Independent Component Analysis

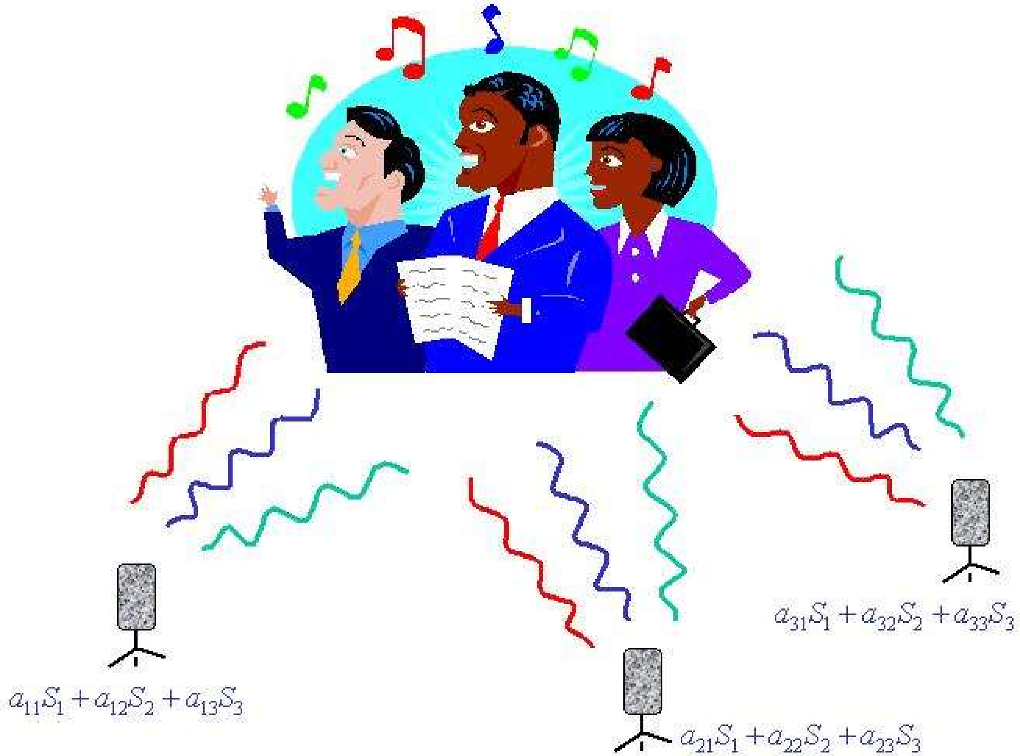
生醫光電所 吳育德

A. Hyvärinen, J. Karhunen, E. Oja (2001) John Wiley & Sons. Independent Component Analysis

The Principle of ICA: a cocktail-party problem

$x_1(t)=a_{11}\, s_1(t) + a_{12}\, s_2(t) + a_{13}\, s_3(t)$
$x_2(t)=a_{21}\, s_1(t) + a_{22}\, s_2(t) + a_{23}\, s_3(t)$
$x_3(t)=a_{31}\, s_1(t) + a_{32}\, s_2(t) + a_{33}\, s_3(t)$

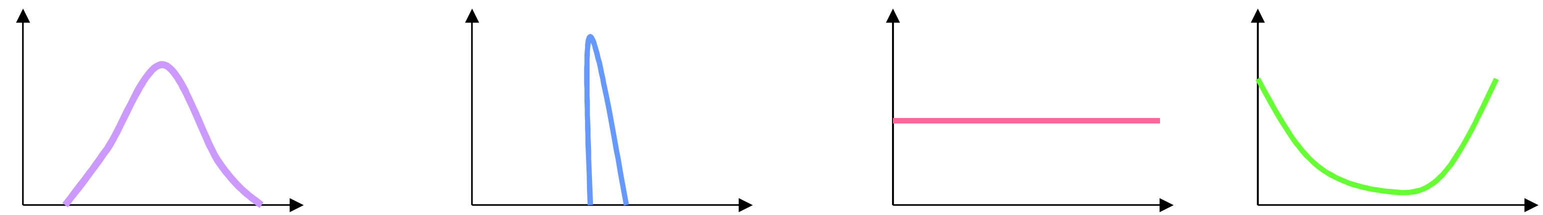
$$\Rightarrow \boldsymbol{x} = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \end{bmatrix}$$



Central limit theorem

The distribution of a sum of independent random variables tends toward a Gaussian distribution

Observed signal = m_1 IC1 + m_2 IC2 ....+ m_n ICn



Toward Gaussian Non-Gaussian Non-Gaussian Non-Gaussian

Central limit theorem

- Partial sum of a sequence $\{z_i\}$ of independent and identically distributed random variables z_i

$$x_i = \sum_{i=1}^{N} z_i$$

- Since mean and variance of x_k can grow without bound as $k \to \infty$, instead of x_k, we consider the standardized variables

$$y_k = \frac{x_k - m_{x_k}}{\sigma_{x_k}}$$

- The distribution of $y_k \to$ a Gaussian distribution with zero mean and unit variance when $k \to \infty$.

How to estimate ICA model

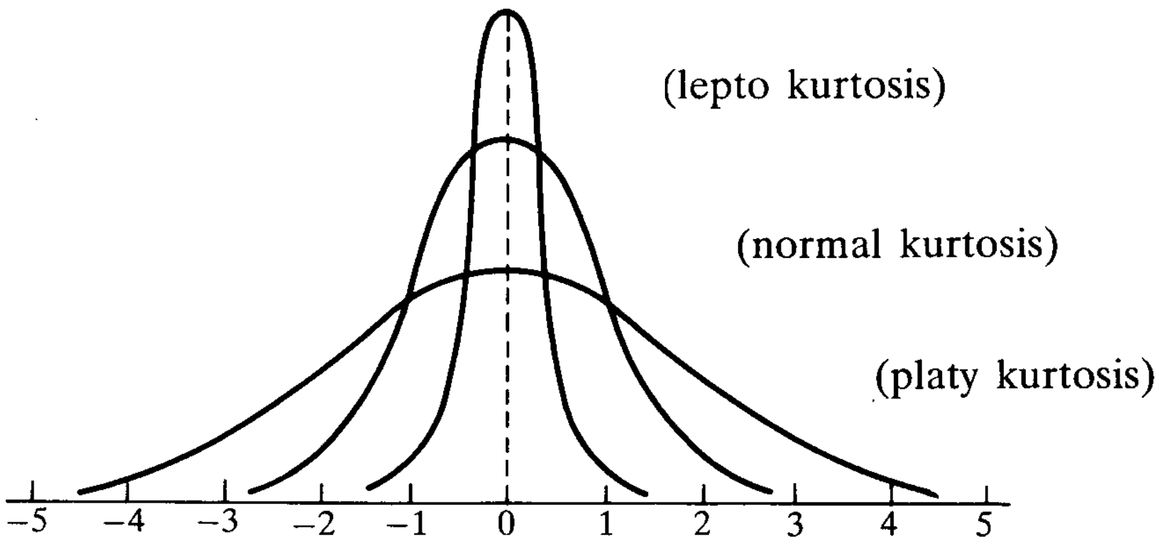
- Principle for estimating the model of ICA

