

MULTIVARIATE SIGNAL PROCESSING

Multivariable signal of dimension M consists of M scalar signals:

$$\{x_1(n), x_2(n), \dots, x_M(n), n = 0, 1, \dots, N\}$$

Examples:

- biomedical signals (MEG using several sensors)
- geophysical signals (several sensors monitoring earthquakes)
- image can be considered as a multivariate signal along the columns (rows)

Problems:

- *data compression*, for example by using redundancies among the individual signals x_i
- *source signal separation*: to find the set of source signals s_j , when the measured signals x_i are mixtures of unknown source signals,

$$x_i(n) = a_{i1}s_1(n) + a_{i2}s_2(n) + \cdots + a_{iM_S}s_{M_S}(n), \quad i = 1, 2, \dots, M$$

Techniques:

Principal Component Analysis (PCA)

- Performs signal decorrelation (for data compression)

Independent Component Analysis (ICA)

- Performs signal separation into independent source signals

PRINCIPAL COMPONENT ANALYSIS (PCA)

Define

$$w_i(n) = x_i(n) - m_i, \quad i = 1, 2, \dots, M$$

where m_i is the mean value,

$$m_i = \frac{1}{N} \sum_{n=0}^{N-1} x_i(n), \quad i = 1, 2, \dots, M$$

Blind signal decorrelation: express signals w_i in the form

$$w_i(n) = a_{i1}s_1(n) + a_{i2}s_2(n) + \dots + a_{iM_S}s_{M_S}(n), \quad i = 1, 2, \dots, M$$

where the source signals s_j are *uncorrelated*,

$$r_{jk} = \frac{1}{N} \sum_{n=0}^{N-1} s_j(n)s_k(n) = 0, \quad j \neq k, j, k = 1, 2, \dots, M_S$$

If source signals s_j are scaled so that

$$r_{jj} = \frac{1}{N} \sum_{n=0}^{N-1} s_j(n)^2 = 1, \quad j = 1, 2, \dots, M_S$$

we have

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} w_i(n)^2 &= \sum_{n=0}^{N-1} \left(a_{i1}s_1(n) + a_{i2}s_2(n) + \dots + a_{iM_S}s_{M_S}(n) \right)^2 \\ &= a_{i1}^2 + a_{i12}^2 + \dots + a_{i1M_S}^2 \end{aligned}$$

Implication for data compression:

Approximating w_i by the first r source signals,

$$w_i^{(r)}(n) = a_{i1}s_1(n) + a_{i2}s_2(n) + \cdots + a_{ir}s_r(n), \quad i = 1, 2, \dots, M$$

we have the error

$$w_i(n) - w_i^{(r)}(n) = a_{i,r+1}s_{r+1}(n) + \cdots + a_{iM_S}s_{M_S}(n), \quad i = 1, 2, \dots, M$$

and

$$\begin{aligned} \frac{1}{N} \sum_{n=0}^{N-1} \left(w_i(n) - w_i^{(r)}(n) \right)^2 &= \sum_{n=0}^{N-1} \left(a_{i,r+1}s_{r+1}(n) + \cdots + a_{iM_S}s_{M_S}(n) \right)^2 \\ &= a_{i,r+1}^2 + \cdots + a_{iM_S}^2 \end{aligned}$$

\Rightarrow

If $a_{i,r+1}, \dots, a_{i1M_S}$ are small, the multivariable signal can be approximated by

$$w_i^{(r)}(n) = a_{i1}s_1(n) + a_{i2}s_2(n) + \dots + a_{ir}s_r(n), \quad i = 1, 2, \dots, M$$

