ADAPTIVE FILTERING FOR NON-GAUSSIAN PROCESSES

Preben Kidmose

Department of Mathematical Modelling, Section for Digital Signal Processing,
Technical University of Denmark, Building 321, DK-2800 Lyngby, Denmark

ABSTRACT

A new stochastic gradient robust filtering method, based on a non-linear amplitude transformation, is proposed. The method requires no a priori knowledge of the characteristics of the input signals and it is insensitive to the signals distribution and to the stationarity of the signals.

A simulation study, applying both synthetic and real-world signals, shows that the proposed method has overall better robustness performance, in terms of modeling error, compared with state-of-the-art robust filtering methods. A remarkable property of the proposed method is that it can handle double-talk in the acoustical echo-cancellation problem.

1. INTRODUCTION

Many signals encountered in practice are decidedly non-Gaussian and the signals are only stationary up to an order less than 2. Recently [6] has investigated audio signals, and demonstrated the usefulness of the α-Stable distribution for modelling of noisy audio signals.

In this work we consider robust adaptive algorithms. The objective is to design adaptive algorithms that are insensitive to the probability distribution and stationarity of the input signals.

1.1. Stochastic Gradient Algorithms

Consider an adaptive FIR filter with input u_n. The estimation error, at time n, of the filter is

$$e_n = d_n - \mathbf{w}^T \mathbf{u}_n \tag{1}$$

where $\mathbf{u}_n$ is the input signal vector, $\mathbf{w}$ is the filter coefficient vector, and d_n is the desired signal. Assume that d_n and u_n are jointly stable processes of order α. The objective is to minimize the dispersion of the error, $\min ||d_n - e_n||_\alpha$. This objective turns out to be intractable, but fortunately proportional to the pth order moment for any $0 < p < \alpha$, [7]. So, an equivalent objective is $J = \mathrm{E}\left\{|d_n - e_n|^p\right\}$. There is no closed-form solution for the set of coefficients that minimize this objective; however the objective is convex for $1 \le p \le \alpha$, and so we may use a stochastic gradient method to solve for the coefficients. The class of stochastic gradient algorithms under consideration has the update

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu h(e_n \mathbf{u}_n) \tag{2}$$

where $h(\cdot)$ is an estimate of the gradient $\hat{\mathrm{E}}\left\{\partial J / \partial \mathbf{w}\right\}$, based on the estimation error and the input.

1.2. Adaptive Filters Based on FLOM

A number of different robust stochastic gradient algorithms, based on a fractional lower order moments (FLOM), has been proposed [1][2][3][5]. One of these is the Symmetric Least Mean P-norm (SLMP), that uses the symmetric norm

$$h(e\mathbf{u}) = (e\mathbf{u})^{\langle p/2 \rangle}$$

as estimate of the gradient, [3]. Another subclass of normalized stochastic gradient algorithms, that has the form of Eq. 2, is proposed in [5]. The proposed filter update uses

$$[h_q(\mathbf{u}_n)]_i = \frac{|u_{i,n}|^{q-1}\mathrm{sign}(u_{i,n})}{\sum_{m=1}^{L}|u_{i,n}|^q} \quad \text{for } 1 \le q < \infty \tag{3}$$

where $[h_q(\mathbf{u})]_i$ denotes the ith element of the vector valued function $h_q(\mathbf{u})$. This update, for $\mu = 1$ and for a valid q, corresponds to minimizing $\|\mathbf{w}_{n+1} - \mathbf{w}_n\|_p$ subject to $d_n - \mathbf{w}_{n+1}^T \mathbf{u}_n = 0$. The relation between the norm, q, of the input signal vector, $\mathbf{u}$, and the q-norm, of the coefficient vector update, $\|\mathbf{w}_{n+1} - \mathbf{w}_n\|_p$, is $1/p + 1/q = 1$, [5]. Thus, the update in Eq. 3 provides the minimum p-norm of the coefficient vector update. For $q = 2$ the algorithm is the classical NLMS algorithm.

