Journal of Statistical Theory and Applications
Volume 10, Number 4, 2011, pp. 539-551 ISSN 1538-7887

Theory for the Simplest Case of ICA

Le Trung Kien
Institute of Mathematics, University of Greifswald, Germany

Abstract

Let X be any non-Gaussian variable with continuously differentiable density and finite variance. Let N be a Gaussian variable independent with X. If denoting $\mathcal{L}_{(X,N)}$ be a linear space generated by X, N, then we prove that (X, N) is a unique local solution of an optimal problem: $\min\{I(Y_1, Y_2) : Y_1, Y_2 \in \mathcal{L}_{(X,N)}, Y_1, Y_2 \neq O\}$, where I be the mutual information operator, and O be the zero-variable, $O = 0$ a.e.. This is the main idea of ICA for the simplest case. To prove this result, we have used some beautiful properties of convolution connecting information theory with estimation theory.

AMS 2000 Subject Classifications. Primary 62B10; Secondary 94A34.

Keywords and phrases. Independent Component Analysis, mutual information, Shannon entropy, Fisher's information, minimum mean square error.

1. Introduction

The goal of Independent Component Analysis (ICA) is finding a new suitable representation of the data, which minimizes the statistical dependence. This is one of central problems in neural network research, statistics and signal processing. The general notations and calculations of ICA can be seen in [2, 3, 4, 9, 10, 16]. According to [4], the ICA problem was actually first introduced and so named by Herault and Jutten around 1983. More than twenty five years, ICA has received a lot of attention in a wide area of science because of the useful information getting from its new representation of the data.

In this paper, we will focus on the theory for the simplest case of ICA. We assume that the data can be presented by random variables and the suitable transformation to make a new representation of the data is a linear transformation. The ICA problem for two random variables is stated as follows. Let's any two random variables X_1, X_2 which not be "purely" dependent (i.e. $\nexists\ a \in \mathbb{R} : X_1 = aX_2$ a.e.). If denoting $\mathcal{L}_{(X_1,X_2)}$ be a linear space generated by X_1, X_2, the ICA of (X_1, X_2) is finding new random variables $Y_1, Y_2 \in \mathcal{L}_{(X_1,X_2)}$ which have the maximum statistical independence. The main result of this paper shows that: If $X_1, X_2 \in \mathcal{L}_{(X,N)}$ (i.e. $\mathcal{L}_{(X_1,X_2)} \equiv \mathcal{L}_{(X,N)}$), where X is a non-Gaussian variable with continuously differentiable density and finite variance, and N is a standard Gaussian variable independent with X, then by using mutual information for measuring the statistical independence of two random variables, a couple variables (X, N) is not only a global solution but also a unique local solution of the optimal ICA problem. This case is called the Simplest case of ICA. Concretely, denoting I be the mutual information operator, the main result can be explained as follows: $\forall\ Y_1, Y_2 \in \mathcal{L}_{(X,N)}$, $(Y_1, Y_2) \notin \{(aX, bN)|a, b \in \mathbb{R}\}\ \Rightarrow\ \forall \epsilon > 0,\ \exists T \in \mathbb{R}^{2\times 2}$, such that $\|T\| \leqslant \epsilon$ and

$$I(Y_1, Y_2) > I[(Y_1, Y_2)T],$$

where $\|T\|$ is the norm of matrix T. This result not only proves the signification of ICA, but also indicates the sufficiency of mutual information operator in the simplest case.

Recently, the properties of mutual information and Shannon entropy of convolution has been studied by many researchers in information theory and estimation theory. This researchers have tried to bring out the concavity of the entropy under the "variance preserving" convolution for proving the popular inequality in information theory, named Entropy Power Inequality, (see in [1, 5, 12, 13, 14, 15]), and describe the structure of the function estimating non-Gaussian variables in Gaussian noise (see in [7, 8, 17]). In their working, some beautiful properties connecting information theory with estimation theory, which are important tools using in this paper, have appeared.

The remainder of the paper is organized as follows: Definition of the simplest case of ICA is introduced in Section 2. The main result is also given in this section. Section 3 is devoted to prove four lemmas, the crucial tools for proving Theorem. The Theorem, main result of the paper, is proven in section 4. Finally, section 5 is discussion.

2. Information and ICA Problem

Information. Let X be any random variable with finite variance. The Shannon entropy H is defined as follows. If X has a density function $f_X(x)$, then $H(X) = -\int f_X(x) \log f_X(x) dx$; otherwise $H(X) = \infty$. Let X, Y be any couple random variables with finite variances. If X has a density function f_X, Y has a density function f_Y, and (X, Y) has a joint density function $f_{X,Y}$, defining the mutual information

$$I(X, Y) = \int f_{X,Y}(x, y) \log \frac{f_{X,Y}(x, y)}{f_X(x) f_Y(y)} dxdy. \tag{1}$$

By the concavity of the logarithm, $I(X, Y)$ is nonnegative and equals zero only if $f_{X,Y} = f_X f_Y$ a.e. Consequently, the mutual information $I(X, Y)$ can be used for measuring the statistical independence of X and Y. The mutual information can be presented from entropy as follows

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \tag{2}$$

where $H(X, Y) = -\int f_{X,Y} \log f_{X,Y}$ is an entropy of random vector (X, Y). From the property of density function: $f_{aX}(x) = a f_X(\frac{x}{a})$ where f_{aX} is a density of variable aX, $H(aX) = H(X) + \log |a|$ and $I(aX, bY) = I(X, Y)$, for all scales $a, b \neq 0$.