$\Downarrow$

$\boxed{\text{Maximization of NonGaussianity}}$

Measures for NonGaussianity

- Kurtosis: $\mathrm{E}\{(x-\mu)^4\} - 3[\mathrm{E}\{(x-\mu)^2\}]^2$



→ Super-Gaussian kurtosis > 0

→ Gaussian kurtosis = 0

→ Sub-Gaussian kurtosis < 0

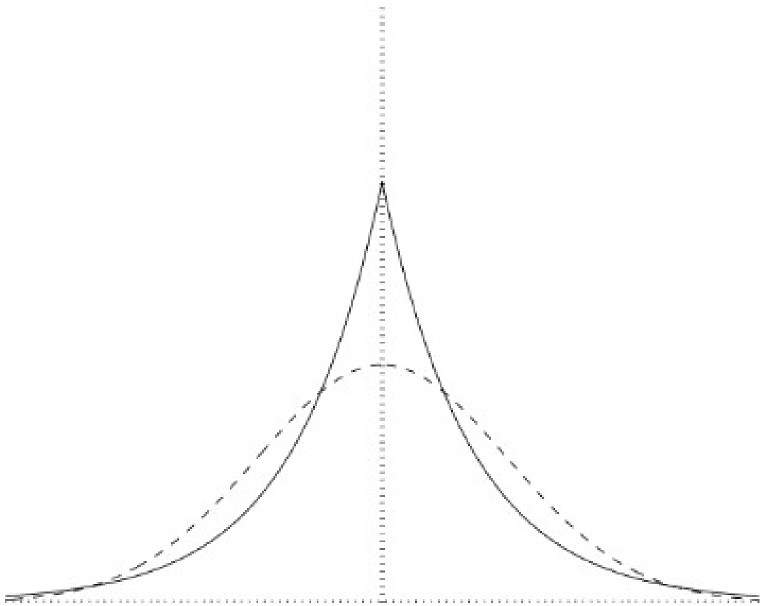
$\mathrm{kurt}(x_1 + x_2) = \mathrm{kurt}(x_1) + \mathrm{kurt}(x_2)$

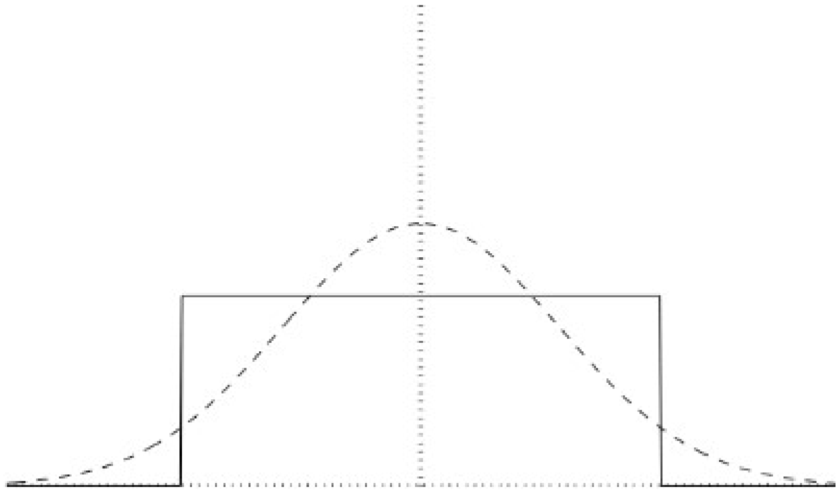
$\mathrm{kurt}(\alpha x_1) = \alpha^4 \mathrm{kurt}(x_1)$

Measures for NonGaussianity

Laplacian distribution (supergaussian): $p(\mathrm{y}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|y|}$

Uniform distribution (subgaussian): $p(\mathrm{y}) = \begin{cases} \frac{1}{2\sqrt{3}}, if\ |y| \leq \sqrt{3} \\ 0, \text{otherwise} \end{cases}$





Whitening process

- Assume measurement $\boldsymbol{x} = \mathbf{As}$ and is zero mean and $E\{\boldsymbol{ss}^T\} = \mathbf{I}$
- Let $\mathbf{D}$ and $\mathbf{E}$ be the eigenvalues and eigenvector matrix of covariance matrix of $\boldsymbol{x}$, i.e. $E\{\boldsymbol{xx}^T\} = \mathbf{EDE}^T, \mathbf{EE}^T = \mathbf{E}^T\mathbf{E} = \mathbf{I}$
- Then $\mathbf{V} = \mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T$ is a whitening matrix

$$\boldsymbol{z} = \mathbf{V}\boldsymbol{x} = \mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T\boldsymbol{x}$$

$\Rightarrow$

$$\begin{aligned} E\{\boldsymbol{zz}^T\} &= \mathbf{V}E\{\boldsymbol{xx}^T\}\mathbf{V}^T \\ &= \mathbf{D}^{-\frac{1}{2}}\mathbf{E}^T\mathbf{EDE}^T\mathbf{ED}^{-\frac{1}{2}} \\ &= \mathbf{I} \end{aligned}$$

Importance of whitening

For the whitened data $\boldsymbol{z}$, find a vector $\boldsymbol{w}$ such that the linear combination y $= \boldsymbol{w}^T\boldsymbol{z}$ has maximum nongaussianity under the constrain $E\{y^2\} = 1$

