

01 Jan 2000

A Robust Proportionate Affine Projection Algorithm for Network Echo Cancellation

T. Gansler

J. Benesty

Steven L. Grant

Missouri University of Science and Technology, sgrant@mst.edu

M. M. Sondhi

Follow this and additional works at: https://scholarsmine.mst.edu/ele_comeng_facwork



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

T. Gansler et al., "A Robust Proportionate Affine Projection Algorithm for Network Echo Cancellation," *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000. ICASSP '00*, Institute of Electrical and Electronics Engineers (IEEE), Jan 2000.

The definitive version is available at <https://doi.org/10.1109/ICASSP.2000.859079>

This Article - Conference proceedings is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Electrical and Computer Engineering Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

A ROBUST PROPORTIONATE AFFINE PROJECTION ALGORITHM FOR NETWORK ECHO CANCELLATION

Tomas Gänsler, Jacob Benesty, Steven L. Gay, and M. Mohan Sondhi

Bell Laboratories, Lucent Technologies
600 Mountain Avenue
Murray Hill, New Jersey 07974-0636
{gaensler, jbenesty, slg, mms}@bell-labs.com

ABSTRACT

Echo cancelers which cover longer impulse responses (≥ 64 ms) are desirable. Long responses create a need for more rapidly converging algorithms in order to meet the specifications for network echo cancelers devised by ITU (International Telecommunication Union). In general, faster convergence implies a higher sensitivity to near-end disturbances, especially “double-talk.” Recently, a fast converging algorithm called Proportionate NLMS (Normalized Least Mean Squares) algorithm (PNLMS) has been proposed. This algorithm exploits the sparseness of the echo path in order to increase the convergence rate. A robust version of PNLMS has also been presented which combines a double-talk detector with techniques from robust statistics to make the algorithm insensitive to double-talk. This paper presents a generalization of the robust PNLMS algorithm to a robust Proportionate Affine Projection Algorithm (APA) called PAPA that converges very fast.

1. INTRODUCTION

There is a need for network echo cancelers (ECs) 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 NLMS inadequate. It will therefore, be desirable to implement fast-converging algorithms in future echo cancelers.

In [2, 3], faster converging algorithms called Proportionate NLMS (PNLMS) and PNLMS++ respectively are proposed. 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 non-zero active coefficients. This is the idea behind the PNLMS algorithm which is a modification of the NLMS algorithm.

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. In PNLMS, an adaptive individual step-size is assigned to each filter coefficient.

The step-sizes are proportional to the magnitude of the latest 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 in the echo path are adjusted faster than non-active coefficients (i.e. small or zero coefficients).

Besides convergence rate, an important aspect of an echo canceler is its performance during “double-talk” (i.e. simultaneous far- and

near-end speech). In general, 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 so that it does not attenuate the echo sufficiently. To inhibit the divergence of the EC 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 undetected large amplitude samples perturb the echo path estimate considerably. The key to the solution presented in this paper is a combination of a DTD with traditional robust statistics and a *delicately tuned scale variable*, s [5, 6].

This work combines the ideas of the proportionate step-size technique and robust statistics with the Affine Projection Algorithm (APA) in order to achieve fast convergence for a wide range of echo paths while maintaining slow divergence when the canceler is exposed to double-talk.

2. ECHO CANCELER ALGORITHMS

In derivations and descriptions the following notations are used (Also see Fig. 1).

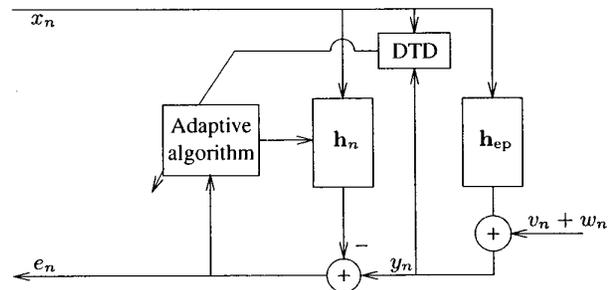


Figure 1: Block diagram of the echo canceler and double-talk detector.

