Double-talk robust fast converging algorithms for network echo cancellation

J. Benesty
T. Gansler
Steven L. Grant
Missouri University of Science and Technology, sgrant@mst.edu
M. M. Sondhi

Follow this and additional works at: http://scholarsmine.mst.edu/faculty_work
Part of the Electrical and Computer Engineering Commons

Recommended Citation
http://scholarsmine.mst.edu/faculty_work/421

This Article is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. For more information, please contact weaverjr@mst.edu.
Double-Talk Robust Fast Converging Algorithms for Network Echo Cancellation
Tomas Gänsler, Member, IEEE, Steven L. Gay, M. Mohan Sondhi, and Jacob Benesty, Member, IEEE

Abstract—There is a need for echo cancelers for echo paths with long impulse responses (≥64 ms). This in turn creates a need for more rapidly converging algorithms in order to meet the specifications for network echo cancelers. Faster convergence, however, in general implies a higher sensitivity to near-end disturbances, especially “double-talk.” Recently, a fast converging algorithm has been proposed called proportionate normalized least mean squares (PNLMS) algorithm. This algorithm exploits the sparseness of the echo path and has the advantage that no detection of active coefficients is needed. In this paper we propose a method for making the PNLMS algorithm more robust against double-talk. The slower divergence rate of these algorithms in combination with a standard Geigel double-talk detector improves the performance of a network echo canceler considerably during double-talk. The principle is based on a scaled nonlinearity which is applied to the residual error signal. This results in the robust PNLMS algorithm which diverges much slower than PNLMS and standard NLMS. Tradeoff between convergence and divergence rate is easily adjusted with one parameter and the added complexity is about seven instructions per sample which is less than 0.3% of the total load of a PNLMS algorithm with 512 filter coefficients. A generalization of the robust PNLMS algorithm to a robust proportionate affine projection algorithm (APA) is also presented. It converges very fast, and unlike PNLMS, is not as dependent on the assumption of a sparse echo path response. The complexity of the robust proportionate APA of order two is roughly the same as that of PNLMS.

Index Terms—Adaptive filter, double-talk detection, echo cancellation, normalized least mean squares (NLMS), proportionate normalized least mean squares (PNLMS), robustness.

I. INTRODUCTION

There is a need for network echo cancelers for echo paths with long impulse responses (≥64 ms). However, longer impulse responses slow down the convergence rate. [1], [2], thus rendering traditional algorithms like normalized least mean squares (NLMS) inadequate. It will therefore, be desirable to implement fast-converging algorithms in future echo cancelers. In [2], [3], faster converging algorithms called proportionate normalized least mean squares (PNLMS) and PNLMS++, respectively, are proposed. These algorithms achieve higher convergence rate by using the fact that the active part of a network echo path is usually much smaller (4–8 ms) compared to 64–128 ms of the whole echo path that has to be covered.

Besides convergence rate and complexity issues, an important aspect of an echo canceler is its performance during “double-talk” (i.e., near-end speech). A high convergence rate is usually accompanied by a high divergence rate in the presence of double-talk. This mode in a conversation perturbs the adaptive filter of the echo canceler (EC) so that it does not attenuate the echo sufficiently for the nonlinear processor (NLP) to eliminate the residual echo. To inhibit the divergence of the EC during double-talk, the standard procedure is to use a level based double-talk detector (DTD) [4]. Whenever double-talk is detected the step-size of the adaptive filtering algorithm is set to zero thus inhibiting the adaptation. Unfortunately, during the time required by the DTD to detect double-talk, the echo canceler often diverges. This is because a few (e.g., <5) undetected large amplitude samples perturb the echo path estimate considerably. Once this happens, the coefficients are frozen in a poor state for at least the hangover time,¹ before adaptation can be resumed.

This work focuses on how to decelerate the divergence of the algorithm due to undetected double-talk while maintaining good convergence rate of the canceler. Our approach has its roots in the theory of robust statistics and is based on introducing a scaled nonlinearity into the adaptive algorithm. The nonlinearity limits the impact of large disturbances on the coefficient setting, but also reduces the convergence rate somewhat. This idea was developed for a subband echo canceler in [5] and showed promising results. However, neither the tradeoff between robustness and convergence rate nor the performance for fullband adaptive filters was studied in that paper. The robust algorithm developed here combines the PNLMS or PNLMS++, [2], [3] algorithm with the appropriate nonlinearity. Another method to improve the overall performance is to allow variability of the global step-size parameter (µ), [6], [7]. However, the algorithm by which the optimal step-size is found is fairly complex and it is difficult to adjust it fast enough when double-talk occurs.

