Open Access

# The Trapezoidal Method of Steepest-Descent and its Application to Adaptive Filtering

T.J. Moir*

*School of Engineering and Advanced Technology, Massey University at Albany, Auckland, New-Zealand*

**Abstract:** The method of steepest-descent is re-visited in continuous time. It is shown that the continuous time version is a vector differential equation the solution of which is found by integration. Since numerical integration has many forms, we show an alternative to the conventional solution by using a Trapezoidal integration solution. This in turn gives a slightly modified least-mean squares (LMS) algorithm.

**Keyword:** Steepest-Descent, Least-mean squares (LMS), Adaptive filters.

## 1. INTRODUCTION

The steepest descent problem is usually quoted in discrete-time. One possible exception is found in [1] where the continuous-time version is explored. Furthermore a continuous-time LMS algorithm is found in [2]. For a scalar cost-function J which minimises some quadratic error, the steepest descent is described as the vector differential equation

$$\frac{d\mathbf{W}(t)}{dt} = -\mu \frac{\partial J}{\partial \mathbf{W(t)}} \tag{1}$$

$\mathbf{W}$ is a vector of unknown weights and $\mu$ is a scalar gain. Now the discrete-time version of (1) is normally quoted and can be found by discrete integration. For example we can approximate the derivative of the vector in (1) as

$$\frac{d\mathbf{W}(t)}{dt} \approx (\mathbf{W}_{k+1} - \mathbf{W}_k) / T \tag{2}$$

where the sampling interval T is normalised to unity and $\mathbf{W}_k$ is the discrete weight vector at some sample interval k. This gives rise to the usual discrete-time steepest descent method by substitution of (2) into (1) thus

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \mu \frac{\partial J}{\partial \mathbf{W}_k} \tag{3}$$

The term $\frac{\partial J}{\partial \mathbf{W}_k}$ is usually referred to as the gradient vector and the steepest descent algorithm is based on the idea that for a given vector $\mathbf{W}_k$ the best direction to go is the one that produces the biggest change in the cost function J. The resultant LMS algorithm has numerous applications in the field of adaptive signal processing. For example in the areas of adaptive filters [3], hearing aids [4], speech processing [5] and image processing [6]. The LMS algorithm is usual favoured over several rival estimation methods (e.g. recursive least-squares (RLS) [7]) due to its superior tracking ability. The LMS algorithm is not considered to be a rival to the more complex RLS when compared with speed of convergence. There are algorithms which are not as complex as RLS that have convergence rates which lie somewhere between LMS and RLS. For example the affine projection algorithm [8] is in this category. The reuse of data in the form of the data vector and error signal is employed in this paper and results in most cases in a faster convergence speed than ordinary LMS. It is interesting to note that the affine projection algorithm also uses the data reuse property.

## 2. MODIFIED STEEPEST DESCENT

It is established in the signal processing literature [9] that the integration method used by (2) above is known as rectangular Euler integration. It is also known that there exists a whole family of better approximations among which is the Trapezoidal method [9]. A Trapezoidal integration algorithm is a closer approximation than that of a Euler integrator and hence must be closer to any potential advantages offered by continuous-time adaptive filters [10].

$$\frac{d\mathbf{W}(t)}{dt} \approx \frac{2\mathbf{W}_{k+1}}{T}(1 - q^{-1}) / (1 + q^{-1}) \tag{4}$$

where $q^{-1}$ is the backward shift operator. Using (4) in (1) gives

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \frac{\mu}{2}\frac{\partial J}{\partial \mathbf{W}_k} - \frac{\mu}{2}\frac{\partial J}{\partial \mathbf{W}_{k-1}} \tag{5}$$

### Application to the Ordinary LMS Algorithm

We first apply the ordinary steepest descent method to a finite-impulse-response filter (FIR) w(z) of length n where $w(z) = w_0 + w_1 z^{-1} + w_2 z^{-2} + \ldots + w_n z^{-n}$ [11]

$$e_k = d_k - \mathbf{X}_k^T \mathbf{W}_k \tag{6}$$

*Address correspondence to this author at the School of Engineering and Advanced Technology, Massey University at Albany, Auckland, New-Zealand; Tel: +64 9 414 0800 ext 9805; Fax: +64 9 443 9774;
E-mail: T.J.Moir@massey.ac.nz