ICA Problem. Let X_1, X_2 be a couple random variables with finite variances. Denote $\mathcal{L}_{(X_1,X_2)}$ is a linear space generated by X_1, X_2. The Independent Component Analysis (ICA) of (X_1, X_2) is finding a new couple random variables $Y_1, Y_2 \in \mathcal{L}_{(X_1,X_2)}$ which minimizes the statistical dependence. If X_1, X_2 are "purely" dependent, i.e. $\exists a \in \mathbb{R} : X_1 = aX_2$ a.e., then $Y_1 = Y_2 = O$, where O is a zero-variable ($O = 0$ a.e.); the ICA problem doesn't have signification. Without loss the signification of ICA problem, we assume that X_1, X_2 are not purely dependent, and ICA solutions $Y_1, Y_2 \neq O$. Using the mutual information operator, the ICA of (X_1, X_2) can be understand as solutions of an optimal problem

$$\min_{Y_1, Y_2 \in \mathcal{L}_{(X_1,X_2)} \setminus \{O\}} I(Y_1, Y_2). \tag{3}$$

Because $I(Y_1, Y_2) = I(aY_1, bY_2)$ for all scales $a, b \neq 0$, we will consider the solution (Y_1, Y_2) be a delegate of the set of solutions $\overline{(Y_1, Y_2)} = \{(aY_1, bY_2) | a, b \neq 0\}$, and stipulate two solutions (Y_1, Y_2) and (Z_1, Z_2) be called different only if $(Y_1, Y_2) \notin \overline{(Z_1, Z_2)}$, and vice versa. By new regulations, the ICA of (X_1, X_2) can be reduced in a sub-space as follows. Defining a relation $\sim$ on $\mathcal{L}_{(X_1,X_2)}$ such that $\forall Y, Z \in \mathcal{L}_{(X_1,X_2)}, Y \sim Z \Leftrightarrow \exists a \in \mathbb{R}, a \neq 0, Y = aZ$ a.e.. It is not difficult to see that $\sim$ is an equivalence relation on $\mathcal{L}_{(X_1,X_2)}$, so we can define a sub-space $\mathcal{L}^*_{(X_1,X_2)} = \mathcal{L}_{(X_1,X_2)} / \sim$. The signification of ICA problem still retains when we consider on the sub-space $\mathcal{L}^*_{(X_1,X_2)}$:

$$\min_{[Y_1],[Y_2] \in \mathcal{L}^*_{(X_1,X_2)} \setminus \{[O]\}} I(Y_1, Y_2), \tag{4}$$

where $[Y] = \{Z \in \mathcal{L}_{(X_1,X_2)} | Z \sim Y\}$. Of course, the ICA of (X_1, X_2) is same with the ICA of $(\widetilde{X}_1, \widetilde{X}_2)$ if $\mathcal{L}_{(X_1,X_2)} \equiv \mathcal{L}_{(\widetilde{X}_1,\widetilde{X}_2)}$.

Main Result. In this paper, we consider a case, which called the Simplest case of ICA, when X_1, X_2 belong to a linear space generated by a non-Gaussian variable X and another standard Gaussian variable N independent with X. We show that, if X is a random variable with continuously differentiable density and finite variance, then the ICA solution of (X_1, X_2) is a couple variables (X, N). Moreover, we prove that $([X], [N])$ is not only a global solution but also a unique local solution of the optimal problem (4). The main result of this paper is stated in following Theorem.

Theorem. *Let any non-Gaussian variable X with continuously differentiable density and finite variance, and a standard Gaussian variable N independent with X. Then the optimal problem*

$$\min_{[Y_1],[Y_2] \in \mathcal{L}^*_{(X,N)} \setminus \{[O]\}} I(Y_1, Y_2) \tag{5}$$

has a unique local solution $([X], [N])$.

Two basic notations in Information theory and Estimation theory are needed in the proof of Theorem. They are *Fisher's information* and *Minimum mean square error* (mmse). Their definitions and some interested properties which using in the proof of Theorem will be introduced in next section.