The robust PNLMS++ algorithm is also generalized to a robust proportionate affine projection algorithm (PAPA)—an algorithm which is a combination of the affine projection algorithm, [8]–[10] and the proportionate step-size technique.

The paper is organized as follows: notations, algorithms (NLMS, PNLMS), and the Geigel DTD are presented in Section II. In Section III, we derive an adaptive scale factor, present robust versions of NLMS and PNLMS, and generalize the robust PNLMS algorithm to a robust proportionate affine projection algorithm. Section IV shows the computational complexity of the robust principle and compares the complexity of the different adaptive algorithms. Simulation results of the robust algorithms are shown in Section V and a discussion and conclusions are given in Section VI.

¹Hangover time is the minimum time for which adaptation is inhibited after detection of double-talk.
Fig. 1. Block diagram of the echo canceler and double-talk detector.

II. NOTATIONS, ADAPTIVE ALGORITHMS AND DOUBLE-TALK DETECTION

In derivations and descriptions, the following notations are used [see also Fig. 1]:
- far-end signal/speech;
- background noise;
- near-end signal/speech (double-talk);
- echo and background noise possibly including near-end signal;
- vector of samples $x_n$;
- excitation vector;
- excitation matrix;
- true echo path vector;
- estimated echo path vector;
- residual error;
- residual error vector;
- $\gamma_{n,l}$ element of the vector $\gamma_n$.

Here, $L$ is the adaptive filter length and $p$ is the projection order of the APA.

A. PNLMS and PNLMS++ Algorithms

The PNLMS algorithm was proposed in [2]. For line echo cancellation, it is reasonable to assume that the echo path is sparse, i.e., many coefficients are zero, and try to identify only the nonzero active coefficients. This is the idea behind the PNLMS algorithm which is a modification of the NLMS algorithm. In PNLMS, an adaptive individual step-size is assigned to each filter coefficient. The step-sizes are calculated from the last estimate of the filter coefficients so that a larger coefficient receives a larger weight, thus increasing the convergence rate of that coefficient. This has the effect that active coefficients are adjusted faster than non active coefficients (i.e., small or zero coefficients). An advantage of this technique compared to other approaches to individual step-size algorithms, e.g., [11] and [12] is that less a priori information is needed. Furthermore, the active regions of the echo path do not have to be detected explicitly. The PNLMS algorithm is described by the following equations:

$$h_{n+1} = h_n - \mu \frac{1}{x_n^T G_n x_n} G_n x_n \gamma_n$$

(1)

$$G_n = \text{diag}\{g_0, \cdots, g_{L-1,n}\}.$$  

(2)

$G_n$ is a diagonal matrix which adjusts the step-sizes of the individual taps of the filter, $\mu$ is the overall step-size parameter, and $\delta$ is a regularization parameter which prevents division by zero and stabilizes the solution when speech is used as input signal. The diagonal elements of $G_n$ are calculated as follows [2]:

$$\gamma_{n,l} = \max\{\rho \max\{\delta_p, |h_{0,n-l}|, \cdots, |h_{L-1,n-l}|\}; 0 \leq l \leq L-1 \}$$

(3)

$$g_{n,l} = L^2 \gamma_{n,l} / \sum_{l=0}^{L-1} \gamma_{n,l} 0 \leq l \leq L-1.$$  

(4)

Parameters $\delta_p$ and $\rho$ are positive numbers with typical values $\delta_p = 0.01, \rho = 5/L$. $\rho$ prevents coefficients from stalling when they are much smaller than the largest coefficient and $\delta_p$ regularizes the updating when all coefficients are zero at initialization.

A variant of this algorithm called the PNLMS++, [3], is the one we will consider here. In this algorithm, for odd-numbered time steps the matrix $G_n$ is derived as above, while for even-numbered steps it is chosen to be the identity matrix

$$G_n = I$$

(5)

which results in an NLMS iteration. The alternation between NLMS and PNLMS iterations has several advantages compared to using just the PNLMS technique, e.g., it makes the PNLMS++ algorithm much less sensitive to the assumption of a sparse impulse response without sacrificing performance.