Then

$$1 = E\{y^2\} = E\{\boldsymbol{w}^T\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{w}\} = \boldsymbol{w}^T E\{\boldsymbol{z}\boldsymbol{z}^T\}\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{w}$$

$\Rightarrow$ Maximize $|\mathrm{kurt}(\boldsymbol{w}^T\boldsymbol{z})|$ under the simpler constraint that $\boldsymbol{w}^T\boldsymbol{w} = 1$

Constrained Optimization

- max $F(\boldsymbol{w})$, $\boldsymbol{w}^T\boldsymbol{w} = 1$
- Using the Lagrangian multiplier λ, the constrained optimization can be rewritten into unconstrained optimization

$$L(\boldsymbol{w}, \lambda) = \boldsymbol{w}^T\boldsymbol{w} = F(\boldsymbol{w}) + \lambda(\boldsymbol{w}^T\boldsymbol{w} - 1)$$

$$\frac{\partial L(\boldsymbol{w}, \lambda)}{\partial \boldsymbol{w}} = \frac{\partial F(\boldsymbol{w})}{\partial \boldsymbol{w}} + \lambda[2\boldsymbol{w}] = 0$$

$$\Rightarrow \quad \frac{\partial F(\boldsymbol{w})}{\partial \boldsymbol{w}} = -2\lambda\boldsymbol{w}$$

- At the stable point, the gradient of $F(\boldsymbol{w})$ must point in the direction of $\boldsymbol{w}$, i.e. equal to $\boldsymbol{w}$ multiplied by a scalar.

Gradient of kurtosis

$$F(\mathbf{w}) = |\text{kurt}(\boldsymbol{w}^T\mathbf{z})| = |\mathrm{E}\{(\boldsymbol{w}^T\mathbf{z})^4\} - 3[\mathrm{E}\{(\boldsymbol{w}^T\mathbf{z})^2\}]^2|$$

$$\frac{\partial F(\boldsymbol{w})}{\partial \boldsymbol{w}} = \frac{\partial|\mathrm{E}\{(\boldsymbol{w}^T\mathbf{z})^4\} - 3[\mathrm{E}\{(\boldsymbol{w}^T\mathbf{z})^2\}]^2|}{\partial \boldsymbol{w}}, \text{recall } E\{\boldsymbol{z}\boldsymbol{z}^T\} = \mathbf{I},$$

$$= \frac{\partial\left|\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{w}^T\mathbf{z}(t))^4 - 3(\boldsymbol{w}^T\boldsymbol{w})^2\right|}{\partial \boldsymbol{w}}, \because\ \mathrm{E}\{\mathrm{y}\} = \frac{1}{T}\sum_{t=1}^{T} y(t)$$

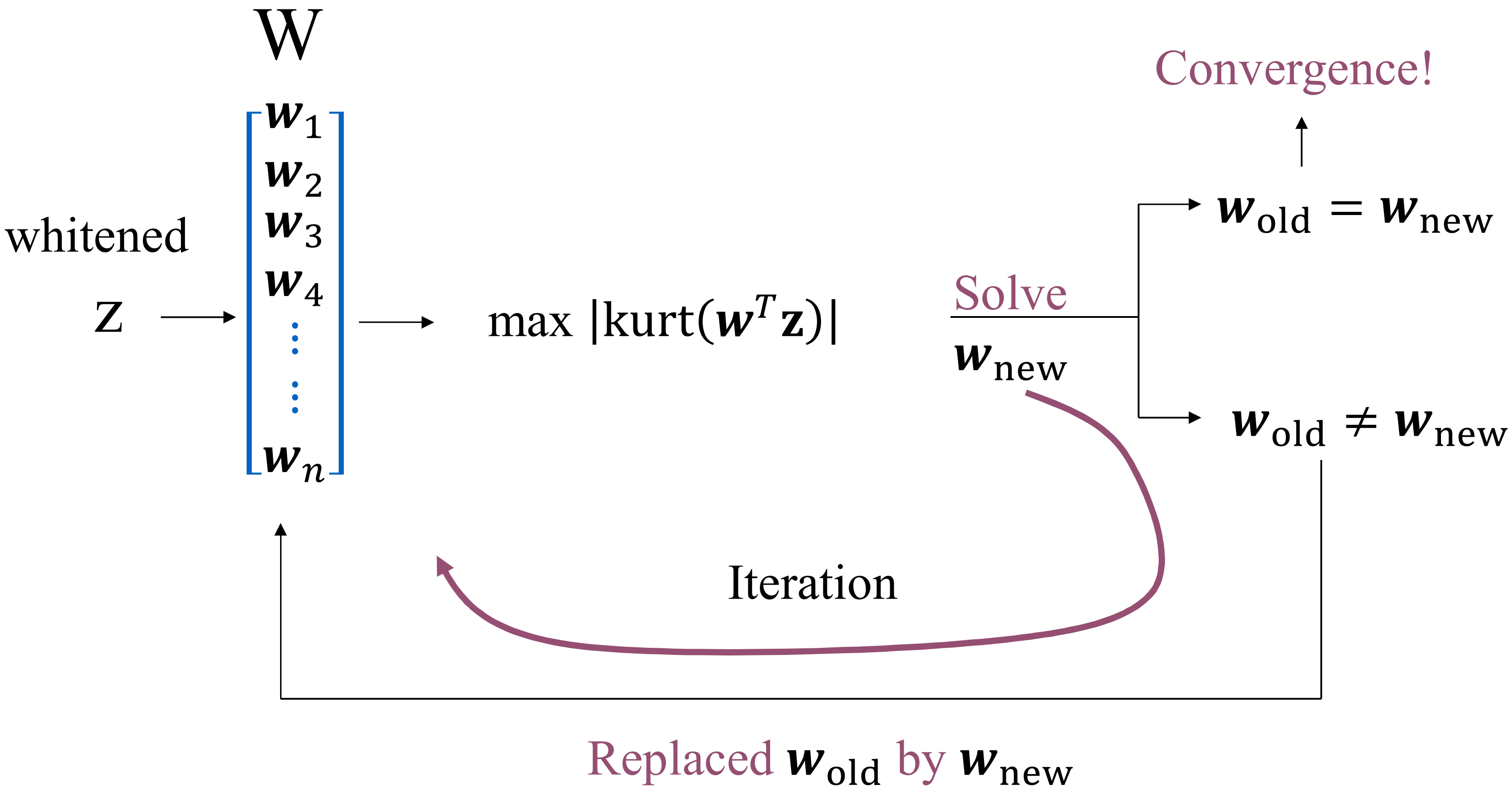
$$= \left|\frac{4}{T}\sum_{t=1}^{T}\boldsymbol{z}(t)\left(\boldsymbol{w}^T\mathbf{z}(t)\right)^3 - 3\times 2(\boldsymbol{w}^T\boldsymbol{w})\,(\boldsymbol{w}+\boldsymbol{w})\right|$$

