ROBUST RLS WITH ROUND ROBIN REGULARIZATION INCLUDING APPLICATION TO STEREO ACOUSTIC ECHO CANCELLATION

Jack W. Stokes and John C. Platt

Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
{jstokes,jplatt}@microsoft.com

ABSTRACT

This paper introduces a new algorithm for implementing subband, adaptive filtering using recursive least squares (RLS) with round robin regularization. We show that modern microprocessors with SIMD (Single Instruction, Multiple Data) instructions can now implement RLS for practical problems thereby avoiding the numerical stability issues associated with fast RLS (FRLS). The desired signal may be multichannel as in the stereo, acoustic echo cancellation (AEC) problem where the separate channels of the playback signals are often highly correlated. In this case, the recursive computation of the inverse correlation matrix in RLS will diverge. To avoid this problem, we extend adaptive subband RLS to include round robin regularization. The new, regularized RLS (RRLS) algorithm has been implemented in real-time on a personal computer (PC) for the stereo AEC problem and performs well in typical PC scenarios.

1. INTRODUCTION

In this paper, we present a new, robust version of recursive, least squares (RLS) and apply the algorithm to the stereo, acoustic echo cancellation (AEC) problem. AEC removes the echo captured by a microphone when a sound is simultaneously played through speakers located in close proximity to the microphone. Previously, many adaptive filter algorithms have been implemented using either LMS (least mean square) or fast RLS (FRLS) due to the limited processing capabilities of older general purpose CPUs (central processing units) and DSPs (digital signal processors). LMS and FRLS are $O(N)$ compared to $O(N^2)$ for RLS. However, LMS and normalized LMS (NLMS) do not work well for the stereo AEC problem, while FRLS can be unstable even with attempts to stabilize the algorithm [1]. With the recent introduction of very fast CPUs with SIMD (single instruction, multiple data) operations and VLIW (very long instruction word) DSPs, implementing real-time RLS is now an option for modern processors. It is well known that adaptive subband filters or frequency domain adaptive filters typically converge more quickly than time domain adaptive filters. The improved convergence is due to the small eigenvalue spread within the subband when compared to processing full bandwidth, time domain signals. In this paper, we implement adaptive subband filters using RLS to solve the stereo AEC problem. However, RLS will still become unstable in the stereo AEC problem when the playback data sent to the speakers is highly correlated across the channels. In this case, the correlation matrix of the input data quickly becomes singular depending on the value of the forgetting factor. To solve this problem, we extend the standard, adaptive subband RLS algorithm to prevent the correlation matrix from becoming singular thereby causing the RLS adaptive filter to diverge. This new method is called regularized RLS (RRLS) and it periodically regularizes the correlation matrix for each subband by adding a small value to the diagonal of the individual correlation matrices. Regularization is often used in machine learning algorithms to choose the most likely solution in the case where the solution is undefined. RLS achieves its efficiency by recursively computing the inverse correlation matrix instead of updating the correlation matrix directly. Since we must regularize the correlation matrix instead of the inverse correlation matrix which involves the computation of two matrix inverses, RRLS uses a round robin scheme to regularize one or more subbands for each new frame of capture and playback data thereby minimizing the CPU consumption.

Stereo AEC has been an active area of research for the past decade [2] [3] [4]. Primarily, previous stereo AEC research has focused on transmitting a stereo speech signal, captured by stereo microphones, during a video conferencing session. This scenario leads to the well known problem that the solution is non-unique [3] [4]. As a result, stereo acoustic echo cancellers must learn and track the acoustic transfer functions from the remote person to the two separate microphones in the far end room in addition to the acoustic transfer functions from each speaker to each microphone in the near end room. To avoid this problem, the stereo speech signals are usually decorrelated by processing each channel with a non-linear transform [3] [4].

