

基于随机变量相似度的 ICA 方法

张旭秀^{①②} 邱天爽^①

^①(大连理工大学电子与信息工程学院 大连 116024)

^②(大连交通大学电气信息学院 大连 116028)

摘要: 从随机变量(微分)熵的概念出发, 定义了随机变量的相似度, 讨论了用求相似度极点的方法实现观测数据线性组合非高斯性最大化, 从而串行估计独立分量分析(ICA)模型中的独立分量的原理和算法。对非多项式矩定理进行了更为一般化的证明, 以此定理为根据阐明了以一般的非二次型光滑偶函数的数学期望近似代替相似度的可行性。给出梯度算法中的符号因子计算公式, 避免了现有的相应算法中符号因子计算公式与目标函数之间的矛盾。通过与极大似然 ICA 方法对比, 表明所定义的相似度就是在预白化条件下单个源变量的极大似然函数。

关键词: 微分熵; 随机变量相似度; ICA 模型的串行估计; 极大似然函数

中图分类号: TN911.7

文献标识码: A

文章编号: 1009-5896(2009)09-2104-05

Random Variable Analogy Based ICA Method

Zhang Xu-xiu^{①②} Qiu Tian-shuang^①

^①(School of Electronics and Information, Dalian University of Technology, Dalian 116024, China)

^②(School of Electronics and Information Engineering, Dalian Jiaotong University, Dalian 116028, China)

Abstract: This paper defines analogy measure of two random variables, and discusses the principle and algorithm of maximizing non-Gaussianity of observed data with a linear transformation to estimate independent components serially. It also proves the non-polynomial moment theorem by a generalized way, and states the feasibility that substitutes the analogy with the expectation of a non-quadratic smooth even function based on the theorem. A formula to compute sign of above algorithm is given. The algorithm overcomes the contradiction between the objective function and the sign computation formula. Comparing with Maximum likelihood ICA, the analogy is Maximum likelihood function of single source under pre-whited.

Key words: Differential entropy; Random variable; Serial estimation of ICA(Independent Component Analysis) model; Maximum likelihood function

1 引言

独立分量分析 ICA 方法大致可分成两大类: 串行 ICA 和并行 ICA。串行 ICA 基于独立分量的非高斯性, 逐个分离出非高斯型独立分量。分离出的各个非高斯型分量之间的相互独立性, 是通过使分离阵(解混阵)中的各个行向量之间保持正交来实现。这种 ICA 方法允许源变量中最多有一个为高斯型的。并行 ICA 通过对观测向量的线性变换, 使变换后的观测向量与一个估计的源向量之间的 Kullback 距离最小化, 从而得到源向量的一个估计。而该估计与实际源向量的相似程度(不考虑各分量标度、正负号和顺序的差异), 与在算法中采用的独立分量的密度函数的近似程度有很大关系^[1-4]。

ICA 方法一般包括目标函数和算法(数值解法)两部分内容。串行 ICA 的目标函数有峰态(绝对值)、负熵的近似以及一般非线性函数的数学期望等等^[5,6]。许多 ICA 方法采用非线性函数的数学期望作为目标函数, 原因在于通过合理地选择非线性函数可以获得良好的韧性^[5]。然而, 关于这种目标函数在 ICA 中作用机理的研究还不够充分, 例如, 在应用中会遇到什么矛盾, 它的产生原因等等都尚待研究。因此有的文献^[7]提出应当把类似上述问题作为 ICA 的一个理论问题加以研究。本文将就上述问题展开讨论。

考虑基本 ICA 问题, 其数学模型为

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1)$$

式中 $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$ 是由 N 个相互独立的随机变量构成的向量, 称为源向量, 其中最多有一个为高斯分布; $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ 称为观测向量, 它是源向量的线性混合; \mathbf{A} 是 $N \times N$ 维满秩矩阵, 称为混

合矩阵。ICA问题的一般描述为:已知 N 个混合信号在其自变量 $t=1,2,\dots,N$ 各点的观测值,要求解 N 个源变量在 $t=1,2,\dots,M$ 各点的值(或者变化曲线)和(或) \mathbf{A} 阵的各元素。对式(1)中的观测向量进行白化处理。变换后得到的向量 \mathbf{z} 的各个分量方差皆为1,且互不相关,并得到式(2)所示模型^[5]

$$\mathbf{z} = \mathbf{A}_z \mathbf{s} \quad (2)$$

式中, $\mathbf{z} = \mathbf{V}\mathbf{x}$, \mathbf{V} 为白化矩阵^[1], $\mathbf{A}_z = \mathbf{V}\mathbf{A}$, \mathbf{A}_z 是一个正交矩阵。

