# Exercices on blind signal separation and ICA course

Christian Jutten
Email: Christian.Jutten@inpg.fr

May 2007

## 1 Change of variables

### 1.1 Random variable

Let us consider the random variable $X$, with probability density function (pdf)$p_X$ and the transformed random variable $Y = f(X)$, where $f$ is an invertible function.

- Express the pdf of $Y$, $p_Y$ as a function of $p_X$.
- The differential entropy of a random variable $Y$ is defined as:

$$H(Y) = -\int_{-\infty}^{+\infty} p_Y(y) \log p_Y(y) dy. \quad (1)$$

  Deduce from the result of the previous question that:

$$H(Y) = H(X) + E \ln |g'(Y)|, \quad (2)$$

  where $g'$ is the derivative with respect of $y$ of $g = f^{-1}$.

### 1.2 Random vector

Let us consider the $N$-dimension random vector $\boldsymbol{X}$, with probability density function (pdf)$p_{\boldsymbol{X}}$ and the transformed random variable $\boldsymbol{Y} = f(\boldsymbol{X})$, where $f$ is an invertible function from $\mathbb{R}^N$ to $Real^N$. The Jacobian matrix of the transform will be denoted $J_f$.

- Denoting $g = f^{-1}$, express the pdf of $\boldsymbol{Y}$, $p_{\boldsymbol{Y}}$ as a function of $p_{\boldsymbol{X}}$ and of the Jacobian matrix $J_g$, i.e.:

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = p_{\boldsymbol{X}}(g(\boldsymbol{y})| \det J_g|. \tag{3}$$

- Show that this expression becomes:

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = p_{\boldsymbol{X}}(\boldsymbol{A}^{-1}\boldsymbol{y})| \det \boldsymbol{A}^{-1}| = \frac{p_{\boldsymbol{X}}(\boldsymbol{y})}{| \det \boldsymbol{A}|}. \tag{4}$$

  in the particular case where $f$ is a linear application represented by an invertible matrix $\boldsymbol{A}$.

- The joint differential entropy of a random vector $\boldsymbol{Y}$ is defined as:

$$H(\boldsymbol{Y}) = -\int \dots \int_{-\infty}^{+\infty} p_{\boldsymbol{Y}}(\boldsymbol{y}) \log p_{\boldsymbol{Y}}(\boldsymbol{y}) d\boldsymbol{y}. \tag{5}$$

  Deduce from the result of the previous question that:

$$H(\boldsymbol{Y}) = H(\boldsymbol{X}) + E \ln |J_g|, \tag{6}$$

  in the general case and

$$H(\boldsymbol{Y}) = H(\boldsymbol{X}) - \ln |\boldsymbol{A}|, \tag{7}$$

  in the linear case.

### 1.3 Mutual information

Consider the random vector $\boldsymbol{X}$ and its transform $\boldsymbol{Y} = f(\boldsymbol{X})$ by an invertible diagonal transform, i.e.:

$$\boldsymbol{Y}_j = f_j(\boldsymbol{X}_j), \tag{8}$$

where $\boldsymbol{X}_j$ denotes the $j$th-component of the random vector $\boldsymbol{X}$. Show that the mutual information, $I(\boldsymbol{X}) = \sum_i H(X_i) - H(\boldsymbol{X})$, is preserved by the transform $f$, i.e.:

$$I(\boldsymbol{X}) = I(\boldsymbol{Y}). \tag{9}$$

## 2 Decorrelation

Consider a linear mixture $\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t)$, with the simplified[1] unknown $2 \times 2$ mixing matrix $\boldsymbol{A}$:

$$\boldsymbol{A} = \begin{pmatrix} 1 & a_{12} \\ a_{21} & 1 \end{pmatrix}. \tag{10}$$

The separating system, which provides estimated sources $\boldsymbol{y}(t) = \boldsymbol{B}\boldsymbol{x}(t)$, is a $2 \times 2$ matrix $\boldsymbol{B}$:

$$\boldsymbol{B} = \begin{pmatrix} 1 & b_{12} \\ b_{21} & 1 \end{pmatrix}. \tag{11}$$

