Machine Learning

Contents

11.1. Problem Statement

- Cocktail Party Problem

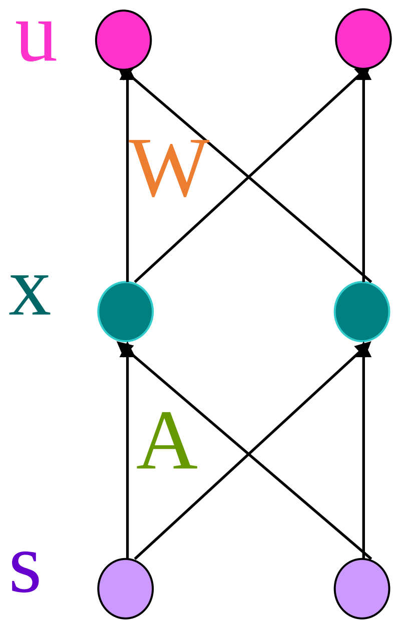
- Unknown mutually independent source signals with zero means

$$s_i(t),\ i = 1, \cdots, n$$

- The model for sensor signals

$$\mathrm{x}(t) = \mathrm{A} \cdot \mathrm{s}(t)$$

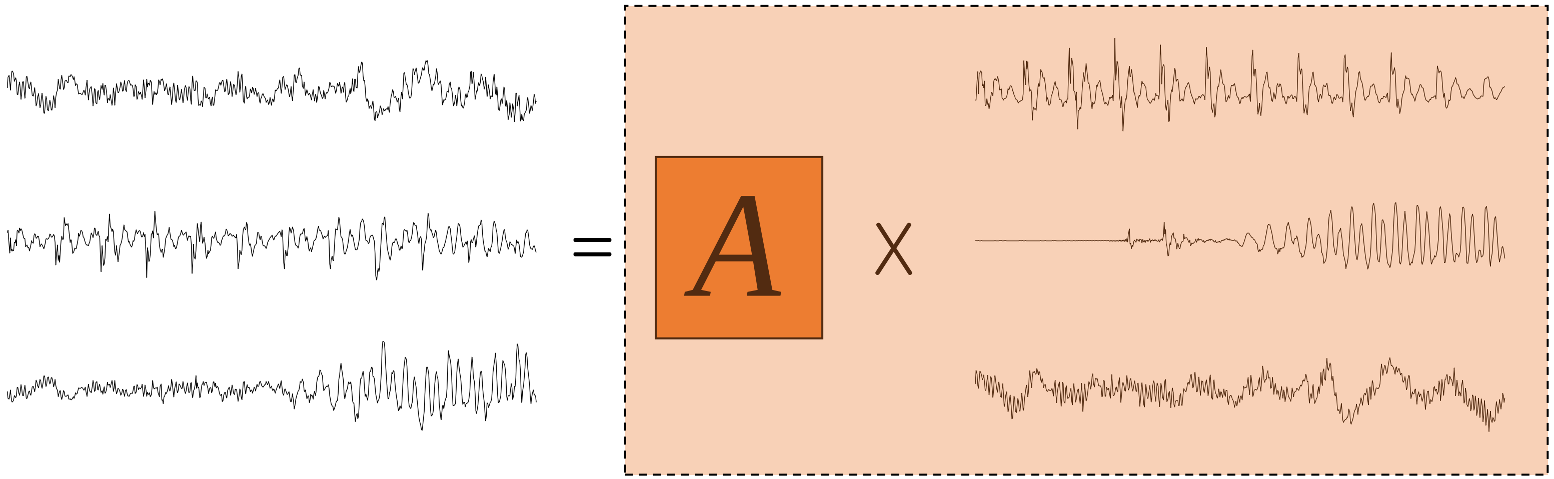
- Problem
  - To recover the original signals except for a permutation of indices and scales
  - Estimation of an unmixing matrix



$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t)$$

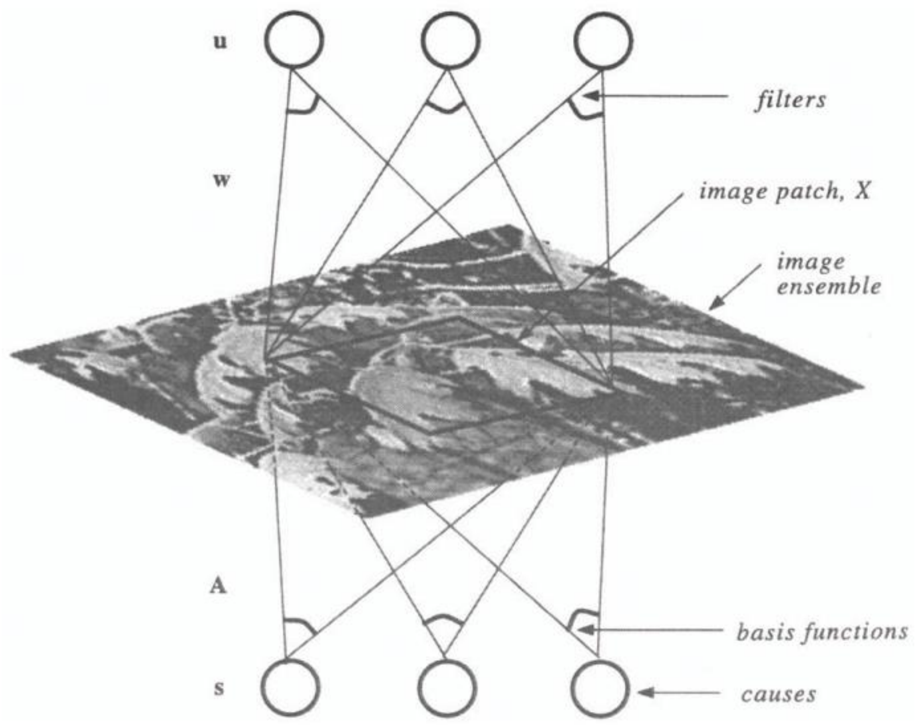
$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t)$$

$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t),$$

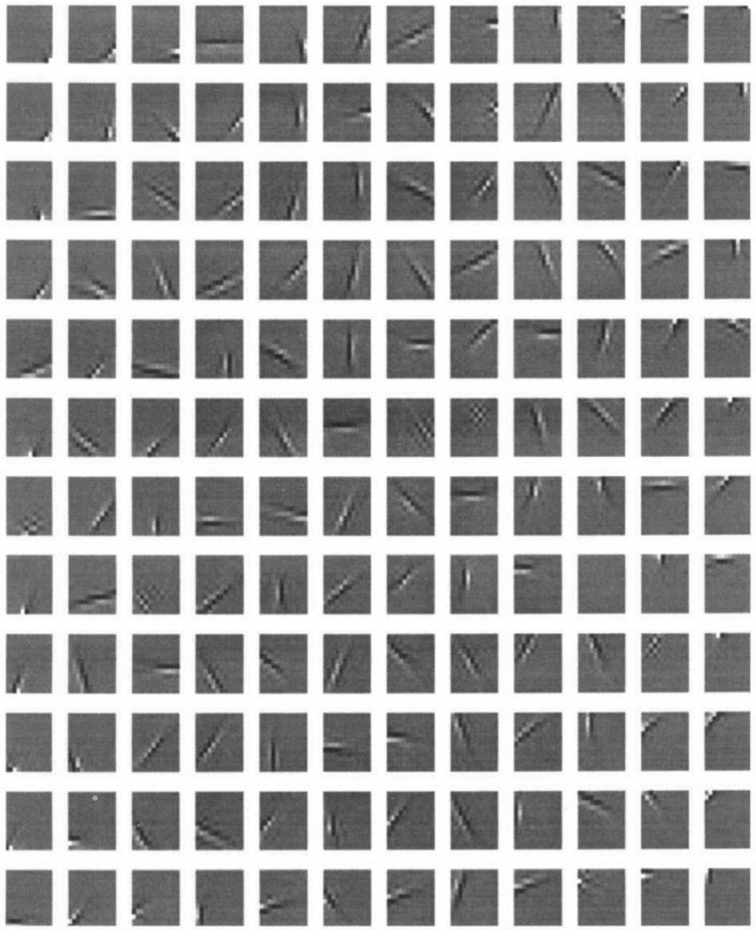


Blind Source Separation by ICA

- Independent Components of Natural Scenes
  (Bell and Sejnowski, Vision Research 1996.[2])



- Independent Components are Edge Filters



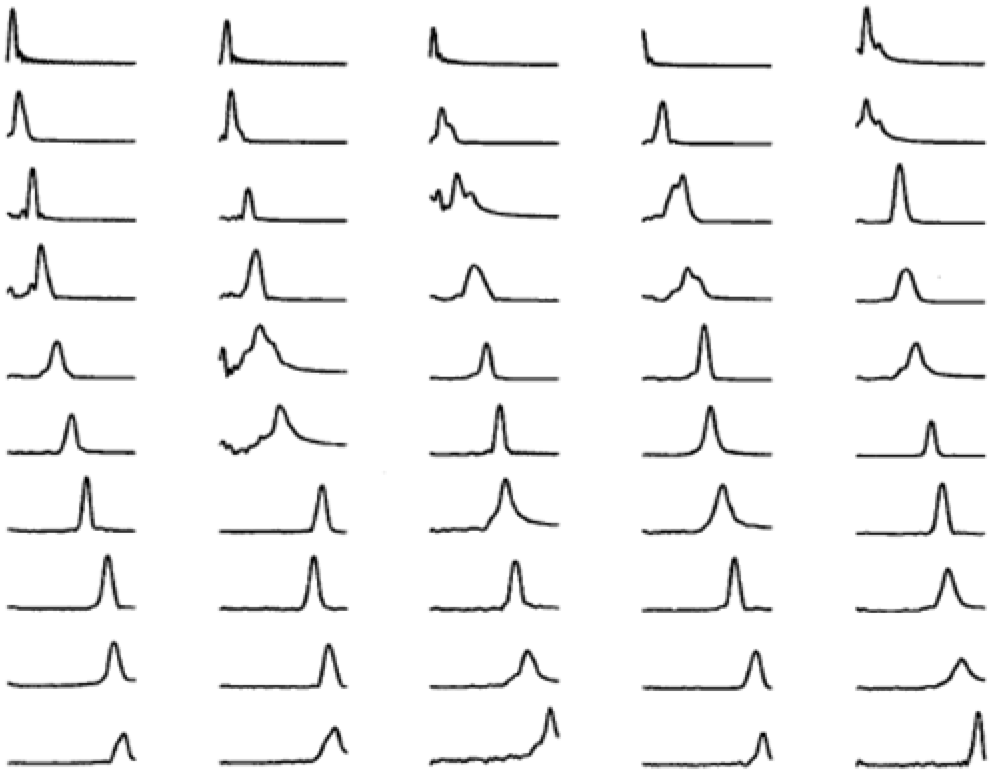


그림 11.3. 음성에서 추출한 독립성분

[41. Jong-Hwan Lee et al. 2002]

Assumptions and limitations on ICA

- Basic assumptions:
  - Individual source signals are statistically independent of the other source signals.

