On the Behavior of the Signed Regressor Least Mean Squares Adaptation with Gaussian Inputs

Sung Ho Cho* *Regular Member*

가우시안 입력신호에 대한 Signed Regressor 최소평균자승 적응 방식의 동작특성

正會員 趙 誠 鎬*

ABSTRACT

The signed regressor (SR) algorithm employs one bit quantization on the input regressor (or tap input) in such a way that the quantized input sequences become +1 or −1. The algorithm is computationally more efficient by nature than the popular least mean square (LMS) algorithm. The behavior of the SR algorithm unfortunately is heavily dependent on the characteristics of the input signal, and there are some inputs for which the SR algorithm becomes unstable. It is known, however, that such a stability problem does not take place with the SR algorithm when the input signal is Gaussian, such as in the case of speech processing. In this paper, we explore a statistical analysis of the SR algorithm. Under the assumption that signals involved are zero-mean and Gaussian, and further employing the commonly used independence assumption, we derive a set of nonlinear evolution equations that characterizes the mean and mean-squared behavior of the SR algorithm. Experimental results that show very good agreement with our theoretical derivations are also presented.

요 약

Signed Regressor 적응 알고리즘은 한 비트 양자화를 이용하여 탭 입력이 +1 또는 −1이 되도록 양자화한다. 따라서 이미 널리 사용되고 있는 Least Mean Square (LMS) 알고리즘에 비하여 계산량 측면에서 효율적이다. 그러나 SR 알고리즘의 동작특성은 입력신호의 특성에 매우 종속적이며, 효율성을 위하여 성능을 약간 희생한다. 본 논문에서는 이 SR 알고리즘의 동작특성에 대하여 통계적 분석을 하였다. 이를 위해, 사용되는 신호가 평균이 제로인 가우시안 신호라는 가정과 이러한 분석에 이미 널리 통용되어 사용되는 독립가정을 이용하여, SR 알고리즘의 평균 및 평균자승 특성을 나타내는 일련의 비선형 관계식을 유도하였다. 그리고 유도된 이론적 결과가 실험적 결과와 매우 일치함을 보였다.

*漢陽大學校 電子工學科
Dept. of Electronics Eng., Han Yang Univ.
論文番號 : 93−105

I. Introduction

The adaptive LMS algorithm [1],[2] has received a great deal of attention during the last two decades and is now widely used in variety of applications due to its simplicity. In very high data rate applications of adaptive filters, however, it is often necessary to reduce the computational requirements of the adaptive mechanism any further. One of the most popular and well known approaches of reducing the computational complexity is to use the sign algorithm [3]-[5], for which the estimation error signal in the coefficient update equation is quantized such that the quantized error becomes $+1$ or -1 according to the sign of the error signal (i.e., one bit quantization). Another such efficient version of the LMS algorithm is the SR algorithm [6]-[8]. This algorithm employs one bit quantization on the input regressor, and thus provides the same complexity as the sign algorithm.

The behavior of the SR algorithm unfortunately is heavily dependent on the characteristics of the input signal, and there are some inputs for which the SR algorithm becomes unstable. It is known, however, that such a stability problem does not take place with the SR algorithm when the input signal is Gaussian, such as in the case of speech processing. Moschner [6] first suggested the SR method, and examined its behavior. It was shown that the algorithm is convergent in the mean sense under the independence assumption and the assumption that the signal $x(n)$ is Gaussian. Sethares, et al. [7] extended the result in [6] by investigating "persistence of excitation" conditions for the SR algorithm. It was proved that the SR algorithm is "exponentially stable" with small enough stepsize if the input $x(n)$ is generated by a zero-mean and white Gaussian random source passed through a stable linear filter, or is independent and identically distributed with finite moments. Other processes were thought possibly to cause divergence. Eweda [8] also studied the performance of the SR algorithm in both stationary and nonstationary environments, and drew several important properties of the algorithm. One of the most important results was a proof of convergence of the SR algorithm with an M-dependence model. (M-dependence means that there exists a positive number M such that for all k and for any random vector process $X(n)$, $\{X(n),\ n \leq k\}$ and $\{X(n),\ n \geq k+M\}$ are independent.)

This paper extends the results in [6]-[8] by analyzing the statistical behavior of the SR algorithm when signals involved are zero-mean, wide-sense stationary, and Gaussian. In the next section, by making use of Price's theorem [9] and further employing the independence assumption [10], a set of nonlinear difference equations that characterizes the mean and mean-squared behavior of the filter coefficients and the mean-squared estimation error is derived. A condition for the mean convergence is also found. Experimental results demonstrating the validity of the analytic results are included in Section III, and the concluding remarks are made in Section IV.

