

Blind Source Separation by Auto-Associative Neural Network with Pruning

Syozo Yasui

Graduate School of Life Science and Systems Engineering
Kyushu Institute of Technology, Iizuka, 820-8502 Japan
E-mail: yasui@ces.kyutech.ac.jp

Abstract

This paper describes a non-information theoretic approach to the Blind Source Separation (BSS) problem by using an auto-associative neural network (AANN) of the conventional type. There is no special computation explicitly intended for BSS, except for a pruning mechanism to deal with the usual case in which the number of the sources is unknown; the nonlinear hidden units that have survived the pruning would recover the source signals separately. A computer simulation study is presented, including an example to show adaptability of the present approach. A mathematical analysis is also made to relate BSS with local minima of the identity transformation error in the AANN.

1 Introduction

Blind Source Separation (BSS) is a task to recover independent source signals from their mixture. In BSS, information about the sources is missing regarding their statistical properties as well as their number. Also, the source-sensor mixing matrix is usually not known either. The only key is the assumption that the sources are statistically independent. The major approaches include informax, maximum likelihood estimation, negentropy maximization, and nonlinear PCA (Principal Component Analysis). These are summarily called as ICA (Independent Component Analysis). The first three are information-theoretic and have been shown to be equivalent (e.g., [1]). The nonlinear PCA can also be viewed from information-theoretic principles [1].

The present approach to BSS uses a conventional type of auto-associative neural network (AANN). A similar AANN architecture was described previously in the context of BSS [2]. In that earlier

work, while a learning rule is obtained by nonlinear PCA to find the sensor-hidden de-mixing matrix that corresponds to the encoder part of AANN, the hidden-output decoder part will become useful if one wishes to know the mixing matrix in addition to the blind sources. The present approach, in contrast, lets the whole AANN (both encoder and decoder together) work for BSS, with the working idea that BSS would be attained as a local minimum associated with the error in the input-output identity transform by the AANN. Thus, there is no special computation intended for BSS *per se*. Another unique aspect is the way to find the unknown number of the source signals. In the ICA approaches, pre-whitening is often helpful to this end. In the present AANN framework, a pruning algorithm is applied so that only the necessary and sufficient hidden units would survive to reproduce the blind sources.

2 Architecture

Figure 1 shows the AANN architecture. A set of M sensors receive a linear mixture $\mathbf{x}^o = [x_1^o, \dots, x_M^o]$ of N statistically independent zero-mean source signals, $\mathbf{s} = [s_1, \dots, s_N]$, $M \geq N$. Thus, $\mathbf{x}^o = \mathbf{A} \mathbf{s}^T$ with $\mathbf{A} = [a_{ij}]$ being the mixing matrix. The sensor signals are each scaled such that $x_i = \kappa_i x_i^o$ to give $\langle x_i^2 \rangle = d$ ($\forall i$) which is a preset constant. Here, $\langle \rangle$ stands for time average. This adjustment is useful for consistency of the BSS performance. $\mathbf{x}^T = [x_1, \dots, x_M]^T = \mathbf{K} \mathbf{A} \mathbf{s}^T$ is the actual input vector to the AANN, where \mathbf{K} is the appropriate diagonal matrix for the sensor signal scaling.

A total of L hidden units are made available for use ($L \geq N$), and their activities are denoted by $[h_1, \dots, h_L] = \mathbf{h}$ and produced by $h_i = f(z_i)$, $\forall i$. Here,

