

# 적응적 학습파라미터를 이용한 독립성분분석의 성능개선

조용현, 민성재\*  
대구가톨릭대학교 공과대학 컴퓨터정보통신공학부  
e-mail:yhcho@cuth.cataegu.ac.kr

## Performance Improvement of Independent Component Analysis by Adaptive Learning Parameters

Yong-Hyun Cho, Seong-Jae Min\*  
School of Computer and Information Comm. Eng.,  
Catholic Univ. of Daegu

### 요약

본 연구에서는 뉴우턴법의 고정점 알고리즘에 적용 조정이 가능한 학습파라미터를 이용한 신경망 기반 독립성분분석기법을 제안하였다. 이는 고정점 알고리즘의 1차 미분을 이용하는 뉴우턴법에서 역혼합행렬의 경신 상태에 따라 학습율과 모멘트가 적응조정되도록 함으로써 분리속도와 분리성능을 개선시키기 위함이다. 제안된 기법을 512x512 픽셀의 10개 영상으로부터 임의의 혼합행렬에 따라 발생하는 영상들의 분리에 적용한 결과, 기존의 고정점 알고리즘에 의한 결과보다 우수한 분리성능과 빠른 분리속도가 있음을 확인하였다.

### 1. 서론

적당한 변환에 의해 알맞은 데이터의 표현을 찾기 위한 연구가 통계학, 신호처리, 그리고 신경망 등에서 활발하게 이루어지고 있다<sup>[1,2]</sup>. 이때 선형변환을 이용하면 문제를 계산적으로나 개념적으로 좀 더 간단하게 만들 수 있으며, 결과의 해석도 용이하게 할 수 있다. 여러 가지 방법과 원리들이 적당한 선형변환을 찾기 위해서 개발되었으나, 그 중에서도 데이터의 차원을 감소시키는 선형변환 기법으로 주요성분분석(principal component analysis : PCA)<sup>[1-3]</sup>이 널리 이용되고 있다. 이는 공분산(covariance)을 이용하는 2차원 방법으로 가우스(Gaussian) 분포의 입력 데이터에 대해서는 적당한 것이다. 하지만 통신이나 신호 및 영상 처리 등에서 입력 데이터는 비가우스 분포로 이러한 경우에도 효과적으로 적용될 수 있는 방법의 연구가 요구된다.

최근 PCA의 제약을 해결하기 위한 독립성분분석(independent component analysis : ICA)이 제안되었으며, 이는 데이터의 통계적 성질을 고려한 선형변환 기법으로 데이터의 분포정보를 이용하는 다차원 방법이다<sup>[4-6]</sup>. 하지만 알려진 ICA 알고리즘들에서는 데이터의 통계적 독립성을 판정하기 위해 확률밀도함수와 역행렬을 구한다든지 고차원의 cumulant나 moment를 구하는 등 복잡한 계산이 요구된다.

따라서 복잡한 계산에 따른 많은 계산시간이 요구되는 수치적 기법에 의한 ICA의 대안으로 간단한 신경망을 이용하여 해결하려는 연구가 활발하게 진행되고 있다. 신경망을 이용한 ICA 방법으로는 고정점(fixed point : FP) 알고리즘과 기울기를 이용하는 알고리즘들이 제안되었다<sup>[5,6]</sup>. 하지만 이들 방법들에서도 신경망을 이용하나 신경망 자체가 가지는 학습에 따른 제약들은 그대로 내포하고 있어 그에 따른 연구가 절실하다. 특히 FP 알고리즘은 빠른 수렴속성을 가지지만 분리성능이 떨어지며, 기울기를 이용하는 방법은 학습 파라미터의 설정에 따른 분리속도가 늦은 제약이 있다.

본 연구에서는 뉴우턴(Newton)법의 고정점 알고리즘에서 적응조정이 가능한 학습 파라미터를 이용한 새로운 고정점 알고리즘의 독립성분분석 기법을 제안한다. 제안된 기법은 기존의 엔트로피 최적화로부터 유도된 기법으로 그 계산을 간략화 하기 위한 뉴우턴법에서 학습율과 모멘트를 추가하고, 이들을 적응조정하도록 한 고정점 알고리즘이다. 제안된 기법을 512x512 픽셀(pixel)의 10개 영상으로부터 임의의 혼합행렬에 따라 발생하는 영상들을 각각 대상으로 시뮬레이션하여 성능을 확인하였으며, 기존의 고정점 알고리즘에 의한 결과와 비교 검토하였다.