3. Lemmas

In this section, four results about the relationships between the first derivative of entropy, the Fisher's information and the minimum mean square error (mmse) of convolution are researched. Let X be a random variable with continuously differentiable density f_X. The Fisher's information of X is defined $\mathcal{F}(X) = \mathbb{E}\rho^2(X)$, where $\rho = f'_X/f_X$ is the score function for X. Let two random variables X, Y with finite variances. The minimum mean square error in estimating X with condition of appearing Y is given as formula $mmse(X|Y) = \mathbb{E}\big\{[X - \mathbb{E}(X|Y)]^2\big\}$, where the expectation is taken over the joint distribution of X and Y.

The first result states a linear relationship between the first derivative of entropy and the Fisher's information in convolution case.

Lemma 1. *Let any non-Gaussian variable X with continuously differentiable density f_X. Let N be a standard Gaussian variable independent with X. Then for all $t \in \mathbb{R}$, the convolution $X + tN$ has an interesting property*

$$\frac{d}{dt}H(X + tN) = t\mathcal{F}(X + tN). \tag{6}$$

Proof. Let f_t be a density of variable $X + tN$. We have

$$f_t(y) = \int f_X(x)\frac{1}{\sqrt{2\pi}t}e^{-\frac{\|y-x\|^2}{2t^2}}dx.$$

$$\begin{array}{llrcl}\Rightarrow & 1. & \frac{\partial}{\partial t}f_t(y) & = & \int f_X(x)\frac{1}{\sqrt{2\pi}t^2}\Big(\frac{\|y-x\|^2}{t^2} - 1\Big)e^{-\frac{\|y-x\|^2}{2t^2}}dx \\ & 2. & \frac{\partial}{\partial y}f_t(y) & = & \int f_X(x)\frac{1}{\sqrt{2\pi}t}\Big(\frac{\|y-x\|}{t^2}\Big)e^{-\frac{\|y-x\|^2}{2t^2}}dx \\ & 3. & \frac{\partial^2}{\partial y^2}f_t(y) & = & \int f_X(x)\frac{1}{\sqrt{2\pi}t^3}\Big(\frac{\|y-x\|^2}{t^2} - 1\Big)e^{-\frac{\|y-x\|^2}{2t^2}}dx.\end{array}$$

$\Rightarrow \quad \frac{\partial}{\partial t}f_t(y) = t\frac{\partial^2}{\partial y^2}f_X(y)$. Hence, the first derivative of entropy in t is computed as follows

$$\begin{aligned}\frac{d}{dt}H(X+tN) &= -\int \log f_t(y)\frac{\partial}{\partial t}f_t(y)dy = -\int t\log f_t(y)\frac{\partial^2}{\partial y^2}f_t(y)dy \\ &= -t\int \frac{\partial}{\partial y}\log f_t(y)\frac{\partial}{\partial y}f_t(y)dy + t\int \Big(\frac{\partial}{\partial y}\log f_t(y)\Big)\Big(\frac{\partial}{\partial y}f_t(y)\Big)dy.\end{aligned}$$

Since the Fisher's information of $X + tN$ is existed, then a value $(\frac{\partial}{\partial y}f_t(y))/\sqrt{f_t(y)}$ is bounded when $\|y\| \to \infty$. By $\sqrt{f_t(y)}\log f_t(y) \to 0$ as $\|y\| \to \infty$, we have

$$\frac{d}{dt}H(X + tN) = t\int \Big(\frac{f_t'(y)}{f_t(y)}\Big)^2 f_t(y)dy = t\mathcal{F}(X + tN).$$

The Lemma 1 is proven. $\square$

A simple relation between Fisher's information and minimum mean square error (mmse) of $X+tN$ is mentioned in Lemma 2. This formula was introduced and proven carefully in [7, 13].

Lemma 2. *Let any non-Gaussian variable X with continuously differentiable density f_X and finite variance. Let N be a standard Gaussian variable independent with X. Then*

1. $t^2\mathcal{F}(X+tN) + mmse(N|X+tN) = 1, \quad \forall t \in \mathbb{R}.$
2. $t^2\mathcal{F}(X+tN)$ *decreases in* $t \in (-\infty, 0]$*, increases in* $t \in [0, +\infty)$*, and bounded by zero and one.*

Proof. Let any random variable Z with a continuously differentiable density f_Z. Denote $S(Z) = f_Z^{'}/f_Z$, which called score function for Z, be a zero mean variable. Blachman, N.M. [1] showed us that $S(X+tN) = \mathbb{E}\big[S(tN)|X+tN\big]$. By the law of total variance, we have

$$\begin{aligned}\mathcal{F}(X+tN) &= Var\big[S(X+tN)\big] = Var\big(\mathbb{E}\big[S(tN)|X+tN\big]\big)\\ &= Var\big[S(tN)\big] - \mathbb{E}\big(Var\big[S(tN)|X+tN\big]\big)\\ &= \mathcal{F}(tN) - mmse\big[S(tN)|X+tN\big],\end{aligned}$$

where $Var(\cdot)$ denotes a variance operator of random variable. tN is Gaussian variable with variance t^2, so $S(tN) = -tN/t^2$ and $\mathcal{F}(tN) = t^{-2}$. Therefore

$$t^2\mathcal{F}(X+tN) = 1 - \frac{1}{t^2}mmse(tN|X+tN) = 1 - mmse(N|X+tN). \tag{7}$$