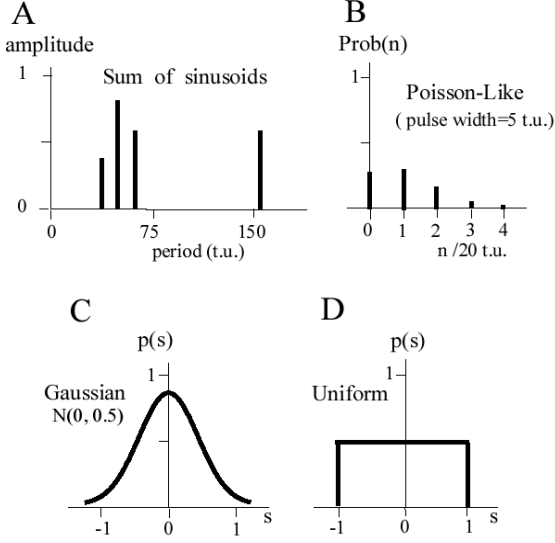


Figure 2: The source signals used for the test.

were used for $N=2$, and $M=5$ and $L=6$ for $N=3$. The results are shown in Table 1. The criteria for the “BSS success” evaluation are explained below.

The effect of the CSDF pruning was continuously evaluated during each simulation run in such a way that the j -th hidden unit was pronounced active if $\max(|w_{ij}|) > 0.1$, and extinct if $\max(|w_{ij}|) < 0.02, \forall i, j$. Also, to monitor the correlation between the sources and hidden-units, the following $L \times N$ correlation matrix $\mathbf{Q} = [q_{ij}]$ was computed every 5,000 t.u.

$$q_{ij} = |\langle z_i s_j \rangle| / (\langle z_i^2 \rangle \langle s_j^2 \rangle)^{1/2} \quad (1)$$

In a report elsewhere[4], u_i is used in place of z_i to compute q_{ij} , but there is no significant difference.

If the number of the active hidden units and that of the extinct hidden units become N and $L - N$, respectively, then \mathbf{Q} deserves to be examined for BSS. The relevant portion of \mathbf{Q} that relates the sources and the active hidden units should form an $N \times N$ square matrix. And this would have to be the identity matrix or its row-or-column permutations, if BSS were perfect. For the practical purpose, however, the following approximate judgment was used. Thus, the higher (θ_H) and lower (θ_L) thresholds were preset for closeness to ones and zeroes, respectively, of the ideal $N \times N$ correlation matrices.

And the simulation run was judged as successful and terminated when this matrix became close enough to one of the ideal matrices (under the θ_H / θ_L criterion) before 150,000 t.u. ($N=2$) or 300,000 t.u. ($N=3$).

The results are shown in Table 1. The BSS success rate was somewhat reduced with $N=3$ in comparison with $N=2$. This was especially so in the “A+B+C” case in which “C” was Gaussian (Fig.2). This might have something to do with the well-known restriction that no more than one Gaussian source is allowed for BSS.

Figure 3 shows how adaptively the AANN dealt with abrupt increase of sources. Namely, “C” was added when BSS for “A” and “B” had been practically attained (clearing the $\theta_H / \theta_L = 0.98/0.10$ criterion) with two surviving hidden units. After some transient increase of the identity transform error and re-arrangement of \mathbf{W} as well as of \mathbf{V} (not shown in Fig.3), the AANN was restored for BSS of the three sources. Correspondingly, one hidden unit became re-activated to represent “C”. The demonstrated set of the recovery records for this $N=3$ example cleared $\theta_H / \theta_L = 0.97/0.15$.

| θ_H / θ_L | A+B | A+B+C | A+B+D |
|-----------------------|------|-------|-------|
| 0.90 / 0.30 | 99.2 | 89.1 | 93.3 |
| 0.95 / 0.25 | 98.6 | 81.9 | 89.1 |
| 0.97 / 0.15 | 94.9 | 60.2 | 73.6 |

Table 1: BSS success rates (% , 1,000 runs for each score) for 2- and 3-source mixtures from Fig.2, under the θ_H / θ_L criterion in the text.