While the multiple microphone, multiple transmitted channel problem described above is a highly important research problem, this scenario is not used in practice on a personal computer today. For internet telephony, voice is either captured using a single microphone, or captured using a microphone array and transmitted via a single channel after beamforming. There are a number of scenarios where using a non-linear transform to decorrelate the playback signals can be problematic. For example, in internet gaming with 3D sounds and voice chat, processing the playback channels with non-linear transforms can destroy the 3D game sounds implemented using HRTFs (head related transfer functions) and cross-talk cancellation. In addition, audiophiles will complain about the added distortion when listening to background music played while using speech recognition or during a video conferencing session. Furthermore, the non-unique solution problem described by [3] does not exist for multiple people speaking simultaneously located in the far end room; it arises when one person speaks in one location and then a second person starts speaking in a second location after the first person has become silent. Since most stereo music involves multiple instruments playing simultaneously and 3D games usually have more than one sound playing at a given time, the two most common sources of multichannel playback signals, we have not found the non-uniqueness problem to be a significant issue on the PC during real-time listening tests with the new RRLS algorithm. In this paper, we show that even for a case where the non-uniqueness problem is constructed, the distortion in the near end captured speech can be minimized by adjusting the regularization constant in the RRLS algorithm.

Adaptive, subband filtering was used for AEC in [5]. In [5], Hätty uses FRLS and a round-robin scheme to *completely* reinitialize each subband periodically. In the RRLS algorithm, we instead use a regularization method to slightly modify the correlation matrix thereby avoiding artifacts generated by completely resetting a subband's adaptive filter. Gay [6] first proposed a dynamically, regularized FRLS (DR-FRLS) algorithm and applied it to the AEC problem. However, the algorithm was implemented in the time domain for the mono AEC problem. We experimented with a stabilized version of FRLS [1] applied to the frequency domain, but unfortunately, it still diverges for the stereo AEC problem where the correlation between the channels is often very high. In [4], an adaptive subband, stereo AEC is proposed using stabilized FRLS. However, this algorithm is still unstable and must use a parallel, adaptive filter architecture to reset individual subbands if they become unstable. *The new RRLS algorithm is completely robust and does not suffer from any instability in the individual subbands.* In addition, it does not require a parallel AEC structure to reinitialize the tap weights thereby saving additional CPU cycles.

This paper is organized as follows. The stereo AEC system architecture is given in section 2. In section 3, we describe the new robust, RLS algorithm with round robin regularization. Finally the algorithmic and computational performance of the RRLS algorithm is analyzed in section 4 and conclusions provided in section 5.

2. STEREO AEC SYSTEM ARCHITECTURE

An adaptive subband AEC system with stereo playback is shown in figure 1. Even though we focus on stereo playback in this paper, the algorithm can handle more than two playback channels. The playback signal $\mathbf{x}$ is composed of two channels $x(0)$ and $x(1)$. The stereo playback signal can be generated in many ways. For example, it may be a true stereo signal (e.g. music, computer sounds), mono speech, silence, or a combination of speech and music. An analog speaker signal is played out through the speakers and produces an echo at the microphone. In addition to the echo from the speakers, the audio signal that is captured by the microphone is also composed of the desired near-end speech and background noise. The audio signal captured by the microphone is given by y. As shown in figure 1, we perform AEC processing by using adaptive subband filtering. The subbands are computed using the modulated complex lapped transform (MCLT) [7], although other frequency domain transforms can also be used (e.g. Fast Fourier Transform). Separate analysis filterbanks convert each of the stereo playback signals from the time domain signals, $x(0)$ and $x(1)$, to the frequency domain signals, $X(0)$ and $X(1)$, respectively. Likewise, a third analysis filterbank converts the mono microphone signal from the time domain to the frequency domain signal Y. A separate adaptive filter is run on each subband independent of the other subbands. After cancelling the echo, the resulting frequency domain signal is transformed back to the time domain using a synthesis filterbank.