B. Geigel DTD

A double-talk detector is used to suppress adaptation during periods of simultaneous far- and near-end speech. A simple and efficient way of detecting double-talk is to compare the magnitude of the far-end and near-end signals and declare double-talk if the near-end magnitude becomes larger than a value set by the far-end speech. A proven algorithm that has been in commercial use for many years is the Geigel DTD [4]. In this algorithm, double-talk is declared if

$$|y_n| \geq \vartheta \max\{|x_n|, |x_{n-1}|, \cdots, |x_{n-L+1}|\}.$$  

(6)

The detector threshold, $\vartheta$, is set to 0.5 if the hybrid attenuation is assumed to be 6 dB and to 0.71 if the attenuation is assumed to be 3 dB. A so-called hangover time, $T_h$, is also specified such that if double-talk is detected, then the adaptation is inhibited for this duration beyond the detected end of double-talk. Although this detector works fairly well, detection errors do occur, and these result in large amounts of divergence of the adapted filter coefficients, which in turn give rise to large amounts of uncanceled echo. One way to model these large disturbances together with the background noise, is to use an outlier-contaminated stochastic process. Note that an outlier-contaminated model is not valid for the residual error in the absence of a Geigel DTD, because without the DTD the residual consists of long-lasting bursts of near-end speech. Hence the DTD is an essential component in the robust PNLMS++ algorithm to be described in the next section.

III. ROBUST NLMS AND PNLMS++ ALGORITHMS

The NLMS and the PNLMS++ algorithm can both be made robust to large disturbances by modification of the criteria on which these algorithms are based. In general, however, such
modifications lower the convergence rate. The difficult problem of robust echo cancellation is to devise an algorithm that diverges slowly in response to double-talk, yet is able to rapidly track changes of the echo path when they occur. These two requirements are contradictory. The key to the solution to this problem is a combination of a DTD with traditional robust statistics and a delicately tuned scale variable, $s$.

Recall that the LMS is an iterative algorithm to adjust the estimated impulse response so as to minimize the cost function, $E\{e_n^2\}$, i.e., the mean square error. Each iteration updates the current estimate of $h_n$ by $\mu x_n e_n$, which is a step in the direction of a stochastic approximation to the gradient of $E\{e_n^2\}$. To make the algorithm insensitive to changes of the level of input signal, $x_n$, the proportionate factor $\mu$ is normalized, resulting in the NLMS algorithm. It is well known, [14], that other gradient algorithms can be derived by changing the cost function to

$$
J = E\{\varrho \left( \frac{|e_n|}{s} \right) \}
$$

where $\varrho(\cdot)$, is any symmetric function with a monotonically nondecreasing derivative (with respect to its argument). $s$ is the very important scale factor. The resulting algorithm, analogous to the steepest-descent method is

$$
h_n = h_{n-1} - \mu \nabla_h J.
$$

The algorithm can be made robust by a proper choice of $\varrho(\cdot)$, which must be chosen such that $\lim_{|e_n| \to \infty} |\nabla_h \varrho(\frac{|e_n|}{s})| < \infty$. Following suggestions in [13], we choose

$$
\nabla_h J = E\left\{ -x_n \text{sign}(e_n) \psi \left( \frac{|e_n|}{s} \right) \left( \frac{1}{s} \right) \right\}
$$

where $\psi(\cdot)$ is a limiter

$$
\psi \left( \frac{|e_n|}{s} \right) = \min \left\{ \frac{|e_n|}{s}, k_0 \right\}.
$$

A. Estimating the Scale Factor

The estimate of the scale factor, $s$, should reflect the background noise level at the near-end, be robust to short burst disturbances (double-talk), and track long term changes of the residual error (echo path changes). To fulfill these requirements we have chosen the scale factor estimate as

$$
s_n = \lambda s_{n-1} + \frac{1 - \lambda}{\beta} s_{n-1} \psi \left( \frac{|e_n|}{s_{n-1}} \right).
$$

where $\beta$ is a normalization constant. The choice of this method of estimating $s$, which is very simple to implement, is justified in the Appendix. With this choice, the current estimate of $s$ is governed by the level of the error signal in the immediate past over a time interval roughly equal to $1/(1-\lambda)$ where the practical range of $\lambda$ is $[0.9, 1]$ and the range of $\beta$ (which depends on $k_0$ in (10)) is approximately $[0.6, 0.74]$. The initial value of the scale factor can be chosen as $s_0 = \sigma_x$ where $\sigma_x$ is set approximately to the average speech level in voice telephone networks (7–8 dBm) which corresponds to the linear value 2000 in a 16 bit representation of data). When the algorithm has not yet converged, $s$ is large. Hence the limiter is in its linear portion and therefore the robust algorithm behaves like the conventional NLMS or PNLMS algorithms. When double-talk occurs, the error is determined by the limiter and by the scale of the error signal during the recent past of the error signal before the double-talk occurs. Thus the divergence rate is reduced for a duration of about $1/(1-\lambda)$. This gives ample time for the DTD to act. If there is a sudden system change, the algorithm will not track immediately. However, as the scale estimator tracks the larger error signal, the nonlinearity is scaled up and the convergence rate accelerates. The tradeoff between robustness and tracking rate of the adaptive algorithm is thus governed by the tracking rate of the scale estimator, which is controlled by a single parameter, $\lambda$. As with the Geigel DTD, it is useful to introduce a hangover time for control of scale updating. When the DTD detects double-talk, adaptation of $s_n$ should be inhibited...
for a specific time, preferably as long as the DTD hangover time, \( T_{\text{hold}} \).