$e_k$ is a scalar error, $d_k$ is the desired output and $\mathbf{X}_k$ is the vector of regressors of the filter input and $\mathbf{W}_k$ has as its elements the coefficients of w(z). Defining a cost-function

$$J = E[e_k^2] \tag{7}$$

where E[.] is the statistical expectation operator and differentiating with respect to the vector $\mathbf{W}_k$ gives the usual Wiener solution

$$\mathbf{W}^* = \mathbf{R}^{-1}\mathbf{P} \tag{8}$$

In (8) above $\mathbf{W}^*$ is the optimal weight vector, $\mathbf{R} = E[\mathbf{X}_k\mathbf{X}_k^T]$ is the covariance matrix of the vector of regressors and $\mathbf{P} = E[d_k\mathbf{X}_k]$ is a vector of cross-correlations. Note also that the desired output is given by $d_k = \mathbf{X}_k^T\mathbf{W}^*$. However, since $\mathbf{R}$ and $\mathbf{P}$ cannot be known *apriori* then they are replaced by their stochastic equivalents [7] (3) giving rise to the ordinary LMS algorithm [12].

$$\mathbf{W}_{k+1} = \mathbf{W}_k + 2\mu\mathbf{X}_k e_k \tag{9}$$

with the error $e_k$ taken from (6).

## 3. THE TRAPEZOIDAL LMS ALGORITHM

We define trapezoidal LMS (TLMS) as the stochastic approximation algorithm obtained from the modified steepest descent equation (5). Using the same approach as ordinary LMS one obtains

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu\mathbf{X}_k e_k + \mu\mathbf{X}_{k-1}e_{k-1} \tag{10}$$

For m=n+1 weights, it can be seen that by comparing (10) with its LMS counterpart (9) that TLMS has only (1+m) extra multiplies and m extra additions. A computational comparison of LMS with other related algorithms is thoroughly studied in [13].

### Convergence of TLMS in the mean

Define a weight error vector

$$\boldsymbol{\varepsilon}_k = \mathbf{W}_k - \mathbf{W}^* \tag{11}$$

and write (10) in the form

$$\begin{aligned}\boldsymbol{\varepsilon}_{k+1} &= [\mathbf{I} - \mu\mathbf{X}_k\mathbf{X}_k^T]\boldsymbol{\varepsilon}_k - \mu\mathbf{X}_{k-1}\mathbf{X}_{k-1}^T\boldsymbol{\varepsilon}_{k-1} + \\ &\mu(d_k\mathbf{X}_k - \mathbf{X}_k\mathbf{X}_k^T\mathbf{W}^*) + \mu(d_k\mathbf{X}_{k-1} - \mathbf{X}_{k-1}\mathbf{X}_{k-1}^T\mathbf{W}^*)\end{aligned} \tag{12}$$

Then by taking expectations [7] we arrive at the mean vector difference equation

$$E[\boldsymbol{\varepsilon}_{k+1}] = [\mathbf{I} - \mu\mathbf{R}]E[\boldsymbol{\varepsilon}_k] - \mu\mathbf{R}E[\boldsymbol{\varepsilon}_{k-1}] + 2\mu(\mathbf{P} - \mathbf{R}\mathbf{W}^*) \tag{13a}$$

and substituting $\mathbf{W}^* = \mathbf{R}^{-1}\mathbf{P}$ from (8) we get the homogeneous equation

$$E[\boldsymbol{\varepsilon}_{k+1}] = [\mathbf{I} - \mu\mathbf{R}]E[\boldsymbol{\varepsilon}_k] - \mu\mathbf{R}E[\boldsymbol{\varepsilon}_{k-1}] \tag{13b}$$

From which the equivalent LMS version is well established [7]. In the above analysis we assume that the weight vector $\mathbf{W}_k$ is statistically independent of the vector $\mathbf{X}_k$ and likewise with vectors $\boldsymbol{\varepsilon}_k$ and $\mathbf{X}_k$.

Define a unitary similarity transformation $E[\mathbf{V}_k] = \mathbf{Q}^T E[\boldsymbol{\varepsilon}_k]$ where $\mathbf{Q}$ has its columns an orthogonal set of eigenvectors associated with the eigenvalues of $\mathbf{R}$. That is $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$ [7] The matrix $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots\lambda_{n+1}\}$.

