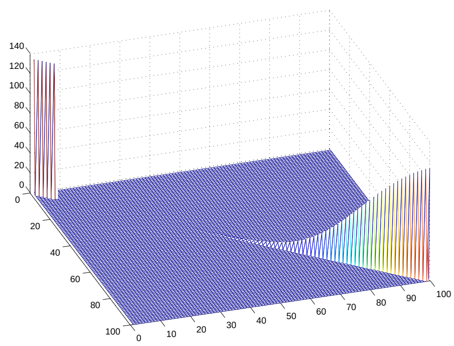

Figure 4: Visualization of regularization matrix based on 3 parameters (horizontal: matrix indices, vertical: matrix element value)

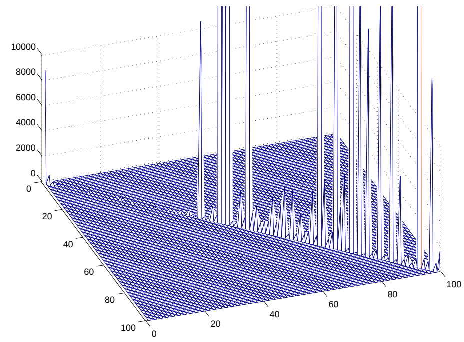

Figure 5: Visualization of regularization matrix based on true RIR (horizontal: matrix indices, vertical: matrix element value)

5. OFF-LINE SIMULATION RESULTS

In a series of preliminary simulations, four different least squares estimators were compared for the AEC application:

- the LS estimator $\hat{\mathbf{f}}_{\mathbf{LS}}$ without regularization,
- the regularized LS estimator $\hat{\mathbf{f}}_{\mathbf{RgLS}}$ with Tikhonov regularization: $\mathbf{P} = \sigma_v^2 \mathbf{I}$,
- the regularized LS estimator $\hat{\mathbf{f}}_{\mathbf{RgLS}}$ with regularization based on the three RIR parameters described above: $\mathbf{P} = \sigma_v^2 \mathbf{R}_{\mathbf{f},synth}^{-1}$, and
- the regularized LS estimator $\hat{\mathbf{f}}_{\mathbf{RgLS}}$ with regularization based on the true RIR: $\mathbf{P} = \sigma_v^2 \mathbf{R}_{\mathbf{f},true}^{-1}$.

For each estimator, 100 simulation runs were performed with a different near-end noise realization (drawn from a stationary white noise process). The normalized distance between the resulting estimates and the true RIR, $\frac{\|\hat{\mathbf{f}}-\mathbf{f}\|}{\|\mathbf{f}\|}$, is plotted in a boxplot to compare the bias and variance of the different estimators (see **Figures 6**, **7**, **8** and **9**). The data record length was 10 times the parameter vector length. The average echo to near-end noise ratio was set to $10dB$. The sampling rate was equal to $f_s = 8kHz$. Two types of loudspeaker signals were used:

- a sum of $n_F - 1$ sinusoids with random frequencies, uniformly distributed in the interval between DC and the Nyquist frequency, and
- a male speech signal.

Two types of acoustic impulse responses were measured in a realistic situation and used in the simulation:

- a hearing aid impulse response with $n_F + 1 = 100$ coefficients, and
- a room impulse response with $n_F + 1 = 1000$ coefficients.

6. CONCLUSIONS AND FUTURE WORK

We have proposed to tackle the poor excitation problem occurring in acoustic echo cancellation by adding a non-identity regularization matrix to the input correlation matrix. In this way a

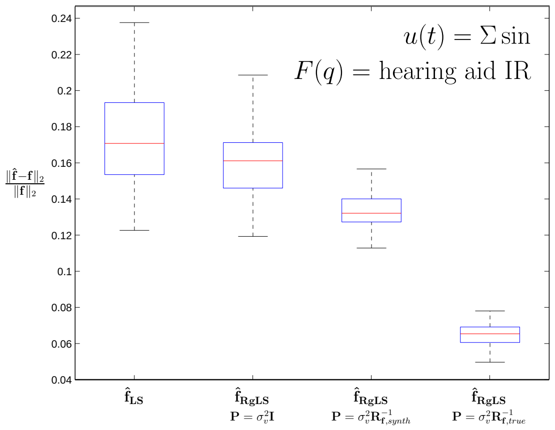

Figure 6: Boxplots for hearing aid impulse response and sum of sinusoids input signal.