$$= 4\text{sign}\left(\text{kurt}(\boldsymbol{w}^T\mathbf{z})\right)\left[\mathrm{E}\{\boldsymbol{z}(\boldsymbol{w}^T\mathbf{z})^3\} - 3\boldsymbol{w}\left|\left|\boldsymbol{w}\right|\right|^{\mathbf{2}}\right], \left|\left|\boldsymbol{w}\right|\right|^{\mathbf{2}} = \boldsymbol{w}^T\boldsymbol{w}$$

Fixed-point algorithm using kurtosis

- $\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \alpha \frac{\partial F(\boldsymbol{w}_k)}{\partial \boldsymbol{w}_k} = \left(\frac{-1}{2\lambda} + \alpha\right) \frac{\partial F(\boldsymbol{w}_k)}{\partial \boldsymbol{w}_k}$, since $\frac{\partial F(\boldsymbol{w})}{\partial \boldsymbol{w}} = -2\lambda \boldsymbol{w}$
- Therefore, $\boldsymbol{w} \leftarrow \mathrm{E}\{\boldsymbol{z}(\boldsymbol{w}^T \mathbf{z})^3\} - 3\boldsymbol{w}\big||\boldsymbol{w}|\big|^{\mathbf{2}}, \boldsymbol{w} \leftarrow \frac{\boldsymbol{w}}{||\boldsymbol{w}||}$
- When algorithm converges, adding the gradient to $\boldsymbol{w}_k$ does not change its direction, since
$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \alpha(-2\lambda \boldsymbol{w}_k) = (1-2\alpha\lambda)\boldsymbol{w}_k$$
- i.e., when it converges : $|<\boldsymbol{w}_{k+1}, \boldsymbol{w}_k>|=1$ since $\boldsymbol{w}_k$ and $\boldsymbol{w}_{k+1}$ are unit vectors

A measure of non-Gaussianity : max $\left|\mathrm{kurt}\left(\boldsymbol{w}^T \mathbf{z}\right)\right|$



Fixed-point algorithm using kurtosis

One-by-one Estimation

Fixed-point iteration

1. Centering: $\boldsymbol{x} = \tilde{\boldsymbol{x}} - \boldsymbol{m}_{\tilde{x}}$
2. Whitening: $\mathbf{z} = \mathbf{V}\boldsymbol{x}, E\{\boldsymbol{z}\boldsymbol{z}^T\} = \mathrm{I}$
3. Choose m, No. of ICs to estimate. Set counter $p \leftarrow 1$
4. Choose an initial guess of unit norm for $\boldsymbol{w}_p$, eg. randomly.
5. Let $\boldsymbol{w}_p \leftarrow \mathrm{E}\left\{\boldsymbol{z}\left(\boldsymbol{w}_p^T\mathbf{z}\right)^3\right\} - 3\boldsymbol{w}_p\left\|\boldsymbol{w}_p\right\|^2$
6. Do deflation decorrelation

$$\boldsymbol{w}_p \leftarrow \boldsymbol{w}_p - \sum_{j=1}^{p-1}(\boldsymbol{w}_p^T\boldsymbol{w}_j)\boldsymbol{w}_j$$

7. Let $\boldsymbol{w}_p \leftarrow \frac{\boldsymbol{w}_p}{\left\|\boldsymbol{w}_p\right\|}$
8. If $\boldsymbol{w}_p$ has not converged ($| < w_p^{k+1}, w_p^k > | \neq 1$), go to step 5.
9. Set p ← p+1. If p ≤ m, go back to step 4.

Fixed-point algorithm using negentropy

- The kurtosis is very sensitive to outliers, which may be erroneous or irrelevant observations

- eg. r.v. with sample size=1000, mean=0, variance=1, contains one value = 10, Kurtosis: $\mathrm{E}\{x^4\} - 3$
  → kurtosis at least equal to $10^4/1000-3=7$

- Need to find a more robust measure for nongaussianity

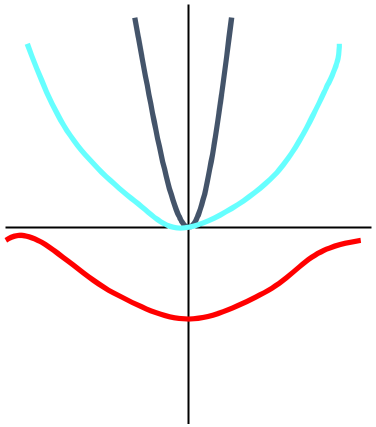
- ⇒ Approximation of negentropy

Fixed-point algorithm using negentropy

- Entropy: $H(\mathrm{y}) = -\int p_y(\eta)\log p_y(\eta)\,\mathrm{d}\eta \ \leq 0$
- Negentropy: $J(\mathrm{y}) = H\left(y_{gauss}\right) - H(\mathrm{y}) \geq 0$
- Approximation of negentropy: $J(\mathrm{y}) \approx [E\{G(y) - E\{G(v)\}]^2, v\sim N(0,1)$
- $G_1(y) = \frac{1}{a_1}\log\cosh a_1 y,\ 1 \leq a_1 \leq 2,$

  $g_1(y) = G_1'(y) = \tanh a_1 y\,, g_1'(y) = a_1[1 - \tanh^2(a_1 y)]$
- $G_2(y) = -\exp(-y^2/2)$