  - # of sources ≤ # of observations
    - $\boldsymbol{x = As} \quad \boldsymbol{s=Wx}$

- Limitations:
  - Scale and sign indeterminacy
    - We cannot determine the variance of sources
    - $\boldsymbol{x = As = (A/\alpha)\,(\alpha s)=A's'}$

  - Permutation indeterminacy
    - We cannot determine the order of sources

  - Sources should be non-Gaussian. At most one source can be Gaussian.

11.2. Information Theory

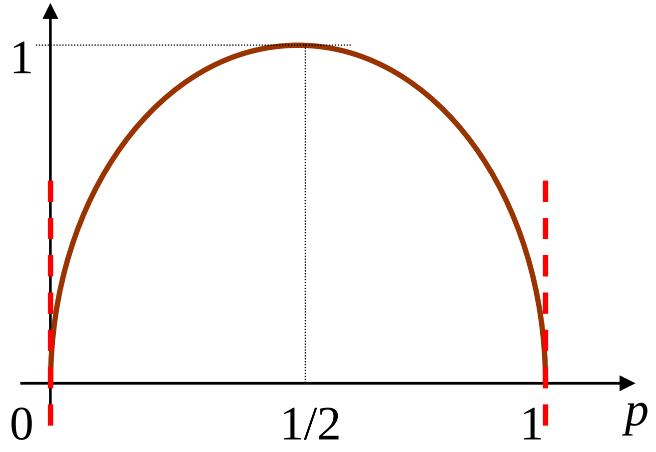
- What is Entropy?
- Definition: The average amount of information of the source.

$$H_r(S) \equiv \sum_{i=1}^{q} p_i I_r(p_i) = \sum_{i=1}^{q} p_i \log_r \frac{1}{p_i}$$

$$H_2(S) \equiv \sum_{i=1}^{q} p_i I_2(p_i) = \sum_{i=1}^{q} p_i \log_2 \frac{1}{p_i} \text{ (bits)}$$

- Ex) Two Symbol Sources

$$H_2(p) \equiv p \log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p} \text{ (bits)}$$



- Differential Entropy $H(X) \equiv -\int p(x) \log p(x) dx$
  - Bounded Case: Uniform Distribution
  - Unbounded Case: Normal Distribution (Gaussian)

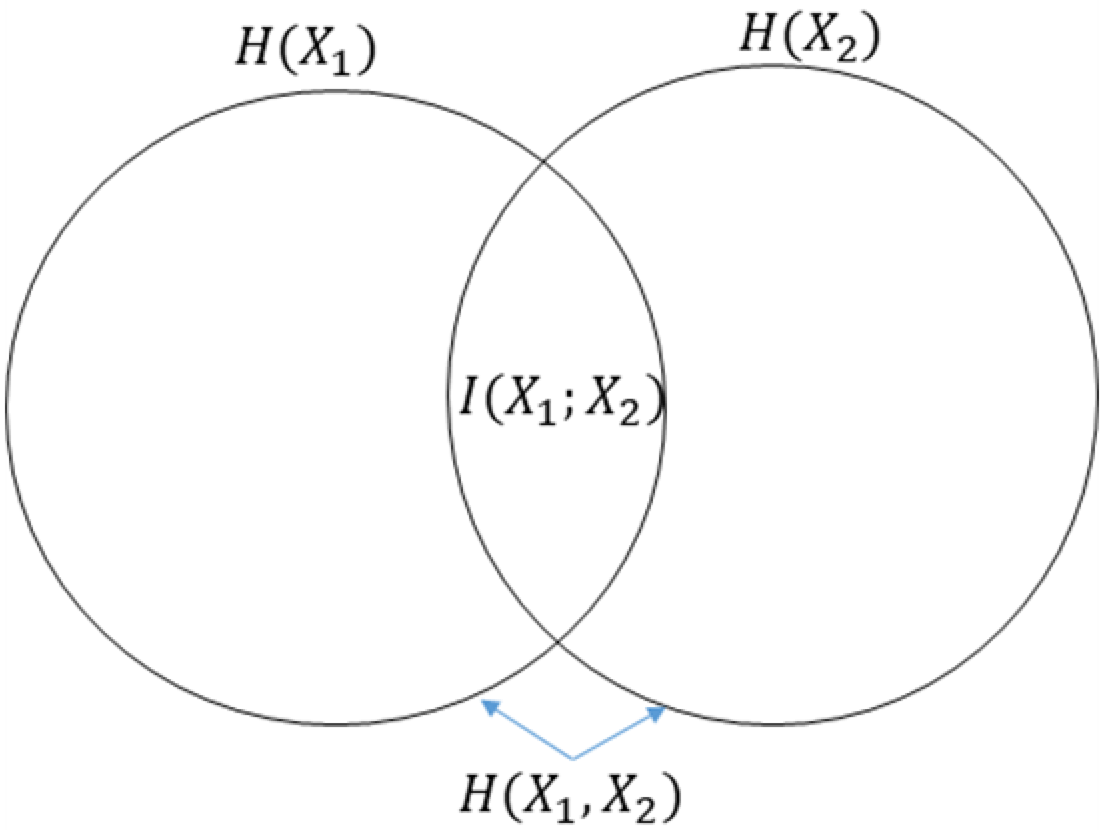
- Joint Entropy

$$H(X_1, X_2) \equiv -\int p(x_1, x_2) \log p(x_1, x_2) dx_1 dx_2$$

- Mutual Information

$$I(X_1; X_2) \equiv \int p(x_1, x_2) \log \frac{p(x_1, x_2)}{p(x_1) p(x_2)} dx_1 dx_2$$

$$H(X_1,X_2) = H(X_1) + H(X_2) - I(X_1;X_2)$$



참조: 상호정보와 엔트로피의 관계유도

상호정보와 엔트로피의 관계를 유도해보자. 식 (11.2.6)으로 주어진 상호정보는

$$I(X_1;X_2) = \int p(x_1,x_2)\log p(x_1,x_2)dx_1dx_2 - \int p(x_1,x_2)\log[p(x_1)p(x_2)]dx_1dx_2 \tag{11.2.8}$$

와 같이 두개의 항으로 나누어진다. 두 번째 항은

$$\int p(x_1,x_2)\log[p(x_1)p(x_2)]dx_1dx_2 = \int p(x_1)\log p(x_1)dx_1 + \int p(x_2)\log p(x_2)dx_2 \tag{11.2.9}$$

로 정리된다. 이때, $p(x_1) = \int p(x_1,x_2)dx_2$을 적용하였다. 식 (11.2.9)를 (11.2.8)에 대입하고 엔트로피의 정의를 적용하면

$$I(X_1;X_2) = -H(X_1,X_2) + H(X_1) + H(X_2) \tag{11.2.10}$$

를 얻게 된다.

11.3. Probability

Probability Density Function

$p_x(x)$: probability density function of x

$p_y(y)$: probability density function of $y = g(x)$

Assume $g(x)$ is a monotonic increasing function, then

$$p_y(y)dy = \Pr\{y < \mathbf{y} < y + dy\}$$

This set consists of the interval $x < \mathbf{x} < x + dx$

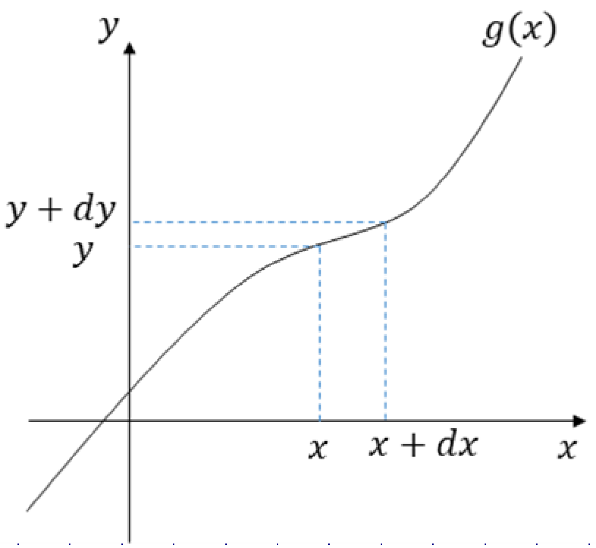
(decreasing : $x + dx < \mathbf{x} < x,\ dx < 0$)

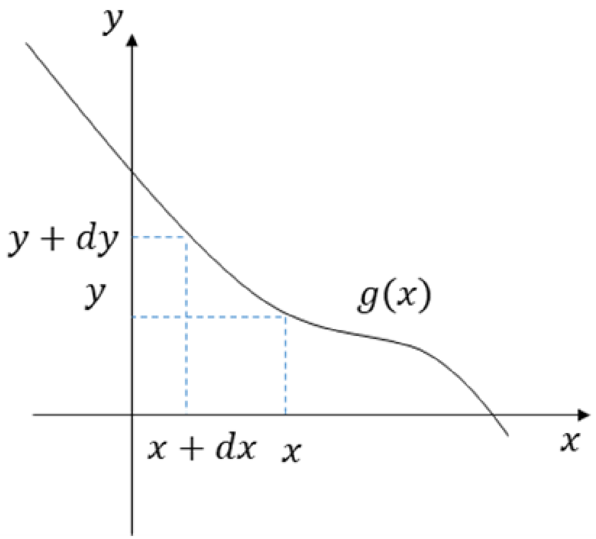
$$p_y(y)dy = \Pr\{y < \mathbf{y} < y + dy\} = \Pr\{x < \mathbf{x} < x + dx\} = p_x(x)dx$$

(decreasing : $p_y(y)dy = \Pr\{x + dx < \mathbf{x} < x\} = p_x(x)|dx|$)