Recently [1] proposed a generalization to the NLMP with the update

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{e_n^{\langle a \rangle}}{\|\mathbf{u}_n\|_{qa}^{qa} + \lambda}\mathbf{u}_n^{\langle (q-1)a \rangle} \tag{4}$$

where, in the presence of an α-stable process, $0 < a \le \alpha - 1$ and $1 \le q$. For any real number v and $p \ge 0$, the convention $v^{\langle p \rangle} = |v|^p \mathrm{sign}(v)$ is used. The update is motivated by observing that it corresponds to a gradient descent adaptation approach to the objective function $J = \mathrm{E}\left\{|e_n|^{a+1}\right\}$. The update in Eq. 4 reduces to the NLMP update if a and q are chosen as $p-1$ and p/a. The update in equation 3 is a special case of Eq. 4 for $a = 1$ and $\lambda = 0$

1.3. Median Orthogonality

If the median of the product, $M(u_1 u_2)$, of two random variables u_1 and u_2, is zero, then u_1 and u_2 are said to be median orthogonal, $u_1 \perp_{\mathrm{MO}} u_2$. For random variables with symmetric probability densities, independence is necessary and sufficient for MO, however MO is necessary but not sufficient for independence. The MO filter criterion, proposed in [2], is that the error should be MO to all elements of the input vector $e \perp_{\mathrm{MO}} \mathbf{u}$. This criterion extends the conventional orthogonality criterion without restricting the distribution of e and $\mathbf{u}$.

2. A NEW ROBUST FILTERING METHOD

In this section we propose a conceptual different approach to robust adaptive algorithms, that are based on a non-linear amplitude transformation. The idea of the amplitude transformation is to ensure the existence of the second order moment.

2.1. Density Transformation

The purpose of the non-linear amplitude transformation, is to transform the probability density of a signal. The proposed transformation, from an arbitrary unknown density to a desired density, is provided by a three-step procedure

i Use the empirical density transformation, to obtain an uniform density.

ii Transform with the inverse function to the indefinite integral of the desired density.

iii Scale the transformed variable.
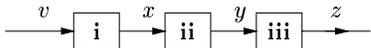
The scaling is necessary since all scale information is lost in the empirical density transformation. The objective of the scaling is to obtain $\|v\|_p = \|z\|_p$, for a sufficient low p.

2.2. Empirical Transformation to Uniform Density

Let $\tilde{v}$ be a random variable, and let $v_1, v_2, \ldots, v_N$ be N observations of the variable. An empirical transformation from the unknown arbitrary probability density random variable, $\tilde{v}$, to the variable $\tilde{x}$ with uniform density, can be done by sorting the observations. Sort the N observations, by the index m, such that $v_{.,1} \leq v_{.,2} \leq \cdots \leq v_{n,m} \leq \cdots \leq v_{.,N}$. The variable $\tilde{x}$ is forced to follow the uniform density

$$p(x) = \begin{cases} \frac{1}{2}r & \text{if} \quad |x-a| \leq r \\ 0 & \text{if} \quad |x-a| > r \end{cases}$$

by assigning x_n the value

$$x_n = (2m - Nr)/N + a$$

where m is the index that corresponds to the nth observation of v.

The transformation is non-parametric and can be interpreted as the inverse cumulated histogram. The transformation, independent of the observations, result in an variable with uniform density. For variables with median equal to zero the transformation is sign conserving, and for symmetric densities the transformation is an odd function. Dependencies between variables are conserved.

2.3. Deterministic Density Transformation

The probability density of a random variable $\tilde{x}$ is denoted $p(x)$. The relation between the probability density for the random variable $\tilde{x}$ and $\tilde{y}$, is determined by the fundamental law of probability, $p(y)dy = p(x)dx$. Consider a transformation of the random variable $\tilde{x}$, following the probability density $p(x)$, to a new variable $\tilde{y}$, following the probability density $p(y)$. Suppose that $\tilde{x}$ follow a uniform distribution, and let $g(y)$ be a positive function with integral equal 1.

The desired transformation, that transform a uniform distributed variable $\tilde{x}$ into a variable $\tilde{y}$, that follows the density $g(y)$, is $y = G^{-1}(x)$, where $G^{-1}(x)$ is the inverse function to the indefinite integral of $g(y)$. The transformation from a uniform distributed variable $\tilde{x}$ into a variable $\tilde{y}$ that follow the probability density $p(y) = g(y)$, is depicted in Fig. 1. Thus, transformation between an arbitrary continuous
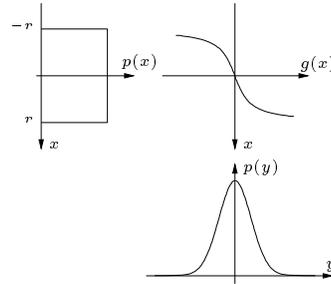

Figure 1: Density transformation from an uniform probability density, $p(x)$, to a probability density, $p(y)$, by the function $g(x)$.

probability density to or from an uniform probability density, is given by the probability distribution function or the inverse probability distribution function respectively.

