A RECURSIVE ESTIMATION OF THE CONDITION NUMBER IN THE RLS ALGORITHM

Jacob Benesty[1] *and Tomas Gänsler*[2]

[1]Université du Québec, INRS-EMT
800 de la Gauchetière Ouest, Suite 6900
Montréal, Québec, H5A 1K6, Canada
[2]Agere Systems
1110 American Parkway
Allentown, Pennsylvania, 18109, USA
e-mail: benesty@inrs-emt.uquebec.ca, gaensler@agere.com

ABSTRACT

The recursive least-squares (RLS) algorithm is one of the most popular adaptive algorithms in the literature. This is due to the fact that it is easily derived and exactly solves the normal equations. In this paper, we present a very efficient way to recursively estimate the condition number of the input signal covariance matrix by utilizing fast versions of the RLS algorithm. We also quantify the misalignment of the RLS algorithm with respect to the condition number.

1. INTRODUCTION

Adaptive algorithms play a very important role in many diverse applications such as communications, acoustics, speech, radar, sonar, seismology, and biomedical engineering [1], [2], [3], [4]. Among the most well-known adaptive filters are the recursive least-squares (RLS) and fast RLS (FRLS) algorithms. The latter is a computationally less complex version of the former. Even though the RLS is not as widely used in practice as the least-mean-square (LMS) algorithm, it has a very significant theoretical interest since it belongs to the Kalman filters family [5].

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

The main objective of this paper is to derive a very simple way to recursively estimate the condition number. The proposed method is very efficient when combined with the FRLS algorithm; it requires only L more multiplications per iteration, where L is the length of the adaptive filter. We also show how the misalignment of the RLS algorithm is affected by the condition number, output SNR, and parameter choice.

2. THE RECURSIVE LEAST-SQUARES (RLS) ALGORITHM

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

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

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

where

$$y(n) = \mathbf{h}_\mathrm{t}^T \mathbf{x}(n) + w(n) \tag{2}$$

is the system output,

$$\mathbf{h}_\mathrm{t} = \begin{bmatrix} h_{\mathrm{t},0} & h_{\mathrm{t},1} & \cdots & h_{\mathrm{t},L-1} \end{bmatrix}^T$$

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

$$\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-L+1) \end{bmatrix}^T$$

is a vector containing the last L samples of the input signal x, w is a white Gaussian noise (uncorrelated with x) with variance σ_w^2,

$$\hat{y}(n) = \mathbf{h}^T(n-1)\mathbf{x}(n) \tag{3}$$

is the model filter output, and

$$\mathbf{h}(n-1) = \begin{bmatrix} h_0(n-1) & \cdots & h_{L-1}(n-1) \end{bmatrix}^T$$

is the model filter of length L.

We also define the popular recursive least-squares error criterion with respect to the modelling filter:

$$J_{\mathrm{LS}}(n) = \sum_{m=0}^{n} \lambda^{n-m} \left[y(m) - \mathbf{h}^T(n)\mathbf{x}(m) \right]^2, \tag{4}$$

where λ ($0 < \lambda < 1$) is a forgetting factor. The minimization of (4) leads to the normal equations:

$$\mathbf{R}(n)\mathbf{h}(n) = \mathbf{r}(n), \tag{5}$$

where

$$\mathbf{R}(n) = \sum_{m=0}^{n} \lambda^{n-m} \mathbf{x}(m)\mathbf{x}^T(m) \quad (6)$$

is an estimate of the input signal covariance matrix and

$$\mathbf{r}(n) = \sum_{m=0}^{n} \lambda^{n-m} \mathbf{x}(m) y(m) \quad (7)$$

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

From the normal equations (5), we easily derived the classical update for the RLS algorithm [1], [3]:

$$e(n) = y(n) - \mathbf{h}^T(n-1)\mathbf{x}(n), \quad (8)$$
$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mathbf{R}^{-1}(n)\mathbf{x}(n)e(n). \quad (9)$$

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

$$\mathbf{k}(n) = \lambda^{-1}\varphi(n)\mathbf{k}'(n), \quad (10)$$

where

$$\varphi(n) = \frac{\lambda}{\lambda + \mathbf{x}^T(n)\mathbf{R}^{-1}(n-1)\mathbf{x}(n)}. \quad (11)$$