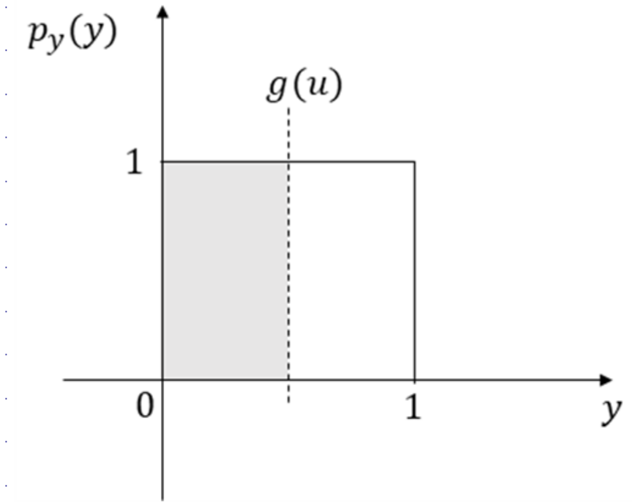
Therefore, for monotonic function of $y = g(x)$

$$p_y(y) = \frac{p_x(x)}{\left|\frac{dy}{dx}\right|} = \frac{p_x(x)}{|g'(x)|}$$





$$F_u(u) = \Pr\{U \leq u\}$$



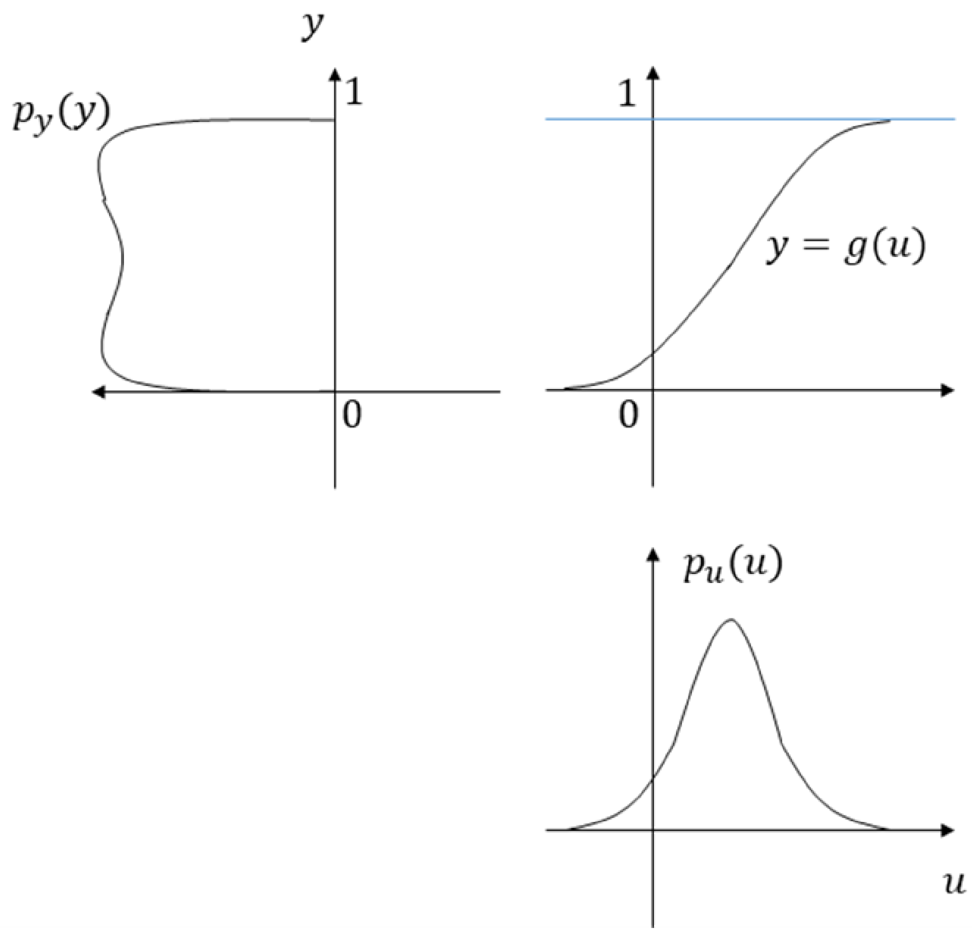


그림 11.11. 확률변수의 CDF 함수 효과

$$F_u(u) = \Pr\{U \leq u\} = \Pr\{g^{-1}(Y) \leq u\} = \Pr\{Y \leq g(u)\} = g(u)$$

예제 11.3-1

연속확률변수 X가 함수 (a) $y=g(x)=ax+b$, (b) $y=g(x)=\frac{1}{x}$로 변환되었다. X의 p.d.f. $p_x(x)$와 Y의 p.d.f. $p_y(y)$의 관계를 각각 유도하라.

풀이

(a) $y=g(x)=ax+b$에서 $g'(x)=a$이다. 또, $x=(y-b)/a$의 관계에 있으므로

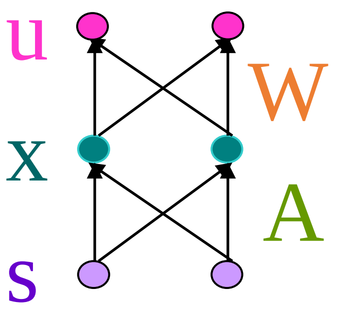
$$p_y(y)\equiv\frac{p_x(x)}{|g'(x)|}=\frac{1}{|a|}p_x\left(\frac{y-b}{a}\right)$$이 된다.

(b) $g'(x)=-\frac{1}{x^2}$이고 $x=\frac{1}{y}$이므로 $p_y(y)=\frac{1}{y^2}p_x\left(\frac{1}{y}\right)$

11.4. InfoMax Algorithm

Objective Function of ICA

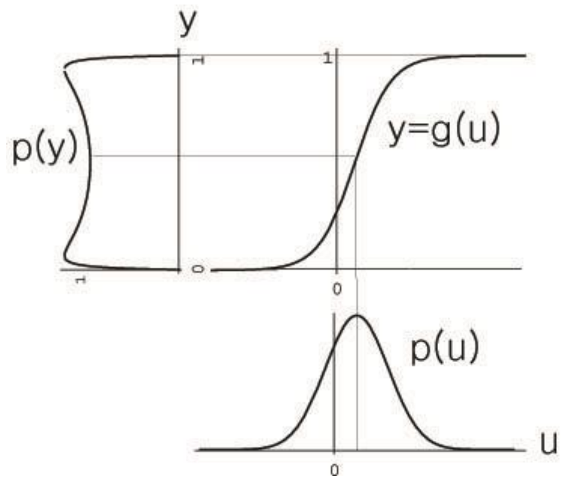
- Objective Function of ICA



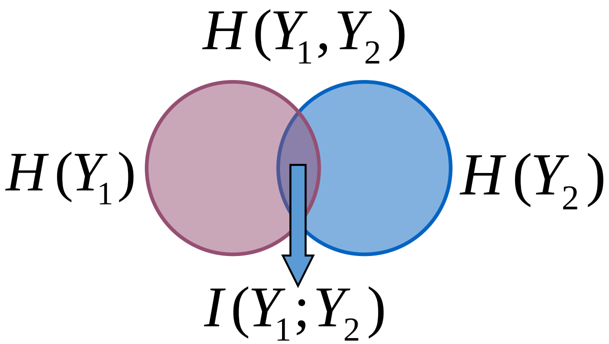
$$\mathbf{x} = \mathbf{As}(?)$$

$$\mathbf{u} = \mathbf{Wx}$$

$$\mathbf{y} = g(\mathbf{u})$$



$g(.)$: c.d.f. of source distribution



- Minimizing Mutual Information
- Maximizing Entropy H(y)

Maximizing Output Entropy: InfoMax

- Joint entropy at the outputs

$$H(y_1,\cdots,y_N)=H(y_1)+\cdots+H(y_N)-I(y_1,\cdots,y_N)$$

where $H(y_i)=-E\{\log p(y_i)\}$

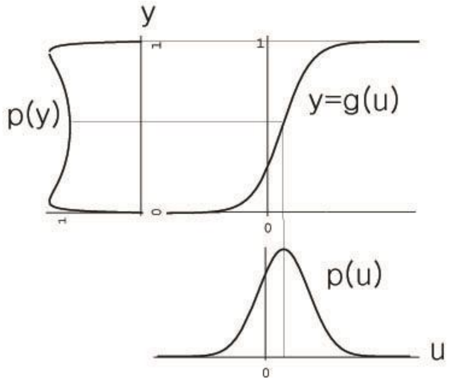
- Pdf of the outputs

$$p(\mathrm{y})=\frac{p(\mathrm{x})}{\left|\mathrm{J}(\mathrm{x})\right|}$$

- Joint entropy at the outputs

$$H(\mathrm{y})=-E\{\log p(\mathrm{y})\}=E\{\log\left|\mathrm{J}(\mathrm{x})\right|\}-E\{\log p(\mathrm{x})\}$$

One Input One Output



- Stochastic gradient ascent learning rule

$$u = wx + w_0, \quad y = \frac{1}{1+e^{-u}} \qquad p_y(y) = \frac{p_x(x)}{\left|\partial y / \partial x\right|}$$

$$H(y) = -E[\log p_y(y)] = -\int p_y(y) \log p_y(y) dy = E[\log\left|\partial y / \partial x\right|] - E[\log p_x(x)]$$

$$\Delta w \propto \frac{\partial H(y)}{\partial w} = \frac{\partial}{\partial w} \log\left|\frac{\partial y}{\partial x}\right| = \left|\frac{\partial y}{\partial x}\right|^{-1} \frac{\partial}{\partial w}\left(\partial y / \partial x\right)$$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} = wy(1-y)$$

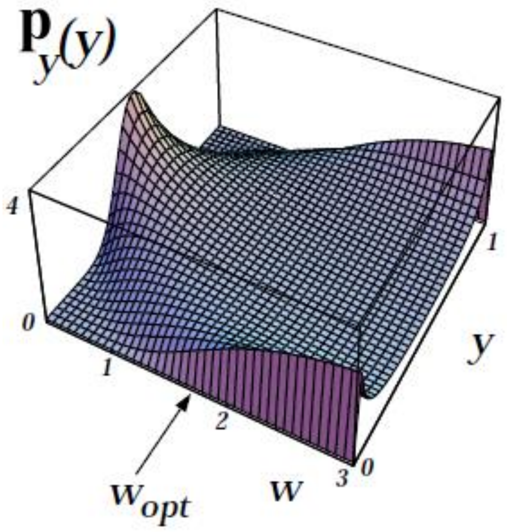
$$\frac{\partial}{\partial w}\frac{\partial y}{\partial x} = y(1-y)(1+wx(1-2y))$$

$$\Delta w \propto \frac{1}{w} + x(1-2y) \qquad \text{Similarly} \quad \Delta w_0 \propto (1-2y)$$

Δw_0 : centre the steepest part of the sigmoid surve on the peak of $p_x(x)$

Δw : scale the slope of the sigmoid to match the variance of $p_x(x)$

$p_y(y)$: flat unit distribution



11.5. InfoMax Algorithm: N→ N Network

- Stochastic gradient ascent learning rule

$$\Delta \mathrm{W} \propto \frac{\partial H(\mathrm{y})}{\partial \mathrm{W}} = \frac{\partial}{\partial \mathrm{W}} \log|\mathrm{J}|$$

- Jacobian of the transform

$$\mathrm{J}(\mathrm{x}) = \det \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \det(\mathrm{W}) \prod_{i=1}^{N} \frac{\partial y_i}{\partial u_i} \qquad \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}$$