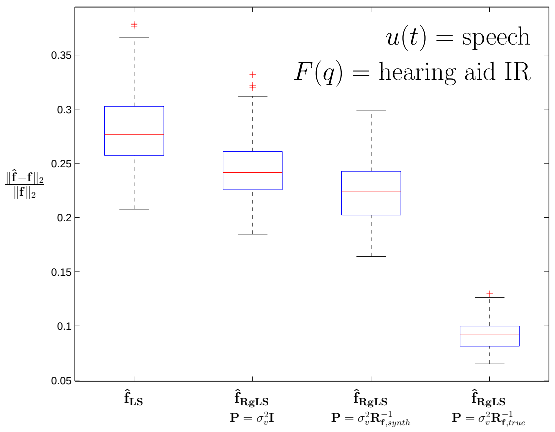

Figure 7: Boxplots for hearing aid impulse response and speech input signal.

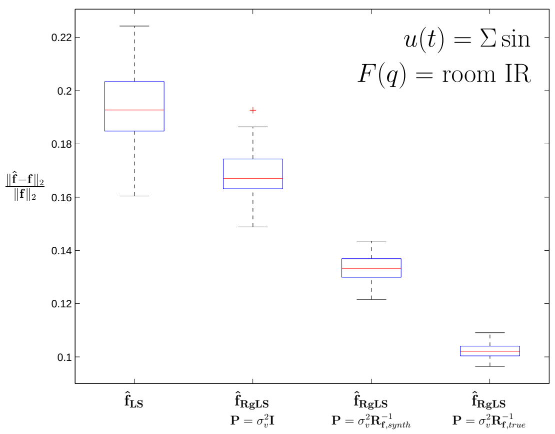

Figure 8: Boxplots for room impulse response and sum of sinusoids input signal.

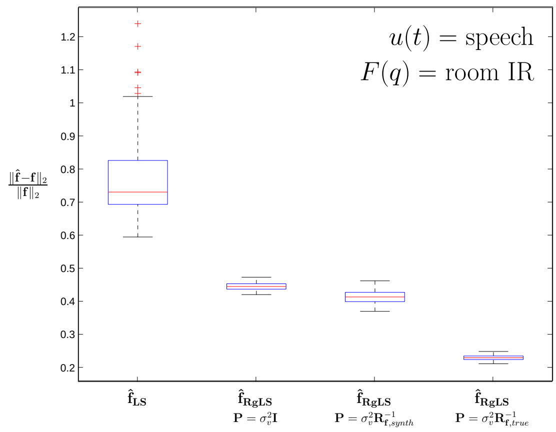

Figure 9: Boxplots for room impulse response and speech input signal.

low-variance estimate of the room impulse response can be obtained which approaches the linear minimum mean square error estimate, depending on the quality of the prior knowledge that is available. We have proposed a three-parameter model to gather prior knowledge on a room impulse response and to construct a diagonal non-identity regularization matrix. Off-line simulations show that for different types of loudspeaker signals and for different acoustic scenarios, the proposed regularization method yields a lower variance and a lower excess error than the unregularized and Tikhonov-like regularized estimators. Future work will focus on developing recursive identification algorithms that incorporate the proposed regularization technique, while avoiding a dramatic increase in computational complexity.

7. REFERENCES

[1] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, 1977.

[2] E. Horita, K. Sumiya, H. Urakami, and S. Mitsuishi, "A leaky RLS algorithm: Its optimality and implementation," *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2924–2932, Oct. 2004.

[3] K. Mayyas and T. Aboulnasr, "Leaky LMS algorithm: MSE analysis for gaussian data," *IEEE Trans. Signal Processing*, vol. 45, no. 4, pp. 927–934, Apr. 1997.

[4] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA, 1993.

[5] 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, Oct. 2004.

[6] W. C. Sabine, *Collected Papers on Acoustics*, Peninsula, Los Altos, California, USA, 1992.