ANALYSIS OF THE SIGN-SIGN ALGORITHM BASED ON GAUSSIAN DISTRIBUTED TAP WEIGHTS

Shin'ichi Koike

NEC Corporation
Tokyo 108-01 Japan
koike@epc.ho.nec.co.jp

ABSTRACT

In this paper, a new set of difference equations is derived for convergence analysis of adaptive filters using the Sign-Sign Algorithm with Gaussian input reference and additive Gaussian noise. The analysis is based on the assumption that the tap weights are jointly Gaussian distributed. Residual mean squared error after convergence and simpler approximate difference equations are further developed. Results of experiment exhibit good agreement between theoretically calculated convergence and that of simulation for a wide range of parameter values of adaptive filters.

I. INTRODUCTION

Sign-Sign Algorithm is said to be a member of a "family" of the LMS Algorithm, in which non-linearities are introduced into the correlation multiplier for the stochastic gradient tap weight adaptation. The algorithm makes use of the signum of the input reference signal in addition to taking the polarity of the error signal, thus requiring only one-bit multiplication or logical EX-OR function. Like his "brother" Sign Algorithm, the Sign-Sign Algorithm is also attractive for robustness against disturbances.

Although the Sign-Sign Algorithm finds many application areas, *e.g.*, adaptive predictive coding [1], there seem to be a limited number of papers which analyze the algorithm [2]-[5]. [5] deals with convergence analysis of the Sign-Sign Algorithm based on the assumption that the input reference and the error signals are jointly Gaussian distributed.

In this paper, we develop a new set of difference equations for repetitive calculation of the mean squared error (MSE) convergence of adaptive filters using the Sign-Sign Algorithm with Gaussian input reference and Gaussian additive noise. In the analysis it is assumed that the tap weights are jointly Gaussian distributed along their adaptation process [6] [7].

In Section II difference equations for the mean and the covariance of "tap weight error" vector are derived, involving integrals of probability generating functions. Section III gives a theoretical expression for the residual MSE after convergence. In Section IV, it is shown that approximation to the equations derived in Section II yields simpler difference equations that coincide with the formulae obtained in [5]. In Section V, results of experiment with some examples are given, comparing the theoretically calculated convergence with that of simulation. Section VI concludes the paper.

II. ANALYSIS

The tap weight update equation for the Sign-Sign Algorithm is given by

$$\boldsymbol{c}^{(n+1)} = \boldsymbol{c}^{(n)} + \alpha_c \, sgn(e_n + v_n) \, sgn(\boldsymbol{a}^{(n)}) \tag{1}$$

where

n time instant, $\boldsymbol{c}^{(n)}$ tap weignt vector (N taps),
$\boldsymbol{a}^{(n)} = [a_n, \cdots, a_{n-N+1}]^T$ input reference vector,
e_n error, v_n additive noise, α_c step size,
$sgn(\bullet)$ signum function, and $(\bullet)^T$ transpose.

Let "tap weight error" vector be defined by $\boldsymbol{\theta}^{(n)} = \boldsymbol{h} - \boldsymbol{c}^{(n)}$, where $\boldsymbol{h}$ is the response vector of the unknown system in system identification application. Then we have

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - \alpha_c \, sgn(e_n + v_n) \, sgn(\boldsymbol{a}^{(n)}) \tag{2}$$

and

$$e_n = \boldsymbol{a}^{(n)T} \boldsymbol{\theta}^{(n)}. \tag{3}$$

For the subsequent analysis, the following Assumptions are made.

(A) Input reference signal a_n is a stationary Gaussian process, colored in general, with zero mean, covariance $\mathbf{Ra} = E\left[\boldsymbol{a}^{(n)}\boldsymbol{a}^{(n)T}\right]$ and variance σ_a^2,

(B) Additive noise v_n is Gaussian with variance σ_v^2,

(C) *Tap weight error* $\boldsymbol{\theta}^{(n)}$ *is jointly Gaussian distributed with mean* $\boldsymbol{m}^{(n)} = E\left[\boldsymbol{\theta}^{(n)}\right]$ *and covariance*
$$\mathbf{R}^{(n)} = E\left[\left(\boldsymbol{\theta}^{(n)} - \boldsymbol{m}^{(n)}\right)\left(\boldsymbol{\theta}^{(n)} - \boldsymbol{m}^{(n)}\right)^T\right], \text{ and}$$

(D) Input reference $\boldsymbol{a}^{(n)}$ and tap weight error $\boldsymbol{\theta}^{(n)}$ are statistically independent.