Multiplying (13b) by $\mathbf{Q}^T$ gives

$$E[\mathbf{V}_k] = [\mathbf{I} - \mu\boldsymbol{\Lambda}]E[\mathbf{V}_{k-1}] - \mu\boldsymbol{\Lambda}E[\mathbf{V}_{k-2}] \tag{14}$$

For the $i^{th}$ natural mode of the TLMS system we have

$$E[\mathbf{v}_k] = (1 - \mu\lambda_i)E[\mathbf{v}_{k-1}] - \mu\lambda_i E[\mathbf{v}_{k-2}] \tag{15}$$

Thus for stability the set of polynomials

$$1 - (1 - \mu\lambda_i)z^{-1} + \mu\lambda_i z^{-2} = 0, i = 1, 2 \ldots (n+1) \tag{16}$$

must have all their roots within ½z½ = 1 and the modes will then all die out. For stability and hence convergence of the mean weight error vector for this quadratic case, it follows from a Jury test [14] that

$$0 < \mu\lambda_{\max} < 1 \tag{17}$$

or alternatively the step size satisfies

$$0 < \mu, < 1/\lambda_{\max} \tag{18}$$

where $\lambda_{max}$ is the largest eigenvalue of $\mathbf{R}$. This result is apparently the same as for ordinary LMS but a simple example will show the difference in performance.

***Example 1***

Consider the identification of a $4^{th}$ order FIR system $w(z) = 1 + 2z^{-1} + 3z^{-2} + 4z^{-3} + 5z^{-4}$ with additive white-noise of variable variance. We consider graphs of only the $5^{th}$ weight for clarity and with $\mu = 0.1$. For small values of $\mu$ it was found that the LMS and TLMS algorithms behaved identically. However, as the step size was increased the TLMS shows improvement over LMS as shown in Fig. (**1**) using $\mu = 0.1$ for all cases. This would be expected since a trapezoidal integrator is a better match to a continuous-time integrator than a Euler (rectangular) integrator.

As the SNR is reduced the TLMS method always gives a smaller mean-square error. This is illustrated for 2 further SNRs of 26dB and 19dB respectively in Figs. (**2 and 3**).

However, operating with higher values of step size has the disadvantage that the stability limit is nearer and hence for non-stationary systems any potential advantages may be outweighed.

Table **1** shows a table of differing SNR values with the stead-state mean and variance of the weight. It can be seen that the TLMS method always wins out and in particular as the SNR reduces the superiority of TLMS over LMS is shown by the reduced variance of the weight estimate.