3. CONDITION NUMBER OF THE INPUT SIGNAL COVARIANCE MATRIX

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

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

$$\mathbf{Q}^T(n)\mathbf{R}(n)\mathbf{Q}(n) = \Lambda(n), \quad (12)$$

where

$$\mathbf{Q}^T(n)\mathbf{Q}(n) = \mathbf{Q}(n)\mathbf{Q}^T(n) = \mathbf{I}, \quad (13)$$
$$\Lambda(n) = \operatorname{diag}\left[\lambda_0(n), \lambda_1(n), \cdots, \lambda_{L-1}(n)\right], \quad (14)$$

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

$$\mathbf{R}^{1/2}(n) = \mathbf{Q}(n)\Lambda^{1/2}(n)\mathbf{Q}^T(n). \quad (15)$$

The condition number of a matrix $\mathbf{R}(n)$ is [6]:

$$\chi\left[\mathbf{R}(n)\right] = \|\mathbf{R}(n)\|\|\mathbf{R}^{-1}(n)\|, \quad (16)$$

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

Consider the following norm:

$$\|\mathbf{R}(n)\|_{\mathrm{E}} = \left\{ \frac{1}{L} \operatorname{tr}\left[\mathbf{R}^T(n)\mathbf{R}(n)\right] \right\}^{1/2}. \quad (17)$$

We can easily check that, indeed, $\|\cdot\|_{\mathrm{E}}$ is a matrix norm. Also, the E-norm of the identity matrix is equal to one.

We have:

$$\|\mathbf{R}^{1/2}(n)\|_{\mathrm{E}} = \left\{ \frac{1}{L} \operatorname{tr}\left[\mathbf{R}(n)\right] \right\}^{1/2} = \left\{ \frac{1}{L} \sum_{l=0}^{L-1} \lambda_l(n) \right\}^{1/2} \quad (18)$$

and

$$\|\mathbf{R}^{-1/2}(n)\|_{\mathrm{E}} = \left\{ \frac{1}{L} \operatorname{tr}\left[\mathbf{R}^{-1}(n)\right] \right\}^{1/2} = \left\{ \frac{1}{L} \sum_{l=0}^{L-1} \frac{1}{\lambda_l(n)} \right\}^{1/2}. \quad (19)$$

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

$$\chi_{\mathrm{E}}\left[\mathbf{R}^{1/2}(n)\right] = \|\mathbf{R}^{1/2}(n)\|_{\mathrm{E}}\|\mathbf{R}^{-1/2}(n)\|_{\mathrm{E}} \geq 1. \quad (20)$$

If $\chi\left[\mathbf{R}(n)\right]$ is large, then $\mathbf{R}(n)$ is said to be an ill-conditioned matrix. Note that this is a norm-dependent property. However, according to [6], any two condition numbers $\chi_\alpha\left[\mathbf{R}(n)\right]$ and $\chi_\beta\left[\mathbf{R}(n)\right]$ are equivalent in that constants c_1 and c_2 can be found for which:

$$c_1\chi_\alpha\left[\mathbf{R}(n)\right] \leq \chi_\beta\left[\mathbf{R}(n)\right] \leq c_2\chi_\alpha\left[\mathbf{R}(n)\right]. \quad (21)$$

For example, for the 1- and 2-norm matrices, we can show [6]:

$$\frac{1}{L^2}\chi_2\left[\mathbf{R}(n)\right] \leq \frac{1}{L}\chi_1\left[\mathbf{R}(n)\right] \leq \chi_2\left[\mathbf{R}(n)\right]. \quad (22)$$

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

$$\chi_2\left[\mathbf{R}(n)\right] = \frac{\lambda_{L-1}(n)}{\lambda_0(n)}. \quad (23)$$

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

$$\operatorname{tr}\left[\mathbf{R}(n)\right]\operatorname{tr}\left[\mathbf{R}^{-1}(n)\right] \geq \frac{\operatorname{tr}\left[\mathbf{R}(n)\right]}{\lambda_0(n)} \geq \frac{\lambda_{L-1}(n)}{\lambda_0(n)}. \quad (24)$$

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

$$\operatorname{tr}\left[\mathbf{R}(n)\right]\operatorname{tr}\left[\mathbf{R}^{-1}(n)\right] \leq L\frac{\operatorname{tr}\left[\mathbf{R}(n)\right]}{\lambda_0(n)} \leq L^2\frac{\lambda_{L-1}(n)}{\lambda_0(n)}. \quad (25)$$

Therefore, we deduce that:

$$\frac{1}{L^2}\chi_2\left[\mathbf{R}(n)\right] \leq \chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right] \leq \chi_2\left[\mathbf{R}(n)\right]. \quad (26)$$

According to the previous expression, $\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right]$ is then a measure of the condition number of the matrix $\mathbf{R}(n)$. In the next section, we will show how to recursively compute $\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right]$.