### B. Generalization of the PNLMS Algorithm to the Affine Projection Algorithm

A PAPA is given by
\[
\mathbf{h}_n = \mathbf{h}_{n-1} + \mu \mathbf{G}_n \mathbf{x}_n (\mathbf{x}_n^T \mathbf{G}_n \mathbf{x}_n + \delta I)^{-1} \mathbf{e}_n
\]  
(18)

where \( \mathbf{G}_n \) is as defined in the Section II. Let
\[
\mathbf{G}_n^{-1} = (\mathbf{x}_n^T \mathbf{G}_n \mathbf{x}_n + \delta I)^{-1}.
\]  
(19)

This matrix “whitens” the input data, \( \mathbf{x}_n \), and thus the convergence rate of the adaptive filter is increased. With \( \mathbf{G}_n = \mathbf{I} \), equation (18) reduces to the standard APA, first introduced in [8]. The regularization parameter in (18) was proposed in [15]. As evident, PAPA is obtained by combining APA with the proportional step-size of PNLMS. We can omit the matrix \( \mathbf{G}_n \) in the definition of \( \mathbf{G}_n^{-1} \) to save computations. Inclusion of the matrix requires a significant \( p^2L \) multiplications per sample, but according to our simulations, the effect on performance and stability is minimal.

A robust version of PAPA (and hence of APA) is obtained straightforwardly, by applying the principles presented previously
\[
\psi_{l,n}(\mathbf{e}_n) = \min \left\{ \frac{|\mathbf{e}_n l|}{s_{n-1}}, k_0 \right\}, \quad l = 1, \ldots, p
\]
(20)

\[
\mathbf{h}_n = \mathbf{h}_{n-1} + \mu \mathbf{G}_n \mathbf{x}_n \mathbf{G}_n^{-1} \psi(e_n) \odot \text{sign}(e_n) s_{n-1}.
\]
(21)

\[
s_n = \lambda s_{n-1} + \frac{(1 - \lambda)}{\beta} \psi_{l,n}(e_n) s_{n-1}.
\]
(22)

In (21), \( \odot \) denotes elementwise multiplications.

Additionally, most of the computational procedures of the fast affine projection (FAP) algorithm [9], [10] can be incorporated in order to reduce the computational complexity of PAPA. Unfortunately, introducing an alternative coefficient vector, as in [9], cannot be done in PAPA, because invariance of the product of the step-size matrix and the excitation vector is destroyed, since \( \mathbf{G}_n \) varies from one iteration to the next.

### IV. COMPLEXITY COMPARISON

It is interesting to compare the complexity of the presented algorithms. We first look at the complexity of the scaled nonlinearity. Reorganize the calculations as,
\[
\psi(c_n) = \min \{|c_n|, s_{n-1}\}
\]
(23)

\[
s_n = \lambda s_{n-1} + \frac{k_0 (1 - \lambda)}{\beta} \psi(c_n)
\]
(24)

which leads to the complexity in instructions per sample shown in Table I.

A standard NLMS algorithm usually requires \( 3L \) multiplications (per sample) which can be reduced to \( 2L \) with (careful) recursive estimation of the normalization factor. Table II shows the complexity including multiplications for the nonlinearity for NLMS, PNLMS and PAPA after some reorganization of (1) and (3). The difference in complexity between PNLMS and PAPA of order two is modest. A low-complexity version of robust PAPA of order two is

\[
\begin{align*}
\psi(c_n) &= \min \{|c_n|, s_{n-1}\} \\
\psi(c_n) &= \frac{1}{\beta} (1 - \mu) s_{n-1} \psi^T(e_n) e_n
\end{align*}
\]
(25)

which can be solved directly solving the 2 \( \times \) 2 system of equations.

### V. SIMULATIONS

A full comparison of every aspect of the robust versus the nonrobust algorithms is beyond the scope of this paper. To reduce the number of simulations we state the following general properties of the algorithms.