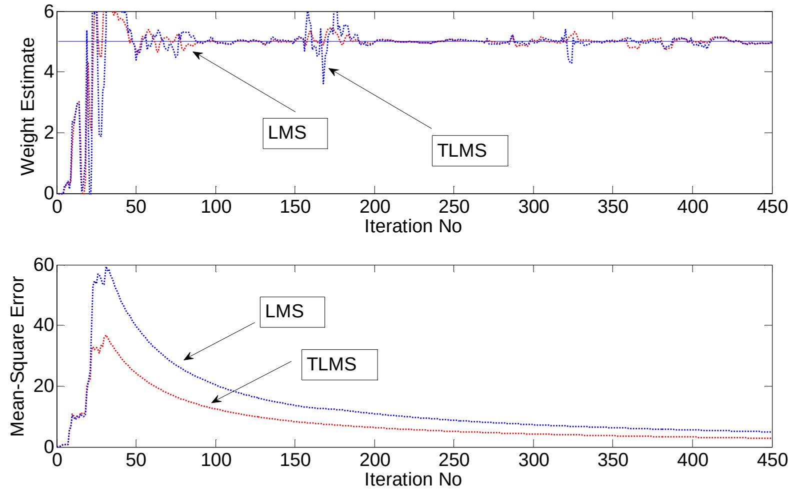


**Fig. (1).** Comparison of LMS and TLMS for a SNR of 40dB.

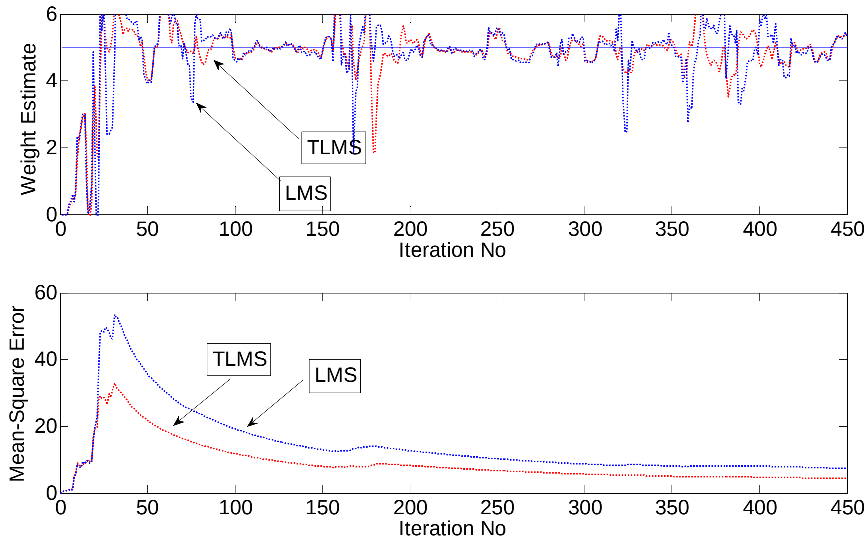


**Fig. (2).** Comparison of LMS and TLMS for a SNR of 26dB.

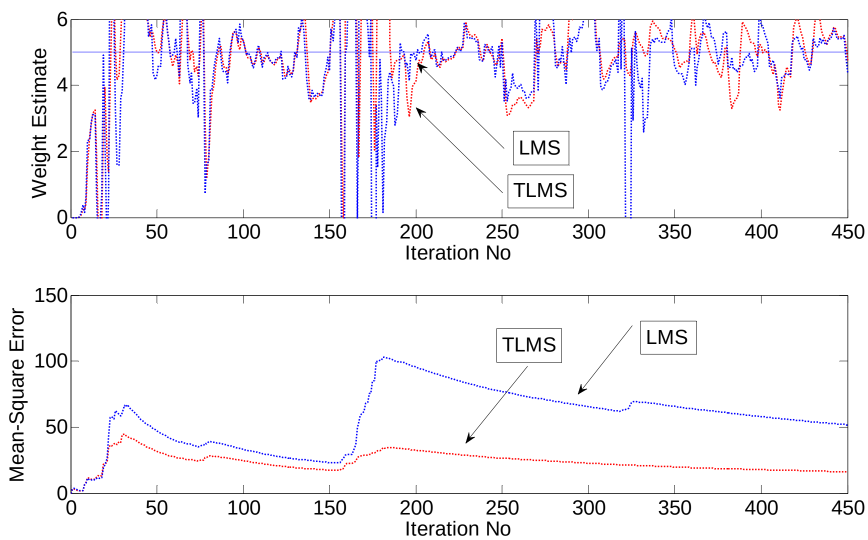


**Fig. (3).** Comparison of LMS and TLMS for a SNR of 19dB.

**Table 1. Mean and Variance of Estimated Weight for Different SNR Values in Example 1**

| SNR | 76dB | 49dB | 40dB | 26dB | 20dB |
|---|---|---|---|---|---|
| Mean weight estimate (LMS) | 5.0016 | 5.009 | 5.02 | 5.049 | 5.122 |
| Mean weight estimate (TLMS) | 5.0018 | 5.004 | 5.0012 | 5.007 | 4.986 |
| Weight variance (LMS) | 5.87e-5 | 0.0199 | 0.306 | 1.98 | 5.97 |
| Weight variance (TLMS) | 2.04e-4 | 0.0011 | 0.006 | 0.34 | 1.1177 |

