

A brief introduction on

**Independent Component Analysis**

and its applications to

Perfusion MR images

Parts 1, 2, and 3 are written by Yi-Hsuan Kao.

Part 4 is written by Yu-Te Wu.

## (1) Independent component analysis

Suppose that there are  $q$  mutually independent random variables described by a vector variable:

$$\mathbf{f}_{q \times 1} = \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ f_i \\ \cdot \\ f_q \end{bmatrix}$$

where  $f_i$  is an independent random variable. Suppose another vector variable  $\mathbf{x}$  is a linear combination of these  $q$  independent random variables described by a vector-matrix multiplication:

$$\mathbf{x}_{p \times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_i \\ \cdot \\ x_p \end{bmatrix} = \mathbf{M}_{p \times q} \mathbf{f}_{q \times 1}$$

where  $\mathbf{M}$  is a mixing matrix,  $x_i$  is the sum of many independent random variables, and  $x_i$  approaches a Gaussian distribution as compared to the latent variables  $f_i$  according to the central limit theorem (19).

The ICA techniques intend to find a  $q \times p$  unmixing matrix,  $\mathbf{W}$ , which converts the vector variable  $\mathbf{x}$  into another vector variable,  $\mathbf{c}$ , composed of  $q$  mutually independent random variables described by:

$$\mathbf{c}_{q \times 1} = \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ c_i \\ \cdot \\ c_q \end{bmatrix} = \mathbf{W}_{q \times p} \mathbf{x}_{p \times 1}$$

The mutual independence of  $c_i$  means that if  $P(c_i)$  represents the probability distribution of the  $i^{\text{th}}$  component, the joint probability distribution for all components can be factorized as:

$$P(c_1, c_2, \dots, c_q) = P(c_1)P(c_2)\dots P(c_q)$$

The ICA techniques use this assumption of mutual independence and an iterative process to calculate the  $\mathbf{W}$  matrix. Ideally,  $\mathbf{c}$  and  $\mathbf{f}$  should describe the same  $q$  independent random variables, except that their order of components may be different and the means and variances of the corresponding components are rescaled.

## (2) Partial-volume mixing on MR images

Signal intensities on MR images are observed from voxels of finite size. Consequently, partial-volume mixing of different tissues on these voxels is inevitable. Suppose that  $q$  tissues are identified on a set of perfusion images composed of  $v$  voxels. Fractional volumes of the  $q$  tissues in the  $v$  voxels can be described by a  $q \times v$  fractional-volume matrix:

$$\mathbf{F}_{q \times v} = \begin{bmatrix} F_{1,1} & F_{1,2} & \cdot & \cdot & \cdot & F_{1,v} \\ F_{2,1} & F_{2,2} & \cdot & \cdot & \cdot & F_{2,v} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & F_{i,k} & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ F_{q,1} & F_{q,2} & \cdot & \cdot & \cdot & F_{q,v} \end{bmatrix}$$

where  $F_{i,k}$  is the fractional volume for  $i^{\text{th}}$  tissue in the  $k^{\text{th}}$  voxel, with a value between zero and one. Using this notation,  $v$  voxels in a two-dimensional image are reordered into a row vector. Each row represents a fractional-volume image for a tissue type.

For a dynamic study with  $p$  temporal images, the signal intensities of voxels occupied by the  $q$  tissues can be expressed using a  $p \times q$  mixing matrix:

$$\mathbf{M}_{p \times q} = \begin{bmatrix} M_{1,1} & M_{1,2} & \cdot & \cdot & \cdot & M_{1,q} \\ M_{2,1} & M_{2,2} & \cdot & \cdot & \cdot & M_{2,q} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & M_{r,i} & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ M_{p,1} & M_{p,2} & \cdot & \cdot & \cdot & M_{p,q} \end{bmatrix}$$

where  $M_{r,i}$  is the signal intensity on the  $r^{\text{th}}$  temporal image for the  $i^{\text{th}}$  pure tissue. Each column represents a signal-time curve on the perfusion images for a voxel occupied by a pure tissue type. The observed signals on the  $p$  temporal images for the  $v$  voxels can be represented using a  $p \times v$  matrix:

$$\mathbf{X}_{p \times v} = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdot & \cdot & \cdot & X_{1,v} \\ X_{2,1} & X_{2,2} & \cdot & \cdot & \cdot & X_{2,v} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & X_{r,k} & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ X_{p,1} & X_{p,2} & \cdot & \cdot & \cdot & X_{p,v} \end{bmatrix}$$

where  $X_{r,k}$  is the observed signal intensity on the  $r^{\text{th}}$  temporal image for the  $k^{\text{th}}$  voxel. Each row represents a temporal image.