Of course, $mmse(N|X+tN)$ increases from zero to one in $t \in (-\infty, 0]$, and decreases from one to zero in $t \in [0, +\infty)$. Then $t^2\mathcal{F}(X+tN)$ decreases from one to zero in $t \in (-\infty, 0]$, and increases from zero to one in $t \in [0, +\infty)$. $\square$

Lemma 3. *Let any non-Gaussian variable X with continuously differentiable density f_X and finite variance. Let N be a standard Gaussian variable independent with X. Then, for all $t \in \mathbb{R}$, $t \neq 0$*

$$mmse(N|X+tN) = \frac{1}{t^2}mmse\big(X|\frac{1}{t}X+N\big). \tag{8}$$

Proof. By the law of total variance,

$$mmse(N|X+tN) = 1 - Var\big[\mathbb{E}(N|X+tN)\big] = 1 - \int \frac{\kappa_1^2(y,t)}{\kappa_0(y,t)}dy,$$

where $\kappa_i(y,t) = \int f_X(y-tx)\frac{x^i}{\sqrt{2\pi}}e^{-\frac{\|x\|^2}{2}}dx,\ i=0,1$. By changing variable $z = \frac{y-x}{t}$, we have

$$\begin{aligned}
\kappa_1(y,t) &= \frac{y}{t^2}\int f_X(z)\frac{1}{\sqrt{2\pi}}e^{-\frac{\|y-z\|^2}{2t^2}}dz - \int f_X(z)\frac{z}{\sqrt{2\pi}t^2}e^{-\frac{\|y-z\|^2}{2t^2}}dz\\
&= \frac{1}{t}\big[yf_t(y) - \mathbb{E}(X|X_t=y)f_t(y)\big]\\
\Rightarrow \quad \int\frac{\kappa_1^2(y,t)}{\kappa_0(y,t)}dy &= \frac{1}{t^2}\Big(Var(X_t) - 2\mathbb{E}\big[X_t\mathbb{E}(X|X_t)\big] + \mathbb{E}\big[(\mathbb{E}(X|X_t))^2\big]\Big)\\
&= \frac{1}{t^2}\Big(1+t^2 - 2\mathbb{E}\big[\mathbb{E}X_tX|X_t\big] + Var\big[\mathbb{E}(X|X_t)\big]\Big)\\
&= 1 - \frac{1}{t^2}mmse(X|X_t),
\end{aligned}$$

where $X_t = X + tN$. It is not difficult to see $mmse(X|X_t) = mmse(X|\frac{1}{t}X+N)$. Thus, $mmse(N|X+tN) = \frac{1}{t^2}mmse(X|\frac{1}{t}X+N)$. □

The function $mmse(N|X+tN)$ decreases from one to zero, and continuous in $t\in[0,+\infty)$, so an equation $mmse(N|X+tN) = 1/2$ always exists a unique solution which denoted t^*. Since, $mmse(N|X) = 1$, $mmse(N|X+tN)\to 0$ as $t\to+\infty$, then $t^*\in(0,+\infty)$. We also define a function $h:[0,+\infty)\cup\{+\infty\}\to[0,+\infty)\cup\{+\infty\}$ satisfied $mmse(N|X+h(t)N) = 1-mmse(N|X+tN)$ $\forall t\in(0,+\infty)$, and $h(0)=+\infty,\ h(+\infty)=0$. The function h is decreasing and continuous. The last result in this section will focus on a structure of $mmse(N|X+tN)$ related to new notations t^* and $h(t)$, which is the important idea for proving Theorem.

Lemma 4. *Let any non-Gaussian variable X with continuously differentiable density f_X and finite variance. Let N be a standard Gaussian variable independent with X. t^* and $h(t)$ are defined as above. Then*

1. *$mmse(N|X+tN)\leqslant 1-\frac{t}{2t^*},\ \forall t\in[0,t^*]$, and $mmse(N|X+tN) > 1-\frac{t}{2t^*},\ \forall t\in(t^*,+\infty)$.*
2. *Equation $tmmse(N|X+tN) = h(t)mmse(N|X+h(t)N)$ has only three solutions $t=0,t^*$, and $+\infty$ on interval $[0,+\infty]$.*