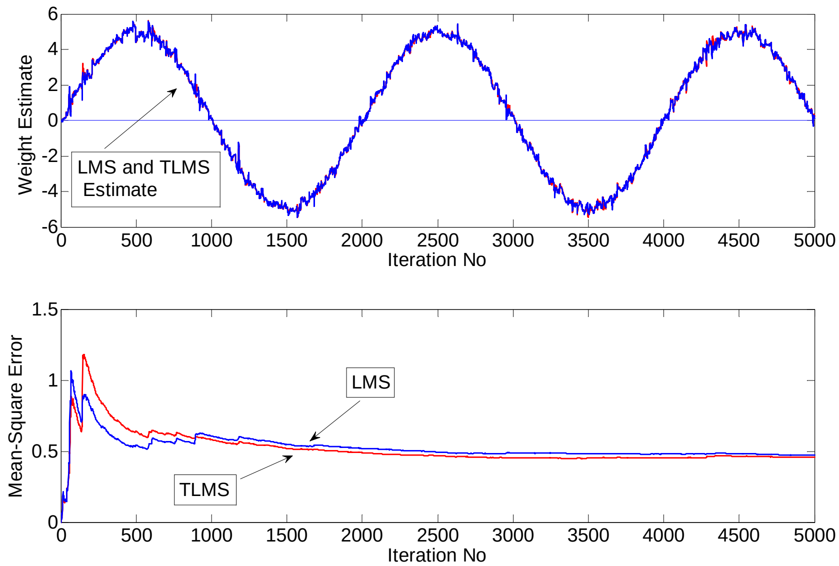


**Fig. (4).**

Fig. (**4**). shows very similar results to the previous example where the weight are fixed. Once again the minimum mean-square error of TLMS is smaller than that of LMS.

### *Example 2*

Consider the identification of a 4th order time-varying FIR system $w(z) = a_1(k) + a_2(k)z^{-1} + a_3(k)z^{-2} + a_4(k)z^{-3} + a_5(k)z^{-4}, k = 0,1...$ with additive white-noise of variable variance. We consider graphs of only the $a_4(k)$ weight for clarity and with $\mu = 0.1$. We vary the weights sinusoidally with increasing amplitude and frequency according to $a_i(k) = i\sin(2\pi ki / f_s), i = 1,2...5, k = 0,1,2...$ where $f_s = 10000$ is the sampling frequency in Hz.

### *Example 3*

Consider a problem closely related to adaptive acoustic beamforming. Three isolated words are spoken at an angle of 45 degrees from two microphones approximately 30cm apart. In some approaches to adaptive beamforming with two microphones, the first stage is to align the two microphones to the “look” direction which in this case was made directly in front of the microphones. In an anechoic chamber there will only be a minor time-difference of arrival (plus perhaps some minor attenuation) in transfer function between the signal (voice) source and each microphone. However, in real-world applications there is reverberation off walls and ceiling and a rapidly time-varying transfer function is needed to align the microphones in the look direction[15]. This transfer function is more than often non-minimum phase in nature. We can examine the performance of the LMS and TLMS algorithms by estimating the transfer function *between* the two microphones. A step size of $\mu = 0.08$ and 200 weights was used. One of the originally recorded speech signals is shown in Fig. (**5**).