- A robust version of an algorithm converges slower than its nonrobust version. This means that the robust NLMS is slower than standard NLMS, etc.
- Different algorithms are affected differently by the scaled nonlinearity, e.g. the robust NLMS is significantly slower than NLMS while the robust PNLMS++ is almost as fast as PNLMS++.
- When double-talk occurs, the divergence rates of the nonrobust algorithms, NLMS, PNLMS++, and PAPA are approximately equal. This was shown in [3] for NLMS and PNLMS++.

Comparisons between nonrobust NLMS and PNLMS(++) with respect to convergence rate can be found in [2] and [3]. Because of the facts above we choose to compare the robust PAPA and PNLMS algorithms with the standard nonrobust NLMS which since the late 1970s is the preferred algorithm in commercial echo cancelers. The purpose of the simulations in this section is to show the excellent performance of robust algorithms during double-talk and the high convergence rate of the robust PAPA.
and PNLMS algorithms. Important factors that affect the results are, e.g., type of excitation signal, far-end to double-talk ratio, and hybrid attenuation. Therefore, experiments using speech as well as the composite source signal (explained below) as excitation signal are shown. Far-end to double-talk ratio is approximately equal to the assumption in the Geigel DTD, which can be regarded as a worst case. Experiments are performed with three different hybrid attenuations. All algorithms incorporate the Geigel DTD in which the settings (\( \theta, \bar{T}_{\text{in}}, k_{\text{G}} \)) are chosen as commonly used in commercial hardware. These settings have been found “optimal” in practice so that the DTD operates well for all different signal situations that may occur in a telephone network.

With a projection order of 2 in PAPA, an input signal having the properties of an AR\{1\} process can be perfectly whitened and maximum improvement of the convergence rate is achieved. Speech, however, is not an AR\{1\} process but can be fairly well modeled by an eighth-order AR process. Choosing the order as two is a compromise between complexity and performance. Moreover, the improvement in performance is quite small when \( p \) is increased beyond two [9].

A. Results with Speech as Excitation Signal

The parameter settings chosen for the following simulations are as follows.

- \( \mu = 0.2 \), \( L = 512 \) (64 ms), \( \delta = 2 \cdot 10^5 \) (NLMS, PNLMS++), \( \delta = 1 \cdot 10^6 \) (PAPA), \( \delta_g = 0.01 \), \( \rho = 0.01 \).
- \( \sigma_x = 1.9 \cdot 10^3 \) (\( \approx -18.6 \) dBm0), SNR = 39 dB (echo-to-noise ratio).
- Average far-end to double-talk ratio is 6 dB.
- Hybrid attenuation: 20 dB.
- Geigel detector assumes 6 dB attenuation (\( \theta = 0.5 \)), \( T_{\text{in}} = 240 \) samples equal to 30 ms at 8 kHz sampling frequency.
- Parameters for the robust algorithm are, \( (\lambda, k_0) = (0.997, 1.1) \). This choice results in \( \beta \approx 0.00665 \).
- \( h_{-1} = 0 \), \( s_{-1} = 1000 \).

The scale estimate in (17), \( s_0 \), is never allowed to become lower than 2. This inhibits bad behavior in low noise situations. All algorithms are tuned to achieve the same asymptotic misalignment in order to fairly compare convergence rate. The misalignment is given by

\[
\varepsilon = ||h - \hat{h}||/||h||.
\]  

The impulse response and corresponding magnitude function of the hybrid is shown in Fig. 2(a) and (b). Fig. 3 shows far-end signal, double-talk, and the misalignment of the three algorithms. Initial convergence rates of PNLMS++ and PAPA are clearly superior to that of NLMS. While the nonrobust NLMS (with Geigel detector) diverges to a misalignment of +5 dB the robust algorithms are much less affected and never perform worse than −10 dB misalignment during double-talk. The slow reconvergence of NLMS and robust PNLMS++ after the double-talk sequence is caused by poor excitation of the speech signal, i.e., this segment is highly correlated.

Fig. 4 shows the behavior after an abrupt system change where the impulse response is shifted 200 samples at 1 s. The robust algorithms outperform NLMS in this case.