Proof. Let $g(t) = mmse(N|X+tN) - 1 + t/(2t^*)$, then $g(t^*)=0$. Denote $M_\delta = Var(X|\sqrt{\delta}X+N)$ for all $\delta\geqslant 0$. In [7, 8], D.Guo, et al. showed that $\frac{d}{d\delta}mmse(X|\sqrt{\delta}X+N) = -\mathbb{E}(M_\delta^2)$. Of course, $mmse(X|\sqrt{\delta}X+N) = \mathbb{E}(M_\delta)$. From the Lemma 3, we know $mmse(N|X+tN) = \delta mmse(X|\sqrt{\delta}X+N)$ with $\delta = 1/t^2$. So

$$\begin{aligned}
g'(t) &= \frac{d}{dt}\Big[\delta mmse(X|\sqrt{\delta}X+N)\Big] + \frac{1}{2t^*} = \frac{2}{t^3}\Big[\delta\mathbb{E}(M_\delta^2) - \mathbb{E}(M_\delta) + \frac{t}{4\delta t^*}\Big]\\
&= \frac{2}{t^3}\Big[\Big(\sqrt{\delta}\mathbb{E}(M_\delta) - \frac{1}{2\sqrt{\delta}}\Big)^2 + \delta\Big(\mathbb{E}(M_\delta^2) - \big[\mathbb{E}(M_\delta)\big]^2\Big) + \frac{1}{4\delta}\Big(\frac{t}{t^*}-1\Big)\Big].
\end{aligned}$$

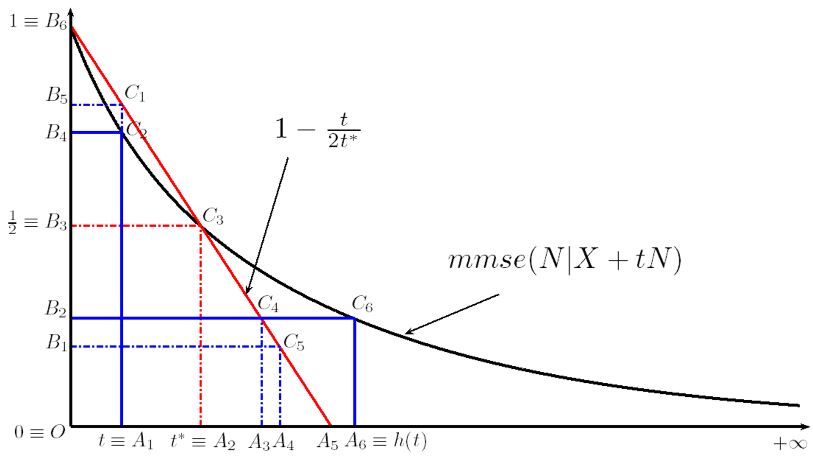

Figure 1: Interpreting a geometric view of the proof of Lemma 4.

Since $\mathbb{E}(M_\delta^2) \geqslant [\mathbb{E}(M_\delta)]^2$, and $t/t^* > 1$, $\forall t > t^* \Rightarrow g$ strictly increases in $t \in (t^*, +\infty)$. Because $g(t^*) = 0$, then $mmse(N|X + tN) > 1 - t/(2t^*)$ for all $t > t^*$. Moreover, if exists $t_1 \in (0, t^*)$ such that $g(t_1) > 0 \;\Rightarrow\; \mathbb{E}(M_{\delta_1}) > 1/\delta_1 - 1/(2\delta_1)\sqrt{\delta^*/\delta_1}$, where $\delta_1 = t_1^{-2}$, $\delta^* = t^{*-2}$. Apply in formula of $g'(t_1)$, we have

$$\begin{aligned} g'(t_1) &\geqslant \left[\sqrt{\delta_1}\mathbb{E}(M_{\delta_1}) - \frac{1}{2\sqrt{\delta_1}}\right]^2 + \frac{1}{4\delta_1}\left(\frac{t_1}{t^*} - 1\right) \\ &> \left[\frac{1}{\sqrt{\delta_1}} - \frac{1}{2\sqrt{\delta_1}}\left(1 + \sqrt{\frac{\delta^*}{\delta_1}}\right)\right]^2 + \frac{1}{4\delta_1}\left(\sqrt{\frac{\delta^*}{\delta_1}} - 1\right) \\ &= \frac{1}{4\delta_1}\left(\sqrt{\frac{\delta^*}{\delta_1}} - 1\right)\left(\sqrt{\frac{\delta^*}{\delta_1}} - 2\right). \end{aligned}$$

$\delta_1 > 0$ and $\delta^*/\delta_1 < 1 \;\Rightarrow\; g'(t_1) > 0 \;\Rightarrow\; \exists \epsilon > 0 : \; g(t_1 + \epsilon) > g(t_1) > 0$. It means that $g(t) > 0$, $\forall t \in [t_1, t^*]$, and $g(t^*) > 0$. Contradict with $g(t^*) = 0$. Thus, the hypothesis $\exists t_1 \in (0, t^*) : \; g(t_1) > 0$ is wrong. Hence, $mmse(N|X + tN) \leqslant 1 - t/(2t^*)$ for all $0 \leqslant t \leqslant t^*$.