- Learning rule

$$\Delta \mathrm{W} \propto \frac{\partial H(\mathrm{y})}{\partial \mathrm{W}} = \frac{\partial}{\partial \mathrm{W}} \log\left|\det(\mathrm{W})\right| + \sum_{i=1}^{N} \frac{\partial}{\partial \mathrm{W}} \log\left|\frac{\partial y_i}{\partial u_i}\right|$$

$$= (\mathrm{W}^T)^{-1} + \frac{\frac{\partial p(\mathrm{u})}{\partial \mathrm{u}}}{p(\mathrm{u})} \mathrm{x}^T$$

$$= (\mathrm{W}^T)^{-1} - \varphi(\mathrm{u})\mathrm{x}^T$$

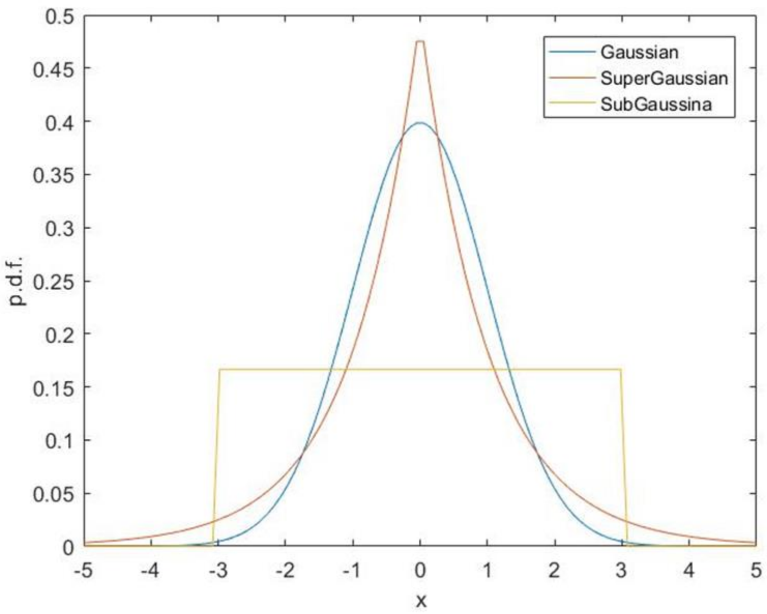
where

$$p(u_i) = \frac{\partial y_i}{\partial u_i} \text{ and } \varphi(u_i) = -\frac{\frac{\partial p(u_i)}{\partial u_i}}{p(u_i)}$$ (Score function)

- Natural Gradient (or Relative Gradient)
  (Amari, Neural Comuptation, 1998. [31])

$$\Delta \mathrm{W} \propto \frac{\partial H(\mathrm{y})}{\partial \mathrm{W}} \mathbf{W}^T \mathbf{W} = [(\mathrm{W}^T)^{-1} - \varphi(\mathrm{u})\mathrm{x}^T]\mathbf{W}^T\mathbf{W}$$

$$= [\mathbf{I} - \varphi(\mathrm{u})\mathrm{u}^T]\mathbf{W}$$

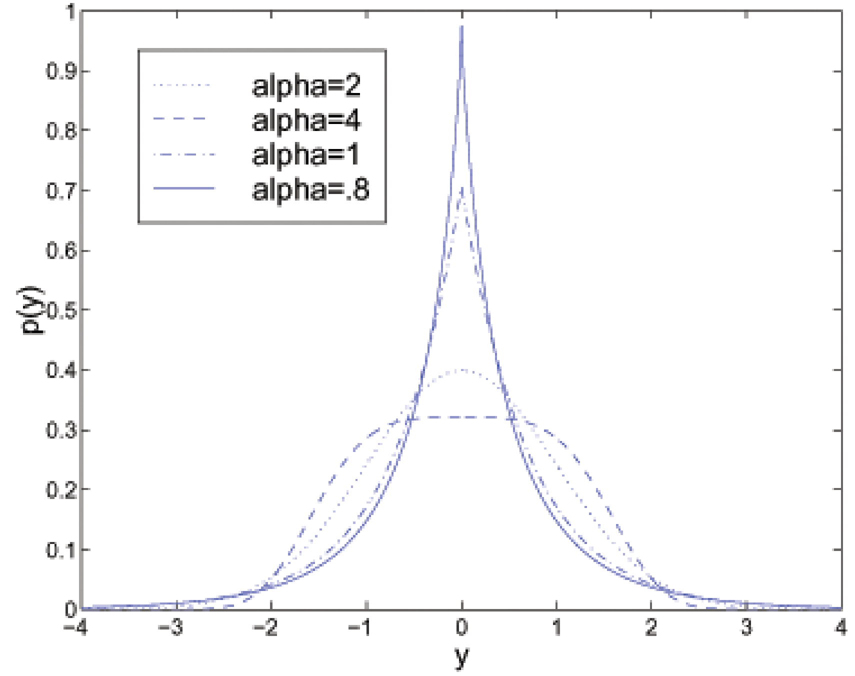


-How can we decide the score ft.:Ext. ICA (T.-W. Lee et al., 1999.[32])

$$\varphi(u_i) = u_i + k_i \tanh(u_i)$$

$$\text{where } k_i = \mathrm{sign}(E\{\mathrm{sech}^2(u_i)\}E\{u_i^2\} - E\{[\tanh(u_i)]u_i\})$$

Flex. ICA: Generalized Gaussian



$\alpha = 2$: Normal

(S. Choi. 2000.[34])

$$p(u;\alpha) = \frac{\alpha}{2\lambda\Gamma(\frac{1}{\alpha})} \exp\left\{ -\left|\frac{u}{\lambda}\right|^{\alpha} \right\}, \quad \varphi(u_i) = |u_i|^{\alpha_i - 1} \operatorname{sgn}(u_i)$$

참조: 행렬의 미분

$m\times n$차원 행렬 $\boldsymbol{W}=(w_{ij})$에 대한 스칼라 함수 g가

$$g=g(\boldsymbol{W})=g(w_{11},\ldots,w_{ij},\ldots w_{mn}) \qquad (11.5.12)$$

로 주어졌다. 행렬식(Determinant)이 이러한 함수의 전형적인 예이다. 그러면, 함수 g의 행렬 $\boldsymbol{W}$에 대한 미분은

$$\frac{\partial g}{\partial \boldsymbol{W}}=\begin{pmatrix} \frac{\partial g}{\partial w_{11}} & \cdots & \frac{\partial g}{\partial w_{1n}} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial g}{\partial w_{m1}} & \cdots & \frac{\partial g}{\partial w_{mn}} \end{pmatrix} \qquad (11.5.13)$$

이다. 이제, $\boldsymbol{W}=(w_{ij})$를 역행렬을 구할 수 있는 $n\times n$ 차원이라고 하자. 그러면,

$$\boldsymbol{W}^{-1}=\frac{1}{\det \boldsymbol{W}}adj(\boldsymbol{W}) \qquad (11.5.14)$$

이 된다. 여기서, $\det \boldsymbol{W}$는 행렬식을 나타내며, $adj(\boldsymbol{W})$는 수반행렬(Adjoint)이다. 수반행렬은

$$adj(\boldsymbol{W})=\begin{pmatrix} W_{11} & \cdots & W_{n1} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ W_{1n} & \cdots & W_{nn} \end{pmatrix} \qquad (11.5.15)$$

이며, 여기서 W_{ij}는 여인수(Cofactor)이다. 이는 행렬 $\boldsymbol{W}$에서 i번째 열과 j번째 행을 제외한 나머지 $(n-1)\times(n-1)$ 행렬을 취한 후 $(-1)^{i+j}$를 곱하여 얻어진다. 행렬식은 여인수의 함수로

$$\det \boldsymbol{W}=\sum_{k=1}^{n} w_{ik}W_{ik} \qquad (11.5.16)$$

와 같이 표현된다. 식 (11.5.16)을 w_{ij}에 대하여 미분하면

$$\frac{\partial \det \boldsymbol{W}}{\partial w_{ij}}=W_{ij} \qquad (11.5.17)$$

이 된다. 그러면, 식 (11.5.15)와 (11.5.16)에 의해

$$\frac{\partial \det \boldsymbol{W}}{\partial W}=adj(\boldsymbol{W})^{T} \qquad (11.5.18)$$

이다. 한편, 식 (11.5.14)에서 $adj(\boldsymbol{W})=(det\ \boldsymbol{W})\boldsymbol{W}^{-1}$이므로

$$\frac{\partial \log|\det \boldsymbol{W}|}{\partial W}=\frac{1}{|\det \boldsymbol{W}|}\frac{\partial|\det \boldsymbol{W}|}{\partial W}=(\boldsymbol{W}^{T})^{-1} \qquad (11.5.19)$$

을 얻게 된다. 이 결과가 식 (11.5.6)의 유도에 사용되었다.

예제 11.5-1

2×2 ICA 회로망에서 측정신호가 $u_1 = w_{11}x_1 + w_{12}x_2$ 와 $u_2 = w_{21}x_1 + w_{22}x_2$에 의해 unmixing 과정을 거친 후, $y_1 = g(u_1)$이고 $y_2 = g(u_2)$로 변환되었다. 이 경우 야코비안 행렬 $J(\boldsymbol{x}) = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}$를 구하여 보아라.

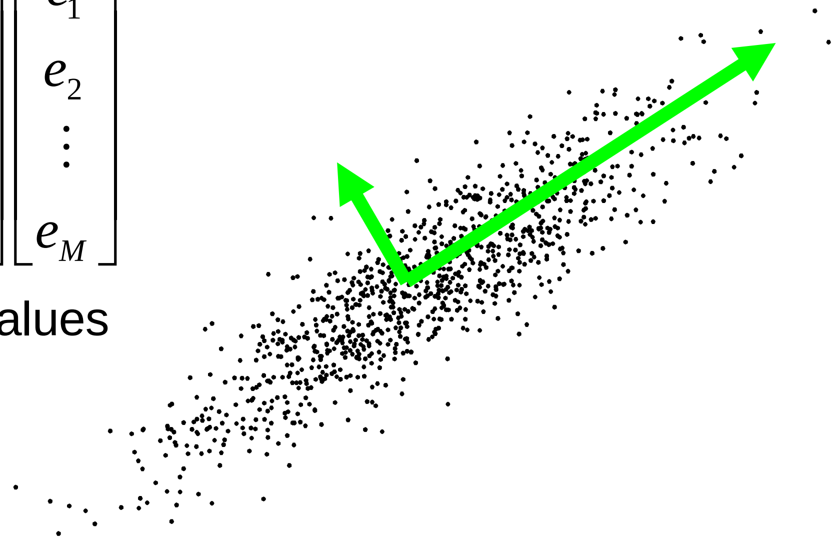
11.6. Comparison with PCA and Applications

- PCA
- Calculate covariance matrix $\mathbf{R}_{\mathbf{x}} = E(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^H$
- Using SVD, find M eigenvalues and eigenvectors.
- Choose L eigen vectors corresponding to L largest eigen values.