4. RECURSIVE COMPUTATION OF THE CONDITION NUMBER

The positive number $\|\mathbf{R}^{1/2}(n)\|_{\mathrm{E}}^2$ can be easily calculated recursively. Indeed, taking the trace of (6), we get:

$$\mathrm{tr}\left[\mathbf{R}(n)\right] = \lambda \mathrm{tr}\left[\mathbf{R}(n-1)\right] + \mathbf{x}^T(n)\mathbf{x}(n). \quad (27)$$

Therefore

$$\|\mathbf{R}^{1/2}(n)\|_{\mathrm{E}}^2 = \lambda \|\mathbf{R}^{1/2}(n-1)\|_{\mathrm{E}}^2 + \frac{\mathbf{x}^T(n)\mathbf{x}(n)}{L}. \quad (28)$$

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

Now we need to determine $\|\mathbf{R}^{-1/2}(n)\|_{\mathrm{E}}^2$. The inverse of $\mathbf{R}(n)$ is:

$$\mathbf{R}^{-1}(n) = \lambda^{-1}\mathbf{R}^{-1}(n-1) - \lambda^{-2}\varphi(n)\mathbf{k}'(n)\mathbf{k}'^T(n). \quad (29)$$

We deduce that:

$$\|\mathbf{R}^{-1/2}(n)\|_{\mathrm{E}}^2 = \quad (30)$$
$$\lambda^{-1}\left[\|\mathbf{R}^{-1/2}(n-1)\|_{\mathrm{E}}^2 - \frac{\lambda^{-1}\varphi(n)\mathbf{k}'^T(n)\mathbf{k}'(n)}{L}\right].$$

By using (28) and (30), we see that we easily compute $\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right]$ recursively with only an order of L multiplications per iteration given that $\mathbf{k}'(n)$ is known. It is easy to combine the estimation of the condition number with a fast RLS (FRLS) algorithm. There exist several methods to compute the *a priori* Kalman gain vector $\mathbf{k}'(n)$ in a very efficient way [1]. Once this gain vector is determined, the estimation of $\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right]$ at each iteration follows immediately.

5. MISALIGNMENT AND CONDITION NUMBER

We define the normalized misalignment in dB as:

$$m_0(n) = 10\log_{10} E\left[\frac{\|\mathbf{h}_{\mathrm{t}} - \mathbf{h}(n)\|_2^2}{\|\mathbf{h}_{\mathrm{t}}\|_2^2}\right], \quad (31)$$

where $\|\cdot\|_2$ denotes the two-norm vector. Equation (31) measures the mismatch between the true impulse response and the modelling filter.

It can easily be shown, under certain conditions, that [7]:

$$E\left[\|\mathbf{h}_{\mathrm{t}} - \mathbf{h}(n)\|_2^2\right] \approx \frac{1}{2}\sigma_w^2 \mathrm{tr}\left[\mathbf{R}^{-1}(n)\right]. \quad (32)$$

We can also show, by using our definition of the condition number, that:

$$E\left[\frac{\|\mathbf{h}_{\mathrm{t}} - \mathbf{h}(n)\|_2^2}{\|\mathbf{h}_{\mathrm{t}}\|_2^2}\right] \approx \frac{(1-\lambda)L}{2}\frac{\sigma_w^2}{\|\mathbf{h}_{\mathrm{t}}\|_2^2\sigma_x^2}\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right],$$

where σ_x^2 is the power of the input signal x. Finally, we have a formula for the normalized misalignment in dB (which is valid only after convergence of the RLS algorithm):

$$m_0(n) \approx 10\log_{10}\frac{(1-\lambda)L}{2} + 10\log_{10}\frac{\sigma_w^2}{\|\mathbf{h}_{\mathrm{t}}\|_2^2\sigma_x^2}$$
$$+ 10\log_{10}\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right]. \quad (33)$$