II. Statistical Convergence Behavior

Consider the problem of estimating the primary input signal $d(n)$ using the reference input $x(n)$. Let $H(n)$ denote the adaptive filter coefficient vector of size N, and $e(n)$ denote the estimation error signal. Define the regressor vector $X(n)$ as

$$X(n) = [x(n),\ x(n-1),\ \cdots,\ x(n-N+1)]^T, \qquad (1)$$

where $[\cdot]^T$ denotes the transpose of $[\cdot]$. The SR algorithm under consideration updates the coefficient vector $H(n)$ using

$$H(n+1) = H(n) + \mu\, sign\{X(n)\}\, e(n), \qquad (2)$$

where μ denotes the adaptation step-size or convergence parameter, $sign\{X(n)\}$ represents the vector of size N that consists of $+1$ or -1 according to the sign of its entries $x(n-i)$ for $0 \leq i$

$\le N-1$, and

$$e(n)=d(n)-X^T(n)\,H(n). \tag{3}$$

As can be seen in (2) and (3), if μ is chosen to be a negative integer power of two, then the SR algorithm requires N multiplications and $2N$ additions for each iteration, while the LMS algorithm needs $2N$ multiplications and $2N$ additions.

Now, before starting the analysis, let us define the following notations : Let H_{opt} denote the optimum coefficient vector given by

$$H_{opt}=R_{XX}^{-1}\,R_{dX}, \tag{4}$$

where

$$R_{XX}=E\{X(n)\,X^T(n)\}, \tag{5}$$

$$R_{dX}=E\{d(n)\,X(n)\}, \tag{6}$$

and $E\{\cdot\}$ denotes a statistical expectation of $\{\cdot\}$. Also, define the coefficient misalignment vector $V(n)$ as

$$V(n)=H(n)-H_{opt}, \tag{7}$$

and its autocorrelation matrix $K(n)$ as

$$K(n)=E\{V(n)\,V^T(n)\}. \tag{8}$$

Using (7) in (2), we get

$$V(n+1)=V(n)+\mu\,sign\{X(n)\}\,e(n). \tag{9}$$

The optimal estimation error $e_{min}(n)$ is given by

$$e_{min}(n)=d(n)-X^T(n)\,H_{opt}. \tag{10}$$

Combining (3), (7), and (10), it follows that

$$e(n)=e_{min}-X^T(n)\,V(n). \tag{11}$$

Finally, let

$$\xi_{min}=E\{e_{min}^2(n)\} \tag{12}$$

denote the minimum mean-squared estimation error.

Convergence analysis of the SR algorithm is much more complicated than that of the LMS algorithm due to existence of the nonlinear "clipping" operation on the input process. We thus make the following assumptions to make the anlysis mathematically more tractable :

Assumption 1 : $d(n)$ and $X(n)$ are zero-mean, wide-sense stationary, and jointly Gaussian random processes.

Assumption 2 : The input pair $\{d(n),\,X(n)\}$ at time n is independent of $\{d(k),\,X(k)\}$ at time k, if $n\neq k$.

A consequence of Assumption 1 is that the estimation error $e(n)$ given in (3) is also a zero-mean and Gaussian process when conditioned on the coefficient vector $H(n)$ (or equivalently, on $V(n)$). Assumption 2 is the commonly employed "independence assumption" and seldom true in practice. It is, however, shown in [10] that the assumption is valid if μ is chosen to be sufficiently small. Also, the analysis using this assumption has produced results in the past that accurately predict the behavior of the adaptive filters even in circumstances where the assumption is grossly violated [4],[5]. One direct consequence of Assumption 2 is that $H(n)$ is independent of the input pair $\{d(n),\,X(n)\}$ since $H(n)$ depends only on inputs at time $n-1$ and before. Note also that Assumption 2 does not restrict the nature of the matrix R_{XX}.