2 随机变量的非高斯性及其度量

绝大多数随机变量的密度函数可以写成如下的指数形式^[4,5]:

$$p(\cdot) = c_1 \exp(c_2 |\cdot|^\alpha) \quad (3)$$

其中 c_1 和 c_2 是常数。 α 称为特征指数,根据 α 的取值,随机变量可分成3种类型: $\alpha=2$ 时,称为高斯型, $0 < \alpha < 2$ 和 $\alpha > 2$ 时,分别称为超高斯型和亚高斯型。

串行ICA的理论基础是中心极限定理。一般地,若干随机变量线性组合的分布趋于高斯型。为了估计式(2)中 \mathbf{s} 向量的一个分量,取向量 \mathbf{z} 的 N 个分量的线性组合 $y = \mathbf{w}^T \mathbf{z}$ 。在 $\|\mathbf{w}\|=1$ 的条件下,调整 \mathbf{w} 的各个分量,可使 y 的非高斯性增强,即 y 对应的 α 远离2,当非高斯性达到某一极值时, y 为一个独立分量的估计。

随机变量的非高斯性可以用峰态及负熵等来度量。根据所采用的非高斯性量度的不同,形成不同的串行ICA方法^[5,6]。本文将要讨论的基于求非线性函数期望值极点的ICA方法,实际上与采用负熵作为非高斯性的量度有关,为此,引入真负熵及相似度两个概念。

随机变量 y 的(微分)熵记作 $H(y)$,定义为

$$H(y) = E\{-\lg p_y(y)\} = -\int p_y(y) \lg p_y(y) dy \quad (4)$$

在方差相等的情况下,高斯型变量的熵最大。

负熵 $J(y)$ ^[5]度量随机变量 y 的非高斯性

$$J(y) = H(y_{\text{gauss}}) - H(y) \quad (5)$$

其中 y_{gauss} 是与 y 具有相同方差的高斯型变量。在保持 y 的方差不变的前提下, $H(y_{\text{gauss}})$ 是一个常数, $H(y)$ 越小,则 $J(y)$ 越大,表明 y 的非高斯性越强。为与负熵 $J(y)$ 区别,本文称 $-H(y)$ 为 y 的真负熵,记作 $N(y)$,

$$N(y) = -H(y) = -E\{-\lg p_y(y)\} = \int p_y(y) \lg p_y(y) dy \quad (6)$$

是 y 的非高斯性的又一种度量。

在下面的讨论中,取

$$y = \mathbf{w}^T \mathbf{z} = \mathbf{w}^T \mathbf{V}\mathbf{x} = \mathbf{w}^T \mathbf{A}_z \mathbf{s} = \mathbf{q}^T \mathbf{s} \quad (7)$$

式中

$$\mathbf{q}^T = \mathbf{w}^T \mathbf{A}_z \quad (8)$$

而

$$\sigma_y^2 = E\{y^T y\} = \mathbf{w}^T E\{\mathbf{z}\mathbf{z}^T\} \mathbf{w} = \mathbf{w}^T \mathbf{I} \mathbf{w} = \|\mathbf{w}\| = 1 \quad (9)$$

在式(9)约束条件下,通过调整 \mathbf{w} 的方向,使 y 的非高斯性极大化, y 成为 \mathbf{s} 的一个分量 s_i 的估计,向量 \mathbf{q} 的第 i 个分量等于1,其余分量皆为0。将式(7)代入式(6)得

$$\begin{aligned} -H(y) &= -H(\mathbf{w}^T \mathbf{z}) = E\{\lg p_{\mathbf{w}^T \mathbf{z}}(\mathbf{w}^T \mathbf{z})\} \\ &= \int p_{\mathbf{w}^T \mathbf{z}}(\mathbf{w}^T \mathbf{z}) \lg p_{\mathbf{w}^T \mathbf{z}}(\mathbf{w}^T \mathbf{z}) d(\mathbf{w}^T \mathbf{z}) \end{aligned} \quad (10)$$