If $r \ll M$, data compression is achieved

Solution using matrix singular value decomposition

Define vectors

$$\mathbf{w}(n) = \begin{bmatrix} w_1(n) \\ w_2(n) \\ \vdots \\ w_M(n) \end{bmatrix}, \quad \mathbf{s}(n) = \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_{M_S}(n) \end{bmatrix}$$

and matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1M_S} \\ \vdots & & \vdots \\ a_{M1} & \cdots & a_{MM_S} \end{bmatrix}$$

we have

$$\mathbf{w}(n) = \mathbf{A}\mathbf{s}(n), \quad n = 0, 1, \dots, N - 1$$

In matrix form:

$$\mathbf{W} = \mathbf{A}\mathbf{S}$$

where

$$\mathbf{W} = [\mathbf{w}(0) \quad \mathbf{w}(1) \quad \cdots \quad \mathbf{w}(N-1)]$$

$$\mathbf{S} = [\mathbf{s}(0) \quad \mathbf{s}(1) \quad \cdots \quad \mathbf{s}(N-1)]$$

NOTE: the orthonormality property on s_j ,

$$r_{jk} = \frac{1}{N} \sum_{n=0}^{N-1} s_j(n)s_k(n) = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases}$$

implies

$$\frac{1}{N}\mathbf{S}\mathbf{S}^T = \mathbf{I}$$

This follows from

$$\begin{aligned}\frac{1}{N}\mathbf{S}\mathbf{S}^T &= \frac{1}{N} (\mathbf{s}(0)\mathbf{s}^T(0) + \mathbf{s}(1)\mathbf{s}^T(1) + \cdots + \mathbf{s}(N-1)\mathbf{s}^T(N-1)) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \begin{bmatrix} s_1(n) \\ s_2(n) \\ \vdots \\ s_{M_S}(n) \end{bmatrix} [s_1(n) \quad s_2(n) \quad \cdots \quad s_{M_S}(n)] \\ &= \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1M_S} \\ \vdots & \vdots & & \vdots \\ r_{M1} & r_{M2} & \cdots & r_{MM_S} \end{bmatrix}\end{aligned}$$

From above we have that the signal decorrelation problem is equivalent to matrix factorization problem:

Given signal matrix \mathbf{W} , find a factorization

$$\mathbf{W} = \mathbf{A}\mathbf{S}$$

such that

$$\frac{1}{N}\mathbf{S}\mathbf{S}^T = \mathbf{I}$$

NOTE:

The signal decorrelation is not unique:

for any matrix \mathbf{Y} such that $\mathbf{Y}\mathbf{Y}^T$, we have

$$\mathbf{W} = \mathbf{A}_Y \mathbf{S}_Y, \text{ with } \mathbf{A}_Y = \mathbf{A}\mathbf{Y}^{-1}, \mathbf{S}_Y = \mathbf{Y}\mathbf{S}$$

and

$$\frac{1}{N} \mathbf{S}_Y \mathbf{S}_Y^T = \frac{1}{N} \mathbf{Y} \mathbf{S} (\mathbf{Y} \mathbf{S})^T = \frac{1}{N} \mathbf{Y} \mathbf{S} \mathbf{S}^T \mathbf{Y}^T = \mathbf{I}$$

\Rightarrow

If $\mathbf{s}(n)$ is a vector of uncorrelated source signals for $\mathbf{w}(n)$, then $\mathbf{s}_Y(n) = \mathbf{Y}\mathbf{s}(n)$ is also a vector of uncorrelated source signals for $\mathbf{w}(n)$.

Optimal signal decorrelation

Optimal signal decomposition with respect to data compression is achieved if the source signals can be selected so that the approximation errors

$$\sum_{n=0}^{N-1} \sum_{i=1}^M \left(w_i^{(r)}(n) - w_i(n) \right)^2$$

where (cf. above)

$$w_i^{(r)}(n) = a_{i1}s_1(n) + a_{i2}s_2(n) + \cdots + a_{ir}s_r(n), \quad i = 1, 2, \dots, M$$

is minimal with respect to all possible source signals $s_j(n)$ and weights a_{ij} , for all r .