L Largest eigen values

$$\mathbf{R}_{\mathbf{x}} = \begin{bmatrix} e_1 & e_2 & \cdots & e_M \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_M \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{bmatrix}$$

M-L small eigen values

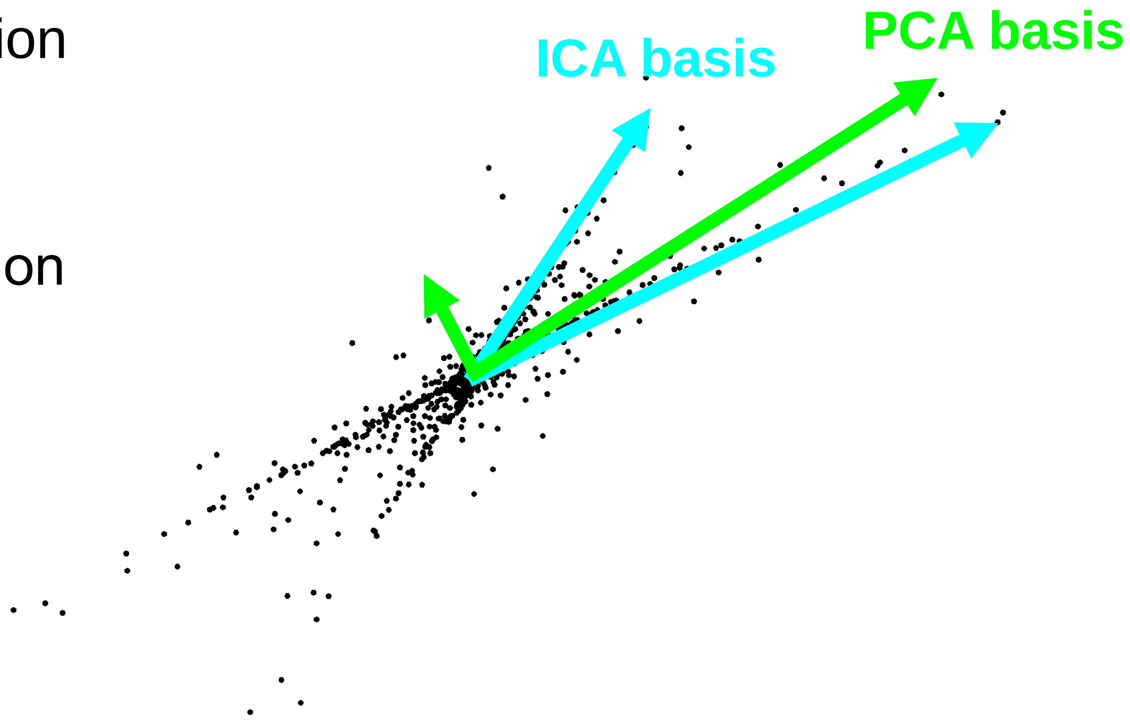


Comparison of PCA and ICA

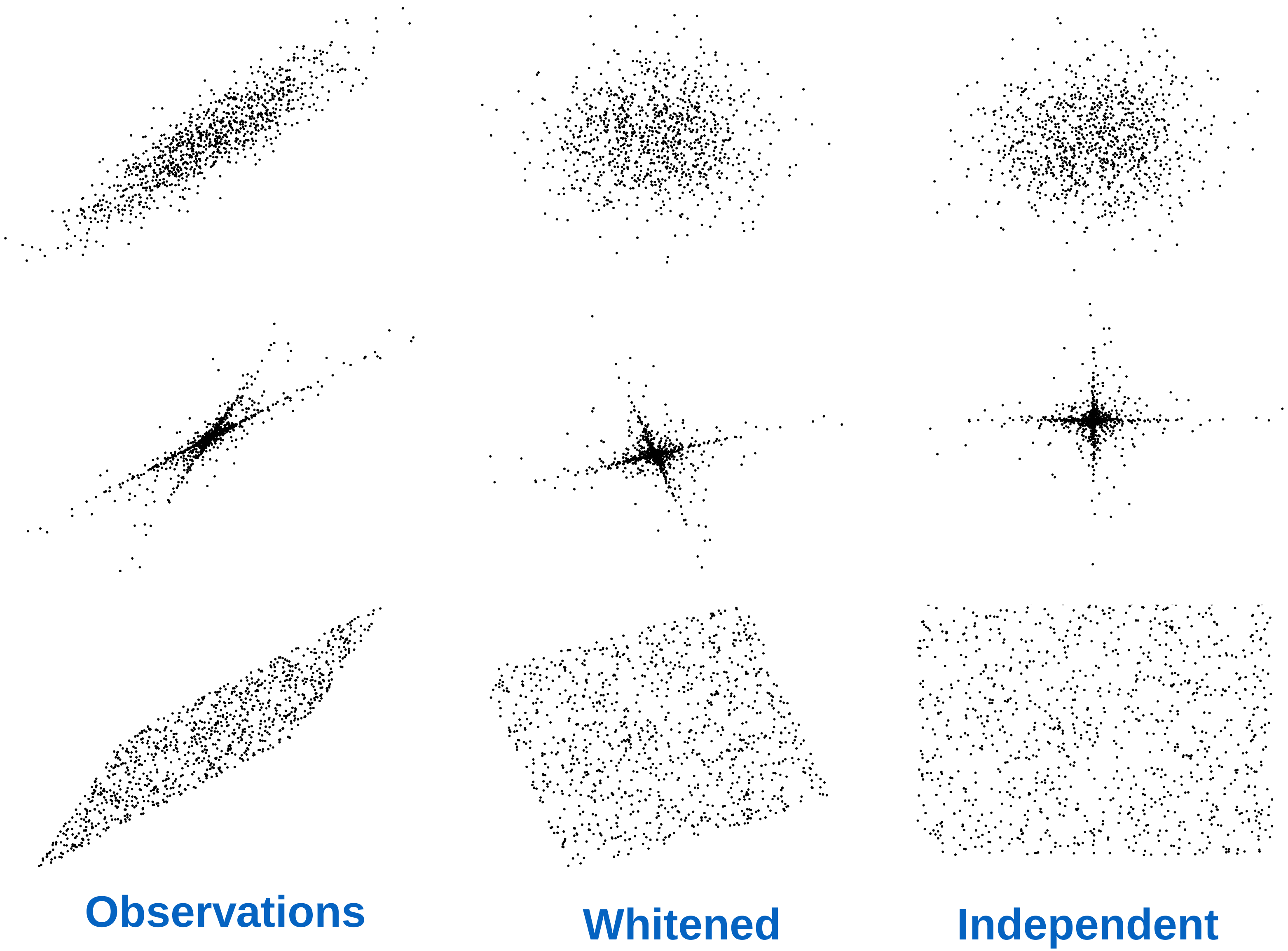
- Independent = uncorrelated ?

- Only for Gaussian data
- dependence = correlation

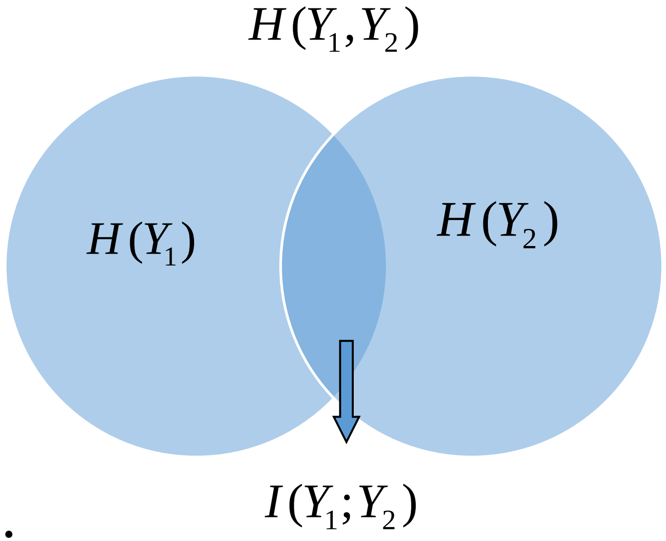
- Generally,
- dependence ≠ correlation



Whitened signal and independent signal



Unifying Information-Theoretic Framework(1)



- Max. H(y): InfoMax
- Minimizing Mutual Information
- Negentropy Maximization

$$J(\mathrm{u}) = D(p(\mathrm{u}) \| p_G(\mathrm{u})) = \int p(\mathrm{u}) \log \frac{p(\mathrm{u})}{p_G(\mathrm{u})} d\mathrm{u}$$

Unifying Information-Theoretic Framework(2)

- Maximum Likelihood Estimation

$$p(\mathrm{x}) = \left|\det(\mathrm{W})\right| p(\mathrm{u})$$

$$L(\mathrm{u}, \mathrm{W}) = \log\left|\det(\mathrm{W})\right| + \sum_{i=1}^{N} \log p_i(u_i)$$

$$\Delta \mathrm{W} \propto \frac{\partial L(\mathrm{u}, \mathrm{W})}{\partial \mathrm{W}} = (\mathrm{W}^T)^{-1} + \frac{\frac{\partial p(\mathrm{u})}{\partial \mathrm{u}}}{p(\mathrm{u})} \mathrm{x}^T$$

$$= (\mathrm{W}^T)^{-1} - \varphi(\mathrm{u}) \mathrm{x}^T$$

Unifying Information-Theoretic Framework(3)

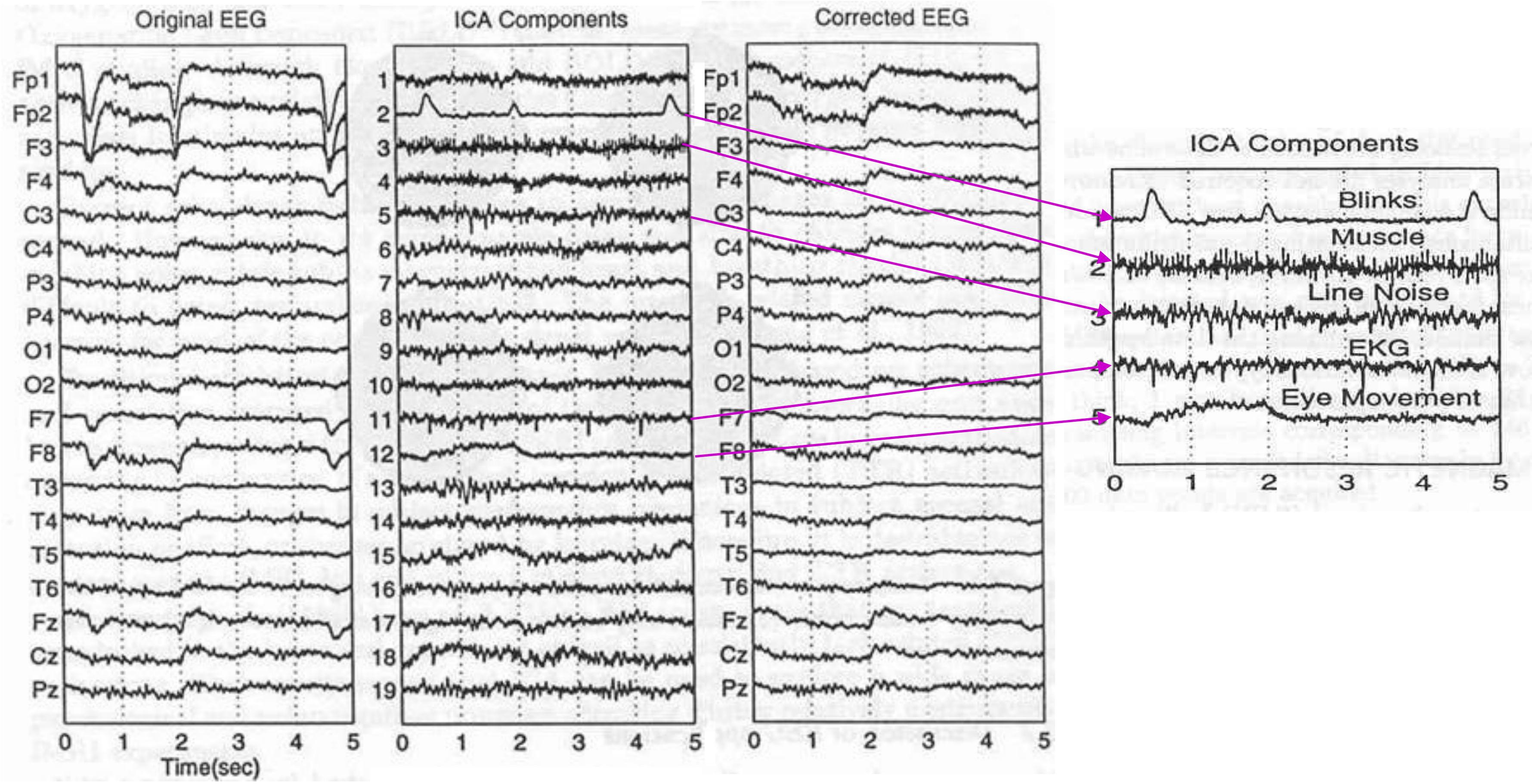
- High-Order Moments and Cumulants
- Nonlinear PCA
- Bussgang Algorithms