可见,在改变 \mathbf{w} 的方向时,随机变量 $\mathbf{w}^T \mathbf{z}$ 的密度函数 $p_{\mathbf{w}^T \mathbf{z}}(\cdot)$ 的表达式要随之改变,这使得式(10)难以计算。为此有以下定义。

定义 设 $p_{s_G}(\cdot)$ 为已知随机变量 s_G 的密度函数, y 是与 s_G 具有相同方差的随机变量,定义 y 与 s_G 的相似度 $N_{s_G}(y)$ 为

$$\begin{aligned} N_{s_G}(y) &= \int p_y(y) \lg p_{s_G}(y) dy \\ &= E\{\lg p_{s_G}(y)\} = E\{G(y)\} \end{aligned} \quad (11)$$

y 的密度函数 $p_y(\cdot)$ 与 $p_{s_G}(\cdot)$ 越接近,相似度越大,反之亦然。

3 基于相似度的串行ICA

假定 $\sigma_{s_G}^2=1$,并记 $G(\cdot) = \lg p_{s_G}(\cdot)$, $G(\cdot)$ 称作参照函数。 $y = \mathbf{w}^T \mathbf{z}$ 与 s_G 之间的相似度 $N_{s_G}(y) = \int p_y(y) \lg p_{s_G}(y) dy = E\{G(y)\} = E\{G(\mathbf{w}^T \mathbf{z})\}$ 。通过求 $E\{G(\mathbf{w}^T \mathbf{z})\}$ 的极大/极小点,可逐个估计所有的独立分量。

3.1 源向量只有亚高斯型独立分量的情形

假设ICA模型中只有亚高斯型独立分量 s_1 , s_2 和 s_3 ,它们的密度函数对应的 α 分别记作 α_1 , α_2 和 α_3 ,且均大于2,但具体数值未知。假设选取的 s_G 对应的 α 值为 α_G 。 α_1 , α_2 , α_3 及 α_G 的相对大小如图1(a)所示。

为估计 s_1, s_2, s_3 ,应当使 $y = \mathbf{w}^T \mathbf{z}$ 对应的 α 值远离 $\alpha=2$ (使 $\mathbf{w}^T \mathbf{z}$ 的非高斯性增强),靠近 α_G 。为此,第1次寻优时,在 $\|\mathbf{w}\|=1$ 的条件下,通过梯度升算法求 $E\{G(\mathbf{w}^T \mathbf{z})\}$ 的极大点。可导出梯度算法为^[5,6]

$$\Delta \mathbf{w} \propto E\{\mathbf{z}g(\mathbf{w}^T \mathbf{z})\} \quad (12)$$

$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \quad (13)$$

式中 $g(\cdot)$ 表示 $G(\cdot)$ 的导数。当算法收敛时,有 $\mathbf{w}^T \mathbf{z} = \mathbf{w}_1^{*T} \mathbf{z} = +s_i$ (或 $-s_i$)。这里 \mathbf{w}_1^* 是 $E\{G(\mathbf{w}^T \mathbf{z})\}$ 的一个局部极大点,下标1表示是第1次寻优。 s_i 是 s_1, s_2, s_3 中的一个。

假设把与 w_1^* 对应的 q 记作 q_1^* (见式(8)), 则 q_1^* 的分量只能有一个非零元素 1(或-1)。否则, $w_1^{*T}z = q_1^{*T}s$ 将是几个源信号的组合, 它不是“远离”高斯型的, 也不是与非高斯的 s_G 局部最近的, 这与 w_1^* 是 $E\{G(w^Tz)\}$ 的一个局部极大点的题设相矛盾。至于 s_i 是 s_1, s_2, s_3 中的哪一个, 则是随机的, 因为算法中初始化向量 w 是随机的。

第 2 次寻优, 需增加一个约束条件 $w^T w_1^* = 0$, 得到的极大点为 w_2^* , 对应的 q 记作 q_2^* , 其分量只能有一个元素为 1(或-1), 它对应的 $w_2^{*T}z$ 是另一个独立分量的估计。同理, 依次得到其余独立分量的估计。

