APPLYING A NEURAL NETWORK FOR STEPSIZE CONTROL IN ECHO CANCELLATION

Christina Breining *

Fachgebiet Theorie der Signale, Darmstadt University of Technology, Germany

breining@nesi.e-technik.th-darmstadt.de

ABSTRACT

In this paper, we derive a new optimum time-variant stepsize for the adaptation of an echo cancellation filter. The optimum stepsize is then used in the form of cost functions to evaluate the performance of stepsize control methods. Finally, we show first results of combining a number of estimators for stepsize control by means of a neural network trained on this stepsize and a suitable cost function.

INTRODUCTION

One main problem in acoustic echo cancellation is the determination of a time-variant stepsize, given the measurable signals. This problem consists of two parts: first, an optimal stepsize must be determined theoretically. This stepsize depends on the misadjustment of the echo cancellation filter and on the background noise level. Since none of these values can be measured directly, several estimation methods [1, 3, 4, 6] have been proposed for that purpose. Those methods for stepsize control were derived on condition of assumptions which do not necessarily hold, so that the methods are not completely reliable. In order to obtain a reliable control algorithm, combinations of the methods should be used [1, 2]. These combinations have usually been found and optimized by trial-and-error on the basis of one specific application and adaptation system. Whenever characteristics of the system are changed, such as prewhitening filters, delay, sampling rate, or the characteristics of the environment in which the telephone set operates, optimization has to be carried out anew.

The adjustment of the parameters for the combination of the criteria is difficult and timely when done empirically. Their optimization may be crucial, though, for the comparison of different control methods. In section 1 of this paper, we therefore derive an optimum stepsize which can be used for automatic evaluation of a given stepsize control algorithm. Some cost functions based on this stepsize will be discussed in section 2 before they are used to train a neural network in order to optimize the combination of stepsize control criteria. Some results obtained with those networks are given in section 3, followed by a conclusion.

*The author is supported by the Graduiertenkolleg Intelligente Systeme in der Informations- und Automatisierungstechnik

1. THEORETICALLY OPTIMUM STEPSIZE

In this paper, we concentrate on the NLMS algorithm, since it is widely used and easy to implement. Its adaptation equation is

$$\underline{c}(k+1) = \underline{c}(k) + \alpha(k)\,\frac{e(k)\underline{x}(k)}{||\underline{x}(k)||^2} \tag{1}$$

with $\underline{c}(k)$ the echo cancellation filter vector, $\underline{x}(k)$ the incoming signal vector, $e(k)$ the distorted error, and $\alpha(k)$ the stepsize at time k.

The optimum stepsize at time k is defined here as the stepsize that minimizes the Euclidean distance $||\underline{g} - \underline{c}(k+1)||^2$, $\underline{g}$ representing the room impulse response. If we substitute $\underline{c}(k+1)$ as in equ. 1, and calculate the partial derivative in the direction of $\alpha(k)$, the condition reads

$$2E\left\{\frac{e(k)\varepsilon(k)}{||\underline{x}(k)||^2} - \frac{\alpha_{opt}(k)e(k)^2}{||\underline{x}(k)||^2}\right\} = 0 \tag{2}$$

$$\Rightarrow \alpha_{opt,1}(k) = \frac{E\left\{\frac{e(k)\varepsilon(k)}{||\underline{x}(k)||^2}\right\}}{E\left\{\frac{e(k)^2}{||\underline{x}(k)||^2}\right\}}, \tag{3}$$

with $\varepsilon(k)$ representing the adaptation error. In most applications, this equation is simplified by assuming the squared Euclidean norm of $\underline{x}(k)$ to be constant, and the joint moment of adaptation error and distortion to be zero [6], which leads to

$$\alpha_{opt,2}(k) = \frac{E\{\varepsilon(k)^2\}}{E\{e(k)^2\}}. \tag{4}$$

This stepsize is easily interpretable: the higher the background noise level is in relation to the adaptation error power, the lower is the stepsize, in the range of $[0,1]$. Thus, we can derive stepsize control mechanisms from noise estimation and double talk detection.