The excitation vector is denoted $\mathbf{x}_n = [x_n, \dots, x_{n-L+1}]^T$ where x_n is the far-end speech signal. v_n is the background noise and w_n is the near-end speech (double-talk). The near-end signal, i.e. echo and noise, possibly including near-end speech, is denoted y_n . The residual echo is $e_n = y_n - \mathbf{h}_n^T \mathbf{x}_n$ where $\mathbf{h}_n =$

$[h_{0,n}, \dots, h_{L-1,n}]^T$ is the estimated echo path. Furthermore, let $\mathbf{y}_n = [y_n \dots y_{n-p+1}]^T$, be a vector of samples y_n and $\mathbf{X}_n = [\mathbf{x}_n, \dots, \mathbf{x}_{n-p+1}]$ the excitation matrix, where p is the projection order. $\mathbf{e}_n = \mathbf{y}_n - \mathbf{X}_n^T \mathbf{h}_n$ denotes the residual echo vector. L is the length of the adaptive filter.

2.1. Double-talk Detection (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 magnitudes 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

$$|y_n| \geq \vartheta \max\{|x_n|, |x_{n-1}|, \dots, |x_{n-L+1}|\}. \quad (1)$$

The detector threshold, ϑ , 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_{hold} , 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 disturbances in the residual echo that is used to update the filter coefficients. Thus divergence of the adjusted filter coefficients occurs, which in turn give rise to large amounts of uncanceled echo. Figure 3c shows an example of these disturbances resulting from double-talk detection errors. This disturbance, which in practice cannot be measured, is made by gating the DTD's decision with the pure double-talk sequence. 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 appropriate for the residual error in the absence of a Geigel DTD. Figure 3b, because without the DTD the residual consists of longer-lasting bursts of near-end speech. Hence, the DTD is an essential component in the robust PAPA algorithm to be described in the next section.

2.2. The Robust Proportionate Affine Projection Algorithm

A robust proportionate affine projection based algorithm (PAPA) is defined by,

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \mu \mathbf{G}_n \mathbf{X}_n \mathbf{R}_{x,n}^{-1} \psi(\mathbf{e}_n). \quad (2)$$

$$\mathbf{G}_n = \text{diag}\{g_{0,n}, \dots, g_{L-1,n}\} \quad (3)$$

where $\mathbf{R}_{x,n}^{-1} = (\mathbf{X}_n^T \mathbf{G}_n \mathbf{X}_n + \delta \mathbf{I})^{-1}$ is a weighted estimate of the inverse correlation matrix of the input signal. δ is a regularization parameter which prevents division by zero and stabilizes the solution for low level input signals when speech is used. $\psi(\mathbf{e}_n) = [\min\{\frac{|e_n|}{s_{n-1}}, k_0\} \odot \text{sign}(\mathbf{e}_n)]_{s_{n-1}}$ where \odot denotes elementwise multiplications and $|\cdot|$ in $|e_n|$ operates on the individual elements. s_{n-1} is an estimated scale factor and k_0 controls the robustness of the algorithm.

With $\mathbf{G}_n = \mathbf{I}$, $\delta = 0$, and $k_0 = \infty$, equation (2) reduces to the standard Affine Projection Algorithm (APA), first introduced in [7]. The regularization parameter in APA was proposed in [8]. As evident, robust PAPA is obtained by combining APA with the proportionate step-size matrix of PNLMS, [2, 3], and a scaled non-linearity ψ . The reasons for introducing these two components are discussed below.

Convergence rate. The purpose of \mathbf{G}_n is to *increase convergence rate* of active coefficients. While μ is the overall step-size parameter, the matrix \mathbf{G}_n is a diagonal matrix which adjusts the step-sizes of the individual taps of the filter. A large value of the element $g_{l,n}$ means that filter coefficient $h_{l,n}$ receives a large step-size, thus increasing its convergence rate. A practical way to calculate the diagonal elements of \mathbf{G}_n is as follows, [2],

$$\begin{aligned} \gamma_{l,n+1} &= \max\{\rho \max\{\delta_p, |h_{0,n}|, \dots, |h_{L-1,n}|\}, |h_{l,n}|\} \\ g_{l,n+1} &= \gamma_{l,n+1} / \sum_{i=0}^{L-1} \gamma_{i,n+1}, \quad 0 \leq l \leq L-1. \end{aligned} \quad (4)$$

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

A variant of the algorithm is to use the step-size matrix \mathbf{G}_n above for odd-numbered time steps, while for even-numbered steps it is chosen to be the identity matrix ($\mathbf{G}_n = \mathbf{I}$) which results in an APA iteration. The alternation between APA and PAPA iterations has some advantages compared to using just the proportionate technique, e.g. it makes the proportionate algorithm less sensitive to the assumption of a sparse impulse response without sacrificing performance. This idea was first proposed for PNLMS in [3].