Now, taking the statistical expectation on both sides of (2) gives

$$E\{H(n+1)\}=E\{H(n)\}+\mu\,E\{sign\{X(n)\}\,e(n)\}. \tag{13}$$

The last expectation of (13) can be computed using either Price's theorem [9] or the following result modified from [11] : for an arbitrary Borel function $G(\cdot)$ and Gaussian X and e,

$$E\{G(X)\,e\}=\frac{E\{Xe\}E\{X^T G\{X\}\}}{E\{X^T X\}} \tag{14}$$

Since $e(n)$ is zero-mean and Gaussian when conditioned on $H(n)$, it follows from (3) and (14) that

$$\begin{aligned} E\{sign\{X(n)\}\,e(n)\} &= E\{E[sign\{X(n)\}\,e(n)\,|\,H(n)]\} \\ &= \frac{E\{X^T(n)\,sign\{X(n)\}\}}{E\{X^T(n)\,X(n)\}} \\ &\quad E\{E[X(n)\,e(n)\,|\,H(n)]\} \end{aligned} \tag{15}$$

Defining $\sigma_x^2=E\{x^2(n)\}$ and making use of the fact that the mean-absolute value of a Gaussian random variable with zero mean and variance σ^2 is $\sigma\sqrt{2/\pi}$, we find under Assumption 1 that

$$E\{X^T(n)\,sign\{X(n)\}\}=\sum_{i=0}^{N-1}E|x(n-i)|=\sqrt{\frac{2}{\pi}}\,N\,\sigma_x, \tag{16}$$

and

$$E\{X^T(n)\,X(n)\}=\sum_{i=0}^{N-1}E\{x^2(n-i)\}=N\,\sigma_x^2, \tag{17}$$

Also, using (3) as well as Assumption 2 gives

$$E\{X(n)\,e(n)\,|\,H(n)\}=R_{dX}-R_{XX}\,H(n). \tag{18}$$

Substituting (16)-(18) in (15), and further using (7) then yields

$$\begin{aligned} E\{sign\{X(n)\}e(n)\} &= \sqrt{\frac{2}{\pi}}\,\frac{1}{\sigma_x}\,[R_{dX}-R_{XX}\,E\{H(n)\}] \\ &= -\sqrt{\frac{2}{\pi}}\,\frac{1}{\sigma_x}\,R_{XX}\,E\{V(n)\}. \end{aligned} \tag{19}$$

Therefore, using (19) in (13), we obtain the mean behavior of the SR algorithm as

$$E\{H(n+1)\}=\left[I_N-\sqrt{\frac{2}{\pi}}\,\frac{\mu}{\sigma_x}\,R_{XX}\right]E\{H(n)\} + \sqrt{\frac{2}{\pi}}\,\frac{\mu}{\sigma_x}\,R_{dX}, \tag{20}$$

where I_N denotes the $N\times N$ identity matrix. The above expression can be rewritten using the coefficient misalignemt vector as

$$E\{V(n+1)\}=\left[I_N-\sqrt{\frac{2}{\pi}}\,\frac{\mu}{\sigma_x}\,R_{XX}\right]E\{V(n)\}. \tag{21}$$

It is easy to show that the mean behavior of the coefficient misalignment vector given in (21) asymptotically converges to the zero vector (or equivalently, $E\{H(n)\}$ is asymptotically convergent to H_{opt}) if the convergence parameter μ is selected to be

$$0<\mu<\frac{\sqrt{2\pi}\,\sigma_x}{\lambda_{max}}, \tag{22}$$

where λ_{max} denotes the maximum eigenvalue of the matrix R_{XX}. Notice that a more restrictive and sufficient, but simpler and more practical condition for the convergence can be given by

$$0<\mu<\frac{\sqrt{2\pi}}{N\sigma_x}. \tag{23}$$

We next derive an expression for the mean-squared estimation error $\sigma_e^2(n)$. Taking the expectation after squaring both sides of (11) yields.

$$\begin{aligned} \sigma_e^2(n) &= E\{e^2(n)\} \\ &= \xi_{min}+E\{V^T(n)\,X(n)\,X^T(n)\,V(n)\} \\ &\quad -2E\{V^T(n)\,X(n)\,e_{min}(n)\}, \end{aligned} \tag{24}$$

where ξ_{min} is obtained by using (10) in (12) so that

$$\xi_{min}=E\{d^2(n)\}-H_{opt}^T\,R_{dX}. \tag{25}$$

The last expectation of (24) becomes zero by orthogonality principle (i.e., $X(n)$ and $e_{min}(n)$ are orthogonal each other). It thus follows under the independence assumption (Assumption 2) that

$$\sigma_e^2(n) = \xi_{min} + tr\{K(n) R_{XX}\}, \quad (26)$$

where $K(n)$ is defined in (8), and $tr\{\cdot\}$ represents the trace of $\{\cdot\}$.