Unfortunately, the assumptions necessary for the calculation of $\alpha_{opt,2}(k)$ are not appropriate for speech signals. In order to examine how the simplification affects convergence speed, we carried out several simulations in which all signals, including the adaptation error, were known. Their statistic characteristics had to be estimated, though, since we applied real speech signals. The expected values were estimated by means of a first order all-pole filter:

$$\widehat{y^2}(k) = \lambda \widehat{y^2}(k-1) + (1-\lambda) y^2(k). \quad (5)$$

We used a FIR room impulse response filter with 2000 coefficients while the echo canceller had 1024. The simulations utilized different values of λ and were carried out with $\alpha_{opt,1}(k)$ and $\alpha_{opt,2}(k)$. Their results during double talk are displayed in fig. 1. For the first stepsize, convergence was achieved in any case, but its speed in double talk depended strongly on λ. With $\alpha_{opt,2}(k)$, the echo cancellation filter did not even converge for any λ in the case of double talk.

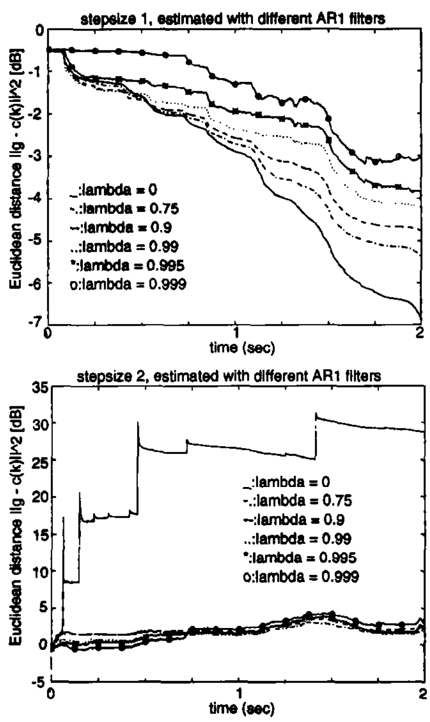

Figure 1: Euclidean distance $||\underline{g}-\underline{c}(k)||^2$ as a function of the all-pole filter coefficient, with speech signals in double talk

In the limit, convergence was fastest when instantaneous values were used for the estimation of the expected values in $\alpha_{opt,1}(k)$, i. e. $\lambda = 0$. In this case, the optimum stepsize reduces to

$$\alpha_{opt,3}(k) = \frac{\varepsilon(k)}{e(k)}. \quad (6)$$

With this stepsize, the distorted error in the adaptation equation of the NLMS algorithm is substituted by the adaptation error. Thus, adaptation speed is the same as for the undisturbed case, which leads to much better convergence in noisy environments. $\alpha_{opt,3}(k)$ can also be derived from equation 2 by setting the argument of the expected value to zero.

Comparison of the two stepsizes can be made by calculating the mean reduction of the misadjustment:

$$E\{\Delta(k)\} = E\{||\underline{g}(k) - \underline{c}(k)||^2 - ||\underline{g}(k) - \underline{c}(k+1)||^2\} \quad (7)$$

If we assume for both cases that $||\underline{x}(k)||^2$ is a good estimation for $N\sigma_x^2(k)$ and can be taken out of the expectation operator, and that the adaptation error $\varepsilon(k)$ is orthogonal to the background noise $n(k)$, equation 7 yields

$$E\{\Delta_1(k)\} = E\{\Delta_2(k)\} = \frac{E\{\varepsilon(k)^2\}}{E\{e(k)^2\}} \frac{E\{\varepsilon(k)^2\}}{E\{||\underline{x}(k)||^2\}} \quad (8)$$

$$\text{and} \qquad E\{\Delta_3(k)\} = \frac{E\{\varepsilon(k)^2\}}{E\{||\underline{x}(k)||^2\}} \quad (9)$$

with $\Delta_i(k)$ calculated by applying the stepsize $\alpha_i(k)$. Since noise and adaptation error are assumed to be orthogonal, the variance of the adaptation error is smaller than the distorted error variance, and thus, convergence speed for $\alpha_{opt,3}(k)$ is higher than for the other two stepsizes.