From (2) and (3), under the Assumptions (A) through (D) above, we can derive the following difference equations.

$$\boldsymbol{m}^{(n+1)} = \boldsymbol{m}^{(n)} - \alpha_c \, \boldsymbol{p}^{(n)} \tag{4}$$

and

$$\mathbf{R}^{(n+1)} = \mathbf{R}^{(n)} - \alpha_c\left(\mathbf{U}^{(n)} + \mathbf{U}^{(n)T}\right) + \alpha_c^2\left(\mathbf{Ta} - p^{(n)}p^{(n)T}\right), \quad (5)$$

where

$$p^{(n)} = E\left[sgn(e_n + v_n)sgn\left(a^{(n)}\right)\right], \quad (6)$$

$$\mathbf{U}^{(n)} = E\left[sgn(e_n + v_n)sgn\left(a^{(n)}\right)\left(\theta^{(n)} - m^{(n)}\right)^T\right], \quad (7)$$

and

$$\mathbf{Ta} = E\left[sgn\left(a^{(n)}\right)sgn\left(a^{(n)}\right)^T\right] = (2/\pi)sin^{-1}\left(\mathbf{Ra}/\sigma_a^2\right). \quad (8)$$

The MSE is calculated as

$$\varepsilon^{(n)} = E\left[e_n^2\right] = tr\left\{\mathbf{Ra}\left(m^{(n)}m^{(n)T} + \mathbf{R}^{(n)}\right)\right\}, \quad (9)$$

where $tr(\bullet)$ is trace of matrix.

Now, with the Assumptions (B) through (D), *given the value of* $a^{(n)}$, $y = e_n + v_n = a^{(n)T}\theta^{(n)} + v_n$ *is a Gaussian random variable, and* y and $\theta^{(n)}$ *are jointly Gaussian disributed.*

Then, from (6) and (7), we find

$$p^{(n)} = (1/\pi)\int_{-\infty}^{\infty}(j/\omega)\,E_a\left[\mathrm{sgn}(a)\Phi_Y(\omega)\right]d\omega, \quad (10)$$

$$\mathbf{U}^{(n)} = \mathbf{W}^{(n)}\mathbf{R}^{(n)}, \quad (11)$$

and

$$\mathbf{W}^{(n)} = (1/\pi)\int_{-\infty}^{\infty}E_a\left[sgn(a)a^T\Phi_Y(\omega)\right]d\omega, \quad (12)$$

where

$$\Phi_Y(\omega) = exp\left(-j\,\omega\,m^{(n)T}a - \omega^2\,a^T\mathbf{R}^{(n)}a/2\right) \times exp\left(-\sigma_v^2\,\omega^2/2\right) \quad (13)$$

is the probability generating function of y, $E_a[\bullet]$ denotes expectation with respect to a, and $j = \sqrt{-1}$.

Using the Gaussian probability density function $p_A(a)$ of a, we calculate

$$E_a\left[sgn(a_{n-k})\Phi_Y(\omega)\right] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}sgn(a_{n-k})\,\Phi_Y(\omega)p_A(a)da = 2erf\left\{-j\,\omega\,\mu_k^{(n)}(\omega)/\sqrt{D_{kk}^{(n)}(\omega)}\right\} \times \Psi^{(n)}(\omega)\,exp\left(-\sigma_v^2\,\omega^2/2\right) \quad (14)$$

and

$$E_a\left[sgn(a_{n-k})a_{n-\kappa}\Phi_Y(\omega)\right] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}sgn(a_{n-k})\,a_{n-\kappa} \times \Phi_Y(\omega)p_A(a)da = (j/\omega)\,\partial E_a\left[\mathrm{sgn}(a_{n-k})\Phi_Y(\omega)\right]/\partial m_\kappa^{(n)}, \quad (15)$$

where

$$\Psi^{(n)}(\omega) = exp\left\{-\omega^2 m^{(n)T}\mu^{(n)}(\omega)/2\right\}/\sqrt{\left|\mathbf{A}^{(n)}(\omega)\right|}, \quad (16)$$