Expression (33) depends on three terms or three factors: the exponential window, the level of noise at the system output, and the condition number. The closer the exponential window is to one, the better is the misalignment but the tracking abilities of the RLS algorithm will suffer a lot. A high level of noise as well as an input signal with a large condition number will obviously degrade the misalignment. With a fixed exponential window and noise, it is interesting to see how the misalignment will degrade by increasing the condition number of the input signal. For example, by increasing the condition number from 1 to 10, the misalignment will degrade by 10 dB. Simulations confirm that.

Usually, we take for the exponential window $\lambda = 1 - \frac{1}{K_0 L}$, where $K_0 \geq 3$. Also, the second term in (33) represents roughly the inverse output signal-to-noise ratio (SNR) in dB. We can then rewrite (33) as follows:

$$m_0(n) \approx -10\log_{10}(2K_0) - \mathrm{oSNR}$$
$$+ 10\log_{10}\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right]. \quad (34)$$

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

$$m_0(n) \approx -49 + 10\log_{10}\chi_{\mathrm{E}}^2\left[\mathbf{R}^{1/2}(n)\right]. \quad (35)$$

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

6. SIMULATIONS

In this section, we present some results on the condition number estimation and how this number affects the misalignment in a system identification context. We try to estimate an impulse response $\mathbf{h}_{\mathrm{t}}$ of length $L = 512$. The same length is used for the adaptive filter $\mathbf{h}(n)$. We run the FRLS algorithm with a forgetting factor $\lambda = 1 - 1/(5L)$. Performance of the estimation is measured by means of the normalized misalignment [eq. (31)]. The input signal $x(n)$ is a speech signal sampled at 8 kHz. The output signal $y(n)$ is obtained by convolving $\mathbf{h}_{\mathrm{t}}$ with $x(n)$ and adding a white Gaussian noise signal with an SNR of 39 dB. In order to evaluate the condition number in different situations, a white Gaussian signal is added to the input $x(n)$ with different SNRs. The range of the input SNR is -10 dB to 40 dB. Therefore, with an input SNR equal to -10 dB (the white noise dominates the speech) we can expect the condition number of the input signal covariance matrix to be close to 1 while with an input SNR of 40 dB (the speech largely dominates the white noise) the condition number will be high. Figures 1-4 show the evolution in time of the input signal, the normalized misalignment (we approximate the normalized misalignment with its instantaneous value), and the condition number of the input signal covariance matrix with different input SNRs (from -10 dB to 40 dB). We can see, as the input SNR increases, the condition number degrades as expected since the speech signal is ill-conditioned. As a result, the normalized misalignment is greatly affected by a large value of the condition number. As expected, the value of the misalignment after convergence in Fig. 1 is equal to -49 dB and the condition number is almost one. Now compare this to Fig. 2. In Fig. 2, the misalignment is equal to -40 dB and the average condition number is 8.2. The higher condition number in this case degrades the misalignment by 9 dB, which is exactly the degradation predicted by formula (33). We can verify the same trend with the other simulations.

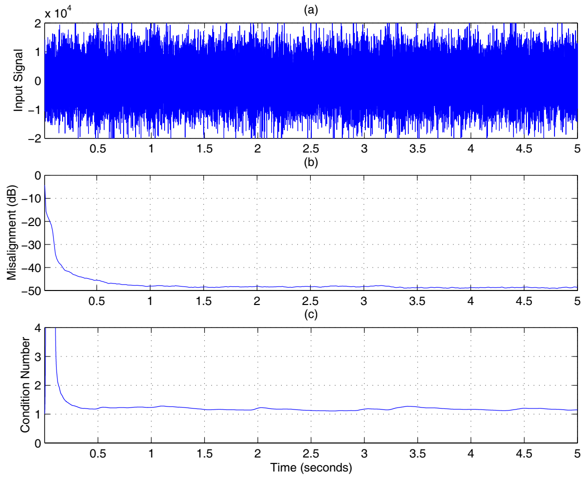

Fig. 1. Evolution in time of the (a) input signal, (b) normalized misalignment, and (c) condition number of the input signal covariance matrix. The input SNR is -10 dB.