2.4. Stochastic Gradient Adaptive Filter Based on non-linear Transformation

The idea of the non-linear amplitude transformation, is to make a density transformation, that ensures the existence of the second order moment. We propose to use a transformation, $g(\cdot)$, that transform an arbitrary density into a Normal distribution. If the estimation error, e_n, and the input signal vector, $\mathbf{u}_n$, is transformed into a Normal distributions by the function $g_e(\cdot)$ and $g_\mathbf{u}(\cdot)$, then the objective $\mathrm{E}\left(|g(d_n - e_n)|^2\right)$ makes sense.

The structure of the adaptive algorithm is depicted in Fig. 2. The proposed amplitude transformation is sign con-
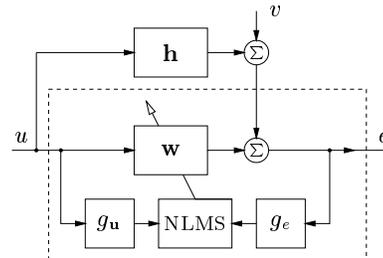

Figure 2: Adaptive filter based on a non-linear transformation of the estimation error.

serving, and that ensures that $g_\mathbf{u}(\mathbf{u}) g_e(e)$ is an ascending direction for the objective function. A straightforward update equation, on the form of Eq. 2, is

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu g_\mathbf{u}(\mathbf{u}_n) g_e(e_n) \tag{5}$$

The median orthogonality (MO) criterion proposed in [2] apply to this update. The algorithm will have a stability point where $e \perp_{MO} \mathbf{u}$ because the integral of the even density function and the odd $g(\cdot)$ is zero.

3. SIMULATIONS

The setup used for evaluation is depicted in Fig. 2, where $\mathbf{h}$ is an 10th order low pass filter. The proposed algorithm is compared to 3 other algorithms, that represents state of the art in robust adaptive algorithms:

$$\text{NSLMP} \qquad \mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{1}{\|\mathbf{u}\|_p^p + a}(e\mathbf{u})^{\langle p/2 \rangle}$$

$$\text{NLMP} \qquad \mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{1}{\|\mathbf{u}\|_p^p + a}e^{\langle p-1 \rangle}\mathbf{u}$$

$$\text{Aydíns method} \qquad \mathbf{w}_{n+1} = \mathbf{w}_n - \mu \frac{\mathbf{u}^{\langle (q-1)a \rangle}}{\|\mathbf{u}\|_{qa}^{qa} + \lambda}e^{\langle a \rangle}$$

The performance of the algorithms are evaluated as the modeling error $10 \cdot \log_{10}\left((\mathbf{w}-\mathbf{h})^T(\mathbf{w}-\mathbf{h})/(\mathbf{h}^T\mathbf{h})\right)$.

3.1. Simulation with synthetic signals

The synthetic signals are modeled as sequences of independent symmetric α-stable distributed random variables. The characteristic function for the random variable is given by $\phi(t) = \exp(-\gamma|t|^\alpha)$. The simulation of the signals are based on the method proposed in [4],[7].

The adaptation constants used in the simulation study, are listed in Table 1. Obviously these constants have considerably influence on the characteristics of the algorithms. The constants are chosen such that all the algorithms are robust and stable in the wide range of signals under consideration in this simulation study. Note that, in the ad-

Algorithm	Adaptation constants
Aydíns method	$\mu = .01$, $p = q = 1.1$, $a = \alpha_v - 1.1$
NSLMP	$\mu = .01$, $p = .9\alpha_v$, $a = 10^{-6}$
NLMP	$\mu = .01$, $p = .9\alpha_v$, $a = 10^{-6}$
Proposed method	$\mu = .05$, $N = 120$

Table 1: Adaptation constants used in the system identification simulation study with α-stable signals

vantage of the algorithms used for comparison, the norms used in these algorithms are adjusted in accordance with the applied signals.

Consider the case where the input signal, u, and the interference signal, v, are identical white Cauchy distributed signals. The convergence of the modeling error is depicted in Fig. 3. All the methods converge to about the same misadjustment level, but the proposed method has the lowest misadjustment. Aydíns method has slower convergence than the other algorithms.