3. ROBUST RLS WITH ROUND-ROBIN REGULARIZATION ALGORITHM DESCRIPTION

RLS [8] is a fast method to solve the normal equation

$$\mathbf{W}^{opt} = \mathbf{R}^{-1}\mathbf{p} \tag{1}$$

where $\mathbf{W}^{opt}$ is the optimal weight vector, $\mathbf{R}$ is the correlation matrix for the multichannel playback data, and $\mathbf{p}$ is the cross-correlation vector between the playback data and the mono capture data. If $\mathbf{R}$ starts to become close to singular, then the values in its inverse become very large and a valid estimate of the weight vector cannot be found. The correlation matrix becomes almost singular if the individual playback channels are highly correlated. To prevent the correlation matrix from becoming almost singular, we propose a new round robin scheme to regularize the correlation matrix for a particular subband. The new RLS algorithm with round-robin regularization is:

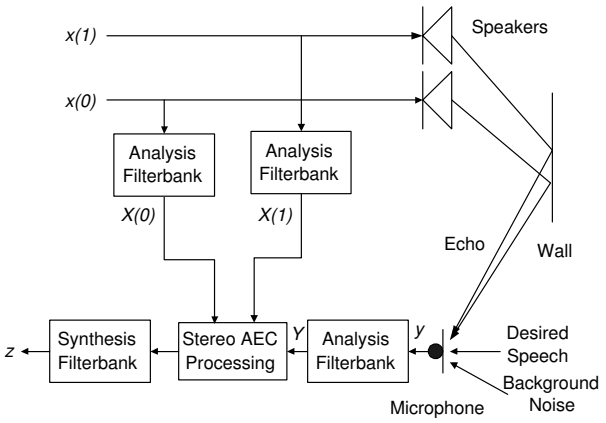

Fig. 1. Adaptive Subband Based Stereo AEC System.

$$\mathbf{P}(0,m) = \delta^{-1}\mathbf{I} \tag{2}$$

$$\hat{\mathbf{W}}(1,m) = 0 \tag{3}$$

$$RoundRobinCount = 0; \tag{4}$$

for each subband $m = 0 \cdots M-1$, compute

$$\mathbf{K}(m) = \frac{\lambda^{-1}\mathbf{P}(n-1,m)\mathbf{X}(n,m)}{1+\lambda^{-1}\mathbf{X}^H(n,m)\mathbf{P}(n-1,m)\mathbf{X}(n,m)} \tag{5}$$

$$\xi(m) = Y(n,m) - \hat{\mathbf{W}}^H(n,m)\mathbf{X}(n,m) \tag{6}$$

$$\hat{\mathbf{W}}(n+1,m) = \hat{\mathbf{W}}(n,m) + \mathbf{K}(m)\xi^*(m) \tag{7}$$

$$\mathbf{P}(n,m) = \lambda^{-1}\mathbf{P}(n-1,m) - \lambda^{-1}\mathbf{K}(m)\mathbf{X}^H(n,m)\mathbf{P}(n-1,m) \tag{8}$$

if (*m == RoundRobinCount*)

$$\mathbf{R}(n,m) = \mathbf{P}^{-1}(n,m) \tag{9}$$

$$\mathbf{R}(n,m) = \mathbf{R}(n,m) + \beta_{RLS}\mathbf{I} \tag{10}$$

$$\mathbf{P}(n,m) = \mathbf{R}^{-1}(n,m) \tag{11}$$

end

end

RoundRobinCount = RoundRobinCount + 1

if (*RoundRobinCount == MaxRoundRobinCount*)