6 Mathematical Discussion and Remarks

The previous section has provided experimental data supporting the AANN approach to BSS. In this section, a preliminary mathematical analysis is made for the two-source case by assuming success of the CSDF pruning, so that $N=L=2$. It suffices to consider only e_1 from $\mathbf{e} = [e_1, \dots, e_M]$ of the identity mapping. Also, without loss of generality, the sensor signal adjustment is not considered ($\mathbf{K}=\mathbf{I}$), so that $\mathbf{x}=\mathbf{x}^o$. Since $E \equiv \langle e_1^2 \rangle = \langle (w_{11}z_1 + w_{12}z_2 - a_{11}s_1 - a_{12}s_2 + b_1)^2 \rangle$, $\partial E / \partial b_1 = 0$ gives $b_1 + w_{11}\langle z_1 \rangle + w_{12}\langle z_2 \rangle = 0$. Also, since $u_i = h_i - \langle h_i \rangle$, it follows that

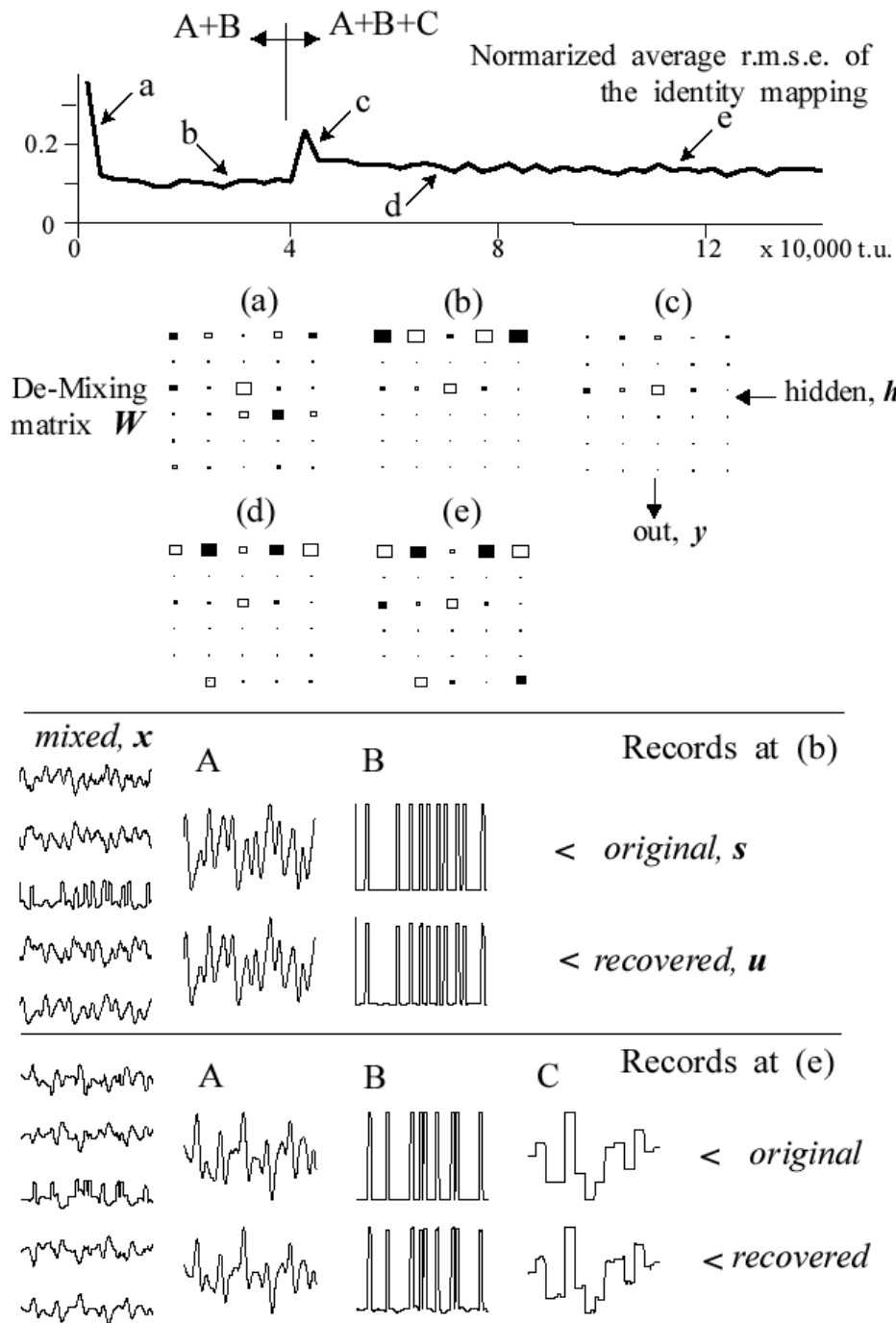


Figure 3: A sample set of simulation records relevant to the situation in which a third source ("C" from Fig.2) suddenly comes to join in the mixing. The top trace plots $[(\langle e_1^2 \rangle / \langle x_1^2 \rangle)^{1/2} + \dots + (\langle e_M^2 \rangle / \langle x_M^2 \rangle)^{1/2}] / M$.

$E = \langle (w_{11} u_1 + w_{12} u_2 - a_{11} s_1 - a_{12} s_2)^2 \rangle$. Let w_{i1}^* ($i=1,2$) be w_{i1} that makes $\partial E / \partial w_{i1} = \langle u_i e_1 \rangle = 0$. Correspondingly, E is denoted by E^* . Then, one finds

$$\begin{aligned} G_1 &\equiv w_{11}^* \langle u_1^2 \rangle + w_{12}^* \langle u_1 u_2 \rangle - \sum_{j=1}^2 a_{1j} \langle s_j u_1 \rangle = 0, \\ G_2 &\equiv w_{11}^* \langle u_1 u_2 \rangle + w_{12}^* \langle u_2^2 \rangle - \sum_{j=1}^2 a_{1j} \langle s_j u_2 \rangle = 0, \end{aligned} \quad (2)$$

and

$$\begin{aligned} E^* &= - (a_{11} \langle s_1 u_1 \rangle + a_{12} \langle s_2 u_1 \rangle) w_{11}^* - (a_{11} \langle s_1 u_2 \rangle \\ &\quad + a_{12} \langle s_2 u_2 \rangle) w_{12}^* + \sum_{i=1}^2 a_{1i}^2 \langle s_i^2 \rangle. \end{aligned} \quad (3)$$