Finally, we chose the stepsize of equ. 6 as the function to be approximated by the neural net, but its range is not limited like for $\alpha_{opt,2}(k)$. Since the training data for the neural network usually do not represent the real-time data properly, the net may become over-adapted to outlyers. Therefore, we limited the optimum stepsize to the interval $[0,1]$ in order to limit sudden decreases in adjustment due to large stepsizes erroneously produced by the network.

In this case, the reduction of misalignment in equ. 7 is calculated as

$$\Delta_3(k) = \begin{cases} \frac{\varepsilon(k)^2}{||\underline{x}(k)||^2} & \alpha_{opt,3}(k) \in [0,1] \\ 0 & \alpha_{opt,3}(k) < 0 \\ \frac{\varepsilon(k)^2 - n(k)^2}{||\underline{x}(k)||^2} & \alpha_{opt,3}(k) > 1 \end{cases} \quad (10)$$

With the assumptions made above and a supplementary request of $\varepsilon(k)$ to be a zero mean process and the density functions of both $\varepsilon(k)$ and $n(k)$ to be symmetric, we can then calculate the mean reduction of misalignment as

$$E\{\Delta_3(k)\} = \frac{1}{N\sigma_x^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_\varepsilon(\varepsilon) f_n(n) \Delta_3(k) d\varepsilon dn$$

$$= \frac{\sigma_\varepsilon^2}{2N\sigma_x^2} + \frac{2}{N\sigma_x^2} \int_0^{\infty} \int_{-\varepsilon}^{0} f_\varepsilon(\varepsilon) f_n(n) (\varepsilon(k)^2 - n(k)^2) d\varepsilon dn$$

(11)

The remaining integral cannot be solved without defining the probability density functions $f_\varepsilon(\varepsilon)$ and $f_n(n)$ of the adaptation error and the background noise, respectively. Solution for Laplace distribution of both $n(k)$ and $\varepsilon(k)$ yields

$$E\{\Delta_3(k)\} = \frac{\sigma_\varepsilon^2}{2N\sigma_x^2}\left(2 - \frac{\sigma_n^2}{(\sigma_\varepsilon + \sigma_n)^2}\right), \qquad (12)$$

whereas with zero-mean uniform distribution, we get

$$E\{\Delta_3(k)\} = \frac{\sigma_\varepsilon^2}{2N\sigma_x^2} + \begin{cases} \frac{\sigma_\varepsilon^4}{4N\sigma_x^2\sigma_\varepsilon\sigma_n} & \sigma_\varepsilon^2 < \sigma_n^2 \\ \frac{\sigma_n^4 + 2\sigma_n\sigma_\varepsilon^3 - 2\sigma_n^3\sigma_\varepsilon}{4N\sigma_x^2\sigma_\varepsilon\sigma_n} & else \end{cases} \qquad (13)$$

Comparison of these expected values with the mean reduction of misalignment obtained when applying $\alpha_{opt,2}(k)$ (see equ. 9) shows that with both Laplace and uniformly distributed processes, better convergence is obtained with the stepsize $\alpha_{opt,3}(k)$ introduced here, even though all the assumptions made for the derivation of $\alpha_{opt,2}(k)$ hold for these distributions.

2. COST FUNCTIONS

Besides, a cost function is needed that is minimized during the training process of the neural network. The simplest cost function is the sum of the squared errors $\delta_\alpha(k) = \alpha_{opt,3}(k) - \widehat{\alpha}_{opt}(k)$ between the optimal stepsize and the stepsize produced by the network, on N_T training samples. But this cost function does not take into account that small stepsizes are especially sensitive to poor estimation.