The second result in this lemma is a consequence of the first result. The explanation is given in Figure 1. Of course $t = 0, t^*, +\infty$ are solutions of equation $tmmse(N|X + tN) = h(t)mmse(N|X + h(t)N)$. Now we prove that if $t \in (0, t^*)$, then $tmmse(N|X + tN) < h(t)mmse(N|X + h(t)N)$. Indeed, denoting $|\cdot|$ be a length of line segment in geometric view as in Figure 1. Denoting $O \equiv (0, 0)$ be a original point, $A_1 \equiv (t, 0)$, $B_4 \equiv (0, mmse(N|X + tN))$, $C_2 \equiv (t, mmse(N|X + tN))$, and other points $A_2, A_3, \ldots, C_6$ as simulating in Figure 1. Because $mmse(N|X + tN) \leqslant 1 - t/(2t^*)$, $\forall t \in (0, t^*) \;\Rightarrow\; |A_1C_2| \leqslant |A_1C_1| \;\Rightarrow\; |B_5B_6| \leqslant |B_4B_6|$. Let's $|OB_1| = |B_5B_6| = t/(2t^*)$, $|OB_2| = |B_4B_6| = 1 - mmse(N|X + tN)$, we have $|OB_1| \leqslant |OB_2|$.

Because $mmse(N|X + tN) > 1 - t/(2t^*)$, $\forall t \in (t^*, +\infty) \Rightarrow |B_2C_4| < |B_2C_6|$. Note that, since C_1, C_3, C_4, C_5 are in straight line B_6A_5, $|B_6B_5| = |OB_1| \leqslant |OB_2| \leqslant |OB_3| = 1/2 \Rightarrow \mathcal{S}_{OA_1C_1B_5} = \mathcal{S}_{OA_4C_5B_1} \leqslant \mathcal{S}_{OA_3C_4B_2} \leqslant \mathcal{S}_{OA_2C_3B_3}$, where $\mathcal{S}$ denotes a area of rectangle. Therefore, $tmmse(N|X + tN) = \mathcal{S}_{OA_1C_2B_4} \leqslant \mathcal{S}_{OA_1C_1B_5} \leqslant \mathcal{S}_{OA_3C_4B_2} < \mathcal{S}_{OA_6C_6B_2} = h(t)mmse(N|X + h(t)N)$. It means $tmmse(N|X + tN) < h(t)mmse(N|X + h(t)N)$ for all $t \in (0, t^*)$. Since, $h[h(t)] = t$, then the function $tmmse(N|X+tN) = h(t)mmse(N|X+h(t)N)$ also doesn't have a solution in $t \in (t^*, +\infty)$. Thus, the equation $tmmse(N|X + tN) = h(t)mmse(N|X + h(t)N)$ does not have any solution excepting $0, t^*, +\infty$. The proof of Lemma 4 is finish. □

4. Proof the Theorem

It is not difficult to see the optimal problem (5) in Theorem is equivalent with the following optimal problem

$$\min_{t_1,t_2\in\overline{\mathbb{R}}} I(X + t_1N, X + t_2N), \tag{9}$$

where $\overline{\mathbb{R}} = \mathbb{R}\cup\{-\infty, +\infty\}$. Note that, in (9), we refer to $I(X+t_1N, X+t_2N) = I(N, X+t_2N)$ when $t_1 = \infty$ and $I(X + t_1N, X + t_2N) = I(X + t_1N, N)$ when $t_2 = \infty$. By equation (2) in section 2, $I(X + t_1N, X + t_2N) = H(X + t_1N) + H(X + t_2N) - \log|t_2 - t_1| - H(X, N)$. Fortunately, the entropy $H(X, N)$ is independent with t_1, t_2. Therefore, if considering a function $L(t_1, t_2) = H(X + t_1N) + H(X + t_2N) - \log|t_2 - t_1|, t_1, t_2 \in \overline{\mathbb{R}}$, then the local solutions of the optimal problem (9) can be determined based on this function as follows: (t_1, t_2) is a local solution of (9) if and only if it is a solution of the following equations

$$\begin{cases} \dfrac{\partial}{\partial t_1}L(t_1, t_2) = \dfrac{\partial}{\partial t_2}L(t_1, t_2) = 0 \\ \\ \Big(\dfrac{\partial^2}{\partial t_i \partial t_j}L(t_1, t_2)\Big)_{i,j\in\{1,2\}} \quad \text{is positive.} \end{cases} \tag{10}$$

From the Lemma 1, we compute the first derivation of L as follows

$$\begin{aligned} \frac{\partial}{\partial t_1}L(t_1, t_2) &= t_1\mathcal{F}(X + t_1N) + \frac{1}{t_2 - t_1} = \frac{a(t_1)(t_2 - t_1) + t_1}{t_1(t_2 - t_1)} \\ \frac{\partial}{\partial t_2}L(t_1, t_2) &= t_2\mathcal{F}(X + t_2N) + \frac{1}{t_1 - t_2} = \frac{a(t_2)(t_1 - t_2) + t_2}{t_2(t_1 - t_2)}, \end{aligned} \tag{11}$$

where $a(t) = t^2\mathcal{F}(X + tN)$. Equation (11) shows us that if $t_1 = t_2$, then $\frac{\partial}{\partial t_1}L(t_1, t_2) = \frac{\partial}{\partial t_2}L(t_1, t_2) = \infty$. Indeed, we only consider $t_1 \neq t_2$. Consider the case $t_1, t_2 \in (0, +\infty]$. If $t_1 < t_2$,