Introducing matrix notation, we have from

$$\mathbf{W} = \mathbf{A}\mathbf{S}$$

that

$$\mathbf{w}(n) = \mathbf{A}\mathbf{s}(n)$$

Let \mathbf{A} and $\mathbf{s}(n)$ be decomposed as

$$\mathbf{A} = [\mathbf{A}_r \quad \tilde{\mathbf{A}}_r]$$

and

$$\mathbf{s}(n) = \begin{bmatrix} \mathbf{s}_r(n) \\ \tilde{\mathbf{s}}_r(n) \end{bmatrix}$$

where \mathbf{A}_r consists of the first r columns of \mathbf{A} , and $\mathbf{s}_r(n)$ contains the first r source signal vectors

Then,

$$\begin{aligned}\mathbf{w}(n) &= \mathbf{A}\mathbf{s}(n) \\ &= [\mathbf{A}_r \quad \tilde{\mathbf{A}}_r] \begin{bmatrix} \mathbf{s}_r(n) \\ \tilde{\mathbf{s}}_r(n) \end{bmatrix} \\ &= \mathbf{A}_r\mathbf{s}_r(n) + \tilde{\mathbf{A}}_r\tilde{\mathbf{s}}_r(n) \\ &= \mathbf{w}^{(r)}(n) + \tilde{\mathbf{A}}_r\tilde{\mathbf{s}}_r(n)\end{aligned}$$

Hence

$$\mathbf{w}^{(r)}(n) = \mathbf{A}_r\mathbf{s}_r(n)$$

and the approximation error is

$$\mathbf{w}(n) - \mathbf{w}^{(r)}(n) = \tilde{\mathbf{A}}_r\tilde{\mathbf{s}}_r(n)$$

In matrix form

$$\mathbf{W} = \mathbf{A}_r \mathbf{S}_r + \tilde{\mathbf{A}}_r \tilde{\mathbf{S}}_r$$

and

$$\mathbf{W} - \mathbf{A}_r \mathbf{S}_r = \tilde{\mathbf{A}}_r \tilde{\mathbf{S}}_r$$

It is straightforward to show that the sum of squared approximation errors, $\left(w_i(n) - w_i^{(r)}(n)\right)^2$ is the sum of the squares of the elements of the M -by- N matrix $\mathbf{W} - \mathbf{A}_r \mathbf{S}_r$,

$$\begin{aligned} \sum_{n=0}^{N-1} \sum_{i=1}^M \left(w_i(n) - w_i^{(r)}(n)\right)^2 &= \sum_{n=1}^N \sum_{i=1}^M [\mathbf{W} - \mathbf{A}_r \mathbf{S}_r]_{in}^2 \\ &= \sum_{n=1}^N \sum_{i=1}^M \left[\tilde{\mathbf{A}}_r \tilde{\mathbf{S}}_r\right]_{in}^2 \end{aligned}$$

The optimal decorrelation problem is thus equivalent to finding a factorization of the signal matrix,

$$\mathbf{W} = \mathbf{A}\mathbf{S}, \quad \text{with} \quad \frac{1}{N}\mathbf{S}\mathbf{S}^T = \mathbf{I}$$

such that for any decomposition

$$\mathbf{W} = [\mathbf{A}_r \quad \tilde{\mathbf{A}}_r] \begin{bmatrix} \mathbf{S}_r \\ \tilde{\mathbf{S}}_r \end{bmatrix} = \mathbf{A}_r\mathbf{S}_r + \tilde{\mathbf{A}}_r\tilde{\mathbf{S}}_r$$

where \mathbf{A}_r consists of the first r columns of \mathbf{A} , and \mathbf{S}_r consists of the first r rows \mathbf{S} , the sum of the squares of the elements of $\tilde{\mathbf{A}}_r\tilde{\mathbf{S}}_r$,

$$\sum_{n=1}^N \sum_{i=1}^M \left[\tilde{\mathbf{A}}_r\tilde{\mathbf{S}}_r \right]_{in}^2$$

is smaller than for any other decomposition of the form $\mathbf{W} = \mathbf{A}\mathbf{S}$.

SOLUTION:

Singular-value decomposition (SVD).