For the case where the source signal, u, is a white Gaussian signal, and the interfering signal, v, is white following a Cauchy distribution, the convergence of the modeling error is depicted in Fig. 4. Two important changes are observed: the misadjustment level is about 20 dB higher and the convergence speed is slower. The γ-ratio is not a suitable measure between variables of different distributions, so even though the γ-ratio between the signals is still equal to one, the empirical SNR is very poor.
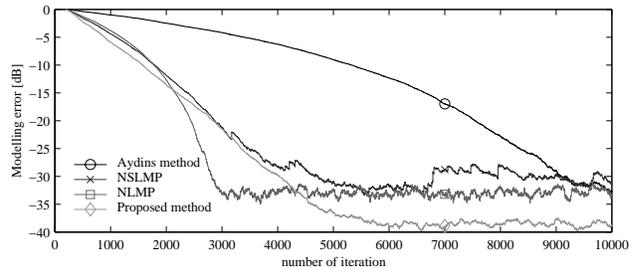

Figure 3: Modeling error convergence for $\alpha_u = \alpha_v = 1$ and $\gamma_u = \gamma_v = 1$. Mean over 10 Monte Carlo runs.
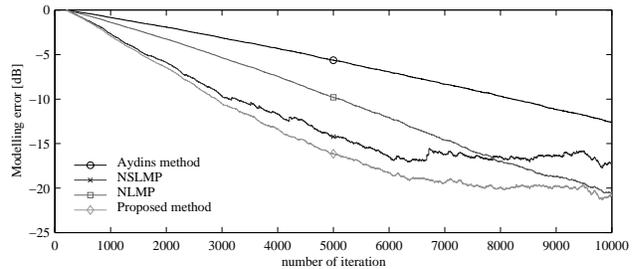

Figure 4: Modeling error convergence for $\alpha_u = 2$, $\alpha_v = 1$ and $\gamma_u = \gamma_v = 1$. Mean over 10 Monte Carlo runs.

In Fig. 5 the modeling error versus signal to interference dispersion ratio is depicted. The dispersion of the
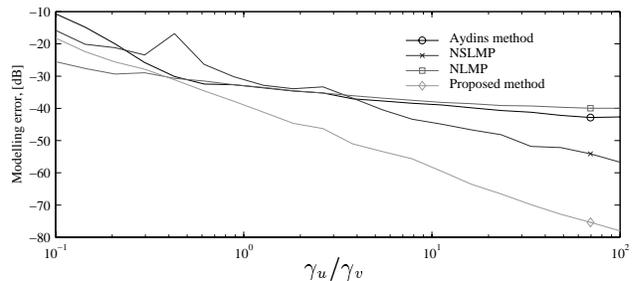

Figure 5: Modeling error versus signal to interference dispersion ratio, $\frac{\gamma_v}{\gamma_u}$. Characteristic exponent $\alpha_u = \alpha_v = 1$.

source signal is varied in the interval $\gamma_u = [0.1, 100]$, and for the interfering signal the dispersion is constant, $\gamma_v = 1$. For low dispersion ratios the performance of the algorithms are much the same; for increasing dispersion ratio the algorithms has increasing performance, and the proposed method takes increasing advantages compared to the other algorithms.

In Fig. 6 the modeling error versus the characteristic exponent is depicted. The methods used for comparison are almost independent of α, since in this simulation study the norm is adjusted in accordance with the applied signal. Even though the proposed method does not require any knowledge of the input signals distribution it outperforms all the algorithms used for comparison. This is a remarkably robustness characteristic.
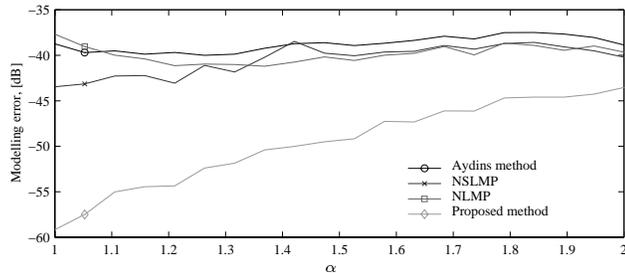
Figure 6: Modeling error versus the characteristic exponent, α. The dispersion is $\gamma_u = 1$ and $\gamma_v = 0.1$. Mean over 10 Monte Carlo runs.