  $g_2(y) = G_2'(y) = y\exp(-\frac{y^2}{2})\,, g_2'(y) = (1 - y^2)\exp(-\frac{y^2}{2})$
- $G_3(y) = \frac{y^4}{4}, g_3(y) = G_3'(y) = y^3, g_3'(y) = 3y^2$



Fixed-point algorithm using negentropy

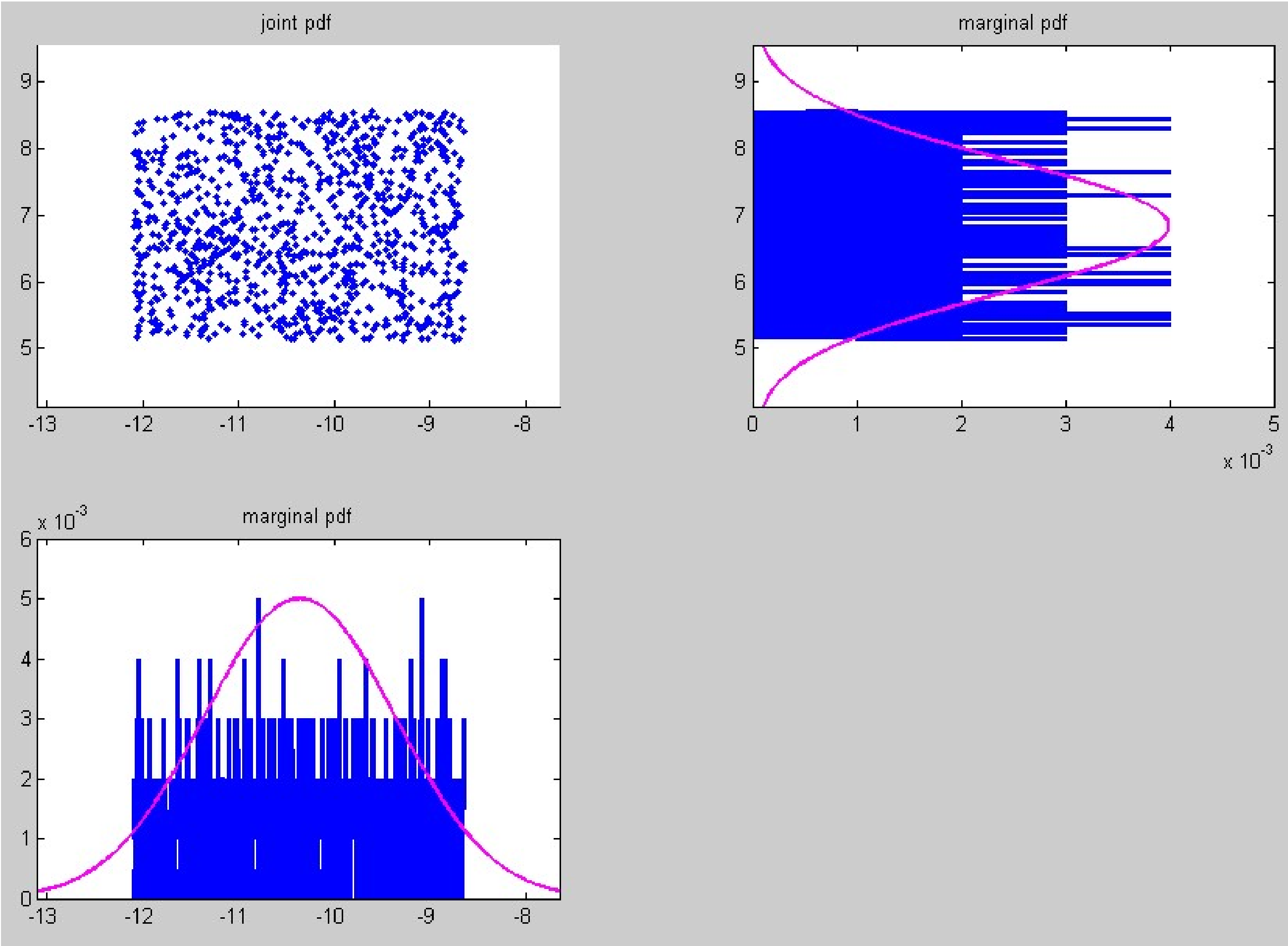
One-by-one Estimation

Fixed-point iteration

1. Centering: $\boldsymbol{x} = \widetilde{\boldsymbol{x}} - \boldsymbol{m}_{\widetilde{\boldsymbol{x}}}$
2. Whitening: $\mathbf{z} = \mathbf{V}\boldsymbol{x}, E\{\boldsymbol{z}\boldsymbol{z}^T\} = \mathrm{I}$
3. Choose m, No. of ICs to estimate. Set counter $p \leftarrow 1$
4. Choose an initial guess of unit norm for $\boldsymbol{w}_p$, eg. randomly.
5. Let $\boldsymbol{w}_p \leftarrow \mathrm{E}\{\boldsymbol{z}\boldsymbol{g}(\boldsymbol{w}_p^T\mathbf{z})\} - \mathrm{E}\{g'(\boldsymbol{w}_p^T\mathbf{z})\}\boldsymbol{w}_p$
6. Do deflation decorrelation
$$\boldsymbol{w}_p \leftarrow \boldsymbol{w}_p - \sum_{j=1}^{p-1}(\boldsymbol{w}_p^T\boldsymbol{w}_j)\boldsymbol{w}_j$$
7. Let $\boldsymbol{w}_p \leftarrow \frac{\boldsymbol{w}_p}{\|\boldsymbol{w}_p\|}$
8. If $\boldsymbol{w}_p$ has not converged ($| < w_p^{k+1}, w_p^k > | \neq 1$), go to step 5.
9. Set p ← p+1. If p ≤ m, go back to step 4.