Consider a real $n \times m$ matrix \mathbf{W} . Let $p = \min(m, n)$. Then there exist an $m \times p$ matrix \mathbf{V} with orthonormal columns

$$\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p], \quad \mathbf{v}_i^T \mathbf{v}_i = 1, \quad \mathbf{v}_i^T \mathbf{v}_j = 0 \text{ if } i \neq j,$$

an $n \times p$ matrix \mathbf{U} with orthonormal columns,

$$\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p], \quad \mathbf{u}_i^T \mathbf{u}_i = 1, \quad \mathbf{u}_i^T \mathbf{u}_j = 0 \text{ if } i \neq j,$$

and a diagonal matrix $\mathbf{\Sigma}$ with non-negative diagonal elements,

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0,$$

such that \mathbf{W} can be written as

$$\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

The factorization $\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is called the *singular-value decomposition* of \mathbf{W} .

The nonnegative scalar σ_i are the *singular values* of \mathbf{W}

The vector \mathbf{u}_i is the i th left singular vector of \mathbf{W} .

The vector \mathbf{v}_j is the j th right singular vector of \mathbf{W} .

Notice that orthonormality of \mathbf{u}_i and \mathbf{v}_j is equivalent to

$$\mathbf{V}^T\mathbf{V} = \mathbf{I}, \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}$$

Moreover, it follows that (see lecture notes for details)

$$\sum_{n=1}^N \sum_{i=1}^M [\mathbf{W}]_{in}^2 = \sum_{n=1}^N \sum_{i=1}^M \left[\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \right]_{in}^2 = \sum_{i=1}^p \sigma_i^2$$

The singular value decomposition has precisely the property that the solution of the optimal decorrelation problem is obtained by taking

$$\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{A}\mathbf{S}$$

with

$$\mathbf{A} = \frac{1}{N}\mathbf{U}\mathbf{\Sigma}, \quad \mathbf{S} = N\mathbf{V}^T$$

Decomposing the SVD as

$$\begin{aligned}\mathbf{W} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ &= [\mathbf{U}_r \quad \tilde{\mathbf{U}}_r] \begin{bmatrix} \mathbf{\Sigma}_r & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Sigma}}_r \end{bmatrix} \begin{bmatrix} \mathbf{V}_r^T \\ \tilde{\mathbf{V}}_r^T \end{bmatrix} \\ &= \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T + \tilde{\mathbf{U}}_r \tilde{\mathbf{\Sigma}}_r \tilde{\mathbf{V}}_r^T\end{aligned}$$

the optimal approximation consisting of r source signals is then

$$\mathbf{W}_r = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T = \mathbf{A}_r \mathbf{S}_r$$

with

$$\mathbf{A}_r = \frac{1}{N} \mathbf{U}_r \mathbf{\Sigma}_r, \quad \mathbf{S}_r = N \mathbf{V}_r^T$$

From above we have that sum of squares of the approximation error is

$$\begin{aligned} \sum_{n=1}^N \sum_{i=1}^M [\mathbf{W} - \mathbf{W}_r]_{in}^2 &= \sum_{n=1}^N \sum_{i=1}^M \left[\tilde{\mathbf{U}} \tilde{\Sigma} \tilde{\mathbf{V}}^T \right]_{in}^2 \\ &= \sum_{i=r+1}^p \sigma_i^2 \end{aligned}$$

Hence the sum of squares of the singular values associated with the discarded singular vectors gives directly the sum of squares of the approximation error.

Matlab computations:

Given the signal matrix \mathbf{W} , the representation $\mathbf{W} = \mathbf{A}\mathbf{S}$ in terms of the source signal matrix \mathbf{S} associated with optimal signal decorrelation, and the reduced signal matrix \mathbf{W}_r based on the first r source signals can be determined as follows:

```
[U,Sigma,V]=svd(W)
```

```
A=U*Sigma/N, S=N*V'
```

```
Wr=U(:,1:r)*Sigma(1:r,1:r)*V(:,1:r)'
```