One can also write $u_1 = f(c_{11} s_1 + c_{12} s_2)$ and $u_2 = f(c_{21} s_1 + c_{22} s_2)$ where $[c_{ij}] \equiv \mathbf{C} = \mathbf{V}\mathbf{A}$, $i, j=1,2$. Let

$$J = E^* + \lambda_1 G_1 + \lambda_2 G_2, \quad (4)$$

where λ_1 and λ_2 are Lagrange multipliers. In calculating $\partial J / \partial c_{11}, \partial J / \partial c_{12}, \partial J / \partial c_{21}$ and $\partial J / \partial c_{22}$ from (2-4), one has terms of the form $\partial u_k / \partial c_{ij}$ ($i, j, k=1,2$). Let u_k' denote $df(\xi)/d\xi$ with $\xi = c_{k1} s_1 + c_{k2} s_2$ ($k=1,2$). Then, it is clear that $\partial u_1 / \partial c_{11} = s_1 u_1'$, $\partial u_1 / \partial c_{12} = s_2 u_1'$, $\partial u_1 / \partial c_{12} = s_2 u_1'$, $\partial u_1 / \partial c_{21} = \partial u_1 / \partial c_{21} = 0$, $\partial u_2 / \partial c_{11} = \partial u_2 / \partial c_{12} = 0$, $\partial u_2 / \partial c_{21} = s_1 u_2'$ and $\partial u_2 / \partial c_{22} = s_2 u_2'$. BSS would be attainable if E^* has a local minimum at $c_{12} = c_{21} = 0$ or at $c_{11} = c_{22} = 0$. It suffices to consider only the former case below. Thus, for instance, $\langle s_1 s_2 u_1' \rangle = \langle s_2 \rangle \langle s_1 u_1' \rangle = 0$. Other ‘‘cross-average terms’’ such as $\langle s_1 u_1' u_2 \rangle, \langle s_2 u_1 u_1' \rangle, \langle s_1 s_2 u_2' \rangle, \langle s_1 u_2 u_2' \rangle, \langle s_1 s_2 u_2' \rangle$, and $\langle s_1 s_2 u_2' \rangle$ vanish as well. Furthermore, note that $w_{11}^* = -a_{11} \langle s_1 u_1 \rangle / \langle u_1^2 \rangle$ and $w_{12}^* = -a_{12} \langle s_2 u_2 \rangle / \langle u_2^2 \rangle$, as is evident from (2). From all this, one can show that, at $c_{12} = c_{21} = 0$,