2. 적응 학습 파라미터를 이용한 고정점 알고리즘의 독립성분분석

ICA는 은닉신호의 분리(blind source separation : BSS)나 특징추출에 널리 이용되고 있다. 여기서는 BSS를 기준으로 ICA를 소개한다.

ICA는 은닉신호의 m개의 입력신호 s로부터 선형적으로 혼합된 n개의 신호 x가 알려져 있을 때, 혼합된 신호로부터 역으로 m개의 독립인 입력신호를 찾는 기법이다. 하지만 입력신호들을 혼합하는데 이용된 혼합행렬 A는 알려져 있지 않으며, 혼합과정에서 잡음 n이 추가 될 수도 있다. 이때 혼합신호와 입력신호와의 관계는

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} = \sum_{i=1}^m s(i)\mathbf{a}(i) + \mathbf{n} \quad (1)$$

로 정의된다. 여기서 n은 보통 입력신호와 구별되지 않기 때문에 생략할 수도 있으며,  $\mathbf{A}=[\mathbf{a}(1), \mathbf{a}(2), \dots, \mathbf{a}(m)]$ 으로  $\mathbf{a}(i)$ 는 ICA의 basis vector이다. 결국 ICA는 알려진 혼합신호로부터 혼합행렬의 역행렬  $\mathbf{A}^{-1}(=\mathbf{W})$ 을 찾는 기법이다. 혼합행렬 A와 역혼합행렬 W에 대하여 상세히 살펴보면 다음 그림 1과 같은 구성도로 나타낼 수 있다. 여기서  $\mathbf{x} = \mathbf{A}\mathbf{s}$ 이고,  $\mathbf{y} = \mathbf{W}\mathbf{x}$ 이다. 그림에서 보면 ICA는 혼합행렬과 일치하는 역혼합행렬을 찾는 과정에서 출력신호가 독립성을 가지도록 하는 기법이다. 따라서 ICA는 알려진 혼합신호 x로부터 출력신호 y를 찾는 기법으로 궁극적으로는 역혼합행렬 W를 찾아서 원신호 s의 근사값을 알아내는 것이다.

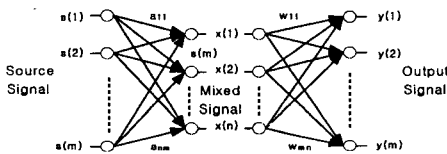


그림 1. 혼합행렬과 역혼합행렬의 상세 설명도

일반적으로 ICA를 좀 더 효과적으로 수행하기 위해 혼합신호의 전처리 과정을 수행한다. 이 과정은 일반적으로 영평균과 whitening 과정이며, 이는 신호의 1차와 2차의 통계성을 고려한 정규화로 학습 알고리즘을 간략화 하기 위함이다. 신호의 영평균은 신호 벡터 x에서 평균값  $\bar{x}$ 를 빼 차를 구함으로써 간단하게 구할 수 있다. 즉, 영평균된 새로운 신호벡터 x는

$$\mathbf{x} = \mathbf{x} - \bar{\mathbf{x}} \quad (2)$$

이다.

한편, whitening은 데이터 벡터 x의 공분산행렬이

단위행렬값을 갖도록 함으로써 구할 수 있다. 즉,  $E(\mathbf{x}\mathbf{x}^T)=\mathbf{I}$ 가 되도록 한다. whitening된 벡터의 성분들은 상호간의 상관성이 줄어들어 비상관성이 더 강한 독립조건을 위해 요구되는 전처리 기법이다. 이러한 whitening의 기법으로 PCA가 널리 사용된다. 일반적인 PCA whitening 행렬 V는

$$\mathbf{V} = \mathbf{D}^{(-1/2)}\mathbf{E}^T \quad (3)$$

이다. 여기서 D는 고유값  $\lambda$ 로 구성되는 대각행렬이며, E는 고유벡터 c로 구성되는 직교행렬이다.