$$\mu^{(n)}(\omega) = \mathbf{D}^{(n)}(\omega)\,m^{(n)} = \left[\mu_0^{(n)}(\omega), \mu_1^{(n)}(\omega), \ldots, \mu_{N-1}^{(n)}(\omega)\right]^T, \quad (17)$$

$$\mathbf{D}^{(n)}(\omega) = \mathbf{A}^{(n)}(\omega)^{-1}\,\mathbf{Ra} = \left[D_{k\kappa}^{(n)}(\omega)\right], \quad (18)$$

$$\mathbf{A}^{(n)}(\omega) = \mathbf{I} + \omega^2\,\mathbf{Ra}\,\mathbf{R}^{(n)}, \quad (19)$$

and

$$erf(x) = \left(1/\sqrt{2\pi}\right)\int_0^x exp\left(-t^2/2\right)dt. \quad (20)$$

From (10) with (14) substituted, one can calculate the k-th elememt of the vector $p^{(n)}$ as

$$p_k^{(n)} = \sqrt{2/\pi}(1/\pi)\int_{-\infty}^{\infty}\rho_k^{(n)}(\omega)G\left(\omega\rho_k^{(n)}(\omega)\right) \times \Psi^{(n)}{}_{(k)}(\omega)exp\left(-\sigma_v^2\,\omega^2/2\right)d\omega, \quad (21)$$

where defined are

$$\rho_k^{(n)}(\omega) = \mu_k^{(n)}(\omega)/\sqrt{D_{kk}^{(n)}(\omega)}, \quad (22)$$

$$\Psi^{(n)}{}_{(k)}(\omega) = exp\left\{-\omega^2\left(m^{(n)T}\mu^{(n)}(\omega) - \left(\rho_k^{(n)}(\omega)\right)^2\right)/2\right\} / \sqrt{\left|\mathbf{A}^{(n)}(\omega)\right|}, \quad (23)$$

$$G(x) = F(x)/x = \int_0^1 exp\left\{-x^2(1-\lambda)^2/2\right\}d\lambda, \quad (24)$$

and

$$F(x) = exp\left(-x^2/2\right)\int_0^x exp\left(u^2/2\right)du \quad (25)$$

is called *Dawson's Integral* [8]. Fig.1 shows the graph of $G(x)$.

The (k,κ)-th element of the matrix $\mathbf{W}^{(n)}$ can also be calculated from (15) as

$$\mathrm{W}_{k\kappa}^{(n)} = \sqrt{2/\pi}(1/\pi) \times \int_{-\infty}^{\infty}\left\{\mathrm{D}_{k\kappa}^{(n)}(\omega) - \omega^2 G\left(\omega\rho_k^{(n)}(\omega)\right)\mu_k^{(n)}(\omega)\mu_\kappa^{(n)}(\omega)\right\} / \sqrt{\mathrm{D}_{kk}^{(n)}(\omega)}\,\Psi^{(n)}{}_{(k)}(\omega)exp\left(-\sigma_v^2\,\omega^2/2\right)d\omega. \quad (26)$$

Thus $p^{(n)}$ and $\mathbf{W}^{(n)}$ are obtained in (21) and (26), respectively, as integrals with respect to ω for which numerical integration is usually required. The vector $p^{(n)}$ and the matrix $\mathbf{U}^{(n)}$ (see (11)) complete the difference equations (4) and (5), further enabling the calculation of the MSE convergence.

III. RESIDUAL MSE AFTER CONVERGENCE

In this section, assuming the filter convergence, we will solve the residual MSE theoretically. If the filter converges as $n \to \infty$, then clearly $m^{(\infty)} = p^{(\infty)} = \mu^{(\infty)}(\omega) = 0$ and $\rho_k^{(\infty)}(\omega) = 0$.