$$\begin{aligned} \partial J / \partial c_{11} &= -a_{11} \langle s_1^2 u_1' \rangle \{ a_{11} \langle s_1 u_1 \rangle / \langle u_1^2 \rangle - \lambda_1 [2 \langle s_1 u_1 \rangle \langle s_1 u_1 u_1' \rangle / (\langle u_1^2 \rangle \langle s_1^2 u_1' \rangle) - I] \} \\ \partial J / \partial c_{12} &= -\langle s_2^2 \rangle \langle u_1 \rangle \{ a_{11} a_{12} \langle s_1 u_1 \rangle / \langle u_1^2 \rangle - a_{12} \lambda_1 [\langle s_2 u_2 \rangle^2 / (\langle s_2^2 \rangle \langle u_2^2 \rangle) - I] - a_{11} \lambda_2 \langle s_1 u_1 \rangle \langle s_2 u_2 \rangle / (\langle s_2^2 \rangle \langle u_1^2 \rangle) \} \\ \partial J / \partial c_{21} &= -\langle s_1^2 \rangle \langle u_2 \rangle \{ a_{11} a_{12} \langle s_2 u_2 \rangle / \langle u_2^2 \rangle - a_{12} \lambda_1 \langle s_1 u_1 \rangle \langle s_2 u_2 \rangle / (\langle s_1^2 \rangle \langle u_2^2 \rangle) - a_{11} \lambda_2 [\langle s_1 u_1 \rangle^2 / (\langle s_1^2 \rangle \langle u_1^2 \rangle) - I] \} \\ \partial J / \partial c_{22} &= -a_{12} \langle s_2^2 u_2' \rangle \{ a_{12} \langle s_2 u_2 \rangle / \langle u_2^2 \rangle - \lambda_2 [2 \langle s_2 u_2 \rangle \langle s_2 u_2 u_2' \rangle / (\langle u_2^2 \rangle \langle s_2^2 u_2' \rangle) - I] \} \end{aligned} \quad (5)$$

By using the approximation $\langle s_i u_i \rangle \langle s_i^2 u_i' \rangle / (\langle s_i^2 \rangle \langle s_i u_i u_i' \rangle) \approx I$ ((A3) in Appendix) and noting $\langle s_i^2 u_i' \rangle = \langle s_i^2 \rangle \langle u_i' \rangle \neq 0$ for both $i=j$ and $i \neq j$, and letting $\zeta_i = \langle s_i u_i \rangle / \langle u_i^2 \rangle$ for $i, j=1,2$, putting $\partial J / \partial c_{11} = \partial J / \partial c_{12} = \partial J / \partial c_{21} = \partial J / \partial c_{22} = 0$ in (5) gives

$$\begin{aligned} a_{11} \zeta_1 - \lambda_1 (2 \pi_1 \zeta_1 - I) &\approx 0, \\ a_{12} \zeta_2 - \lambda_2 (2 \pi_2 \zeta_2 - I) &\approx 0, \\ a_{11} a_{12} \zeta_1 - \lambda_1 a_{12} (\pi_2 \zeta_2 - I) - \lambda_2 a_{11} \pi_2 \zeta_1 &\approx 0, \\ a_{11} a_{12} \zeta_2 - \lambda_1 a_{12} \pi_1 \zeta_2 - \lambda_2 a_{11} (\pi_1 \zeta_1 - I) &\approx 0. \end{aligned} \quad (6)$$