$a(t_1), t_2 - t_1, t_1 > 0 \;\Rightarrow\; \frac{\partial}{\partial t_1} L(t_1,t_2) > 0$. If $t_1 > t_2$, $a(t_2), t_1 - t_2, t_2 > 0 \;\Rightarrow\; \frac{\partial}{\partial t_2} L(t_1,t_2) > 0$. Similarly for the case $t_1, t_2 \in [-\infty, 0)$. Hence, (t_1,t_2) is a local solution of (9) only if $t_1 t_2 \leqslant 0$. Without loss of generality, we assume $t_1 \in [-\infty, 0]$ and $t_2 \in [0, +\infty]$. Next, we see that

$$\begin{cases} \dfrac{\partial}{\partial t_1} L(t_1,t_2) = 0 \\ \\ \dfrac{\partial}{\partial t_2} L(t_1,t_2) = 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} a(t_1) = \dfrac{t_1}{t_1 - t_2} \\ \\ a(t_2) = \dfrac{t_2}{t_2 - t_1} \end{cases} \tag{12}$$

$$\Rightarrow \quad a(t_1) + a(t_2) = 1.$$

Lemma 2 shows that $a(t) = t^2 \mathcal{F}(X+tN) = 1 - mmse(N|X+tN)$ is a continuous function and decreasing from one to zero in $t \in \mathbb{R}^+$. We define a value t^* and a function $h : [0, +\infty] \to [0, +\infty]$ satisfy $a(t^*) = 1/2$ and $tmmse(N|X+tN) + h(t)mmse\big[N|X + h(t)N\big] = 1$ as in Lemma 4. Because N is a standard Gaussian variable, the functions $H(X+tN)$ and $mmse(N|X+tN)$ will be symmetric at $t = 0$, i.e $H(X+tN) = H(X-tN)$, $mmse(N|X+tN) = mmse(N|X-tN)$, $\forall\, t \in \overline{\mathbb{R}}$. Therefore the equation $a(t_1) + a(t_2) = t_1 mmse(N|X+t_1N) + t_2 mmse(N|X+t_2N) = 1$ with $t_2 \geqslant 0 \geqslant t_1$ infers $t_1 = -h(t_2)$. So, the function $L(t_1,t_2)$ can rewrite according parameter $t \equiv t_2$ as follows

$$\begin{aligned} L(t_1,t_2) \equiv L(t) &= H(t) + H[-h(t)] - \log[t + h(t)] \\ &= H(t) + H[h(t)] - \log[t + h(t)], \end{aligned}$$

where $H(t) = H(X+tN)$. Note that, $tH'(t) + h(t)H'[h(t)] = 1$ comes from $a(t) + a[h(t)] = 1$, and h is a nonnegative decrease function. Then

$$\begin{aligned} L'(t) = 0 \;&\Leftrightarrow\; H'(t) + H'\big[h(t)\big]h'(t) - \frac{1 + h'(t)}{t + h(t)} = 0 \\ &\Leftrightarrow\; \frac{1 - h(t)H'\big[h(t)\big]}{t} + H'\big[h(t)\big]h'(t) - \frac{1 + \frac{h(t)}{t}}{t + h(t)} - \frac{h'(t) - \frac{h(t)}{t}}{t + h(t)} = 0 \\ &\Leftrightarrow\; \Big(h'(t) - \frac{h(t)}{t}\Big)\Big(h^2(t)\mathcal{F}\big[X + h(t)N\big] - \frac{h(t)}{t + h(t)}\Big) = 0 \\ &\Leftrightarrow\; mmse\big[N|X + h(t)N\big] = 1 - \frac{h(t)}{t + h(t)} = \frac{t}{t + h(t)} \\ \\ &\Leftrightarrow\; tmmse(N|X+tN) = h(t)mmse(N|X + h(t)N). \end{aligned}$$

The second result in Lemma 4 states that the equation $tmmse(N|X+tN) = h(t)mmse(X|X+h(t)N)$ has only three solutions $t = 0,\; t^*,\; +\infty$ on $[0, +\infty]$. By the continuousness property,

$L(t)$ has two local minimum points $\{0,\ +\infty\}$, and one local maximum point $\{t^*\}$. It infers the function $L(t_1, t_2)$ has only two local minimum points $(t_1, t_2) \equiv (-\infty, 0)$ and $(t_1, t_2) \equiv (0, +\infty)$. In other words, $([X], [N])$ is a unique local solution of the optimal problem (5). Theorem is proven.