B. Results with the Composite Source Signal as Excitation Signal

The standard ITU-T G.168 [16] recommends certain test procedures for evaluating the performance of echo cancelers. Test signals used are the so-called composite source signals (CSS) that have properties similar to those of speech with both voiced and unvoiced sequences as well as pauses. This section presents
results from test situations evaluating the performance during double-talk, the so-called Test 3B. The mean square error is defined as performance index and is given by

$$\text{MSE} = \frac{P_{y-w}}{P_{y-w}}$$  \hspace{1cm} (34)

$$P_{y-w} = \text{LPF}\{(e_{t1} - w_{n})^2\}$$  \hspace{1cm} (35)

where the LPF denotes a lowpass filter. In this case, it has a single pole at 0.999. This choice does not significantly affect the convergence rate in the figures. $P_{y-w}$ is analogously calculated.

Test 3B evaluates the performance of the echo canceler for a high level double-talk sequence. The double-talk level in this case is about the same as that of the far-end signal, thus a fixed threshold Geigel DTD assuming 3 or 6 dB attenuation is able to detect this double-talk. False alarms and failures to detect double-talk are influenced by the chosen threshold in the DTD and the attenuation in the hybrid. Results from two hybrid attenuations are therefore evaluated. What differs in parameter and initial value settings from the previous simulation is:

- $\sigma_x = 1.3 \times 10^{3}$, SNR $\approx$ 37 dB (echo-to-PCM quantization noise ratio);
- hybrid attenuations: 6, 11 dB;
- Geigel detector assumes 3 dB attenuation ($\beta = 0.71$), $T_{\text{loc}} = 240$.

The tests are made with a sparse echo path with 6 dB attenuation and a multireflection echo path with 11 dB attenuation. These are shown in Fig. 2(c)–(f).

Far- and near-end signals used in Test 3B are shown in Figs. 5(a) and 6(a) for the 11 dB and 6 dB hybrid, respectively. Double-talk starts after about 12 s. Mean-square errors of the three algorithms are shown in Figs. 5(b) and 6(b). The latter case with a 6 dB hybrid is considered to be very difficult in practice. The two robust algorithms handle double-talk without severe degradation of MSE, while the nonrobust NLMS, despite the Geigel detector, diverges as much as 30 dB. This divergence occurs when the DTD fails to detect the double-talk and in the first case [Fig. 5(b)], as few as three samples drive the hybrid estimate far from optimum. In general, the lengths of undetected bursts in these simulations range from a few up to a couple of hundred samples. A change of the threshold, $\beta$, to a 6 dB assumption would reduce the divergence of NLMS by only about 3–5 dB, while the false alarms would increase and slow down convergence rate significantly, especially for a 6 dB hybrid.

VI. CONCLUSIONS

A scaled nonlinearity combined with a Geigel DTD increases the robustness of the echo canceler. The scaled nonlinearity operates in the same manner as varying the step-size ($\mu$). That is, bounding the error signal can be interpreted as a reduction of the step-size parameter. What differentiates our approach is that while traditional variable step-size methods [6], [7], try to detect periods of double-talk and then take action, in our robust technique we use the signal before double-talk in order to be prepared for it. Due to this fact, the robust technique is faster and more efficient. Another major advantage is that only a few instructions and little memory is required to implement the robust principle.

The complexity of PNLMS and PAPA is on the order of $4L$ multiplications per sample which is about twice that of NLMS.
However, the convergence rate of PNLMS and PAPA is considerably higher. It is shown in this paper that the robust version of PNLMS and PAPA converge faster than NLMS and perform significantly better, up to 30 dB higher echo attenuation, during double-talk in the ITU test 3B. Several other simulations, not shown here, confirm reliable performance of the robust algorithms for different double-talk situations. The principle of robustness works at all stages of convergence of the robust algorithms. They resist divergence during double-talk, even in situations when they have not yet fully converged. It should also be mentioned, that the performance loss due to the use of the nonlinearity for robustness, is only minor. The robust PAPA converges faster than the nonrobust PNLMS+++ for different tested speech sequences.

APPENDIX

DERIVATION OF THE SCALE ESTIMATE

The purpose of the scale factor is multifold. Traditionally, the scale is used to make a robust algorithm, e.g., (16) invariant to the background noise level. In this case, the noise and the echo path to be identified are assumed stationary. In echo cancellation, however, it is desirable that the scale factor should track nonstationary background noise. It should also scale the nonlinearity appropriately so that the canceler is robust to double-talk bursts. Consequently, the tracking rate of the scale estimator controls both convergence rate and robustness of the algorithm.