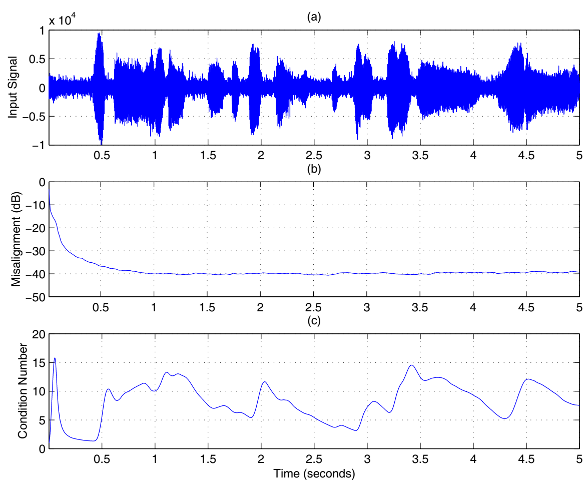

Fig. 2. Presentation the same as in Fig. 1. The input SNR is 10 dB.

7. CONCLUSION

The RLS algorithm plays a major role in adaptive signal processing. We proposed a simple and an efficient way to estimate the condition number of the input signal covariance matrix. We have shown that this condition number can be easily integrated in the FRLS structure at a very low cost from an arithmetic complexity point of of view. Finally, we have shown how the misalignment of the RLS depends on the condition number.

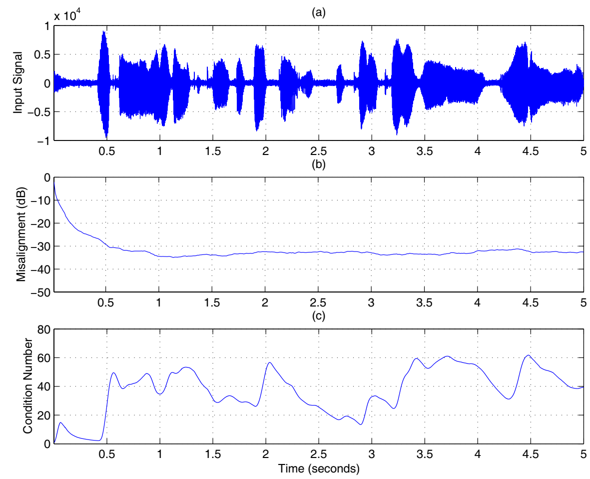

Fig. 3. Presentation the same as in Fig. 1. The input SNR is 20 dB.

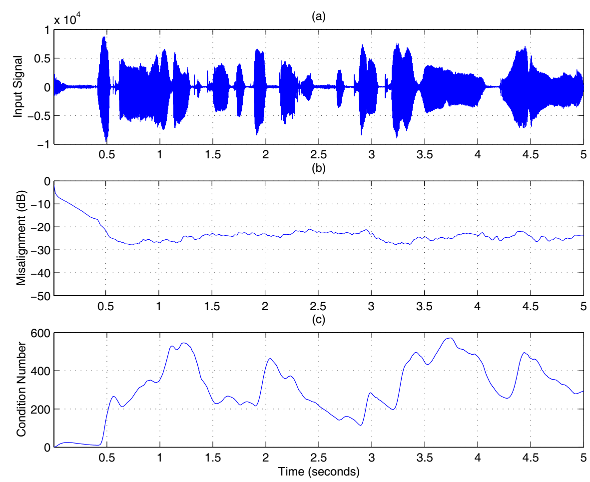

Fig. 4. Presentation the same as in Fig. 1. The input SNR is 40 dB.

8. REFERENCES

[1] M. G. Bellanger, *Adaptive Digital Filters and Signal Analysis*. Marcel Dekker, New York, 1987.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, N.J., 1985.

[3] S. Haykin, *Adaptive Filter Theory*. Fourth Edition, Prentice Hall, Upper Saddle River, N.J., 2002.

[4] J. Benesty and Y. Huang, editors, *Adaptive Signal Processing: Applications to Real-World Problems*. Springer-Verlag, Berlin, 2003.

[5] A. H. Sayed and T. Kailath, “A state-space approach to adaptive RLS filtering,” *IEEE Signal Processing Magazine*, vol. 11, pp. 18–60, July 1994.

[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, 1996.

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