We can omit the matrix \mathbf{G}_n in the definition of $\mathbf{R}_{x,n}$ to save computations. Inclusion of the matrix requires a significant $p^2 L$ multiplications per sample, but according to our simulations, the effect on performance and stability is minimal. Additionally most of the computational procedures of the Fast Affine Projection (FAP) algorithm, [9], can be incorporated in order to reduce the computational complexity of PAPA.

Robustness. The purpose of $\psi(\cdot)$ is to *decrease divergence rate* of all coefficients when the returned echo contains near-end speech bursts in spite of the DTD. 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 squared error. Each iteration updates the current estimate of \mathbf{h}_n by $\mu \mathbf{x}_n e_n$, 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 μ is normalized, resulting in the NLMS algorithm. Furthermore, NLMS is a special case of the APA with $p = 1$. It is well known, [10], that other gradient algorithms can be derived by changing the cost function to

$$J = E\left\{\varrho\left(\frac{|e_n|}{s}\right)\right\} \quad (5)$$

where $\varrho(\cdot)$ is any even symmetric function with a monotonically non-decreasing derivative (with respect to its argument) and s is its scale factor. Using a stochastic approximation of the gradient and normalizing the step-size in the same manner as in standard NLMS we get the robust NLMS algorithm,

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \frac{\mu \mathbf{x}_n}{\mathbf{x}_n^T \mathbf{x}_n + \delta} \psi\left(\frac{|e_n|}{s}\right) \text{sign}\{e_n\} s \quad (6)$$

where $\psi(\cdot)$ is the derivative of $\varrho(\cdot)$. The non-linearity is chosen as the limiter, [11],

$$\psi\left(\frac{|e_n|}{s}\right) = \min\left\{\frac{|e_n|}{s}, k_0\right\}. \quad (7)$$

Details of the derivation of (6) can be found in [6]. Generalizing this idea to the affine projection algorithm combined with the proportionate step-size matrix we obtain the robust PAPA, (2).

2.3. 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,

$$s_n = \lambda s_{n-1} + \frac{1-\lambda}{\beta} s_{n-1} \psi\left(\frac{|e_n|}{s_{n-1}}\right), \quad (8)$$

where $s_{-1} = \sigma_x$. The choice of this method of estimating s is justified in [6]. 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 $1/(1-\lambda)$. 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 PAPA 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 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 system change, the algorithm will not track immediately. However, as the scale estimator tracks the larger error signal the nonlinearity is scaled up and convergence rate accelerates. The trade off between robustness and tracking rate of the adaptive algorithm is thus governed by the tracking rate of the scale estimator which is controlled by one single parameter, λ . 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_{hold} .

2.4. A "Fast" Robust PAPA for $p = 2$

A fast version of the algorithm is given by

$$r_{22,n} = r_{11,n-1} \quad (9)$$

$$r_{11,n} = r_{11,n-1} + x_n^2 - x_{n-L}^2 \quad (10)$$

$$r_{12,n} = r_{12,n-1} + x_n x_{n-1} - x_{n-L} x_{n-L-1} \quad (11)$$

$$e_n = y_n - \mathbf{x}_n^T \mathbf{h}_{n-1} \quad (12)$$

$$\psi(|e_n|) = \min\{|e_n|, s_n k_0\} \quad (13)$$

$$\psi(\mathbf{e}_n) = \begin{bmatrix} \text{sign}(e_n) \psi(|e_n|) \\ (1-\mu) \text{sign}(e_{n-1}) \psi(|e_{n-1}|) \end{bmatrix} \quad (14)$$

$$\mathbf{R}_{xx,n}^{-1} = \begin{bmatrix} r_{11,n} & r_{12,n} \\ r_{12,n} & r_{22,n} \end{bmatrix}^{-1} \quad (15)$$

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \mu \mathbf{G}_n \mathbf{X}_n \mathbf{R}_{xx,n}^{-1} \psi(\mathbf{e}_n) \quad (16)$$

if (no double-talk detected)

$$s_{n+1} = \lambda s_n + \frac{(1-\lambda)}{\beta} \psi(|e_n|) \quad (17)$$

else

$$s_{n+1} = \lambda s_n + (1-\lambda) s_{\min} \quad (18)$$

end

where s_{\min} is a preset constant. s_{\min} is the assumed minimum background power level of the noise. Inversion of the correlation matrix $\mathbf{R}_{xx,n}$ is preferably made with Gaussian elimination. Using this technique the robust APA requires $2L + 13$ multiplies and 3 divisions and robust PAPA requires $4L + 12$ multiplies and 3 divisions.