Assuming that signal intensities of different tissues in a voxel can be added linearly, the observed signal is equal to the signal intensities of the  $q$  pure tissues weighted by their fractional

volumes in the voxel. The process of fractional-volume mixing can be expressed using a matrix representation described by:

$$\mathbf{X} = \mathbf{M} \mathbf{F}$$

$p \times v \quad p \times q \quad q \times v$

If an inversion matrix can be found, the fractional-volume images are calculated as follows:

$$\mathbf{F} = \mathbf{M}^{-1} \mathbf{X}$$

$q \times v \quad q \times p \quad p \times v$

However, usually only the observed signals are known, and it is not possible to calculate  $\mathbf{M}^{-1}$  based on  $\mathbf{X}$  alone. The present study applies ICA to decompose the  $\mathbf{X}$  matrix into an independent-component matrix that is related to the  $\mathbf{F}$  matrix.

### (3) Application of ICA to MR images

To apply ICA to perfusion images, the meanings of the  $\mathbf{F}$  and  $\mathbf{X}$  matrices are modified as follows. The sample space for  $f$  is represented by the  $\mathbf{F}$  matrix. Each row in the  $\mathbf{F}$  matrix is treated as the  $v$  measurements for an independent random variable,  $f_i$ . Here we assume a spatial independence for the sample space of  $f$ . This implies that the fractional-volume images are statistically independent with respect to each other. The sample space of  $x$  is represented by the  $\mathbf{X}$  matrix. There are  $v$  measurements for each  $x_i$  and the  $\mathbf{X}$  matrix is the only known information in this study.

Similarly, the sample space for  $c$  is calculated as:

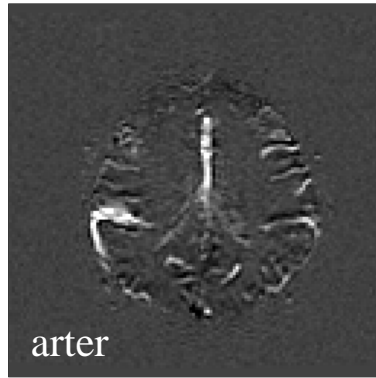
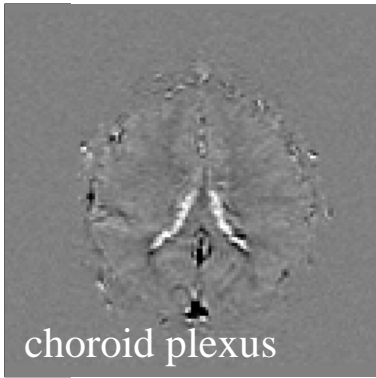
$$\mathbf{C} = \begin{bmatrix} C_{1,1} & C_{1,2} & \cdot & \cdot & \cdot & C_{1,v} \\ C_{2,1} & C_{2,2} & \cdot & \cdot & \cdot & C_{2,v} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & C_{i,k} & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ C_{q,1} & C_{q,2} & \cdot & \cdot & \cdot & C_{q,v} \end{bmatrix} = \mathbf{W} \mathbf{X}$$

$q \times v \quad q \times p \quad p \times v$

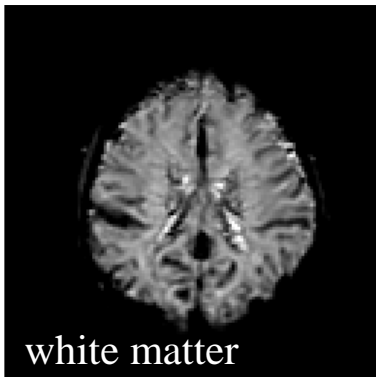
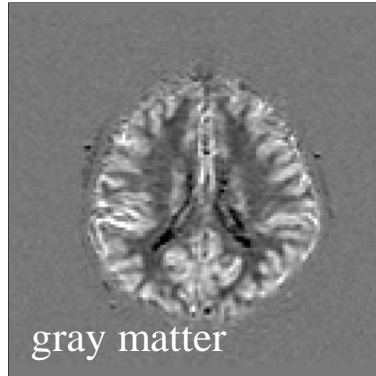
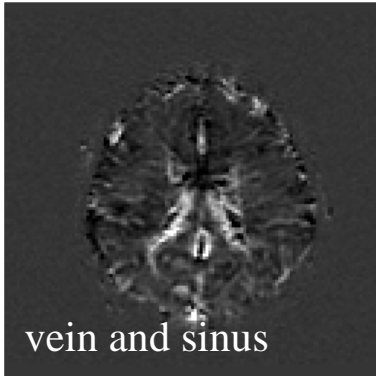
where the matrix element  $C_{i,k}$ , for  $k = 1, 2, \dots, v$ , is the sample space for the independent variable  $c_i$ . The  $i^{th}$  row in  $\mathbf{C}$  is treated as the  $i^{th}$  independent-component image, which is related to one of the fractional-volume images described by the rows of  $\mathbf{F}$  matrix. Because the  $\mathbf{C}$  matrix is the calculated sample space for the vector variable  $c$ , it follows that the independent-component images are uncorrelated and their inner products are expected to be zero:

$$E\{c_i c_j\} \approx \frac{1}{v} \sum_{k=1}^v C_{ik} C_{jk} = 0, \quad \text{if } i \neq j$$