In order to find a more suitable cost function, we calculate the misalignment $\Delta(k+1)$ as the misalignment obtained with the optimum stepsize and an additional positive term,

$$\Delta(k+1) = \Delta_{opt}(k) + \frac{\delta_\alpha(k)^2 e(k)^2}{||\underline{x}(k)||^2} \qquad (14)$$

The cost function is constructed by summing up all the weighted error terms.

$$C_1 = \sum_{k=0}^{N_T} \frac{\delta_\alpha(k)^2 e(k)^2}{||\underline{x}(k)||^2} \qquad (15)$$

These error terms may additionally be devided by the previous misalignment, so as to emphasize the cases when the error is already small. For the last part of the equation, this leads to

$$\sum_{k=0}^{N_T} \frac{\delta_\alpha(k)^2 e(k)^2}{\Delta(k)||\underline{x}(k)||^2} \approx \sum_{k=0}^{N_T} \frac{\delta_\alpha(k)^2}{\alpha_{opt,3}(k)^2 + a} \qquad (16)$$

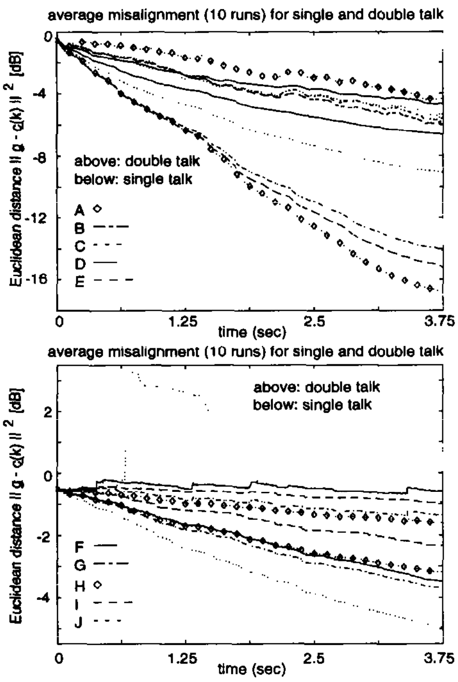

Figure 2: Adaptation with the delay coefficients method (above) and the correlation coefficient method (below) in various parametrizations

with $\Delta(k)||\underline{x}(k)||^2 \approx \varepsilon(k)^2$ and a a small number so as to avoid division by zero.

Since the sensitivity of the human ear is proportional to the logarithm of the signal power, one is likely to judge the echo relative to earlier echo signals. On the other hand, loss of convergence in the beginning may be very hard to compensate for later on, so that the absolute measure might be more efficient. Furthermore, the simplifications used to obtain the cost functions of equ. 16 might have negative effects. Therefore, we tested the cost functions with respect to their capability of ranking a set of stepsize control algorithms in a satisfying way. We controlled an adaptation sequentially by two algorithms, namely a simplified application of the correlation coefficient [3], and the delay coefficient method [6], with various sets of parameters. Their mean adaptation results (ensemble average over ten runs for both single talk and double talk) are shown in fig. 2. Some results for the cost functions of equ. 14 and 16 during the adaptation with the respective stepsize control methods are presented in table 2.

Although the performance of the criteria in single talk was comparable, only the cost function C_1 correctly classified the double talk cases controlled by the correlation coefficients as worse than the ones controlled by the delay method. We therefore chose C_1 for the neural network experiments.

	$\sum_k \frac{\Delta\alpha(k)^2}{\alpha_{opt,3}(k)^2}$		C_1		$C_2 = \frac{C_1}{\|\|\Delta\underline{c}(k)\|\|^2}$	
delay method	s	d	s	d	s	d
A	147.6	$7 \cdot 10^{11}$	0.14	28.0	0.48	46.8
B	7541	$8 \cdot 10^{8}$	0.38	23.1	1.32	48.0
C	19205	$6 \cdot 10^{7}$	1.41	18.6	3.65	38.7
D	24300	$4 \cdot 10^{8}$	2.89	15.8	6.08	30.3
E	4897	$2 \cdot 10^{10}$	0.33	21.6	1.12	46.5
correlation coefficient	s	d	s	d	s	d
F	23363	$3 \cdot 10^{7}$	18.1	65.2	27.6	67.6
G	25022	$3 \cdot 10^{8}$	15.3	47.9	23.6	57.1
H	27611	$6 \cdot 10^{6}$	15.2	36.9	22.4	46.5
I	27090	$5 \cdot 10^{6}$	25.4	48.4	34.4	55.4
J	22951	$1 \cdot 10^{7}$	11.1	382.7	19.1	85.8