독립성분분석을 위한 FP 알고리즘은 신경망이 가지는 병렬성과 분산성, 그리고 더 작은 메모리 요구 등의 제약을 해결하기 위해 제안된 알고리즘이다<sup>[6]</sup>. FP 알고리즘은 엔트로피 최적화 방법으로부터 유도된 기법으로 지금까지 알려진 기법 중 가장 빠른 학습속도를 가진다. 이는 신호 내에 포함된 상호정보를 최소화함으로써 ICA의 해를 구하는 기법이다.

상호정보는 랜덤변수 상호간의 독립성의 정보이론적 척도로 ICA 변환을 찾기 위한 판별로서 자연스럽게 이용될 수 있다. 일반적으로 변수들 상호간의 정보를 표현하는 상호정보는 비가우시성의 척도로 이용되는 negentropy로 정의된다. n개의 랜덤변수 벡터  $\mathbf{y} = (y_1, \dots, y_n)^T$ 를 위한 negentropy J(y)는

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gauss}}) - H(\mathbf{y}) \quad (4)$$

이다. 여기서 H(y)는 랜덤변수 y의 엔트로피(entropy)이며,  $\mathbf{y}_{\text{gauss}}$ 는 가우시 랜덤벡터이다. 이때 negentropy J(y)를 이용한 상호무관한 n개 랜덤변수  $y_i(i=1, \dots, n)$  사이의 상호정보  $I(y_1, y_2, \dots, y_n)$ 는

$$I(y_1, y_2, \dots, y_n) = J(\mathbf{y}) - \sum_{i=1}^n J(y_i) \quad (5)$$

로 정의된다. 결국 ICA에서 독립성을 최대화하기 위해 식 (5)로 정의되는 상호정보  $I(y_1, y_2, \dots, y_n)$ 를 최소화하는 것은 negentropy를 최대화하는 것이다. negentropy의 최대화는 식 (5)에서 우측 2번째 항의 최대화로 가능하다. 이를 위한 근사식으로 최대 엔트로피 원리를 이용하면

$$J(y_i) \approx c[E\{G(y_i)\} - E\{G(v)\}]^2 \quad (6)$$

와 같이 정의된다. 여기서 G와 c는 각각 비2차(non-quadratic)함수와 상수이고, v는 평균이 0이고 분산이 1인 표준화된 가우시 변수이며, 랜덤변수  $y_i$ 도 평균이 0이고 분산이 1인 변수로 가정한다. negentropy의 근사화를 위한 식 (6)은 ICA 변환을

추정하는 목적함수로 이용될 수 있다. 따라서 ICA에서 독립변수들의 상호간 무관성을 고려한 하나의 독립성분  $y_i = \mathbf{W}^T \mathbf{x}$ 를 찾기 위한 목적함수  $J_G(\mathbf{W})$ 는 식 (6)에 따라

$$J_G(\mathbf{W}) = [E\{G(\mathbf{W}^T \mathbf{x})\} - E\{G(v)\}]^2 \quad (7)$$

로 표현된다. 결국 식 (7)로 정의된 ICA를 위한 목적함수를 최대화하기 위한 방법이 요구된다. 이러한 방법으로 FP 알고리즘이 제안되었다<sup>[5]</sup>.

FP 알고리즘은 알고리즘의 한 단계에 많은 수의 데이터들이 이용되는 batch mode로 계산을 수행하며, 신경망으로 수행될 수 있는 알고리즘이다. 데이터 벡터  $\mathbf{x}$ 의 상관행렬  $E\{\mathbf{x}\mathbf{x}^T\}=\mathbf{I}$ 로 whitening되어 있다고 가정할 때, 하나의 독립성분을 위한 식 (7)에서 FP 알고리즘을 유도하면 다음과 같다.  $J_G(\mathbf{W})$ 의 최대화는 반드시  $E\{G(\mathbf{W}^T \mathbf{x})\}$ 의 최적에서 얻어진다. 이때 Kuhn-Tucker 조건으로부터,  $E\{(\mathbf{W}^T \mathbf{x})^2\} = \|\mathbf{W}\|^2 = 1$  조건하에서  $E\{G(\mathbf{W}^T \mathbf{x})\}$ 의 최적은 다음의 값