A scale factor, $s$, can be found by defining an implicit function, denoted $\xi(n)$, which is a weighted sum of an even function $\chi(\cdot)$ of $s$, given by [13]

$$\xi(n) = \sum_{k=0}^{n} \lambda_k \chi\left(\frac{|e_k|}{s}\right) = 0$$

(A1)

where $s$ is the scale factor and $\lambda_k$ is a forgetting factor. For a bounded $\chi(\cdot)$, it is obvious that $s$ will scale the value of $\xi(n)$. We have chosen $\chi(\cdot)$ as

$$\chi(\cdot) = \psi(\cdot) - \beta.$$  

(A2)

The reason for using $\psi(\cdot)$ here as well as in (16) is purely for reducing computational complexity. Other bounded even functions would also do since the choice of $\chi(\cdot)$ turns out to be not very critical. For normalization, $\beta$ is chosen such that for a zero-mean, unit-variance Gaussian process, $z$

$$E\{\chi(z)\} = 0.$$

This gives for $\psi(\cdot)$ according to (10)

$$\beta = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} \psi(z) e^{-(1/2)z^2} \, dz$$

$$= \sqrt{\frac{2}{\pi}} \left(1 - e^{-(1/2)k_0^2}\right) + k_0 \text{erfc}\left(\frac{k_0}{\sqrt{2}}\right)$$

(A3)

where

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} \, dt.$$  

(A4)

This leads to a scale estimate $s = \sigma_n$ for Gaussian noise, upon convergence.

A recursive scale estimator is derived with a Gauss–Newton technique in the same fashion as for the robust PNLMS+++ algorithm. Let

$$s_n = s_{n-1} - \left[\nabla s \xi(n)\right]^{-1} \xi(n).$$

(A5)

The gradient of (A1) is

$$\nabla s \xi(n) = \frac{-\lambda_k^{-1} |e_k|}{2s^2} \chi'\left(\frac{e_k}{s}\right)$$

(A6)

which can be written as

$$\nabla s \xi(n) = -\frac{1}{s} \lambda_k b_{n-1} = -\frac{1}{s} \left[\frac{|e_k|}{s} \chi'\left(\frac{e_k}{s}\right) + \lambda_k b_{n-1}\right].$$

(A7)

Introduce the following approximations

$$\hat{\xi}(n) = \frac{|e_k|}{s_n}, \quad \nabla s \hat{\xi}(n) = \nabla s \xi(n-1).$$

A recursive least squares type of algorithm using (A5) and (A7) then becomes

$$s_n = s_{n-1} + \frac{s_{n-1}}{b_{n-1}} \chi'\left(\frac{|e_k|}{s_{n-1}}\right)$$

(A8)

$$b_n = \lambda_k b_{n-1} + \frac{|e_k|}{s_{n-1}} \chi'\left(\frac{|e_k|}{s_{n-1}}\right).$$

(A9)

The complexity of this algorithm is reduced by using the following assumptions: $\nabla s \xi(n)$ is considered stationary and $s_n$ converges to the background standard deviation of the noise (Gaussian). Then

$$E\{\nabla s \xi(n)\} = \sum_{k=0}^{n} -\lambda_k^{-1} \frac{|e_k|}{s^2} \chi'(\frac{e_k}{s})$$

$$= \sum_{k=0}^{n} -\lambda_k^{-1} \frac{1}{s} \chi'(\frac{e_k}{s})$$

(A10)

$$= \sum_{k=0}^{n} -\lambda_k^{-1} \frac{1}{s} \sqrt{\frac{2}{\pi}} \int_{0}^{k_0} z e^{-(1/2)z^2} \, dz$$

$$= \sum_{k=0}^{n} -\lambda_k^{-1} \frac{1}{s} \sqrt{\frac{2}{\pi}} \left(1 - e^{-(1/2)k_0^2}\right)$$

$$\rightarrow -\frac{1}{s} \frac{\alpha}{1 - \lambda_k}, \quad n \rightarrow \infty$$

(A12)

$$\alpha = \sqrt{\frac{2}{\pi}} \left(1 - e^{-(1/2)k_0^2}\right).$$

(A13)

Combining (A2), (A5), and (A12) leads to

$$s_n = s_{n-1} + \frac{1-\lambda_k}{\alpha} s_{n-1} \psi\left(\frac{|e_k|}{s_{n-1}}\right)$$

$$= \lambda s_{n-1} + \frac{1-\lambda}{\beta} s_{n-1} \psi\left(\frac{|e_k|}{s_{n-1}}\right)$$

(A14)

$$\lambda = 1 - \frac{\beta}{\alpha} (1 - \lambda_k).$$

(A15)
GÄNSLER et al.: DOUBLE-TALK ROBUST FAST CONVERGING ALGORITHMS

ACKNOWLEDGMENT

The authors would like to thank J. Cezanne, Lucent Technologies, Holmdel, NJ, for supplying test data and giving instructive comments.