3.2. Simulation with speech signals

Consider the adaptive algorithms used for acoustical echo canceling. The setup is as depicted in Fig. 2 where v is the local speaker and u is the remote speaker. The acoustical path, modeled by the filter $\mathbf{h}$, is the same 10th order filter as in the previous simulation study. The adaptation constants used in this example are listed in Table 2. For the NSLMP, NLMP and Aydins method the p-norm and μ are set to relative small values to obtain stability. The signals are speech

Algorithm	Adaptation constants
Aydíns method	$\mu = .01$, $p = q = 1.1$, $a = .1$
NSLMP	$\mu = .01$, $p = 1.44$, $a = 10^{-6}$
NLMP	$\mu = .01$, $p = 1.44$, $a = 10^{-6}$
Proposed method	$\mu = .1$, $N = 1000$

Table 2: Adaptation constants used in the acoustical echo cancellation for speech signals

signals sampled at 8kHz. The input signals and the modeling error result of the simulation is shown in Fig. 7. The
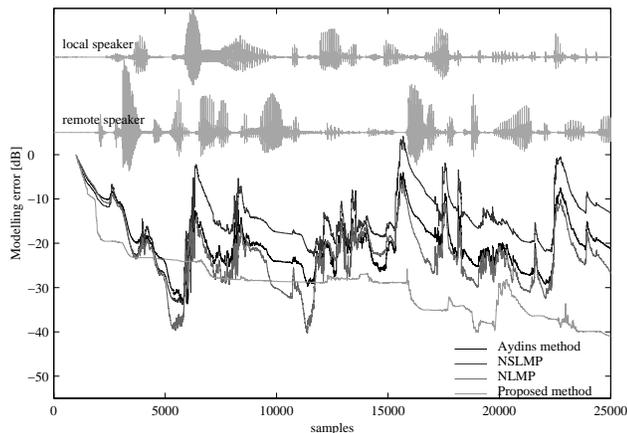

Figure 7: Modeling error, in acoustical echo-canceler in cross-talk environment, for the 4 methods.

simulation example contains several difficulties: the signals are non-stationary and contains temporal correlation, the amplitude density is non-Gaussian, and the dispersion ratio

between the signals are strongly varying. It appears, from Fig. 7, that the proposed method has slower convergence but apart from that, considerably better performance compared with the other methods. In the intervals where the local speaker is active, which corresponds to a poor SNR, the algorithms, except for the proposed method, loss the tracking of the filter. For these methods a control strategy is necessary. Contrary to the methods used for comparison, the proposed method has the ability to handle double-talk situations and environments with high background noise levels.

4. CONCLUSION

A new robust adaptive filtering method, based on a nonlinear amplitude transformation, is proposed. The method is compared with state-of-the-art robust filtering methods and a simulation study with synthetic and real-world signals is carried out.

The proposed method requires no a priori knowledge of the input signals, it is insensitive to the signals density and to the stationarity of the signals.

The proposed method has an overall better robustness performance, compared with state-of-the-art methods. The method is computational more expensive than the algorithms used for comparison, but these methods require either the use of a low p-norm or the estimation of a valid p, in order to ensure stability.

A remarkable byproduct is that the proposed amplitude transformation method can be used in a variety of situations for signal conditioning.

5. REFERENCES

[1] Gül Aydín, Orban Aríkan, and A. Enis Çetin. Robust Adaptive Filtering Algorithms for α-Stable Processes. *IEEE transactions on Circuits and Systems - II*, 46(2):198–202, February 1999.

[2] John S. Bodenschatz and Chrysostomos L. Nikias. Recursive Local Orthogonal Filtering. *IEEE Transactions on Signal Processing*, 45(9):2293–2300, 1997.

[3] John S. Bodenschatz and Chrysostomos L. Nikias. Symmetric Alpha-Stable Filter Theory. *IEEE Transactions on Signal Processing*, 45(9):2301–2306, 1997.

[4] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[5] Scott C. Douglas. A Family of Normalized LMS Algorithms. *IEEE Signal Processing Letters*, 1(3):49–51, March 1994.

[6] Panayiotis G. Georgiou, Panagiotis Tsakalides, and Chris Kyriakakis. Alpha-Stable Modeling of Noise and Robust Time-Delay Estimation in the Presence of Impulsive Noise. *IEEE Transactions on Multimedia*, 1(3):291–301, September 1999.

[7] Chrysostomos L. Nikias and Min Shao. *Signal Processing with Alpha-Stable Distribution and Applications*. Wiley, 1 edition, 1995.