RoundRobinCount = 0

end

where $n \geq 1$ is the frame of audio data, δ is a small constant, $\mathbf{I}$ is the identity matrix, M is the number of subbands,

$$\mathbf{K}(m) = [K(m,0)\cdots K(m,C*L-1)]^T$$

is the multichannel Kalman gain vector, C is the number of channels, L is the subband filter length,

$$\mathbf{X}(n,m) = [X(n,m,0)\cdots X(n,m,C-1) \; X(n-1,m,0)\cdots X(n-L+1,m,C-1)]^T$$

is the multichannel speaker input vector, $\mathbf{P}(n,m)$ is the inverse of $\mathbf{R}(n,m)$, $\mathbf{P}^{-1}(n,m)$ is the inverse of the inverse correlation matrix, β_{RLS} is the regularization factor,

$$\hat{\mathbf{W}}(n,m) = [\hat{W}(n,m,0)\cdots\hat{W}(n,m,C-1) \; \hat{W}(n-1,m,0)\cdots\hat{W}(n-L+1,m,C-1)]$$

is the weight vector, λ is the forgetting factor, and ξ^* is the complex conjugate of the error.

3.1. Round Robin Regularization

The regularization process corresponds to equations (9) to (11). To regularize the correlation matrix $\mathbf{R}$, we can either add a small value to each term on the diagonal or set the terms along the diagonal to some threshold value if they become smaller than the threshold. β_{RLS} can be a small constant (e.g. 5000 with 16-bit input data), or β_{RLS} can be chosen using other non-uniform techniques.

By regularizing each subband in a round robin scheme, the CPU consumption involved with the two matrix inverses can be minimized. A round robin scheme is where a single band or several bands are regularized per frame or one band is regularized every several frames. Depending on the actual round robin scheme, the counter, $RoundRobinCount$, which points to the current band to be regularized is updated to point to the next band or group of bands to be regularized and reset back to the first band if necessary. Selecting a value for $MaxRoundRobinCount$ depends upon the value of the exponential forgetting factor. The value of λ is usually chosen as a trade off between the convergence accuracy of the RLS solution and the tracking speed. If λ is very close to 1, the RLS algorithm will obtain very accurate tap weights and hence cancel most of the echo provided nothing moves in the near end room. However for large values of λ, if someone moves, then the RLS algorithm cannot track the changes in the acoustic environment quickly. The value of $MaxRoundRobinCount$ should be chosen such that the correlation matrix of the speaker signal is regularized often enough so that $\lambda^{MaxRoundRobinCount}$ does not reach too small a value. Such a small value would produce a correlation matrix that is almost singular.

4. PERFORMANCE ANALYSIS

This section presents an analysis of the algorithmic and computational performance of the new RRLS algorithm. For the following experiments, stereo playback wave files were played on a PC and captured in real-time in a standard 10' x 10' x 8' office. In order to isolate the performance of the RRLS algorithm, both the capture and playback sampling rates are 16 kHz. The AEC algorithm can be extended to higher playback sampling rates using frequency domain interpolation [9]. In this paper, we use the MCLT to implement the adaptive subband stereo AEC algorithm [7] in order to minimize to end-to-end latency when combined with a modified G.722.1 encoder. As mentioned in the introduction, stereo AEC with a mono source located in the far end room leads to a non-unique solution. RRLS does not solve this problem, but by varying the regularization constant, the amount of residual echo due to a change in the far end room can be minimized. To investigate the behavior of the RRLS algorithm, we played a mono speech file recorded from one person with a gain of 0.9220 on one channel and 0.3873 on the second channel. At sample 2.4e5 which occurs 15 seconds after the first person started speaking, we switch the mono speech file to a second speaker and reverse the gains on each channel. This setup could represent a simple video conference call with mono speech transmission and constant power panning to simulate the location of different speakers. In figure 2, we show how choosing the regularization constant, β_{RLS}, affects the residual of the RRLS algorithm. The plots show that by increasing the value of β_{RLS}, we can significantly decrease the amount distortion in the envelope of the processed near end speech. Without regularization, we show in figure 3 that the algorithm quickly becomes unstable and diverges for the experiment as expected. Figure 4 compares the echo return loss