Here, $\pi_1 = k_1(c_{11})$ and $\pi_2 = k_1(c_{22})$ are the coefficient of the first-order term in the orthogonal expansion of u_1 and u_2 , respectively (see Appendix). It follows from (6) that $\lambda_1 \approx a_{11} \zeta_1 = w_{11}^*$, $\lambda_2 \approx a_{12} \zeta_2 = w_{12}^*$, and

$$\pi_i \zeta_i = \langle s_i u_i \rangle^2 / (\langle u_i^2 \rangle \langle s_i^2 \rangle) \approx I, \quad i=1,2. \quad (7)$$

Notice that $\pi_i \zeta_i$ is closely related to the correlation matrix \mathbf{Q} defined by (1) for the BSS success criterion, *i.e.*, $\pi_i \zeta_i = q_{ii}^2$. If $f(\bullet)$ were purely linear, then one would have exactly $\pi_i \zeta_i = I$. In such linear case, however, $\mathbf{C} (= \mathbf{V}\mathbf{A})$ needs not be diagonal (*i.e.*, BSS) in order for $\mathbf{W}\mathbf{V}$ to be the identity matrix. In the nonlinear case, one actually has a Schwarz inequality, $\pi_i \zeta_i = \langle s_i u_i \rangle^2 / (\langle u_i^2 \rangle \langle s_i^2 \rangle) < I$ for $i=1,2$, rather than (7).

Thus, a separate simulation test was undertaken to examine how close to I the foregoing quantity of $\pi_i \zeta_i$ would be for each source signal of Fig. 2 if $\tanh(\bullet)$ is employed for the nonlinearity. Thus, $q^2 = \pi \zeta = \langle s u \rangle^2 / (\langle u^2 \rangle \langle s^2 \rangle)$ was computed with $u = \tanh(cs)$, with s and c being the source and parameter, respectively. The results are shown in Table 2. Note that ‘‘Source B’’ gives precisely $q^2 = I$ regardless of the value of c . This is not surprising, since this source uses only two values of $\tanh(\bullet)$ and hence the nonlinearity is irrelevant. Aside from this, the obvious general property that $q^2 \rightarrow I$ as $c \rightarrow 0$ can be seen.

What all this indicates is a tentative conclusion that local minima for the identity transformation exist near the exact BSS state of $\{c_{11} \neq 0, c_{12} = 0, c_{21} = 0, c_{22} \neq 0\}$. And, the values of the relevant parameters used in the present simulation study are thought to be appropriate enough to prevent the values of c_{11} and c_{22} from becoming too small for the nonlinearity to be effective in selecting a nearly

diagonal form of C that makes $WV \approx I$. Thus, as in the existing major methods [1], the nonlinearity plays an essential role for BSS. Furthermore, if the values of c_{11} or c_{22} are too small (*i.e.*, almost linear), then the activity of the relevant hidden unit would be correspondingly very small. This would usually demand the relevant elements of W to have huge values, in an attempt to attain $WV \approx I$. In fact, such simulation runs were occasionally encountered in this study, and they went unstable causing BSS to fail. In any event, further analysis remains to be made, especially for mathematical elucidation of the present AANN framework for BSS.

| C ↓ | A | B | C | D |
|--------|-------|-------|-------|-------|
| 0.5 | 0.994 | 1.000 | 0.998 | 1.000 |
| 1.0 | 0.960 | 1.000 | 0.984 | 0.995 |
| 1.5 | 0.917 | 1.000 | 0.959 | 0.983 |
| 2.0 | 0.879 | 1.000 | 0.932 | 0.967 |

Table 2: The fraction of the equivalent linearity, $q^2 = \langle su \rangle^2 / (\langle u^2 \rangle \langle s^2 \rangle)$ with $u = \tanh(cs)$, for each source signal of Fig. 2.