$$E\{\mathbf{x}g(\mathbf{W}^T \mathbf{x})\} - \beta \mathbf{W} = 0 \quad (8)$$

로 얻어진다. 여기서  $g(\cdot)$ 는 비2차 함수  $G$ 의 미분이고,  $\beta = E\{\mathbf{W}_0^T \mathbf{x}g(\mathbf{W}_0^T \mathbf{x})\}$ 인 상수이며  $\mathbf{W}_0$ 는  $\mathbf{W}$ 의 최적값이다. 따라서 방정식 (8)을 풀기 위해 빠른 수렴 속도를 가지는 뉴우턴법을 도입한다. 이를 위해 식의 좌변항을  $F(\mathbf{W})$ 로 나타내고, 그것의 Jacobian matrix  $JF(\mathbf{W})$ 를 구하면 다음과 같다. 즉,

$$JF(\mathbf{W}) = E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{W}^T \mathbf{x})\} - \beta \mathbf{I} \quad (9)$$

이다. 한편 데이터 벡터  $\mathbf{x}$ 가 이미 whitening되어 있어 뉴우턴법에 따른 해를 구하는데 이용되는  $JF(\mathbf{W})$ 의 역행렬을 간단하게 계산하기 위해 식 (9)의 우변 첫째 항을 다음과 같이 근사화할 수 있다. 즉,

$E\{\mathbf{x}\mathbf{x}^T g'(\mathbf{W}^T \mathbf{x})\} = E\{\mathbf{x}\mathbf{x}^T\}E\{g'(\mathbf{W}^T \mathbf{x})\} = E\{g'(\mathbf{W}^T \mathbf{x})\}\mathbf{I}$  (10)이다. 식 (10)을 식 (9)에 대입하면 Jacobian matrix는 대각선 행렬이 되어 쉽게 역행렬을 계산할 수 있다. 따라서  $\beta$ 를  $\mathbf{W}_0$  대신에  $\mathbf{W}$ 의 현재값으로 나타내면 근사화된 반복기법의 역혼합행렬  $\mathbf{W}$ 를 구하는 뉴우턴법은

$$\begin{aligned} \mathbf{W}^* &= \mathbf{W} - [E\{\mathbf{x}g(\mathbf{W}^T \mathbf{x})\} - \beta \mathbf{W}] / [E\{g'(\mathbf{W}^T \mathbf{x})\} - \beta], \\ \mathbf{W}^* &= \mathbf{W}^* / \|\mathbf{W}^*\| \end{aligned} \quad (11)$$

와 같이 정의된다. 여기서  $\mathbf{W}^*$ 는  $\mathbf{W}$ 의 새롭게 경신된 값이고,  $\beta = E\{\mathbf{W}^T \mathbf{x}g(\mathbf{W}^T \mathbf{x})\}$ 이다. 결국 식 (11)은 뉴우턴법에 기초를 둔 ICA를 위한 FP 알고리즘이다. 또한 식 (11)의 첫째식 양쪽에  $\beta - E\{g'(\mathbf{W}^T \mathbf{x})\}$

를 곱해 구해지는 더욱 간단한 뉴우턴법의 FP 알고리즘은

$$\mathbf{W}^* = E\{\mathbf{x}g(\mathbf{W}^T \mathbf{x})\} - E\{g'(\mathbf{W}^T \mathbf{x})\}\mathbf{W}, \mathbf{W}^* = \mathbf{W}^* / \|\mathbf{W}^*\| \quad (12)$$

으로 된다.

한편 식 (11)이나 식 (12)의 우측항의 계산을 보면 뉴우턴법에서 점선을 구하기 위한 기울기로 함수  $g(\mathbf{W}^T \mathbf{x})$ 의 1차 미분이 요구된다. 하지만 미분에 따라  $\mathbf{W}$ 가 최적해로 수렴하는 성능을 달라진다. 따라서 수렴에 따른 학습 파라미터를 이용하면 좀 더 빠르고 정확하게 최적해로 수렴시킬 수 있을 것이다. 이를 위해 본 논문에서는  $\mathbf{W}$ 의 경신과정에서 빠른 계산을 위해 학습율과 이전 변화의 속성을 반영하는 모멘트를 추가하고, 이를 출력 변화식  $g'(\mathbf{W}^T \mathbf{x})$ 의 값에 따라 적용조정되도록 하였다. 이렇게 하면 뉴우턴법이 가지는 빠른 수렴속성을 그대로 살리면서도 최적해로의 수렴특성에 따른 더욱 효과적인 독립성분분석이 가능하다. 제안된 적용조정 속성을 가지는 학습율  $\eta$ 와 모멘트  $\nu$ 를 이용한 역혼합행렬  $\mathbf{W}$ 의 최종경신식은 식 (12)를 변형한