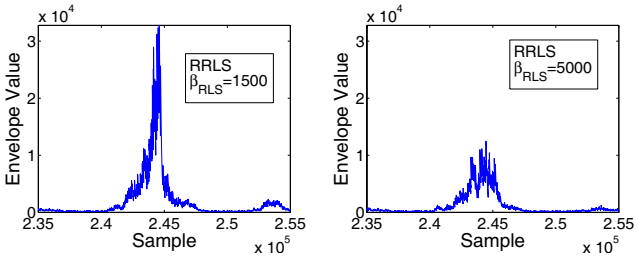

Fig. 2. Signal Envelope for Panning Change with $\beta_{RLS} = 1500$ in the left figure and $\beta_{RLS} = 5000$ in the right figure.

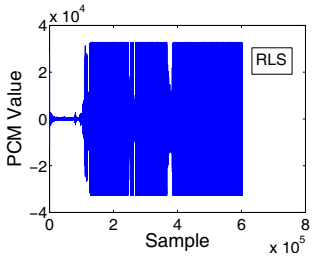

Fig. 3. Microphone Output with No Regularization.

enhancement (ERLE) for RRLS, RLS and NLMS where the ERLE is defined in [9]. The wave file played by the speakers is a light jazz track. As shown in figure 4, both the RRLS and standard RLS perform much better than NLMS. To better understand how RRLS and RLS compare, we plot the ERLE difference between RRLS and RLS algorithm in figure 5 for the two plots shown in figure 4. During the initial convergence, the standard RLS performs better than the new RRLS. During this portion of the music, a solo percussion track is panned from the right speaker to the left and the standard RLS is able to adapt more quickly. However, figure 5 shows that after convergence, the RRLS usually has better ERLE than the standard RLS algorithm by up to 6 dB depending on the correlation between the two playback channels.

Finally, we investigate the real-time performance of the new

RRLS algorithm. The accuracy results, above, reflected RRLS processing for 280 bands of the 320 bands produced by the MCLT for 16 kHz data using 20 msec frames. The lower 72 bands are processed using adaptive filters with 7 taps providing 140 msec of cancellation. The upper 208 bands are processed with 4 taps. The algorithm consumes about 22.1% of a 2.4 GHz Intel Pentium 4. To achieve the result, the core RLS algorithm is implemented using single precision floating point SSE instructions. The inverse functions are implemented in standard C using double precision arithmetic. Furthermore, the CPU consumption drops to 13.4% of the same processor when the upper 208 bands are processed using NLMS instead of RRLS [4]. Historically, algorithmic complexities are often

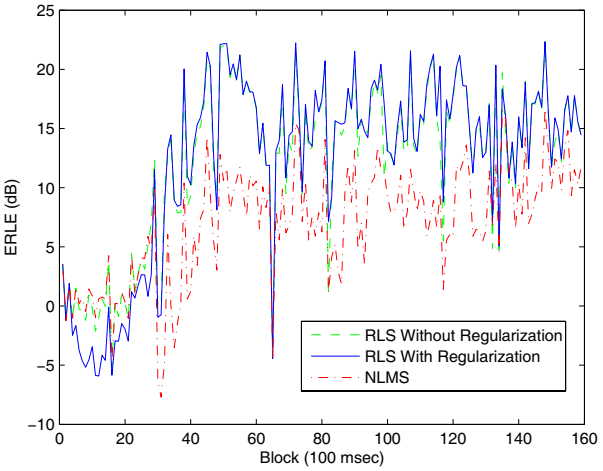

Fig. 4. Comparison of ERLE for AEC processing with RRLS, RLS, and NLMS.