Principal components

The singular value decomposition gives

$$\begin{aligned}\mathbf{W} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ &= [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_p] \begin{bmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & \sigma_p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_p^T \end{bmatrix} \\ &= \mathbf{u}_1\sigma_1\mathbf{v}_1^T + \mathbf{u}_2\sigma_2\mathbf{v}_2^T + \dots + \mathbf{u}_p\sigma_p\mathbf{v}_p^T\end{aligned}$$

As $\mathbf{W} = [\mathbf{w}(1) \quad \mathbf{w}(2) \quad \dots \quad \mathbf{w}(N)]$ it follows that

$$\mathbf{w}(n) = \mathbf{u}_1\sigma_1v_1(n) + \mathbf{u}_2\sigma_2v_2(n) + \dots + \mathbf{u}_p\sigma_pv_p(n)$$

where $v_i(n), i = 1, 2, \dots, p$ is the n th element of the right singular vector \mathbf{v}_i

\Rightarrow

The vector-valued signal $\mathbf{w}(n)$ can be represented as a linear combination of the p vectors $\mathbf{u}_i \sigma_i, i = 1, \dots, p$. These are called *principal components*.

In particular, the solution of the optimal approximation problem discussed above is equivalent to representing the signal with its first r principal components.

Example – Image compression

One application of PCA is in image compression. The idea is to represent the array $\mathbf{X}(i, j)$ associated with an image as a data matrix \mathbf{W} , and compression is achieved by approximating \mathbf{W} with its dominating principal components. The matrix \mathbf{W} can be constructed in various ways, for example by:

- letting \mathbf{W} consist of the columns (or rows) of the array $\mathbf{X}(i, j)$ (after subtraction by mean values)
- letting the columns of \mathbf{W} consist of the elements of sub-blocks of $\mathbf{X}(i, j)$ obtained by stacking the sub-block columns (rows) after each other (using 8 by 8 sub-blocks, \mathbf{W} will have 64 rows).

Example – Eigenfaces

An image consists of an N_{row} -by- N_{col} dimensional array which can be represented as an $N_{\text{row}}N_{\text{col}}$ -dimensional vector \mathbf{w} by stacking the rows (or columns) after each other.

A sequence of images $\mathbf{w}(0), \mathbf{w}(1), \dots, \mathbf{w}(N - 1)$ can be considered as vector-valued signal sequence, and which can be approximated by its principal components.

This is the idea of *eigenfaces*, where principal component representations are used to

- compress, and
- recognize

images of human faces.

INDEPENDENT COMPONENT ANALYSIS (ICA)

Objective:

Given a vector-valued signal sequence $\{\mathbf{w}(\mathbf{n})\}_{\mathbf{n}=0}^{N-1}$, find a decomposition in terms of source signals

$$\mathbf{w}(n) = \mathbf{A}\mathbf{s}(n)$$

such that the source signals s_i, s_j , all $i \neq j$ are independent.

SOLUTION:

The problem can be defined quantitatively in a statistical framework:

Two random variables y_1 and y_2 are independent if knowledge of the value of y_1 does not give any information about the value of y_2 and vice versa.

Joint probability density function $p(y_1, y_2)$ of two independent random variables can be factored as

$$p(y_1, y_2) = p(y_1)p(y_2)$$

Remark:

If y_1 and y_2 are uncorrelated,

$$E[y_1 y_2] = 0$$

that does not imply that they are independent (cf. Example 5.4).

Uncorrelated source signals can be found using PCA.

Exception:

Normally distributed (gaussian) variables are independent if and only if they are uncorrelated (remark 5.3).

As decorrelation is not unique, it follows that normally distributed signals cannot be uniquely separated into independent components.

Measure of information, entropy, mutual information

Information associated with an event with prior probability p :

$$I = \log_2(1/p) = -\log_2 p$$

For $p = 1/2$: $I = -\log_2(1/2) = 1$ bit of information.

Entropy $H(Y)$ of a random variable Y is the expected information obtained when making an observation of the random variable:

$$H(Y) = E[-\log_2(Y)]$$

Among all random variables with the same variance, a normally distributed (gaussian) variable has the largest entropy.