3. SIMULATIONS

The purpose of these simulations is to show the performance of robust PAPA in comparison to NLMS when speech is used as excitation signal. The order p of PAPA is chosen as 2 in order to compromise between complexity and performance.

The parameter settings chosen for the simulations are: $\mu = 0.2$, $L = 512$ (64 ms), $\delta = 2 \cdot 10^5$ (NLMS), $\delta = 1 \cdot 10^6$ (PAPA), $\delta_p = 0.01$, $\rho = 0.01$. $\sigma_x = 1900$, SNR = 39 dB (echo-to-noise ratio). Average far-end to double-talk ratio is 6 dB. The hybrid attenuation is 20 dB and the Geigel detector assumes 6 dB attenuation. Parameters for the robust algorithm are, $(\lambda, k_0) = (0.997, 1.1)$. This choice results in $\beta \approx 0.60665$, see [6]. $\mathbf{h}_{-1} = \mathbf{0}$. $s_{-1} = 1000$. The scale estimate in (8), s_n , is never allowed to become smaller than 2. This inhibits bad behavior in low noise situations. All algorithms are tuned to achieve the same minimum misalignment in order to fairly compare convergence rate. The misalignment is given by, $\varepsilon = \|\mathbf{h} - \mathbf{h}_{\text{ep}}\| / \|\mathbf{h}_{\text{ep}}\|$ where \mathbf{h}_{ep} is the true echo path. The impulse response and corresponding magnitude function of the hybrid is shown in Fig. 2a, b.

Convergence rate. Figure 4a shows the behavior after an abrupt echo path change where its impulse response is shifted 200 samples (25 ms) at 1 second. Initial and reconvergence rate of PAPA is considerably higher than for NLMS. The robust NLMS and more evidently PAPA, have very similar convergence rate to their non-robust counterparts.

Robustness. Figure 4b shows far-end signal, double-talk and the misalignment of the three algorithms. While the non-robust NLMS and PAPA (with Geigel detector!) diverges to a misalignment of +5 dB the robust versions are much less affected. The robust PAPA never perform worse than -10 dB misalignment during double-talk. The slow reconvergence of NLMS after the double-talk sequence is caused by poor excitation of the speech signal, i.e. this segment is highly correlated.

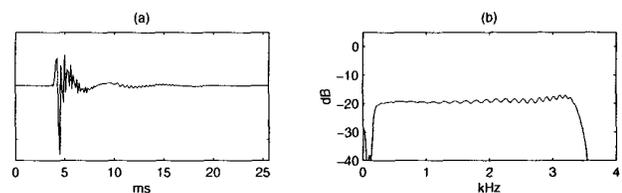


Figure 2: Impulse response (a) and magnitude (b) respectively of the frequency response of the hybrid in the simulation. Hybrid attenuation: 20 dB.

4. CONCLUSIONS

Increased convergence rate and robustness of the APA has been achieved by combining the algorithm with a proportionate step-size matrix, a scaled non-linearity and a DTD. The complexity of the PAPA ($p = 2$) is on the order of $4L$ multiplications per sample which is about twice that of NLMS. However, the convergence