Finally, we derive an expression for $K(n)$ to complete the analysis. Substituting (9) in (8) leads to

$$K(n+1) = K(n) + \mu^2 E[sign\{X(n) X^T(n)\} e^2(n)] + \mu E[V(n) sign\{X^T(n)\} e(n)] + \mu E[sign\{X(n)\} V^T(n) e(n)]. \quad (27)$$

Before simplifying (27) any further, we define the followings:

$$r_{i-j} = E\{x(n-i) x(n-j)\}, \quad (28)$$

$$\Lambda = E[sign\{X(n) X^T(n)\}], \quad (29)$$

$$\Omega(n) = E[V^T(n)\} X(n) X^T(n) V(n) sign\{X(n) X^T(n)\}], \quad (30)$$

and for $0 \le k, \ell \le N-1$,

$$T(k, \ell) = E\{x(n-k) x(n-\ell) sign\{X(n) X^T(n)\}\}. \quad (31)$$

Also, for $1 \le i, j \le N$, let Λ_{ij}, $\Omega_{ij}(n)$, $T_{ij}(k, \ell)$, and $K_{ij}(n)$ denote the (i,j)-th entries of the matrices $\Lambda, \Omega(n)$, $T(k, \ell)$, and $K(n)$, respectively. Note that Λ and $T(k, \ell)$ are constant matrices, while the matrix $\Omega(n)$ is time-varying.

The following results can be derived by invoking the Gaussian assumption and employing Price's theorem [9]. These are to be used to evaluate (27).

- Let x_1 and x_2 be zero-mean, jointly Gaussian random variables with covariance matrix

$$R = \begin{bmatrix} \sigma_x^2 & r \\ r & \sigma_x^2 \end{bmatrix}$$

Then,

$$E\{sign\{x_1\} sign\{x_2\}\} = \frac{2}{\pi} \sin^{-1} \frac{r}{\sigma_x^2}. \quad (32)$$

- Let x_1, x_2, x_3, and x_4 be zero-mean, jointly Gaussian random variables with covariance matrix

$$R = \begin{bmatrix} \sigma_x^2 & r_{12} & r_{13} & r_{14} \\ r_{12} & \sigma_x^2 & r_{23} & r_{24} \\ r_{13} & r_{23} & \sigma_x^2 & r_{34} \\ r_{14} & r_{24} & r_{34} & \sigma_x^2 \end{bmatrix},$$

and let $|r_{34}| < \sigma_x^2$. Then,

$$E\{x_1 x_2 sign\{x_3\} sign\{x_4\}\} = \frac{2}{\pi} r_{12} \sin^{-1} \frac{r_{34}}{\sigma_x^2} + \frac{2}{\pi} \frac{1}{\sqrt{\sigma_x^4 - r_{34}^2}} \left\{[r_{14} r_{23} + r_{13} r_{24}] - \frac{r_{34}}{\sigma_x^2} [r_{13} r_{23} + r_{14} r_{24}]\right\}. \quad (33)$$

It is shown in Appendix that the mean-squared behavior $K(n)$ in (27) can be expressed as

$$K(n+1) = K(n) + \mu^2[\xi_{min} \Lambda + \Omega(n)] - \sqrt{\frac{2}{\pi}} \frac{\mu}{\sigma_x} [K(n) R_{XX} + R_{XX} K(n)]. \quad (34)$$

The matrices Λ and $\Omega(n)$ are also evaluated in Appendix.

Ⅲ. Experimental Results

Here, we present the experimental results for which the SR algorithm is used in the third-order adaptive predictor to demonstrate the validity of our derivations. The primary input $d(n)$ is modeled as an autoregressive process given by

$$d(n) = \zeta(n) + 0.9d(n-1) - 0.1d(n-2) - 0.2d(n-3), \quad (35)$$

where $\zeta(n)$ is a pseudorandom white Gaussian process with zero-mean and variance such that $d(n)$ has unit variance. The reference input $x(n)$ to the predictor is

$$x(n) = d(n-1). \quad (36)$$

Note that the ratio of the maximum and minimum eigenvalues of the autocorrelation matrix R_{XX} is approximately 16.3 in this case.

The results are produced by taking the ensemble averages over 100 independent runs using 10,000 samples each. The parameters μ are selected to be 0.005. Figure 1 illustrates the theoretical and empirical results of the mean behavior $E\{h_i(n)\}$ for the i-th element of the adaptive filter coefficient vector $H(n)$, and Figure 2 shows those of the mean-squared behavior $K_{ii}(n)$ of the i-the coefficient by plotting the three diagonal elements of $K(n)$. It can be seen that the theoretical curves agree with simulation ones fairly well.