Mutual information between random variables $y_i, i = 1, 2, \dots, M$:

$$I(y_1, y_2, \dots, y_M) = \sum_{i=1}^M H(y_i) - H(\mathbf{y})$$

Difference between the sum of the entropy of the random variables y_i considered individually and the entropy of the random vector \mathbf{y} , where dependence between the individual random variables y_i are taken into account.

The mutual information is non-negative, and zero if and only if the variables are statistically independent.

Mutual information gives a quantitative measure of the (in)dependence of the random variables.

Independent component analysis

Decompose vector-valued signal $\{\mathbf{w}(n)\}$ into independent components $\{s_i(n)\}$ such that

$$\mathbf{w}(n) = \mathbf{A}\mathbf{s}(n)$$

Solution:

Minimize the mutual information of the signals $\{s_i(n)\}$!

Assuming $\dim(\mathbf{s}) = \dim(\mathbf{w})$:

$$\mathbf{s}(n) = \mathbf{B}\mathbf{w}(n), \quad \mathbf{B} = \mathbf{A}^{-1}$$

Mutual information of sources signals can be minimized by observing that:

1. It can be shown that the entropies of \mathbf{s} and \mathbf{w} are related according to

$$H(\mathbf{s}) = H(\mathbf{w}) + \log_2 \det(\mathbf{B})$$

where $\det(\mathbf{B})$ is the determinant of \mathbf{B} .

2. By normalizing the (independent and uncorrelated) source signals to have unit variance, it follows that the determinant of \mathbf{B} satisfies $\det(\mathbf{B}) = \text{constant}$, where the constant depends on the signals $\mathbf{w}(n)$ only.

\Rightarrow

Entropy $H(\mathbf{s})$ of source signals is constant, and:

Mutual information of source signals:

$$\begin{aligned} I(s_1, s_2, \dots, s_M) &= \sum_{i=1}^M H(s_i) - H(\mathbf{s}) \\ &= \sum_{i=1}^M H(s_i) + C \end{aligned}$$

where $C = \text{constant}$.

Minimizing mutual information of source signals is equivalent to minimizing their sum of entropies:

Equivalent problem:

Find \mathbf{B} which minimizes $\sum_{i=1}^M H(s_i)$ (sum of entropies).

This problem is still somewhat intractable.

Simplified problem:

Maximize

$$\sum_{i=1}^M J(s_i)$$

where $J(s_i)$ is the **negentropy**:

$$J(y) = H(y_{gauss}) - H(y)$$

y_{gauss} : gaussian random variable with the same variance as y .

Maximizing negentropy \approx maximizing the distance from a gaussian distribution

Negentropy can be approximated as

$$J(y) \approx \left(E[G(y)] - E[G(y_{gauss})] \right)^2$$

where $G(y)$ is a non-quadratic function (cf. eqs. (5.73)).

$$G_1(y) = \frac{1}{a} \log \cosh(ay)$$

$$G_2(y) = -e^{-y^2/2}$$

$$G_3(y) = y^4$$

Remark

$G_3(y)$ is related to the *kurtosis* defined for a zero-mean variable y as

$$\text{kurt}(y) = E[y^4]/\sigma^2 - 3$$

where $\sigma^2 = E[y^2]$. Kurtosis is a measure of "peakedness" of a probability distribution compared to a normally distributed variable, which has kurtosis value = 0.

Practical solution of the ICA problem:

Find \mathbf{B} such that the source signals

$$\mathbf{s}(n) = \mathbf{B}\mathbf{w}(n)$$

are uncorrelated and the approximated sum of negentropies

$$\sum_i J(s_i) \approx \sum_i \left(E[G(s_i)] - E[G(s_{i,gauss})] \right)^2$$

is minimized, where the expectation is approximated as

$$E[G(s_i)] \approx \frac{1}{N} \sum_{n=0}^{N-1} G(s_i(n))$$

Iterative algorithm:

FastICA (research.ics.aalto.fi/ica/fastica/)