(Te-Won Lee et al., Computers and Mathematics with Applications, 2000. [1])

Applications of ICA

- Image processing
  - e.g. denoising, feature extraction, etc
- Biomedical data analysis
  - e.g. anaysis of EEG, MEG, ECG, etc…
- Speech and audio processing
  - e.g. speech separation, feature extraction
- Telecommunication
  - e.g. MIMO System, etc.
- Bioinformatics
  - e.g. micro-array data analysis, etc
- Financial data analysis

- Blind Source Separation of EEG Signals



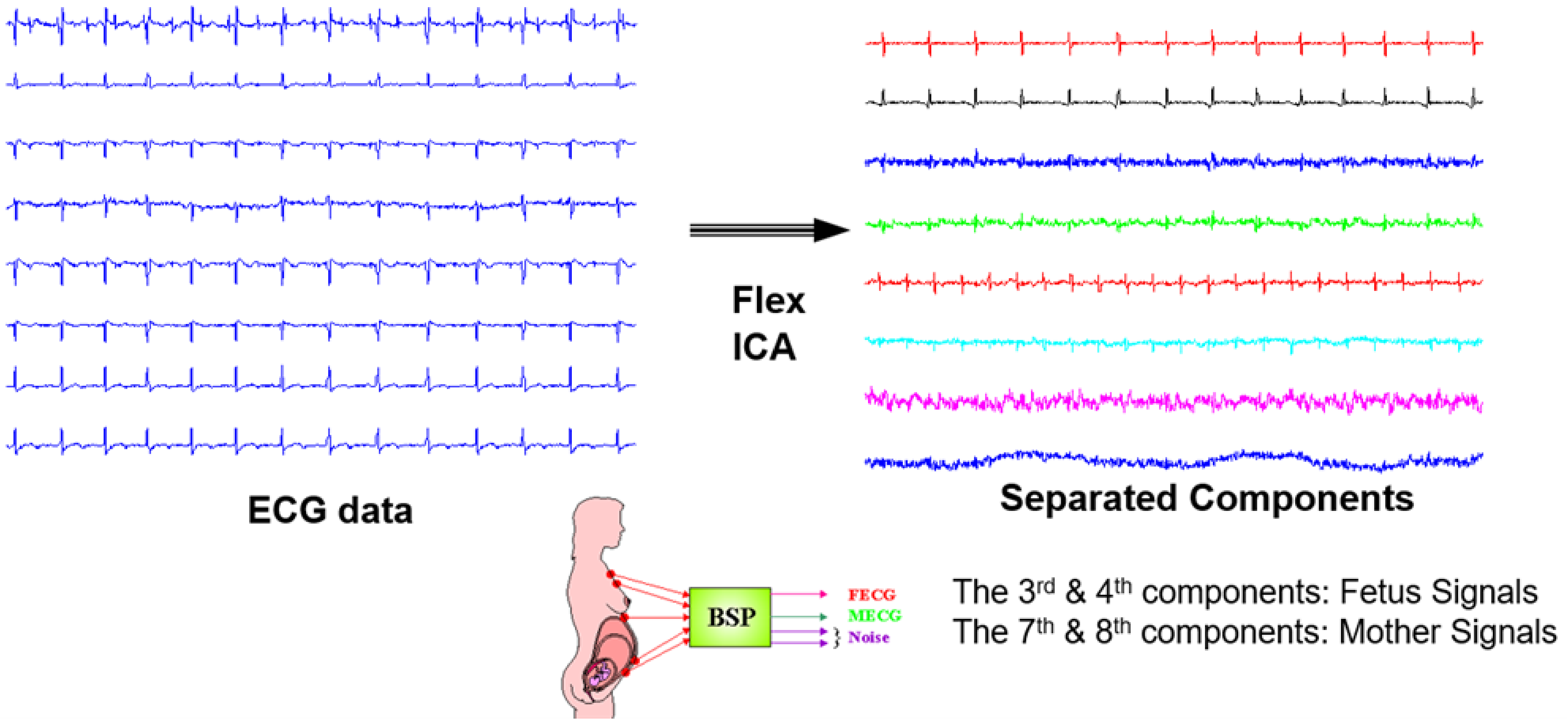


그림 11.16. 유연한 독립성분분석(Flexible ICA)에 의한 임산부와 태아 신호 분리

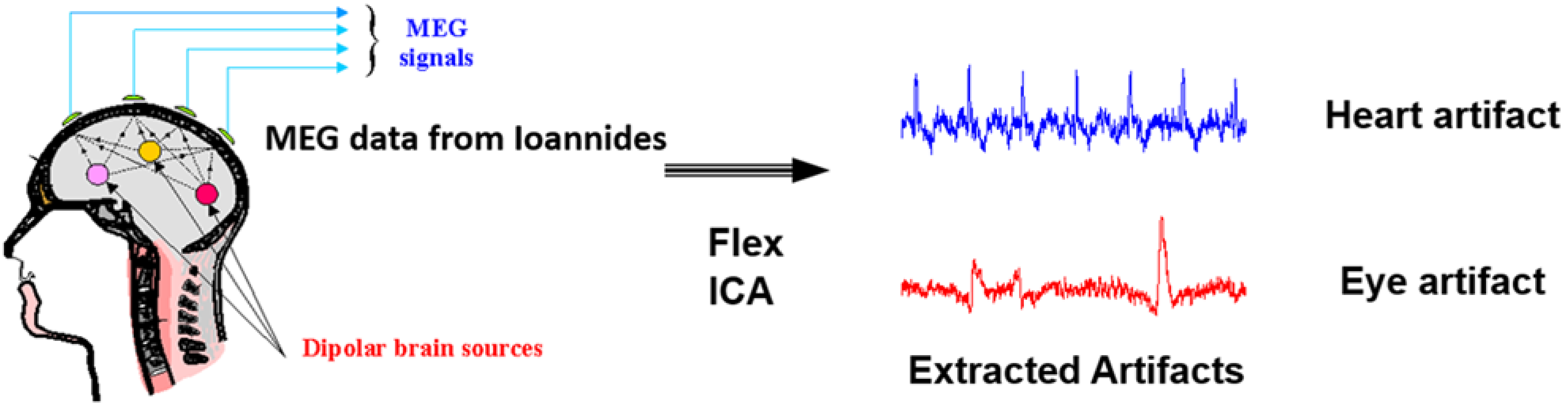


그림 11.17. MEG(Magnetoencephalography) 데이터에서 심장과 눈에 의한 성분 분리

11.7. Projection Pursuit Learning for Latent Structure

Data Model $\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t)$

where $\boldsymbol{s}(t) \in R^M (M \leq N)$ are independent sources.

$p(\boldsymbol{s}) = \Pi_{i=1}^{M} p_i(s_i)$

Objective

$\boldsymbol{u}(t) = \boldsymbol{W}\boldsymbol{x}(t)$

where $\boldsymbol{W} \in R^{P \times N} (P \ll N)$

If P=2, visualization!

Negentropy

$$KL(p||p_G) = \int p_u(u_i)\log\frac{p_u(u_i)}{P_G(u_i)}du_i$$

$$KL(p_F||p_G) = \int p_{\boldsymbol{u}}(\boldsymbol{u})\log\frac{\Pi_{i=1}^{P} p_i(u_i)}{P_G(\boldsymbol{u})}d\boldsymbol{u}$$

$$KL(p_F||p_G) = \frac{1}{2}\log\left((2\pi e)^P \det(\boldsymbol{R_{uu}})\right) + \int p_{\boldsymbol{u}}(\boldsymbol{u})\log\Pi_{i=1}^{P} p_i(u_i)d\boldsymbol{u}$$

$$\boldsymbol{R_{uu}} = E[\boldsymbol{u}\boldsymbol{u}^T] = E[\boldsymbol{W}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{W}^T] = \boldsymbol{W}E[\boldsymbol{x}\boldsymbol{x}^T]\boldsymbol{W}^T = \boldsymbol{W}\boldsymbol{R_{xx}}\boldsymbol{W}^T$$

$$\frac{\partial}{\partial \boldsymbol{W}}\left[\frac{P}{2}\log(2\pi e) + \frac{1}{2}\log(\det(\boldsymbol{W}\boldsymbol{R_{xx}}\boldsymbol{W}^T)) + \sum_{i=1}^{P} E[\log p_i(u_i)]\right]$$

$$\frac{\partial}{\partial \boldsymbol{W}}\left[\frac{1}{2}\log(\det(\boldsymbol{W}\boldsymbol{R_{xx}}\boldsymbol{W}^T))\right] = (\boldsymbol{W}\boldsymbol{R_{xx}}\boldsymbol{W}^T)^{-1}\boldsymbol{W}\boldsymbol{R_{xx}}$$

$$\frac{\partial}{\partial \boldsymbol{W}}\left[\sum_{i=1}^{P} E[\log p_i(u_i)]\right] = E[\boldsymbol{f}(\boldsymbol{u})\boldsymbol{x}^T]$$

$$\frac{\partial}{\partial \boldsymbol{W}} KL(p_F \| p_G) = (\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{x}\boldsymbol{x}} \boldsymbol{W}^T)^{-1} \boldsymbol{W}\boldsymbol{R}_{\boldsymbol{x}\boldsymbol{x}} + E[\boldsymbol{f}(\boldsymbol{u})\boldsymbol{x}^T]$$

$$\Delta \boldsymbol{W}_t = \eta_t \left(\frac{\partial}{\partial \boldsymbol{W}} KL(p_F \| p_G) \right) \boldsymbol{W}_t \boldsymbol{W}_t^T$$

$$\Delta \boldsymbol{W}_t = \eta_t \left(\frac{\partial}{\partial \boldsymbol{W}} KL(p_F \| p_G) \right) (\boldsymbol{W}_t \boldsymbol{W}_t^T + \epsilon_t \boldsymbol{I})$$

$$\Delta \boldsymbol{W}_t = \eta_t (\boldsymbol{I} + \boldsymbol{f}(\boldsymbol{u}_t)\boldsymbol{u}_t^T) \boldsymbol{W}_t + \eta_t \epsilon_t ((\boldsymbol{W}\boldsymbol{R}_{\boldsymbol{x}\boldsymbol{x}} \boldsymbol{W}^T)^{-1} \boldsymbol{W}\boldsymbol{R}_{\boldsymbol{x}\boldsymbol{x}} + \boldsymbol{f}(\boldsymbol{u})\boldsymbol{x}^T)$$

$$\Delta \boldsymbol{W}_t = \eta_t (\boldsymbol{I} + \boldsymbol{f}(\boldsymbol{u}_t)\boldsymbol{u}_t^T) \boldsymbol{W}_t$$