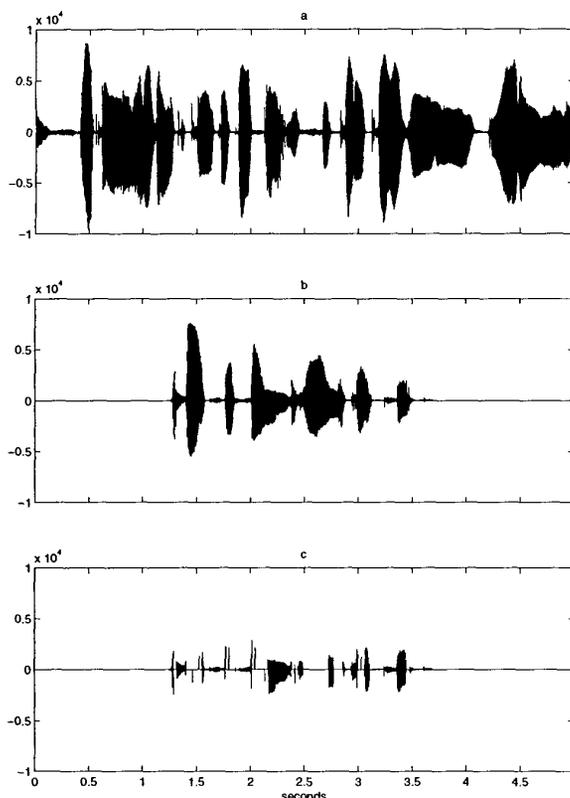


Figure 3: (a) Far-end speech. (b) Near-end speech, i.e. double-talk. (c) Near-end speech gated with the DTD's decision. These are the disturbances that actually enters the adaptive algorithm. Average far- to near-end ratio: 6 dB (1.125-3.625 s).

rate of PAPA is considerably higher. It is shown in this paper that the robust PAPA converges faster than NLMS and perform significantly better during double-talk.

What differentiates our approach for handling double-talk is that the robust technique uses the signal before double-talk in order to be prepared for it, whereas traditional methods try to detect periods of double-talk and then take action. 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. Furthermore, any DTD and adaptive algorithm can be enhanced with this technique resulting in a low complexity high performing echo canceler.

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 though they have not yet fully converged.

5. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Information and System Sciences Series. Prentice-Hall, 3 edition, 1995.

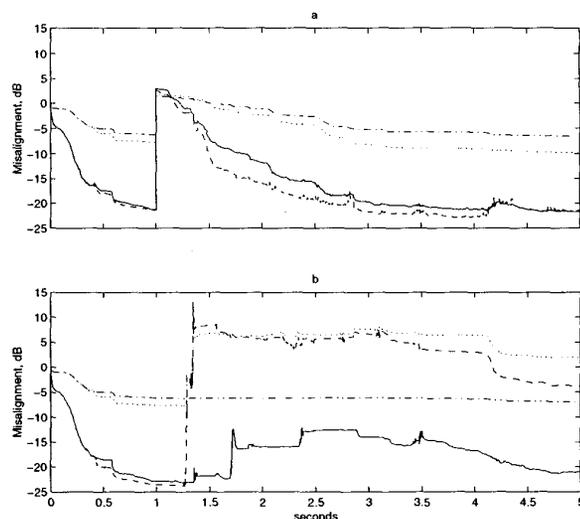


Figure 4: Misalignment for speech signal. (a) Convergence after echo path change at 1 s when the impulse response of the echo path is shifted 200 samples (25 ms). (b) Robustness to double-talk occurring between 1.125-3.625 s, Fig. 3b. The DTD assumes a 6 dB hybrid attenuation. Parameter settings are: $\lambda = 0.997$, $k_0 = 1.1$, hybrid attenuation: 20 dB. Solid line: Robust PAPA, Dashed line: PAPA, Dashed-dotted line: Robust NLMS, Dotted line: NLMS.

[2] D. L. Duttweiler, "Proportionate normalized least mean squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Processing*, Submitted.

[3] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," in *Proc. of Assilomar*, Nov. 1998.

[4] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647-653, May 1978.

[5] T. Gänslér, "A double-talk resistant subband echo canceler," *Signal Processing*, vol. 65, no. 1, pp. 89-101, Feb. 1998.

[6] T. Gänslér, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Processing*, Submitted.

[7] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electr. and Commun. in Japan*, vol. 67-A, no. 5, 1984.

[8] S. G. Kratzer and D. R. Morgan, "The partial-rank algorithm for adaptive beamforming," *Proc. SPIE Int. Soc. Optic. Eng.*, vol. 564, no. 5, pp. 9-14, 1985.

[9] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. IEEE ICASSP*, 1995, pp. 3023-3026.

[10] M. M. Sondhi, "An adaptive echo canceler," *The Bell Syst. Techn. J.*, vol. XLVI, no. 3, pp. 497-510, March 1967.

[11] P. J. Huber, *Robust Statistics*, pp. 68-71, 135-138, New York: Wiley, 1981.