From (5) and (11), an equation

$$\mathbf{W}^{(\infty)}\mathbf{R}^{(\infty)} + \mathbf{W}^{(\infty)}\mathbf{R}^{(\infty)} = \alpha_c\,\mathbf{Ta} \quad (27)$$

results, where

$$\mathrm{W}_{k\kappa}^{(\infty)} = 2/\pi/\sqrt{2\pi}\int_{-\infty}^{\infty}\mathrm{D}_{k\kappa}^{(\infty)}(\omega)/\sqrt{\mathrm{D}_{kk}^{(\infty)}(\omega)} / \sqrt{\left|\mathbf{A}^{(\infty)}(\omega)\right|}\,exp\left(-\sigma_v^2\,\omega^2/2\right)d\omega. \quad (28)$$

If we approximate $\mathbf{R}^{(\infty)} \cong \left(\varepsilon^{(\infty)}/\sigma_a^2/N\right)\mathbf{I}$, $1/\sqrt{\left|\mathbf{A}^{(\infty)}(\omega)\right|} \cong 1 - \varepsilon^{(\infty)}\omega^2/2$, and $\mathrm{D}_{kk}^{(\infty)}(\omega) \cong \sigma_a^2$, then

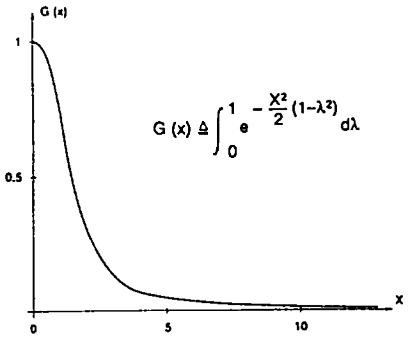

Fig.1 Graph of function $G(x)$.

$$\mathbf{W}^{(\infty)} \cong 2/\pi/\sigma_a/\sqrt{\varepsilon^{(\infty)}+\sigma_v^2}$$
$$\times\left\{\mathbf{Ra}-\varepsilon^{(\infty)}/\left(\varepsilon^{(\infty)}+\sigma_r^2\right)\mathbf{Ra}^2/\sigma_a^2/N\right\}. \quad (29)$$

From (27) with (29) one obtains approximate theoretical expression for the residual MSE after convergence as follows.

$$\varepsilon^{(\infty)} \cong (\alpha_c/2)(\pi/2)\,\sigma_a N\sqrt{\varepsilon^{(\infty)}+\sigma_v^2}$$
$$\times\left\{1+(\alpha_c/2)(\pi/2)(\sigma_a N/\sigma_v)\,tr\left(\mathbf{Ra}^2\right)/\left(\sigma_a^2 N\right)^2\right\}$$
$$\cong(\alpha_c/2)(\pi/2)\,\sigma_a N\sigma_v, \quad (30)$$

for a sufficiently small step size or for $\varepsilon^{(\infty)} \ll \sigma_v^2$.

Theoretical value of the residual MSE often helps us determine the step size which meets the requirement for the estimation accuracy in the adaptive filter design.

IV. APPROXIMATE DIFFERENCE EQUATIONS

In this section, we try to integrate (21) and (26) in Section II analytically, through approximation, to obtain simpler expressions for $p^{(n)}$ and $\mathbf{W}^{(n)}$.

First, let us define

$$\mathbf{Sa}^{(n)} = \mathbf{Ra}\ \mathbf{R}^{(n)}, \quad (31)$$

and assume $\left|\mathbf{Sa}^{(n)}/\sigma_v^2\right| \ll 1$. Then, from (19)

$$\mathbf{A}^{(n)}(\omega)^{-1} \cong \mathbf{I}-\omega^2\,\mathbf{Sa}^{(n)}. \quad (32)$$

By approximation using (32), (21) becomes analytically integrable as follows.

$$p_k^{(n)} \cong 2/\pi/\sqrt{2\pi}\int_{-\infty}^{\infty}\left(\mu_{ak}^{(n)}/\sigma_a\right)G\left(\omega\mu_{ak}^{(n)}/\sigma_a\right)$$
$$\times\, exp\left\{-\omega^2\left(\varepsilon^{(n)}-\left(\mu_{ak}^{(n)}/\sigma_a\right)^2+\sigma_v^2\right)/2\right\}d\omega, \quad (33)$$

where

$$\mu_a^{(n)} \triangleq \mathbf{Ra}\ \boldsymbol{m}^{(n)}=\left[\mu_{a0}^{(n)},\mu_{a1}^{(n)},\ldots,\mu_{aN-1}^{(n)}\right]^T. \quad (34)$$