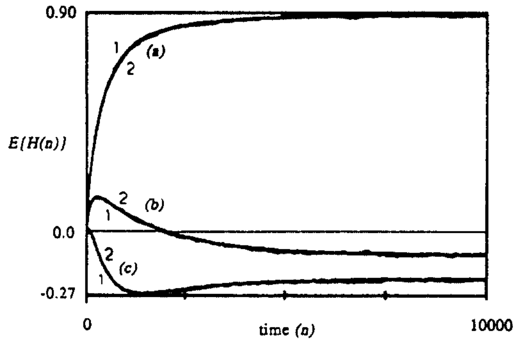
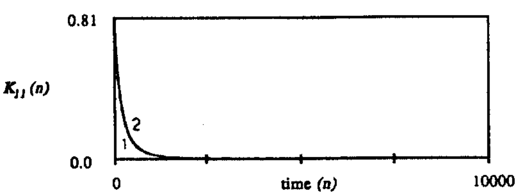
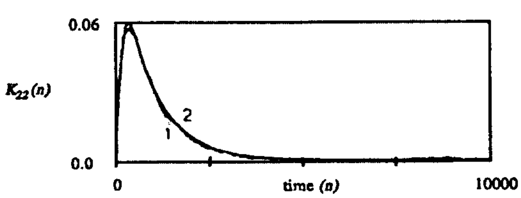


Figure 1. Mean behavior of the three coefficients ; (a) $E\{h_1(n)\}$, (b) $E\{h_2(n)\}$, (c) $E\{h_3(n)\}$; 1 = simulation result, 2 = theoretical result.





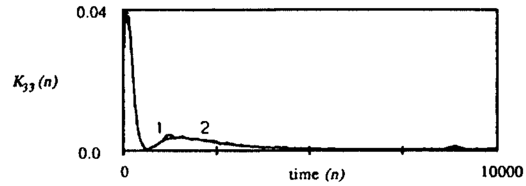


Figure 2. Mean-squared behavior of the three coefficients ; 1 = simulation result, 2 = theoretical result.

Ⅳ. Conclusion

In this paper, we investigate the statistical behavior of the SR algorithm when signals involved are zero-mean, wide-sense stationary, and Gaussian. By making use of Price's theorem and further employing the independence assumption, a set of nonlinear evolution equations that characterizes the mean and mean-squared behavior of the filter coefficients as well as the mean-squared estimation error is derived. A condition for the mean convergence is also found. Experimental results show that our theoretical derivations agree with simulations very well.

We have obtained some empirical results in which the performance of the SR algorithm is quite comparable with that of the LMS or the sign algorithm. We have had, however, difficulties in deriving analytical expressions for the steady-state mean-squared responses of the SR algorithm in order to make quantitative comparisons among other competing algorithms, such as the LMS or the sign algorithms. We are currently working on obtaining these results as well as conditions for the mean-squared convergence of the algorithm.

Appendix Derivation of (34)

From (27), using the result in (19), we have

$$E[V(n)\,sign\{X^T(n)\}e(n)]$$
$$=E\{V(n)\,E[sign\{X^T(n)\}\,e(n)\,|\,V(n)]\}$$

$$= -\sqrt{\frac{2}{\pi}}\,\frac{1}{\sigma_x}\,K(n)\,R_{XX}. \quad (37)$$

Similarly,

$$E[sign\{X(n)\}V^T(n)e(n)] = -\sqrt{\frac{2}{\pi}}\,\frac{1}{\sigma_x}\,R_{XX}\,K(n). \quad (38)$$

Using (11) as well as orthogonality principle, it is easy to simplify the last term of (27) to

$$E[sign\{X(n)\,X^T(n)\}\,e^2(n)] = \xi_{min}\Lambda + \Omega(n), \quad (39)$$

where the matrices Λ and $\Omega(n)$ are defined in (29) and (30), respectively. We now need to compute Λ and $\Omega(n)$. From (29) and (32), it follows that

$$\Lambda_{ij} = E\{sign\{x(n-i+1)\}\,sign\{x(n-j+1)\}\}$$

$$= \frac{2}{\pi}\,sin^{-1}\frac{r_{i-j}}{\sigma_x^2}. \quad (40)$$

Also, since

$$V^T(n)\,X(n)\,X^T(n)\,V(n) = \sum_{k=0}^{N-1}\sum_{\ell=0}^{N-1} v_{k+1}(n)\,v_{\ell+1}(n)\,x(n-k)\,x(n-\ell), \quad (41)$$