- Compute the global matrix $\boldsymbol{G} = \boldsymbol{B}\boldsymbol{A}$.
- Compute the output intercorrelation $E[y_1 y_2]$, and write them as a function of $a_{12}$ and $a_{21}$ depending of the parameter $\sigma_2/\sigma_1$.
- Show that the equation $E[y_1 y_2] = 0$ has an infinite number of solutions $(b_{12}, b_{21})$, and compute the solutions as $b_{21} = f(b_{12})$.
- Deduce from the previous result that the solution of $E[y_1 y_2] = 0$, computed on different time windows where the variance ratios $\sigma_2/\sigma_1$ changes, is the source separation solution.

## 3 Cumulants

Cumulant of order $n$ are function of statistical moments up to order $n$. They are defined from the second characteristic function. The first characteristic function of a random variable $X$, with a pdf $p_X$, is the inverse Fourier transform:

$$\varphi(\nu) = E[\exp(j\nu x)] = \int p_X(x)\exp(j\nu x)dx. \tag{12}$$

It is easy to check that the characteristic function is continuous and equal $\varphi(0) = 1$. We can then define the log of this function: it is the second characteristic denoted:

$$\phi(\nu) = \ln \varphi(\nu). \tag{13}$$

One deduces that this function is continuous too near $x = 0$ and equal to $\phi(0) = 0$. We can then expand it in Taylor series near 0. The cumulants $\kappa_p$

[1] due to indeterminacies

are then defined as the entries of the Taylor series:

$$\phi(\nu) = \ln \varphi(\nu) = \sum_{p=1}^{+\infty} \kappa_p \frac{(j\nu)^p}{p!}. \tag{14}$$

- Writing the Taylor expansion of $\phi(\nu)$ near $\nu = 0$, show that

$$\kappa_p = (-j)^p \frac{d^p \phi(\nu)}{d\nu^p}\Big|_{\nu=0}. \tag{15}$$

- Computing the first terms of the Taylor expansion, compute the first cumulants $\kappa_1$, $\kappa_2$, $\kappa_3$ et $\kappa_4$.

- Deduce the 4 first cumulants for a zero-mean random variable $X$.

- If $X$ is a Gaussian random variable, i.e. with pdf $p_X(x) = (\sqrt{2\pi}\sigma_X)^{-1} \exp(-x^2/(2\sigma_X^2))$, compute the 4 first cumulants.

- Compute the kurtosis, i.e. the normalize 4th-order cumulant $kurt = \kappa_4/\kappa_2^2$, in the general case, for a zero-mean variable and for a Gaussian variable.

## 4 Mutual information as an independence criterion

The mutual information:

$$I(\boldsymbol{Y}) = \int p_{\boldsymbol{Y}}(\boldsymbol{y}) \log \frac{p_{\boldsymbol{Y}}(\boldsymbol{y})}{\prod_k p_{Y_k}(y_k)} d\boldsymbol{y} \tag{16}$$

is positive and vanishes if and only if the two distributions are equal. For proving this property, write first, we will show that $-I(\boldsymbol{Y}) \leq 0$.

- Starting from:

$$-I(\boldsymbol{Y}) = \int p_{\boldsymbol{Y}}(\boldsymbol{y}) \log \frac{\prod_k p_{Y_k}(y_k)}{p_{\boldsymbol{Y}}(\boldsymbol{y})} d\boldsymbol{y}, \tag{17}$$

  apply the inequality: $\ln x \leq x - 1$, $\forall x \in \mathbb{R}^{*+}$, and prove $-I(\boldsymbol{Y}) \leq 0$.

- Then, taking into account that the above equality only holds for $x = 1$, show that $I(\boldsymbol{Y}) = 0$ if $p_{\boldsymbol{Y}}(\boldsymbol{y}) = \prod_k p_{Y_k}(y_k)$, i.e. if components of $\boldsymbol{Y}$ are independent.