Performing the integration of (33) using a formula

$$\left(1/\sqrt{2\pi}\right)\int_{-\infty}^{\infty}G(at)\,exp\left(-t^2/2\right)dt$$
$$=(1/a)\,sin^{-1}\left(a/\sqrt{1+a^2}\right) \quad (35)$$

results in

$$p_k^{(n)} \cong (2/\pi)\,sin^{-1}\left(\mu_{ak}^{(n)}/\sigma_a/\sqrt{\varepsilon^{(n)}+\sigma_v^2}\right), \quad (36)$$

and further differentiating (36) with respect to $m_k^{(n)}$ yields

$$w_{kk}^{(n)} \cong (2/\pi)\mathrm{Ra}_{kk}/\sigma_a/\sqrt{\varepsilon^{(n)}-\left(\mu_{ak}^{(n)}/\sigma_a\right)^2+\sigma_v^2}\,. \quad (37)$$

(4) and (5) combined with (11), (36) and (37) form approximate difference equations *(ADEs)* for calculating filter convergence with less computing time. Note that (36) and (37) are exactly the same formulae derived in [5].

V. EXPERIMENT

In order to examine the accuracy of theoretically calculated convergence of adaptive filters using the proposed difference equations, we have done experiment with some examples, in which simulation results and theoertical ones are compared. Also compared are convergence curves calculated with the formulae in [5], or *(ADEs)* with (36) and (37). In the experiment the simulation result is an ensemble average over *1000* independent runs of the filter convergence.

Three examples are carefully prepared. In these examples, input reference signal is *AR1* modelled with regression coefficient η.

Example #1 $\quad N=4,\quad \sigma_a^2=1,\quad \eta=.5$

$$\boldsymbol{h}=[.05,\ .994,\ .01,\ -1]^T$$
$$\sigma_v^2=.01\ (-20\,dB),\ \alpha_c=2^{-11}$$
$$\rightarrow\quad \varepsilon^{(\infty)}\cong -38\,dB\quad \ll\quad \sigma_v^2$$

Example #2 $\quad N=2,\quad \sigma_a^2=1,\quad \eta=.8$

$$\boldsymbol{h}=[\,.995,\ -1]^T$$
$$\sigma_v^2=.0001\ (-40\,dB),\quad \alpha_c=2^{-4}$$
$$\rightarrow\quad \varepsilon^{(\infty)}\cong -21\,dB\quad \gg\quad \sigma_v^2$$

Example #3 $\quad N=8,\quad \sigma_a^2=1,\quad \eta=0$ (W&G input)

$$\boldsymbol{h}=[.01,\ .881,\ .05,\ -.572,\ -.1,\ -.05,\ -.02,\ -.01]^T$$
$$\sigma_v^2=.001\ (-30\,dB),\quad \alpha_c=2^{-8}$$
$$\rightarrow\quad \varepsilon^{(\infty)}\cong -30\,dB\quad \cong\quad \sigma_v^2$$

Fig.2 shows the convergence curves for Example #1, where excellent agreement is observed between the simulation result and the ones theoretically calculated with the proposed difference equations and with the formulae in [5] *(ADEs)*.

Fig.3 illustrates the convergence for Example #2 in which the step size is chosen fairly large. While the convergence calculated with the *ADEs* exhibits poor accuracy, the proposed method gives a theoretical curve in good agreement with the empirical one.

In Fig.4 the convergence curves are shown for Example #3. The proposed difference equations estimate the theoretical convegence that well agrees with the simulation result. However, the convergence curve based on the *ADEs* slightly deviates from the above one.

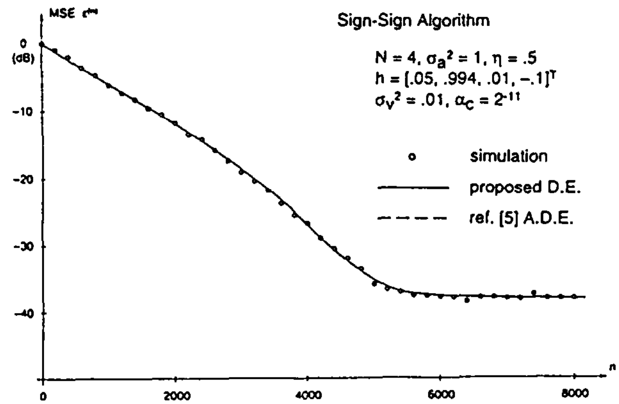

Fig.2 Filter convergence (Example #1, $N=4$, $\varepsilon^{(\infty)} \ll \sigma_v^2$).