by the independence assumption once again, we have

$$\Omega(n) = \sum_{k=0}^{N-1}\sum_{\ell=0}^{N-1} E\{v_{k+1}(n)\,v_{\ell+1}(n)\}\,T(k,\ell), \quad (42)$$

or

$$\Omega_{ij}(n) = \sum_{k=0}^{N-1}\sum_{\ell=0}^{N-1} K_{ij}(n)\Big|_{\substack{i=k+1\\ j=\ell+1}}\,T_{ij}(k,\ell). \quad (43)$$

where $v_i(n)$ denotes the i-th element of the vector $V(n)$, and $T(k,\ell)$ is defined in (31). In order to evaluate (42), we have to compute $T(k,\ell)$ or its elements $T_{ij}(k,\ell)$. Using the result in (33), it is not difficult to compute $T_{ij}(k,\ell)$: i.e., for $0 \leq k, \ell \leq N-1$, if $i=j$

$$T_{ij}(k,\ell) = r_{k-\ell}, \quad (44)$$

and if $i \neq j$

$$T_{ij}(k,\ell) = \frac{2}{\pi}\,r_{k-\ell}\,sin^{-1}\frac{r_{i-j}}{\sigma_x^2} + \frac{2}{\pi}\,\frac{1}{\sqrt{\sigma_x^4 - r_{i-j}^2}}\,[r_{k-j+1}\,r_{\ell-i+1} + r_{k-i+1}\,r_{\ell-j+1}] - \frac{2}{\pi}\,\frac{r_{i-j}}{\sigma_x^2\sqrt{\sigma_x^4 - r_{i-j}^2}}\,[r_{k-i+1}\,r_{\ell-i+1} + r_{k-j+1}\,r_{\ell-j+1}]. \quad (45)$$

Therefore, using (37)-(45), we obtain (34), and are able to completely evaluate the mean-squared behavior of the coefficient misalignment vector.

References

1. B. Widrow, et al., "Adaptive noise cancelling : Principles and applications," *Proc. of IEEE*, pp. 1692-1716, Dec. 1975.
2. B. Widrow, et al., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. of IEEE*, pp.1151-1162, Aug. 1976.
3. T.A.C.M. Claasen and W.F.G. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Acoust., Speech, and Signal Proc.*, pp.670-678, Dec. 1979.
4. V.J. Mathews and S.H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, and Signal Proc.*, pp.450-454, Apr. 1987.
5. S.H. Cho and V.J. Mathews, "Tracking analysis of the sign algorithm in nonstationary environments," *IEEE Trans. Acoust., Speech, and Signal Proc.*, pp.2046-2057, Dec. 1990.
6. J.L. Moschner, "Adaptive filter with clipped input data," Tech. Rep. 6796-1, Stanford Univ.

Information Syatems Lab., June 1970.
7. W.A. Sethares, et al., "Excitation conditions for signed regressor least mean squares adaptation," *IEEE Trans. Circuits and Systems*, pp. 613-624, June 1988.
8. E. Eweda, "Analysis and design of a signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data," *IEEE Trans. Circuits and Systems*, pp.1367-1374, Nov. 1990.
9. R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, pp.69-72, Jun. 1958.
10. J.E. Mazo, "On the independence theory of the equalizer convergence," *Bell Sys. Tech. J.*, pp.963-993, May-June 1979.
11. T. Koh and E.J. Powers, "Efficient methods to estimate correlation functions of Gaussian processes and the performance analysis," *IEEE Trans. Acoust., Speech, and Signal Proc.*, pp.1032-1035, Aug. 1985.

趙 誠 鎬(Sung Ho Cho) 종신회원
1960년 2월 21일생
1982년 2월 : 한양대학교 전자공학과 졸업(B.S.)
1984년 12월 : Univ. of Iowa 전기 및 컴퓨터 공학과 졸업(M.S.)
1989년 8월 : Univ. of Utah 전기공학과 졸업(Ph.D)
1989년 8월 ~ 1992년 8월 : 한국전자통신연구소 선임연구원
1992년 8월 ~ 현재 : 한양대학교 안산캠퍼스 전자공학과 조교수
※주관심분야 : 신호처리, 적응필터, 디지틀 통신 확률이동