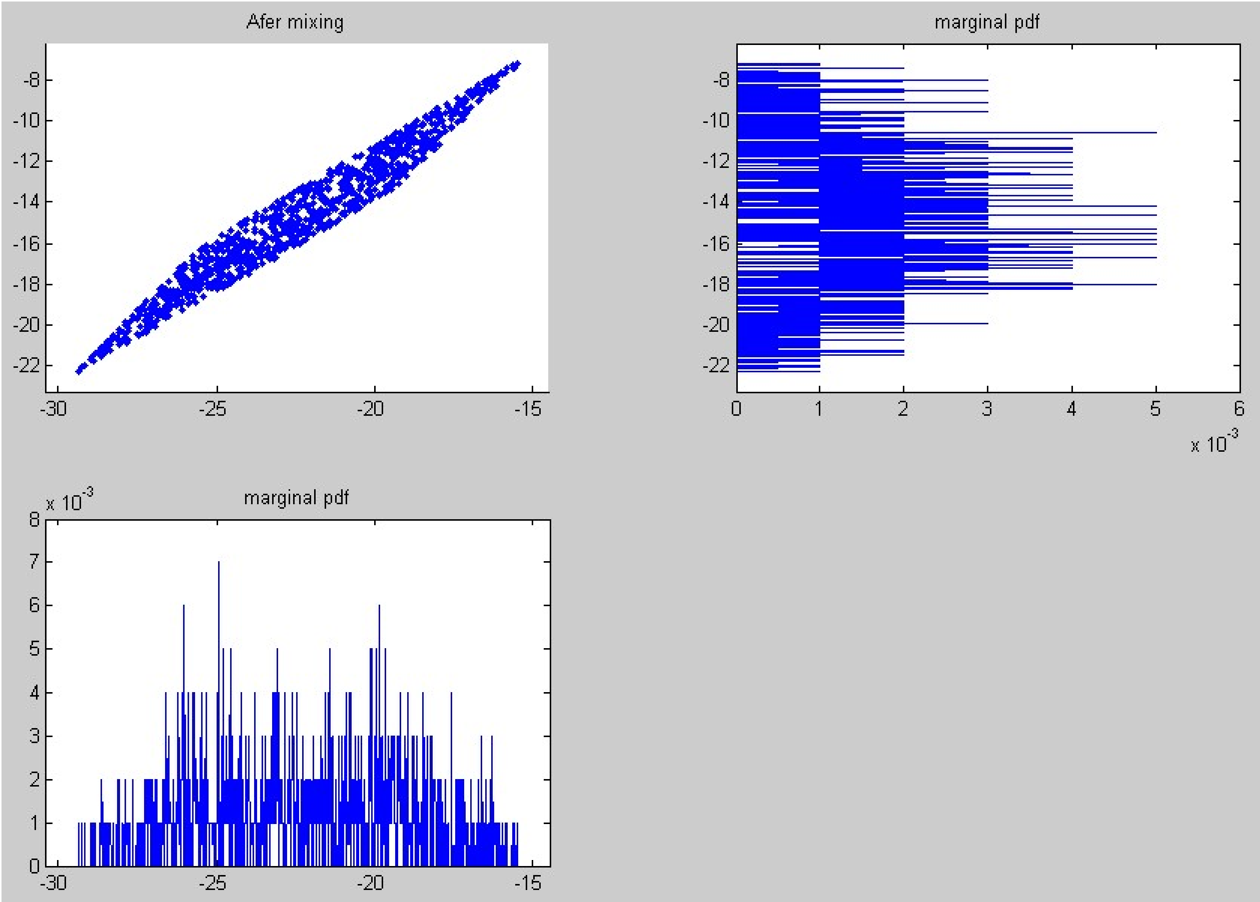
Example

- Create two uniform sources



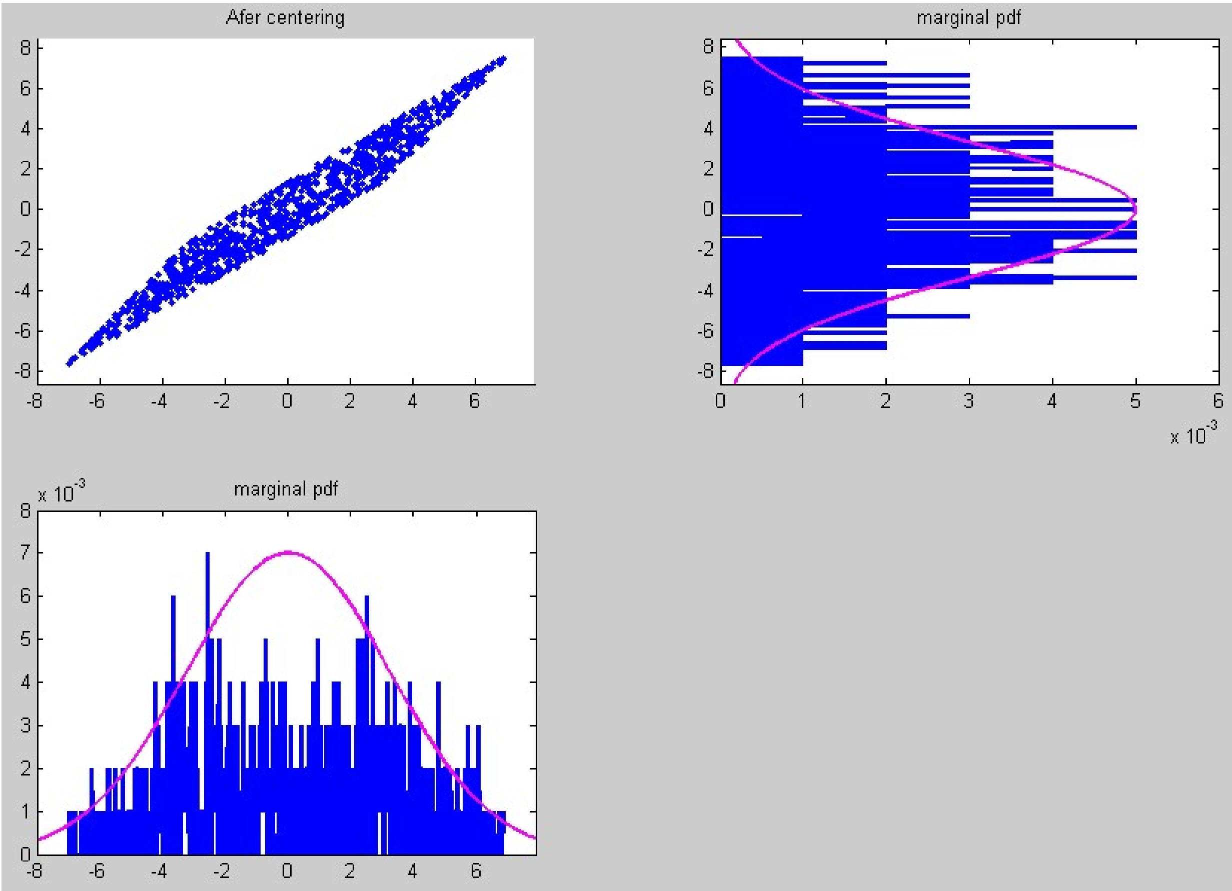
Example

- Two mixed observed signals



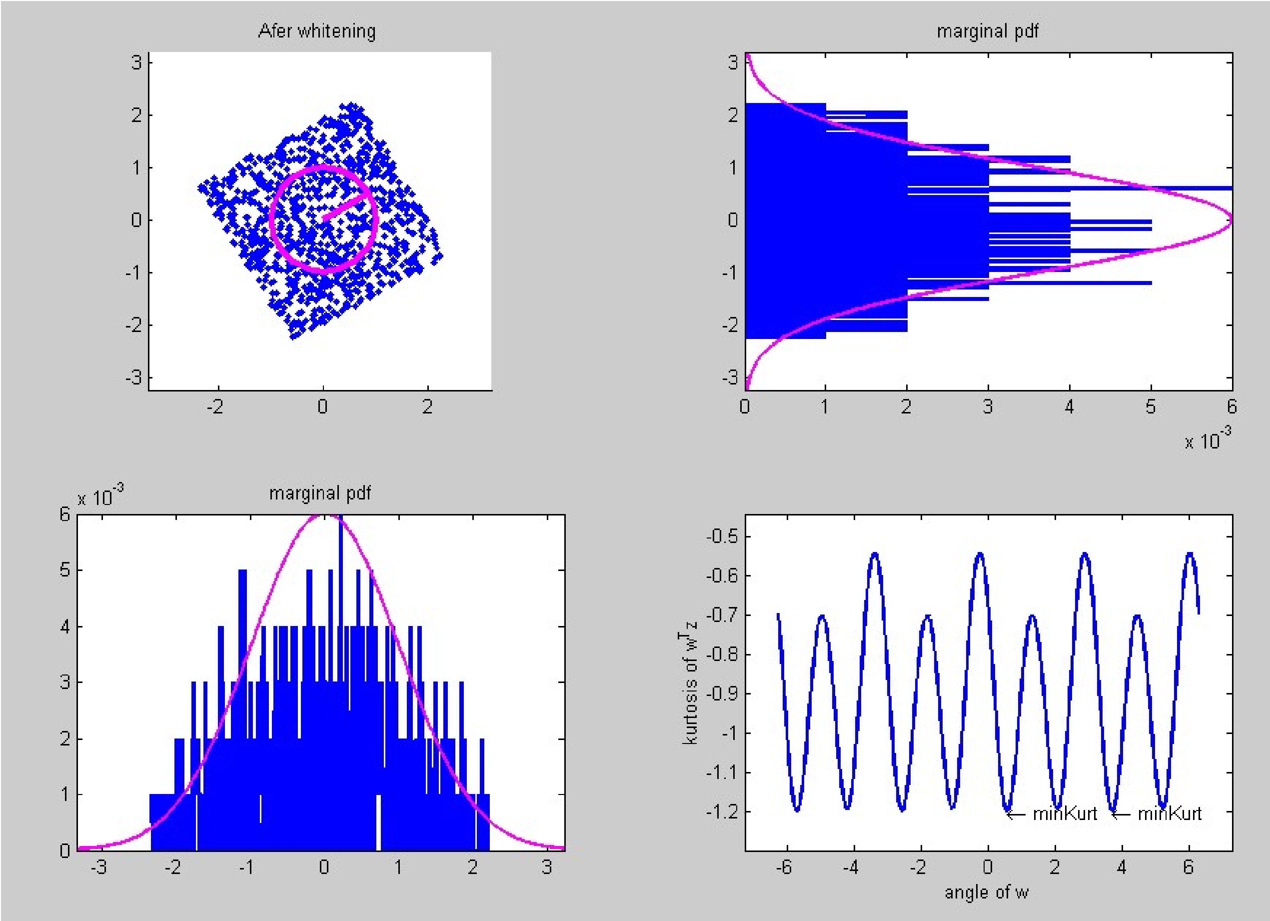