5. Discussion

Much more recently, Guo et al. in [7], and Payaró et al. in [11] have showed that: Let any non-Gaussian variable X with continuously differentiable density and finite variance, and any standard Gaussian variable N independent with X. Then the function $I(X, \sqrt{\delta}X + N)$ is a concave function with respect to δ. This result is very beautiful, but it can not solve the Simplest case of ICA. The Theorem in this paper is not a generalization of above result. However, it is a sufficient result for proving the Simplest case of ICA.

Moreover, the theory for the simplest case of ICA has a considerable signification in the way to improve the theory for general linear case of ICA (i.e. the suitable transformation to make a new representation of data is still a linear transformation). It is not difficult to see the theory for general linear case of ICA will be done if we finish to solve the following problem: Finding a new measure $\widetilde{I}$ for the statistical dependence of two random variables such that the Theorem can be expand for the case "N be any random variable independent with X". The problem can be stated more clearly as follows: Finding a new measure for evaluating the statistical dependence of any two random variables, $\widetilde{I}$, which satisfies: Let X, Y be any independent random variables with continuously differentiable density and finite variances. For any couple random variables $Y_1, Y_2 \in \mathcal{L}_{(X,Y)}$, $Y_1, Y_2 \notin \{(aX, bY)|a, b \in \mathbb{R}\}$, then $\forall \epsilon > 0,\ \exists T \in \mathbb{R}^{2\times 2}$, such that $\|T\| \leqslant \epsilon$ and

$$\widetilde{I}(Y_1, Y_2) > \widetilde{I}[(Y_1, Y_2)T].$$

Some information measures have been studied carefully in many papers (see [5, 6]). We trustfully believe that the results and technical proving in this paper is very useful for finding measure $\widetilde{I}$.

Acknowledgements

I would like to thank Professor Christoph Bandt for fruitful discussions and useful comments. I am also grateful the Ministry of Education and Training of Vietnam for supporting my work.

References

[1] Blachman, N.M. (1965). *The convolution inequality for entropy powers.* IEEE Transactions on Information Theory, vol. 11, no. 2, pp. 267-271.

[2] Boscolo, R., Pan, H. and Roychowdhury, V.P (2004). *Independent Component Analysis Based on Nonparametric Density Estimation.* IEEE Transactions on Neural Networks, vol.15, no.1, pp.55-65.

[3] Cardoso, J.-F. (1999). *High-order contrasts for Independent Component Analysis.* Neural Computation, vol. 11, no. 1, pp. 157-192.

[4] Comon, P. (1994). *Independent component analysis, a new concept?.* Signal Processing, vol. 36, no. 3, pp. 287-314.

[5] Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory.* 2nd ed. New York: Wiley.

[6] Ebrahimi, N., Soofi, E.S. and Soyer, R. (2010). *Information Measure in Pespective.* International Statistical Review, 78, 383-412.

[7] Guo, D., Shamai, S., and Verdú, S. (2005). *Mutual information and minimum mean-square error in Gaussian channels.* IEEE Transactions on Information Theory, vol. 51, no. 4, pp. 1261-1282.

[8] Guo, D., Shamai, S. and Verdú, S. (2008).*Estimation of non-Gaussian random variables in Gaussian noise: Properties of the MMSE.* in IEEE International Symposium on Information Theory, (ISIT'08), pp. 1083-1087.

[9] Hyvärinen, A. and Oja, E. (2001). *Independent Component Analysis.* New York: John Wiley & Sons.

[10] Lee, T.-W., Girolami, M. and Sejnowski, T.J. (1999). *Independent component analysis using an extended Infomax algorithm for mixed sub-Gaussian and super-Gaussian sources.* Neural Computation, vol. 11, no. 2, pp. 417-441.

[11] Payaró, M. and Palomar, D.P. (2009). *Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels.* IEEE Trans. Inf. Theory, vol. 55, pp. 3613-3628.

[12] Rioul, O. (2007). *Information Theoretic Proofs of Entropy Power Inequalities.* arXiv: 0704.1751v1 [cs.IT].

[13] Rioul, O. (2007). *A simple proof of the entropy-power inequality via properties of mutual information.* Proc. IEEE Int. Symp. Inform. Theory, Nice. France, pp. 46-50.

[14] Shannon, C.E. (1948). *A mathematical theory of communication.* Bell System Technical Journal, vol. 27, pp. 623-656.

[15] Stam, A.J. (1959). *Some inequalities satisfied by the quantities of information of Fisher and Shannon.* Information and Control, vol. 2, pp. 101-112.

[16] Stone, J.V. (2005). *A Brief Introduction to Independent Component Analysis.* in Encyclopedia of Statistics in Behavioral Science, vol. 2, pp. 907-912, Editors Everitt, B.S. and Howell, D.C. John Wiley & Sons.

[17] Wu, Y. and Verdú (2010). *Functional properties of MMSE.* Proc. IEEE Int. Symp. Inform. Theory, Austin, TX, pp. 13-18.