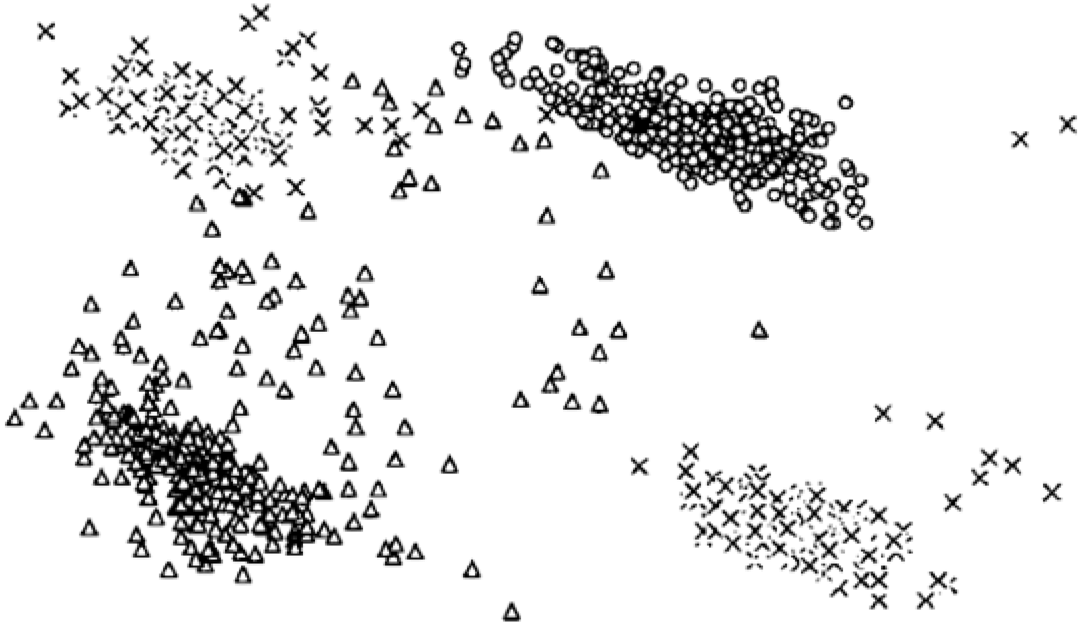


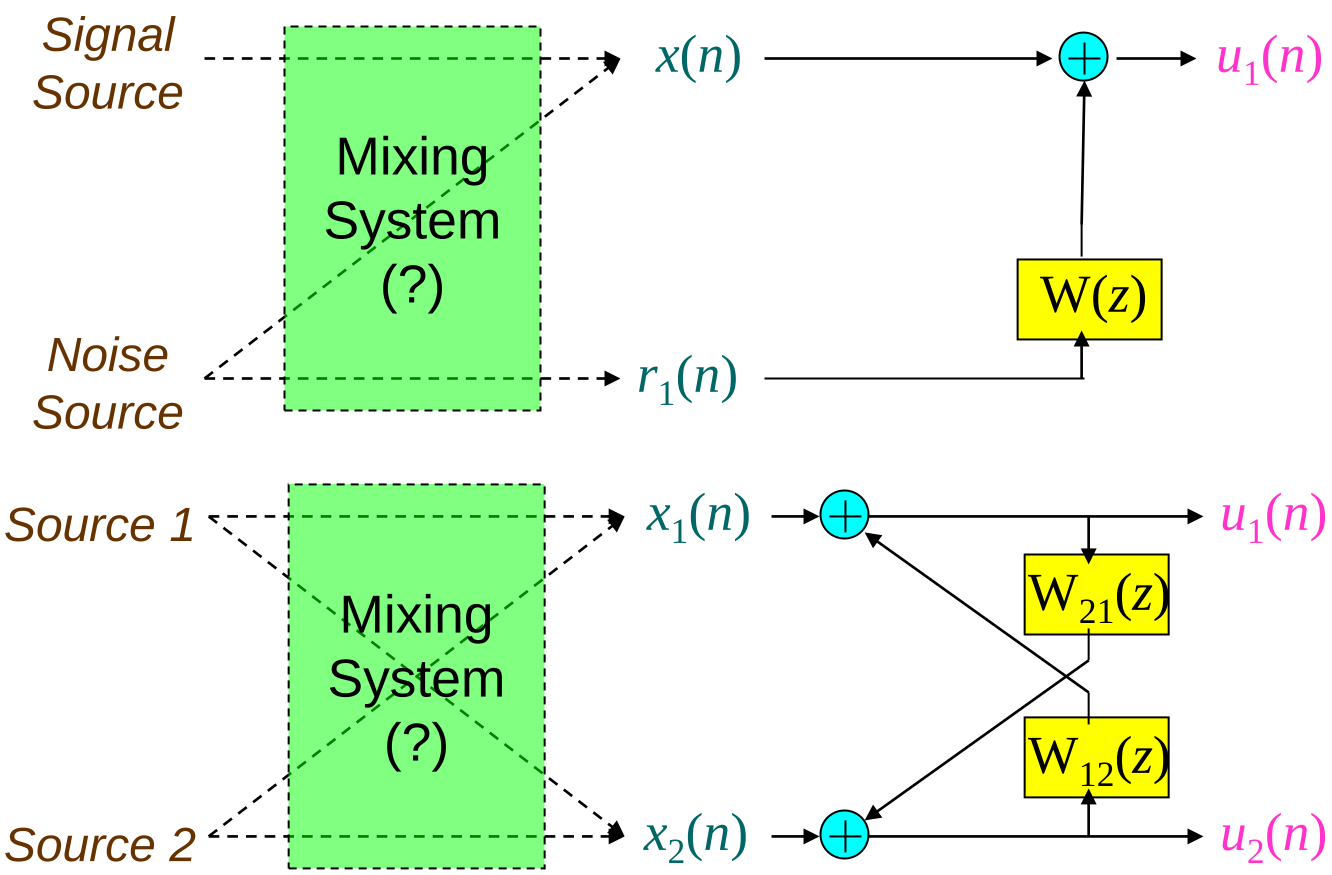
그림 11.18. ICA에 의한 데이터 시각화 예

(× 층류, △고리류, ○균일류)

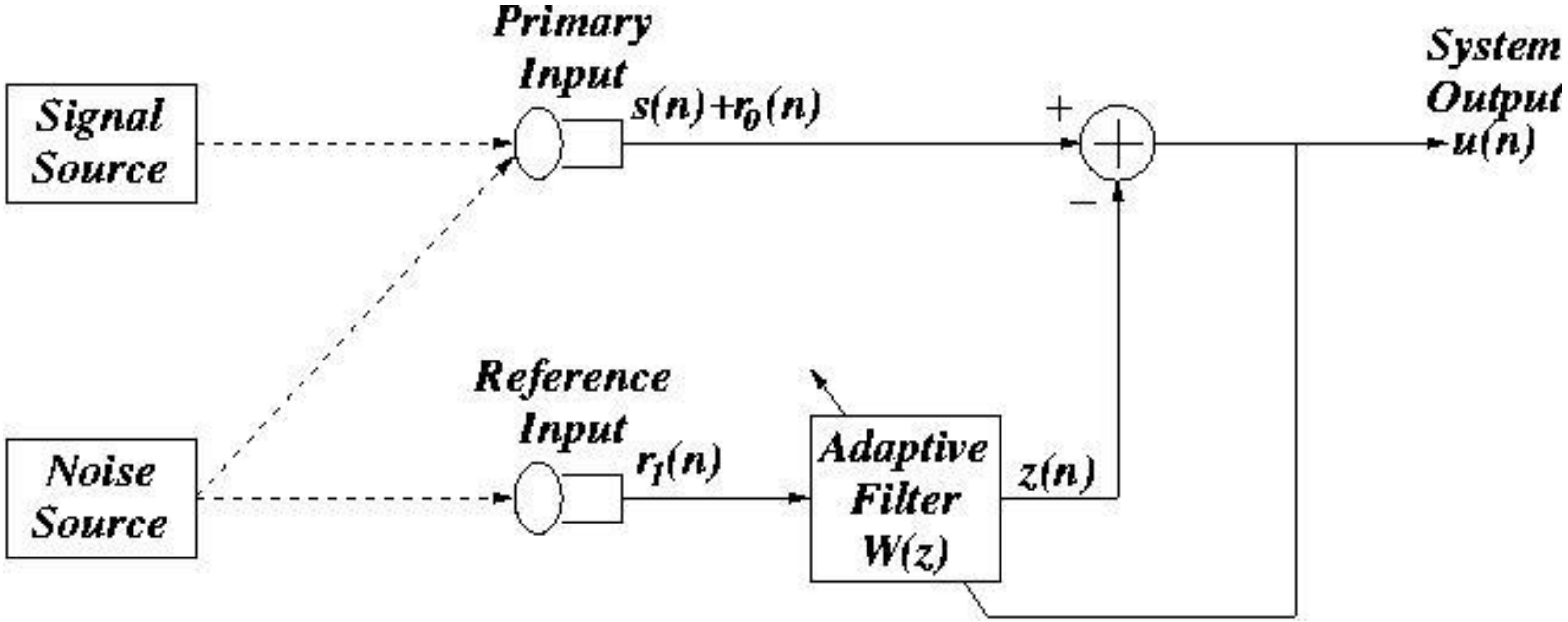
[33. Mark Girolami et al. 1998]

11.8 Convolved Mixtures: ANC vs. BSS

- ANC
- BSS



Adaptive Noise Canceling(ANC)



- System output $u(n) = s(n) + r_0(n) - \sum_{k=1}^{L_a} w(k) r_1(n-k)$
- Minimize the output power $\min E[u^2] = E[s^2] + \min E[(r_0 - z)^2]$
- The LMS algorithm $\Delta w(k) \propto u(n) r_1(n-k)$