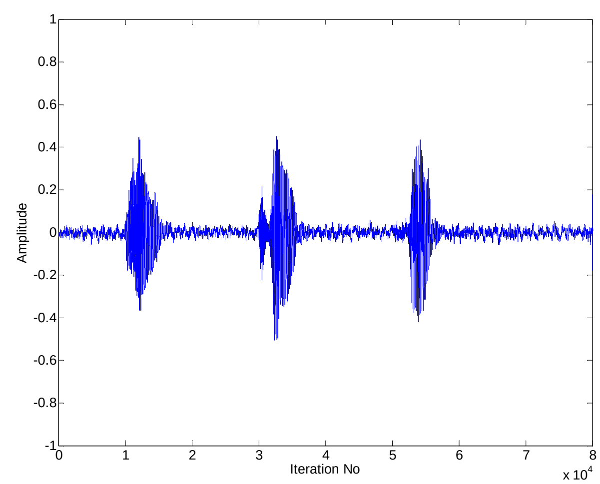


**Fig. (5).** A recording at one of the microphones of the speech signal.

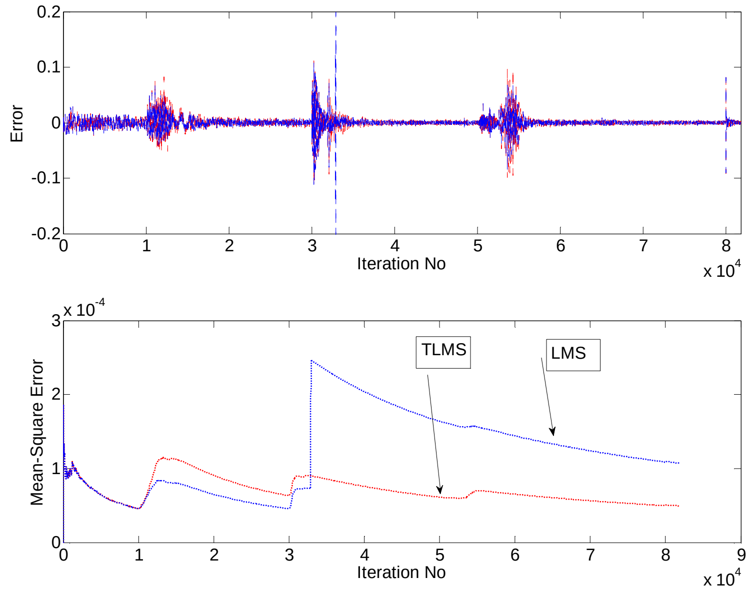


**Fig. (6).** Comparison of LMS and TLMS for a real-recorded speech signal.

Fig. (**6**) shows that ordinary LMS gives a lower mean-square error to begin with, but when a sudden change occurs at the second word, TLMS achieves a lower mean-square error. The TLMS algorithm maintains this minimum through to the third word. Fig. (**6**) also shows the error signals for both cases showing that the speech has been severely attenuated. In an acoustic beamformer this error signal would be heard as residual noise. It should be noted that for all of these examples, for smaller step-size values the two algorithms yield identical results. This is because the two types of integrators only differ significantly at higher frequencies which corresponds to when the step-size is bigger.

## 4. CONCLUSIONS

The continuous-time steepest-descent problem has been re-visited. It has been shown that the method of solving the continuous-time problem by finding a discrete-time solution is not in fact unique. In fact the conventional approach when applied to the estimation of weights in an FIR filter uses Euler integration giving rise to ordinary LMS. Consequently, by using a slight modification of a more accurate Trapezoidal integration method, a modified LMS algorithm was found that offers some advantages and is of some theoretical and practical interest.

## REFERENCES

[1] P. Mitter, and C. W. Ueberhuber, "The continuous method of steepest descent and its discretizations," *Lecture Notes in Control and Information Sciences*. vol. 7, pp. 94-100, 1978.

[2] S. Karni, and G. Zeng, "The analysis of the continuous-time LMS algorithm," *IEEE Transactions Acoustics, Speech and Signal Processing*, vol. 37, pp. 595-597, 1989.

[3] S. Haykin, and B. Widrow, *Least-Mean-Square Adaptive Filters (Adaptive and Learning Systems for Signal Processing, Communications and Control)*: Wiley Interscience, 2003.

[4] J. V. Berghe, and J.Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones," *The Journal of the Acoustical Society of America*, vol. 103, pp. 3621-3626, June 1998.

[5] J. E. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceller," *IEEE Transactions on Speech and Audio Processing*, vol. 6, pp. 338-351, July 1998.

[6] C. P.-Chi, Y. C.-Sheng, and L. T.-Hsu, "Hybrid LMS-MMSE inverse halftoning technique," *IEEE Transactions on Image Processing*, vol. 10, pp. 95-103, 2001.

[7] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs New Jersey: Prentice Hall, 1986.

[8] V. Yoganathan, and T. J. Moir, "Switched Griffiths-Jim beamformer using the affine projection algorithm," In: *Mechatronics and Machine Vision in Practice*, 2008. M2VIP 2008. 15th International Conference, 2008, pp. 391-396.

[9] A. Ambardar, *Analog and Digital Signal Processing*, Brooks/Cole, CA, 1999.

[10] D. A. Johns, W. M. Snelgrove, and A. S. Sedra, "Continuous-time LMS adaptive recursive filters," *IEEE Transactions on Circuits and Systems*, vol. 38, pp. 769-778, 1991.

[11] P. S. R. Dinez, *Adaptive filtering: Algorithms and Practical Implementation*, USA, Springer, 2008.

[12] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *IRE Wescon Convention Record*, pp. 96-104, 1960.

[13] R. C. de Lamare, and R. Sampaio-Neto, "Reduced-rank adaptive filtering based on joint iterative optimization of adaptive filters," *IEEE Signal Processing Letters*, vol. 14, pp. 980-983, 2007.

[14] E. I. Jury, *Theory and Application of the Z-Transform Method*, New-York: Wiley, 1964.

[15] T. J. Moir, "A z-domain transfer function solution to the non-minimum phase acoustic beamformer," *International Journal of Systems Science*, vol. 38, pp. 563-575, 2007.

Received: November 16, 2009 Revised: December 14, 2009 Accepted: December 31, 2009