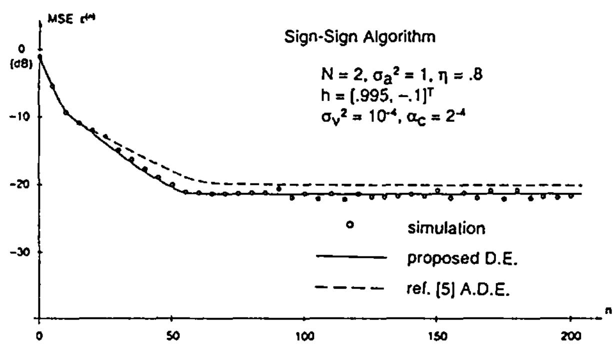

Fig.3 Filter convergence (Example #2, $N=2$, $\varepsilon^{(\infty)} \gg \sigma_v^2$).

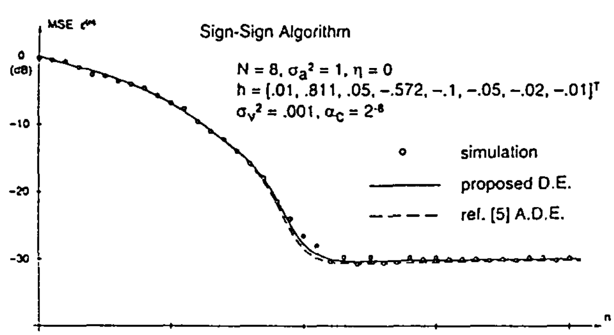

Fig.4 Filter convergence (Example #3, $N=8$, $\varepsilon^{(\infty)} \cong \sigma_v^2$).

In summary, we observe that the proposed set of difference equations has sufficient accuracy in calculating the filter convergence for a wide range of parameter values. However as was mentioned, the equations require numerical integration, and therefore a large amount of computation.

VI. CONCLUSION

In this paper, a new set of difference equations has been proposed for convergence analysis of adaptive filters using the Sing-Sing Algorithm with Gaussian input reference and additive Gaussian noise. The analysis is based on the assumption that the tap weights are jointly Gaussian distributed, and yields equations expressed as integrals involving probability generating functions.

Theoretical formula for the residual MSE after convergence has been solved for use in the filter design, and simpler approximate difference equations have been further developed.

Experimental results show that the theoretically calculated convergence with the proposed diffence equations exhibits better agreement with that of simulation than the approximate difference equations, and has sufficient accuracy for a wide range of parameter values of adaptive filters.

Although the proposed difference equations are proven useful for practical design, further improvement may be required to reduce the computing time.

ACKNOWLEDGEMENT

The author would like to thank his colleagues in C&C Media Laboratory, NEC Central Research Laboratories for their helpful comments and suggestions.

REFERENCES

[1] ITU-T *Recommendation G.726* (former CCITT *Recommendation G.721*), ITU, 1994.

[2] N.J.Bershad, "On the Optimum Data Nonlinearity in LMS Adaptation," *IEEE Trans. on ASSP*, Vol.34, No.2, pp.69-76, February 1986.

[3] C.E.Rohrs, C.R.Johnson, and J.D.Mills, "A Stability Problem in Sign-Sign Adaptive Algorithms," *IEEE Proc. ICASSP86*, pp.2999-3001, Tokyo, April 1986.

[4] W.A.Sethares, "Adaptive Algorithms with Nonlinear Data and Error Functions," *IEEE Trams. on SP*, Vol.40, No.9, pp.2199-2206, September 1992.

[5] S.Koike, "New Recurrence Formulae for Convergence Analysis of Adaptive Filters Using the Sign-Sign Algorithm," *IEEE Proc. ISPACS94*, pp.247-251, Seoul, Korea, October 1994.

[6] S.Koike, "Convergence Analysis of a Data Echo Canceller with a Stochastic Gradient Adaptive FIR Filter Using the Sign Algorithm," *IEEE Trans. on SP*, Vol.43, No.12, pp.2852-2862, December 1995.

[7] S.Koike, "Analysis of an Echo Canceller with an Adaptive FIR Filter Using the Sign Algorithm for Gaussian Transmit and Receive Signals," *IEEE Proc. ICASSP94*, Vol.III, pp.361-364, Adelaide, Australia, April 1994.

[8] M.Abramowitz and I.A.Stegun, *Handbook of Mathematical Functions*, Chap.7, New York: Dover, 1965.