Appendix

Consider the following orthogonal expansion and the Parseval relation ($i=1,2$).

$$u_i = f(c_{ii}s_i) = l.i.m._{N \rightarrow \infty} \sum_1^N k_n(c_{ii}) H_n(s_i). \quad (A1)$$

$$\langle u_i^2 \rangle = \sum_1^\infty k_n^2(c_{ii}) \langle H_n^2(s_i) \rangle. \quad (A2)$$

Here, $\{H_n\}$ is the orthogonal basis that can be constructed by the Gram-Schmidt procedure and $k_n(c_{ii})$ is the expansion coefficient. Noting $\langle s_i u_i \rangle = \langle u_i H_1(s_i) \rangle = k_1 \langle s_i^2 \rangle$, it follows that $\langle s_i u_i u_i' \rangle / \langle s_i^2 u_i' \rangle = [(1/2)d \langle u_i^2 \rangle / dc_{ii}] / [d \langle s_i u_i \rangle / dc_{ii}] = k_1(c_{ii}) + \Sigma_2^\infty \{ [dk_n(c_{ii}) / dc_{ii}] / [dk_1(c_{ii}) / dc_{ii}] \} (\langle H_n^2 \rangle / \langle H_1^2 \rangle)$. If the higher order terms are negligible, then $\langle s_i u_i u_i' \rangle / \langle s_i^2 u_i' \rangle \approx k_1(c_{ii}) = \langle s_i u_i \rangle / \langle s_i^2 \rangle$, so that

$$\langle s_i u_i \rangle \langle s_i^2 u_i' \rangle / (\langle s_i^2 \rangle \langle s_i u_i' \rangle) \approx 1, i=1,2 \quad (A3)$$

The approximation above, which is exploited in Section 6, is made by ignoring $\Sigma_2^\infty [(dk_n(c_{ii}) / dc_{ii}) /$

$(dk_1(c_{ii}) / dc_{ii})] (\langle H_n^2 \rangle / \langle H_1^2 \rangle)$ which reflects the nonlinearity involved in the activation function $f(\bullet)$. Putting it in another way, $\pi_i \zeta_i$ in Section 6 represents the fraction of the equivalent linearity in the Parseval formula (A2), since $\pi_i \zeta_i = \langle s_i u_i \rangle^2 / (\langle u_i^2 \rangle \langle s_i^2 \rangle) = k_1^2(c_{ii}) / [\Sigma_1^\infty k_n^2(c_{ii}) \langle H_n^2(s_i) \rangle]$.

References

- [1] T-W. Lee, M. Girolami, A.J. Bell, and T.J. Sejnowski, "A Unifying Information-Theoretic Framework for Independent Component Analysis," *Computers and Mathematics with Applications*, vol. 39, pp.1-21, 2000.
- [2] E. Karhunen, L. Oja, R. Wang, R. Vigarito, and J.A. Joutsensalo, "A Class of Neural Networks for Independent Component Analysis," *IEEE Trans. on Neural Networks*, vol.8, pp.487-504, 1997.
- [3] S. Yasui, "Convergence Suppression and Divergence Facilitation: Minimum and Joint Use of Hidden Units by Multiple Outputs," *Neural Networks*, vol.10,no.2, pp.353-367,1997.
- [4] S. Yasui, "Conventional Auto-Associative Neural Network Separates Blind Sources without adding Intentional Algorithms other than Pruning (Submitted).