$$\begin{aligned} \mathbf{W}^* &= E\{\mathbf{x}g(\mathbf{W}^T \mathbf{x})\} - \eta E\{g'(\mathbf{W}^T \mathbf{x})\}\mathbf{W} + \nu(\mathbf{W}^* - \mathbf{W}), \\ \mathbf{W}^* &= \mathbf{W}^* / \|\mathbf{W}^*\| \end{aligned} \quad (13)$$

로 정의된다. 이때 학습율과 모멘트의 적용조정을 위한 기준으로  $g'(\mathbf{W}^T \mathbf{x}) < \sigma$  이면  $\eta = 1 + 1/(\text{itr})$ 과  $\nu = \log(\text{itr})$ 로 하고, 그렇지 않으면  $\eta = 1$ 과  $\nu = 1/\log(\text{itr})$ 로 하였다. 이는 출력의 변화가 설정치  $\sigma$  이상이면 학습율은 줄이고 모멘트는 증가시키며, 그렇지 않으면 반대의 값으로 조정되도록 한 것이다. 여기서 itr은 학습반복수이다.

역혼합행렬의 변화량  $\varepsilon$ 은 1보다 작은 임의의 값으로 실험에서는 0.0001로 하여 알고리즘을 종료 하였다. 특히 경신의 식에서  $g(\cdot)$ 는 비선형 함수이며 일반적으로  $(\cdot)^3$ 과  $\tanh(\cdot)$ 의 함수값을 가진다. 여기서는  $\tanh(\cdot)$  함수를 적용하였다.

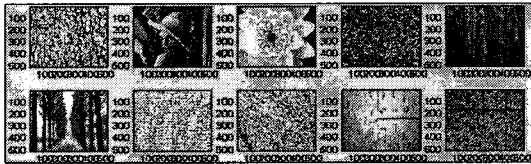
### 3. 시뮬레이션 결과 및 분석

제안된 ICA 알고리즘의 타당성과 성능을 평가하기 위해서 영상으로부터 임의의 혼합행렬에 따라 발생하는 혼합영상들을 각각 대상으로 시뮬레이션 하였다. 시뮬레이션에서는 512x512 픽셀을 가지는 10개의 독립된 영상을 대상으로 임의의 값을 가진 혼합행렬로부터 생성된 영상들을 대상으로 제안된 알고리즘을 수행하였다. 복원 성능을 평가하기 위해 정합도 A를 정의하여 원 영상과 복원된 영상을 비교하였다. 이때 정합도는

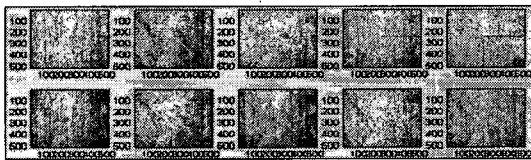
$A = 100 - \frac{1}{512 \times 512} \left[ \sum_{i=1}^{512} |s_{ij} - y_{ij}| \right]$ 로 하였다. 또한 시뮬레이션은 펜티엄 IV-1.5G 컴퓨터 상에서 Matlab 5.1로

실행하였다.

그림 2는 10개의 원 영상과 임의의 혼합행렬로부터 생성된 10개의 혼합영상들만을 나타낸 것이다. 그림에서 보면 혼합영상 각각에서는 원 영상들의 정보들을 발견할 수 없다. 실제에서는 혼합영상들은 알려져 있으나 원 영상들은 미지의 영상들이다. 시뮬레이션에서는 혼합된 영상들만 입력으로 하여 ICA를 수행하였다.