ICA-Based Approach to ANC [36]

- Maximizing entropy
  - Set dummy output $v = r_1$

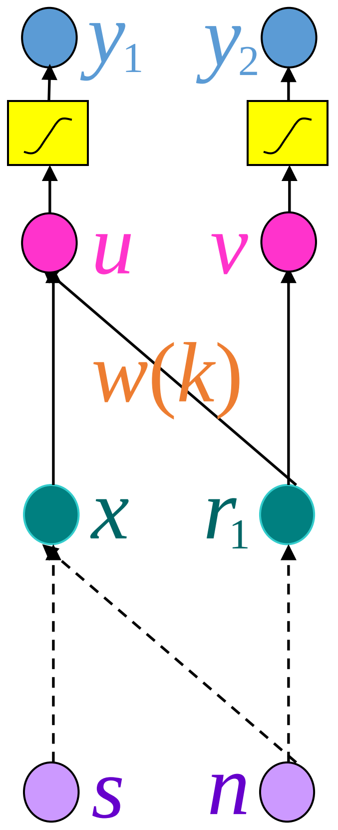
$$H(y_1, y_2) = -E[\log(p(y_1, y_2))] = -E[\log(\frac{p(x, r_1)}{|J|})]$$

where $$J = \frac{\partial y_1}{\partial x}\frac{\partial y_2}{\partial r_1} - \frac{\partial y_1}{\partial r_1}\frac{\partial y_2}{\partial x} = \frac{\partial y_1}{\partial u}\frac{\partial y_2}{\partial v}$$

- Learning rules of adaptive filter coefficients in ANC (Park et al. 2002)

$$\Delta w(k) \propto \frac{\partial H(y_1, y_2)}{\partial w(k)} = \frac{\partial}{\partial w(k)}\log|J| = \left(\frac{\partial y_1}{\partial u}\right)^{-1}\frac{\partial}{\partial w(k)}\left(\frac{\partial y_1}{\partial u}\right)$$

$$= \varphi(u(n))r_1(n-k)$$ where $$\varphi(u) = -\left(\frac{\partial y_1}{\partial u}\right)^{-1}\frac{\partial^2 y_1}{\partial u^2}$$



Known Noise Source

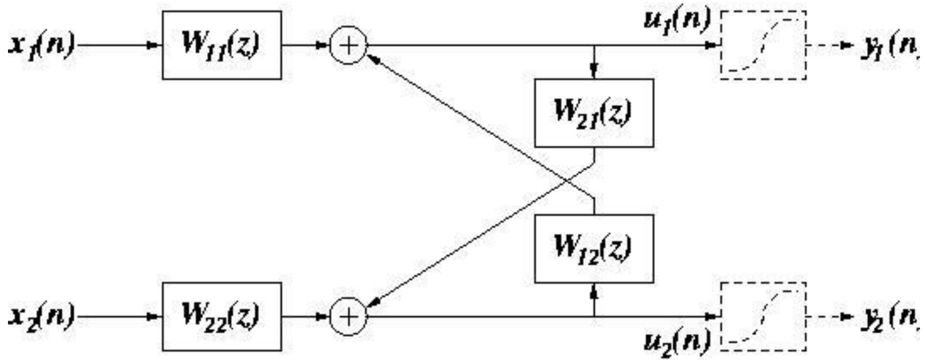
Input Signal with Noise

Output Signal after Process

Time Domain Approach to BSS

- Convolved mixtures

$$x_i(n) = \sum_{j=1}^{N} \sum_{k=0}^{K-1} a_{ij}(k) s_j(n-k)$$



- Feedback architecture

$$u_i(n) = \sum_{k=0}^{K} w_{ii}(k) x_i(n-k) + \sum_{j=1,\, j\neq i}^{N} \sum_{k=1}^{K} w_{ij}(k) u_j(n-k)$$

- Learning rules

$$\Delta w_{ii}(0) \propto 1/w_{ii}(0) - \varphi(u_i(n)) x_i(n),$$
$$\Delta w_{ii}(k) \propto -\varphi(u_i(n)) x_i(n-k),\ k \neq 0,$$
$$\Delta w_{ij}(k) \propto -\varphi(u_i(n)) u_j(n-k),\ i \neq j,$$

$$\varphi(u_i(n)) = -\frac{\dfrac{\partial p(u_i(n))}{\partial u_i(n)}}{p(u_i(n))}$$

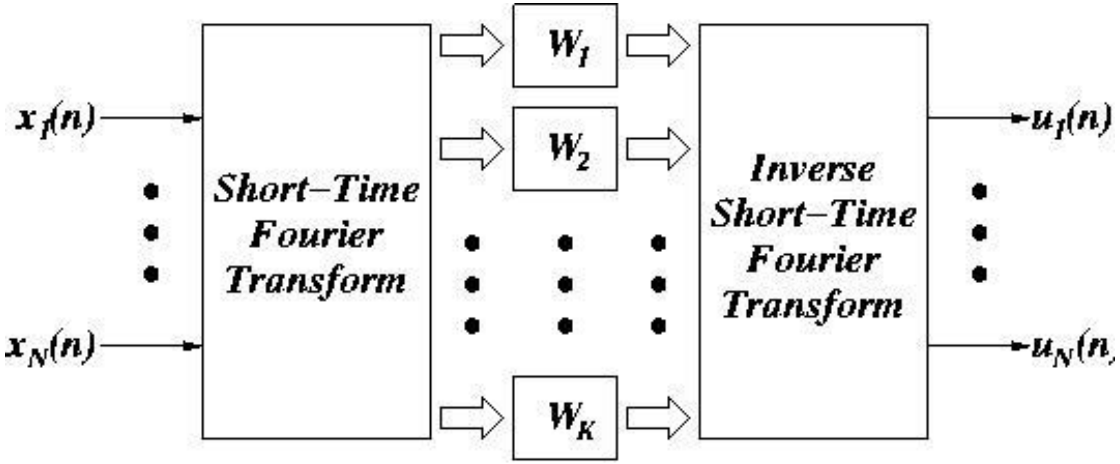
Frequency Domain Approach to BSS

- In the frequency domain

$$X_f(n) = A_f \cdot S_f(n), \quad \forall f$$

- Complex score function

$$\varphi(u_i) = -\frac{\dfrac{\partial p(|u_i|)}{\partial(|u_i|)}}{p(|u_i|)} \exp(j \cdot \angle u_i)$$



- Learning rule

$$\Delta W \propto [I - \varphi(\mathrm{u})\, \mathrm{u}^H]W$$

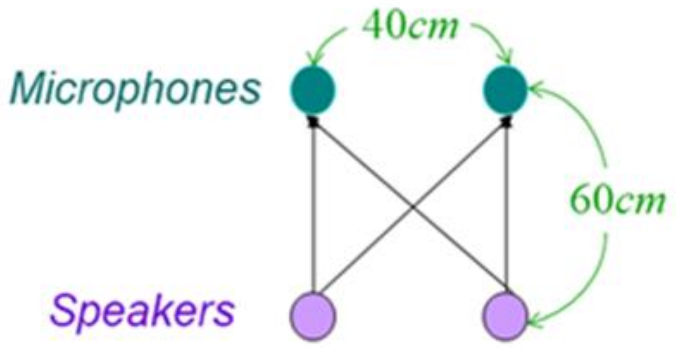


그림 11.24. 2×2 실험환경

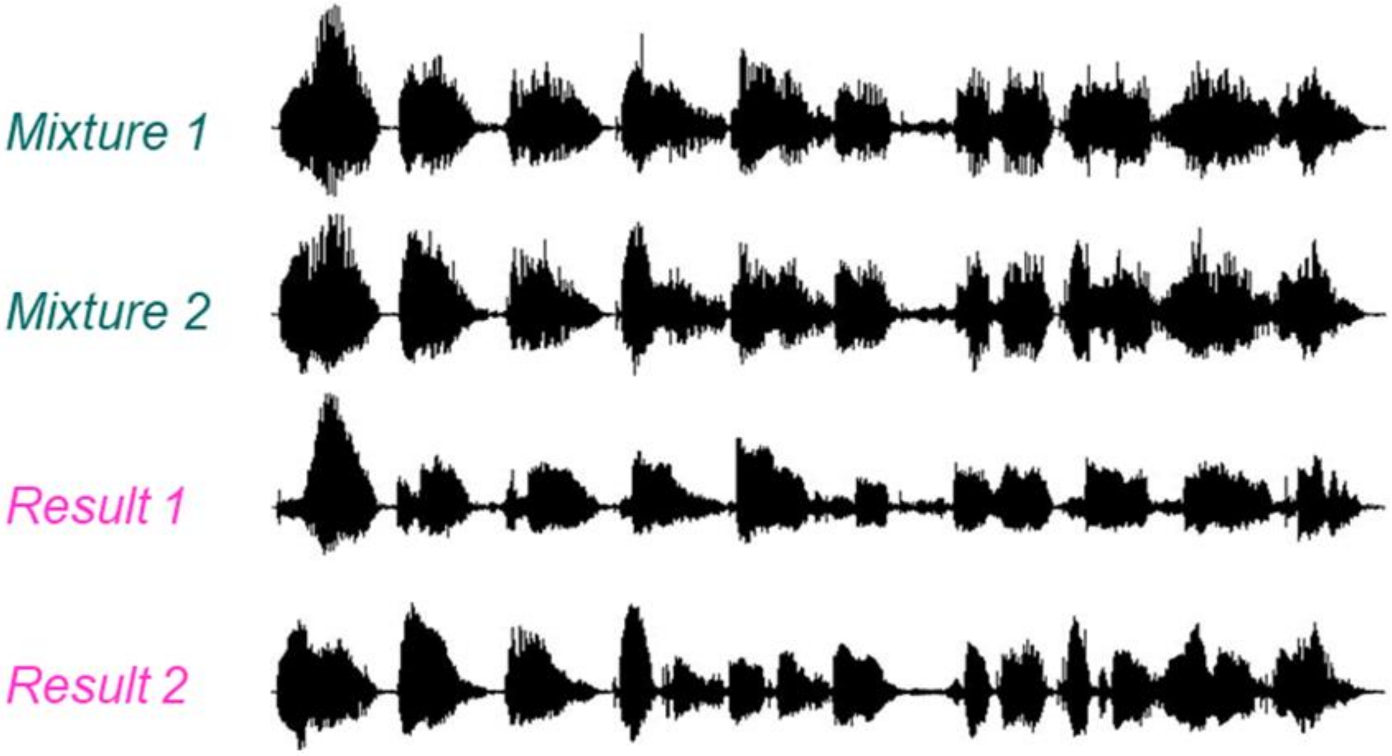


그림 11.25. 2×2 실험환경에서 필터뱅크 방법으로 암묵신호분리 한 결과

[37. Hyung-Min Park et al. 2006]

예제 11.8-1

식 (11.8.10)과 같이 주어진 암묵신호분리 필터에서 입력이 2이고 출력이 2인 경우에 야코비안 행렬을 구하여 보아라.