Table 1: cost functions of equ. 14 and 16, averaged for 10 runs as shown in fig. 2, in single talk (s) and double talk (d)

All of the criteria ranked the algorithms with low maximum stepsize (i. e. C, D, and H) over the other ones, although their performance was worse. This means that rather smooth convergence is prefered by the cost functions. As far as the training of neural networks is concerned, this may even be favourable, since it helps to avoid instability problems.

3. APPLICATION IN A NEURAL NETWORK

Our theoretical results were applied for the training of a multilayer perceptron (MLP) type neural network with backpropagation [5] in the modified version derived from minimizing equation 16. It consisted of the input layer of 2 neurons, two hidden layers of 5 neurons each, and one output neuron. As input of the network, we used the parts that form the delay method, i. e. the estimated misalignment, and the mean squared error devided by the squared norm of the input vector. We trained the network with cost function C_1 on about ten times the amount of single talk samples as of double talk samples, so as to compensate for the emphasis of double talk by the cost functions. Training was carried out for roughly 600,000 samples, corresponding to 75 seconds realtime. The MLP was initialized randomly and trained during adaptation of the echo canceller with the optimal stepsize. First results are shown in fig. 3. It can be shown that the network is capable of adjusting the stepsize, but that fast convergence in the beginning is prefered.

4. CONCLUSION

In this contribution, we discussed several ways to determine the optimum stepsize for an echo cancellation algorithm, given the misalignment and the far-end and near-end signals in the hands-free telephone set. We presented an optimum stepsize that can be calculated without knowing the statistical properties

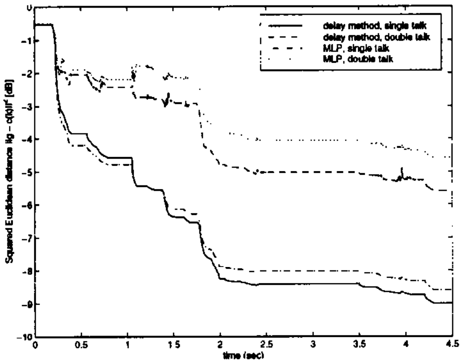

Figure 3: Adaptation with speech and stepsize controlled by the neural network in comparison to the artificial delay control method of [6]

of the signals, and that nevertheless performs better than the statistically motivated optimum stepsizes. We could evaluate different stepsize control algorithms automatically by calculating a corresponding cost function. This cost function and optimum stepsize were then used to train a multilayer perceptron which mapped the input signals onto a stepsize. Further studies have to be done to incorporate time delay and to select a representative database or set of training situations.

REFERENCES

[1] R. FRENZEL: *Freisprechen in gestörter Umgebung*, VDI-Fortschritt-Berichte, Reihe 10, Nr. 228, Düsseldorf, Germany, 1992

[2] C. BREINING: *Control of a Hands-free Telephone Set*, Signal Processing, vol. 61, no. 1, Elsevier, Holland, 1997

[3] P. HEITKÄMPER: *Optimization of an acoustic echo canceller combined with adaptive gain control*, Proc. ICASSP-95, Detroit, Michigan, 1995, pp. 3047–3050

[4] J. MARX: *Akustische Aspekte der Echokompensation in Freisprecheinrichtungen*, VDI-Fortschritt-Berichte, Reihe 10, Nr. 400, Düsseldorf, Germany, 1996

[5] D. E. RUMELHART, J. L. MCCLELLAND (EDS.): *Parallel Distributed Processing*, vol. 1, MIT Press, Cambridge, USA, 1986

[6] S. YAMAMOTO, S. KITAYAMA: *An Adaptive Echo Canceller with Variable Step Gain Method*, Trans. IECE of Japan, vol. E 65, no. 1, 1982, pp. 1–8