(a) 원영상



(b) 혼합된 영상

그림 2. 512x512 픽셀의 10개 영상

그림 3은 출력의 변화량  $\sigma$ 의 변화에 따른 학습 반복수  $n$ 을 나타낸 것이다. 그림에서 보면 학습파라미터의 적용조정 판정기준으로 이용되는 출력의 변화량의 값이 0.4에서 0.7까지에서 학습반복수가 낮음을 알 수 있다. 또한 변화량이 증가할수록 학습반복수가 증가함을 확인할 수 있다.

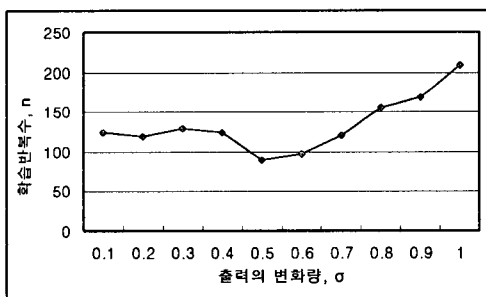


그림 3. 출력의 변화량에 따른 학습반복수의 변화

한편 표1은 출력의 변화량  $\sigma$ 를 0.5로 하였을 때, 10개의 혼합영상을 대상으로 10회 시행에 따른 평균 학습반복수  $n$ , CPU 시간인 분리시간  $t$ , 그리고 정합도 ASE를 각각 나타낸 것이다. 표에서 보면 제안된 알고리즘의 경우, 기존의 학습파라미터를 사용하지 않은 고정점 알고리즘에 비해 학습반복수와 분리시간에서 각각 약 1.4배 정도 개선되었으며, 원영상과

분리영상의 픽셀단위 정합도에서도 우수한 결과를 보임을 알 수 있다. 이는 제안된 고정점 알고리즘의 독립성분석이 기존의 단순 뉴우턴법에만 기초한 고정점 알고리즘에 비해 빠른 분리시간과 개선된 분리성능이 있음을 확인할 수 있다.

표 1. 기존과 제안된 고정점 알고리즘의 분리성능

	학습반복수, n	분리시간 t	정합도, ASE
기존 알고리즘	111.8	143.1	99.68
제안된 알고리즘	80	102.2	99.73

#### 4. 결 론

본 연구에서는 뉴우턴법의 고정점 알고리즘에서 적용조정이 가능한 학습 파라미터를 이용한 새로운 고정점 알고리즘의 독립성분석 기법을 제안한다. 제안된 기법은 기존의 엔트로피 최적화로부터 유도된 기법으로 그 계산을 간략화 하기 위한 뉴우턴법에서 학습율과 모멘트를 추가하고, 이들을 적용조정하도록 한 고정점 알고리즘이다.

제안된 기법을 512x512 픽셀의 10개 영상으로부터 임의의 혼합행렬에 따라 발생하는 영상의 분리에 적용하여 실험한 결과, 제안된 알고리즘은 기존의 알고리즘보다 우수한 추출성능과 빠른 추출시간이 있음을 확인하였다.

향후 다양한 규모의 문제를 대상으로 비교연구가 이루어져야 할 것이다.

#### 참 고 문 헌

- [1] K.I.Diamantaras and S. Y. Kung, Principal Component Neural Networks : Theory and Applications, Adaptive and learning Systems for Signal Processing, Communications, and Control, John Wiley & Sons, Inc., 1996
- [2] S. Haykin, Neural Networks : A Comprehensive Foundation, Prentice-Hall, 2ed, London, 1999
- [3] J. Karhunen and J. Joutsensalo, "Generation of Principal Component Analysis, Optimization Problems, and Neural Networks," Neural Networks, Vol. 8, No. 4, pp. 549-562, 1995
- [4] P. Comon, "Independent Component Analysis A New Concept?", Signal Processing, vol.36, No.3, pp.287-314, Apr.1994
- [5] T.W. Lee, Independent Component Analysis : Theory and Applications, Kluwer Academic Pub., Boston, 1998
- [6] J. Karhunen, "Neural Approaches to Independent Component Analysis and Source Separation", "4th European Symp., Artificial Neural Network, ESANN96, Burges, Belgium, pp. 249-266, Apr. 1996