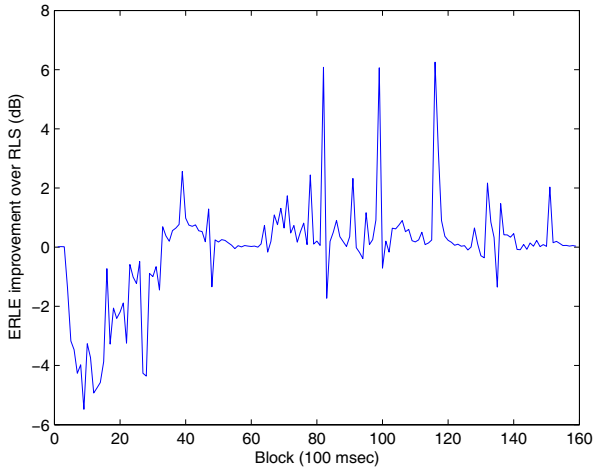

Fig. 5. ERLE improvement for AEC processing with RRLS compared to RLS.

measured by the number of multiplies. However, modern general purpose CPUs with SIMD instruction sets as well as current DSPs are much more complex than the older DSPs originally used to run signal processing algorithms. These modern processors use on-chip cache and virtual memory managers to significantly improve performance. The authors have found that signal processing algorithms implemented on general purpose CPUs are often cache bound instead of CPU bound. In addition, algorithms implemented using SIMD instruction sets perform much better for simpler algorithms since updating large numbers for pointers and reloading internal variables significantly reduce the efficiencies of SIMD implementations. The RLS algorithm falls into this category of simple algorithms and is well suited to SIMD implementations. Using the SIMD instructions on an 800 MHz Pentium III to implement just the four core RLS equations, we noted a speed up of approximately 3x for the entire adaptive subband, stereo AEC system as compared to implementing the four equations in C. With the recent improvements in CPU clock rates along with the addition of SIMD instruction sets, many complex algorithms previously thought to be too slow for real world applications are now practical.

5. CONCLUSIONS

In this paper, we have presented a new robust RRLS algorithm with round robin regularization and applied it to the stereo AEC problem. This new algorithm has been implemented in real-time on a standard PC and exhibits good performance for cancelling the echo from stereo playback signals without adding distortion from non-linear transforms. By regularizing the inverse of the correlation matrix, we can minimize the distortion in the near end speech. In addition, the new RRLS algorithm can also be combined with adding non-linear transforms [3] to the playback signal to make additional trade-offs between distortion in the stereo playback signal and the captured near end speech.

6. REFERENCES

[1] D. Slock, and T. Kailath, "Numerically Stable Fast Transversal Filters for Recursive Least Squares Adaptive Filtering", IEEE Trans. Signal Processing, vol. 39, no. 1, pp. 92-114, Jan 1991.

[2] J. Benesty, et. al., "Adaptive Filtering Algorithms for Stereophonic Acoustic Echo Cancellation", Proc. ICASSP'95, vol. 5, 9-12 May, pp.3099-3102, 1995.

[3] J. Benesty, D. Morgan, M. Sondhi, "A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation", Proc. ICASSP'97, pp. 303-306, 1997.

[4] P. Eneroth, T. Gansler, S. Gay, J. Benesty, "Studies of a Wideband Stereophonic Acoustic Echo Canceler", Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on, pp. 207-210, 17-20 Oct 1999.

[5] B. Hätty, "Recursive Least Squares Algorithms using Multirate Systems for Cancellation of Acoutical Echos", Proc. ISCAS'90, pp. 1145-1148, 1990.

[6] S. Gay, "Dynamically regularized fast RLS with application to echo cancellation", Proc. ICASSP'96, pp. 957-960, 1996.

[7] H. Malvar, "A Modulated Complex Lapped Transform and Its Applications to Audio Processing", Proc. ICASSP'99, pp. 1421-1424, Mar 1999.

[8] S. Haykin, "Adaptive Filter Theory", Prentice Hall, Upper Saddle River, NJ, 3th ed., 1995.

[9] J. Stokes, and H. Malvar, "Acoustic Echo Cancellation With Arbitrary Playback Sampling Rate", Proc. ICASSP'04, vol. IV, pp. 153-157, May 2004.