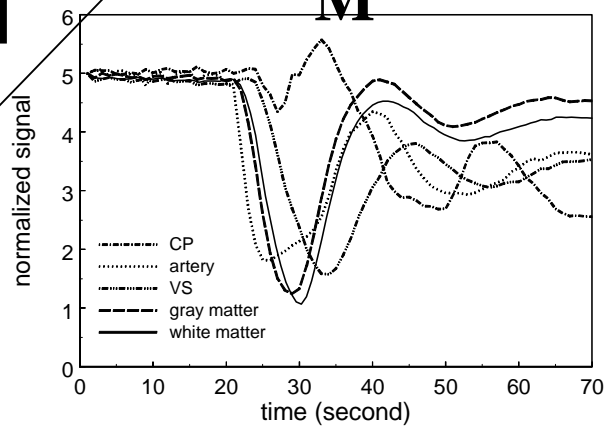
where  $E\{.\}$  is the expected value and  $k$  is the location of a voxel in an image. The signal-time curves for the independent-component images can be obtained from the columns of  $\mathbf{M}$ , which is calculated from the pseudo-inverse of  $\mathbf{W}$ .



$C_i$ , rows of  $c$



columns of  
 $M$



#### (4) The FastICA algorithm

In brief, the salient features of the FastICA algorithm are: (1) cubic convergence, meaning high speed; (2) no need to choose the step size parameters or learning rate as in gradient-based algorithms, which makes FastICA easy to use and more reliable.

The FastICA technique first removes means of the row vectors in the  $\mathbf{X}$  matrix such that each variable  $x_i$  has zero mean. The new data is whitened by a linear transformation  $\mathbf{U}$  such that the covariance matrix of the whitened data,  $\mathbf{z}=\mathbf{U}\mathbf{x}$ , is an identity matrix. The whitening process is mathematically similar to principal component analysis. The uncorrelated variances of the transformed variables provided by the whitening process is a necessary prerequisite for the stricter independence condition and makes the subsequent separation task easier. Another advantage of the whitening process is that the sources or independent components may be estimated from the first  $N$  ( $N \leq n$ ) largest principal components, which interpret the greatest amount of variance. After whitening, only the first  $N$  most significant terms are preserved in the FastICA calculation.

The next step is to look for an un-mixing matrix such that the transformed data  $\mathbf{c}=\mathbf{V}\mathbf{z}$  is mutually independent. Mutual information is a natural measure of the independence of random variables and can be used as the criterion for finding the ICA transformation. Hyvarinen showed that the mutual information could be expressed in terms of negentropy, an important measure of non-Gaussianity. Therefore, the problem of finding the independent components ( $\mathbf{c}$ ) and the transform matrix ( $\mathbf{W}$ ) can be translated into a search for linear combinations of the data ( $\mathbf{z}$ ) that maximize the negentropy of the distributions of  $c_i$ . More specifically, to determine the un-mixing matrix  $\mathbf{W}$  so that the transformed components  $c_i$  are as statistically independent as possible, the FastICA technique minimizes the negentropy defined as follows:

$$J(\mathbf{c}) = H(\mathbf{c}_{Gauss}) - H(\mathbf{c})$$

where  $H(\cdot)$  denotes the differential entropy:

$$H(\mathbf{c}) \equiv -\int p(\mathbf{c}) \log_2 [p(\mathbf{c})] d\mathbf{c}$$

where  $p(\cdot)$  is the probability density function and  $\mathbf{c}_{Gauss}$  is a Gaussian-distributed random vector having the same covariance matrix as  $\mathbf{c}$ . The negentropy is always non-negative and it is zero if and only if all the random variables in  $\mathbf{c}$  are Gaussian distributed.

Because it is computationally difficult to estimate  $p(\mathbf{c})$  in Eq. [13], simpler approximations of negentropy were proposed to use higher moments, for example, kurtosis or other non-quadratic functions. Kurtosis is zero for a Gaussian random variable, positive for a super-Gaussian random variable and negative for a sub-Gaussian random variable. An approximation of negentropy in the FastICA technique is as follows:

$$J(\mathbf{c}) \propto [E\{G(\mathbf{c})\} - E\{G(\boldsymbol{\mu})\}]^2$$

where  $E\{\cdot\}$  is the expected value,  $G(\cdot)$  is a nonquadratic function, for example,  $G(c_i) = \log \cosh(c_i)$ , and  $\boldsymbol{\mu}$  is a Gaussian-distributed random variable with zero mean and unit variance. The FastICA technique maximizes this approximation for the  $N$  random variables  $c_i$ , using batch algorithms based on a fixed-point iteration. A more detailed description of the FastICA iteration process can be found in the reference.

Reference: Hyvarinen A, Karhunen J, Oja E. Independent Component Analysis, John Wiley & Sons, Inc., New York, 2001.