REFERENCES


Steven L. Gay was born in Kansas City, MO, in 1958. He received the B.S. degree in electrical engineering from the University of Missouri, Rolla, in 1979, the M.S. degree in electrical engineering from the California Institute of Technology, Pasadena, in 1981, and the Ph.D. degree in electrical engineering from Rutgers, The State University of New Jersey, New Brunswick, in 1994.

He joined Bell Labs, Whippany, NJ, in 1980. After working on speech coding with the ITT-Defense Communications Division from 1982 to 1984, he returned to Bell Labs to work in speech coding, echo cancellation, and other digital signal processing research and applications in various development departments. In 1991, he joined the Bell Labs Acoustics and Audio Research Department, Murray Hill, NJ. His main research topics include network and acoustic echo control, efficient adaptive filtering algorithms, and real-time implementations. He co-edited the book Acoustic Signal Processing for Telecommunication.

Dr. Gay co-chaired the 1999 International Workshop on Acoustic Echo and Noise Control and continues to serve on its standing committee.

M. Mohan Sondhi received the B.Sc. (Hons.) degree in physics in 1950, from Delhi University, Delhi, India, the D.I.I.Sc. degree in communications engineering in 1953 from the Indian Institute of Science, Bangalore, and the M.S. degree in electrical engineering and the Ph.D. degree, both from the University of Wisconsin, Madison, in 1955 and 1957, respectively.

Before joining Bell Laboratories, Murray Hill, NJ, he was with the Avionics Division of John Oster Manufacturing Co., Racine, WI, and with the Central Electronics Engineering Research Institute, Pilani, India, and taught for one year at Toronto University, Toronto, ON, Canada. He was a Guest Scientist from 1971 to 1972 at the Royal Institute of Technology, Stockholm, Sweden, and for shorter periods at CNET, Lannion, France, and at NTT Human Interface Laboratory, Musasashino, Japan. His research interests have included speech signal processing, echo cancellation, modeling of auditory, speech, and visual processing by humans, acoustical inverse problems, speech recognition, and analysis and synthesis of speech using articulatory models. He has authored or co-authored more than 95 published papers and is co-editor of Advances in Speech and Signal Processing. He is currently on the Editorial Board of Speech Communication and of the International Journal of Imaging Systems and Technology.

Dr. Sondhi was Associate Editor of the IEEE TRANSACTIONS ON ACOUSTICS IN SPEECH AND SIGNAL PROCESSING for several years. He was a Distinguished Lecturer of the IEEE Acoustics, Speech, and Signal Processing Society for two years. He is a Bell Labs Fellow.

M. GÄNSLER was born in Sweden in 1966. He received the M.S. degree in electrical engineering and the Ph.D. degree in signal processing from Lund University, Lund, Sweden, in 1990 and 1996, respectively.

From 1997 to 1999, he was an Assistant Professor at Lund University. During 1998, he was with Bell Labs, Lucent Technologies, Murray Hill, NJ, as a Consultant and since October 1999 as Member of Technical Staff. His research interests include robust estimation, adaptive filtering, and echo cancellation.

Tommaso Benesty (M'98) was born in Marrakesh, Morocco, in 1963. He received the Master degree in microwaves from Pierre and Marie Curie University, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, Orsay, France, in 1991. During the Ph.D. studies, he worked on adaptive filters and fast algorithms at the Centre National d’Etudes des Telecommunications (CNET), Paris, France.

From January 1994 to July 1995, he was with Telecom Paris working on multichannel adaptive filters and acoustic echo cancellation. He joined Bell Labs, Lucent Technologies (formerly AT&T) in October 1995, first as a Consultant and then as a Member of Technical Staff. Since then, he has been working on stereophonic acoustic echo cancellation, adaptive filters, source localization, robust network echo cancellation, and blind deconvolution. He is co-editor/co-author of the book "Acoustic Signal Processing for Telecommunication."

Dr. Benesty was the co-chair of the 1999 International Workshop on Acoustic Echo and Noise Control.

Tomas Gänssler (S'91–M'96) was born in Sweden in 1966. He received the M.S. degree in electrical engineering and the Ph.D. degree in signal processing from Lund University, Lund, Sweden, in 1990 and 1996, respectively.

From 1997 to 1999, he was an Assistant Professor at Lund University. During 1998, he was with Bell Labs, Lucent Technologies, Murray Hill, NJ, as a Consultant and since October 1999 as Member of Technical Staff. His research interests include robust estimation, adaptive filtering, and echo cancellation and subband signal processing.