# 5 Givens and Jacobi rotation

## 5.1 Givens representation of an orthogonal matrix

Consider a 3-dimension rotation in $\mathbb{R}^3$, associated in a orthogonal basis and to a rotation (orthogonal) matrix $\boldsymbol{U}$:

$$\boldsymbol{U} = \begin{pmatrix} 1 & sin(\theta_1) & sin(\theta_1) \\ -sin(\theta_1) & & sin(\theta_3) \\ -sin(\theta_1) & -sin(\theta_3) & \end{pmatrix}. \tag{18}$$

Show that this matrix can be written as the product of 3 Givens rotation (plane rotation) matrices:

$$\boldsymbol{U}_1 = \begin{pmatrix} cos(\theta_1) & sin(\theta_1) & 0 \\ -sin(\theta_1) & cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{19}$$

$$\boldsymbol{U}_2 = \begin{pmatrix} cos(\theta_2) & 0 & sin(\theta_2) \\ 0 & 1 & 0 \\ -sin(\theta_2) & 0 & cos(\theta_2) \end{pmatrix}. \tag{20}$$

$$\boldsymbol{U}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & cos(\theta_3) & sin(\theta_3) \\ 0 & -sin(\theta_3) & cos(\theta_3) \end{pmatrix}. \tag{21}$$

- Compute the product of the 3 matrices.
- Check the product is an orthogonal matrix.

## 5.2 Diagonalisation by Jacobi rotation

Consider the $n$-dimension symetric matrix:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}, \tag{22}$$

i.e. satisfying $a_{ij} = a_{ji}$, $\forall i, j$.

We then apply a Jacobi rotation by doing the multiplication $\boldsymbol{A}' = \boldsymbol{Q}(k,l,\theta)^T \boldsymbol{A}\boldsymbol{Q}(k,l,\theta)$, with the matrix $\boldsymbol{Q}(k,l,\theta)$:

$$\boldsymbol{Q}(k,l,\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & \cos\theta & \dots & \sin\theta & 0 \\ \vdots & \dots & 0 & 1 & 0 & 0 \\ 0 & 0 & -\sin\theta & \dots & \cos\theta & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \end{pmatrix}, \tag{23}$$

associated to a planar rotation in the plane $(k,l)$. For simplifying the notations, we denote $c = \cos\theta$ and $s = \sin\theta$.

- For computing $\boldsymbol{A}' = \boldsymbol{Q}(k,l,\theta)^T \boldsymbol{A}\boldsymbol{Q}(k,l,\theta)$, first show that the only entries of $\boldsymbol{A}$ modified are the entries of rows $k$ and $l$ and columns $k$ and $l$.

- Show especially, for $h \neq k$ and $h \neq l$,

$$\begin{aligned} a'_{hk} &= a'_{kh} = ca_{hk} - sa_{hl} \\ a'_{hl} &= a'_{lh} = ca_{hl} + sa_{hk} \end{aligned} \tag{24}$$

- Show now that :

$$\begin{aligned} a'_{kk} &= c^2 a_{kk} - s^2 a_{ll} - 2sca_{kl} \\ a'_{ll} &= c^2 a_{kk} - s^2 a_{ll} + 2sca_{kl} \end{aligned} \tag{25}$$

- Show finally, using the symmetry assumption of matrix $\boldsymbol{A}$ that:

$$a'_{kl} = a'_{lk} = (c^2 - s^2)a_{kl} + sc(a_{kk} - a_{ll}). \tag{26}$$

  $a_{kk}$, $a_{kl}$, $a_{lk}$ and $a_{ll}$

- For partly diagonalizing the matrix $\boldsymbol{A}$ with the matrix $\boldsymbol{Q}(k,l,\theta)$, one wants that the off-diagonal elements of matrix $\boldsymbol{A}'$ become equal to zero, i.e. $a'_{kl} = a'_{lk} = 0$. Show that, denoting $t = \tan\theta$ and $2\beta = (a_{ll} - a_{kk})/a_{kl}$, this condition leads to the 2nd degree equation:

$$t^2 + 2\beta t - 1 = 0. \tag{27}$$

Once $t$ is computed, using $c = 1/\sqrt{t^2+1}$ and $s = ct$, we can compute all the modified terms of $\boldsymbol{A}'$ in one shot, by applying the above set of equations. It then leads to a very simple and efficient algorithm.