Example

- Centering



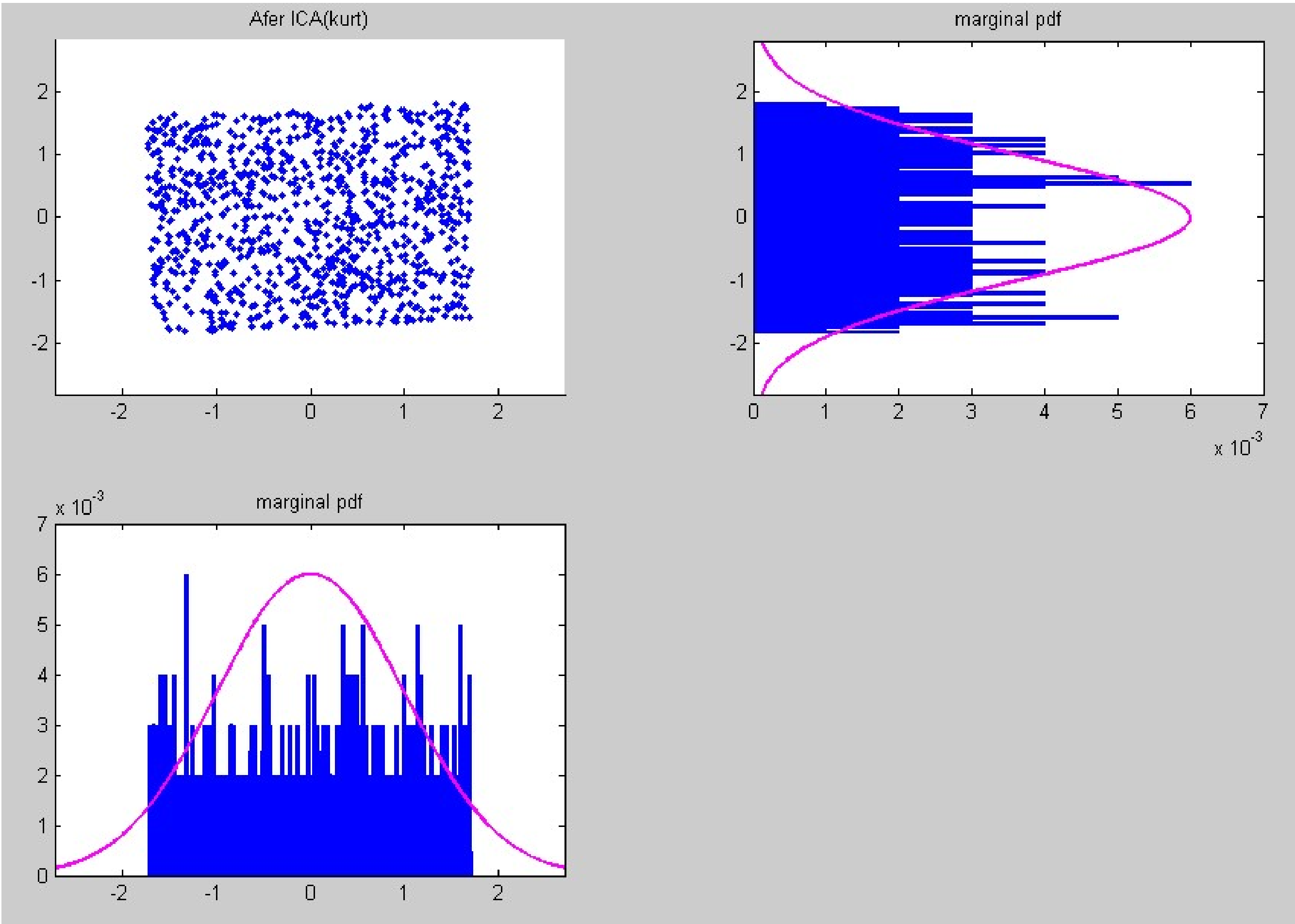
Example

- Whitening



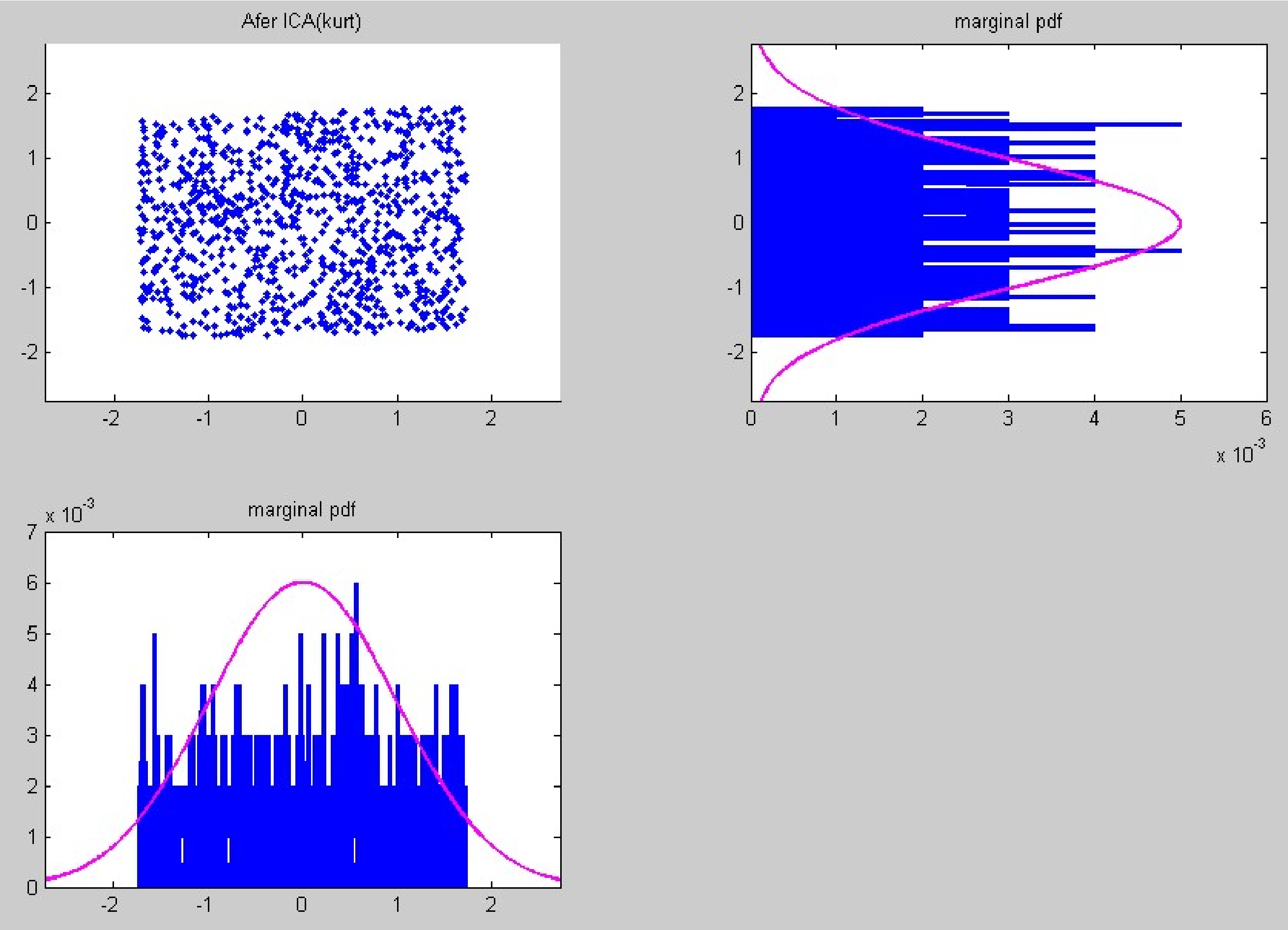
Example

- Fixed-point iteration using kurtosis



Example

- Fixed-point iteration using negentropy



Homework # 3

Rewrite the code Lecture 03_ICA.jpynb into the object-oriented form.

- Hint: see Homework 3_ICA_oop_to_do.jpynb

Deadline of Homework #3: 2022/10/17 3:30pm