3.2 源向量只有超高斯型独立分量的情形

设 s 中只有两个超高斯型的独立分量 s_1 和 s_2 , 它们密度函数的 α 值分别为 α_1, α_2 。参照函数 $G(\cdot)$ 的选取同上。 α_1, α_2 同 α_G 的相对位置见图 1(b)。

这里采取梯度降算法求 $E\{G(w^Tz)\}$ 的极小点, 式(12)修正成

$$\Delta w \propto -E\{z g(w^Tz)\} \quad (14)$$

约束条件同 3.1 节。

3.3 源向量中为混合型独立分量的情形

见图 1(c)。设 s 中含有 2 个超高斯型独立分量 s_1, s_2 以及 3 个亚高斯型独立分量 s_3, s_4 和 s_5 , 对应的 α 值分别为 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 和 α_5 。参照函数 $G(\cdot)$ 同上。

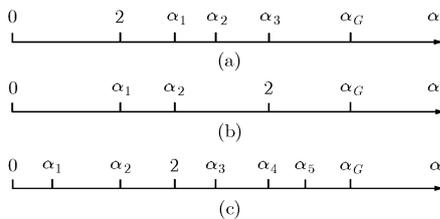


图 1 参照函数的 $\alpha_G > 2$ 时的 ICA

可以看出, 估计 s_3, s_4 和 s_5 时, 采用梯度升算法, 估计 s_1, s_2 时, 采用梯度降算法, 有

$$\Delta w \propto \gamma E\{z g(w^Tz)\} \quad (15)$$

式中的 γ 代表正负号。

在寻优过程中, 需判断当前 w 决定的变量 w^Tz 所对应的 α 值是大于 2 还是小于 2。

3.4 参照函数 $G(\cdot)$ 对应于超高斯分布的情形

前面的讨论中, 所取用的 $G(\cdot)$, 对应于亚高斯分布, $\alpha_G > 2$ 。现在取 $G(\cdot)$ 对应于超高斯分布, 即 $\alpha_G < 2$ 。见图 2。这时, 估计超高斯型独立分量 s_1 和 s_2 时, 应采用梯度升算法; 估计亚高斯型独立分量

s_3, s_4, s_5 时, 应采用梯度降算法。

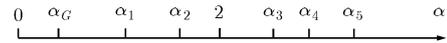


图 2 参照函数的 $\alpha_G < 2$ 时的 ICA

3.5 梯度算法中正负号的估算

式(15)中的 γ 可以按下式计算^[5]

$$\gamma = \text{sgn}(E\{G(w^Tz)\} - E\{G(v)\}) \quad (16)$$

式中 v 是方差为 1 的高斯型随机变量。

基于式(16)的一种在线估计 γ 的“自适应”算法为^[5]

$$\left. \begin{aligned} \Delta C &\propto G(w^Tz) - E\{G(v)\} - C \\ \gamma &= \text{sgn}(C) \end{aligned} \right\} \quad (17)$$

3.6 讨论

前面的讨论实际上是假定参照函数 $G(\cdot)$ 对应的 α_G 满足

$$\alpha_G > \max\{\alpha_1, \alpha_2, \dots, \alpha_N\} > 2 \quad (18)$$

或者

$$\alpha_G < \min\{\alpha_1, \alpha_2, \dots, \alpha_N\} < 2 \quad (19)$$

如果不符合上述假定, 则会出现下述情况。

例如, 见图 3, 假设 $\alpha_G > 2$, 但 $\alpha_G < \alpha_3 = \max\{\alpha_1, \alpha_2, \dots, \alpha_N\}$ 。为了估计 α_3 对应的独立分量 s_3 , 应当使 w^Tz 对应的 α 值远离 $\alpha = 2$, 采用梯度降算法, 但若将在 α_3 附近的 α 对应的 w^Tz 代入符号判别式(16), 则算得 $\gamma = 1$, 这与要求恰好相反, 结果与 α_3 对应的独立分量 s_3 无法估计出来。类似地, 当 $\alpha_G < 2$, 但 $\alpha_G > \min\{\alpha_1, \alpha_2, \dots, \alpha_N\}$ 时, 也会出现符号因子 γ 的计算值与要求相反的情况。这种矛盾的解决见 4.2 节。



图 3 $2 < \alpha_G < \max\{\alpha_1, \alpha_2, \alpha_3\}$ 的情形

前面所举的例子中, 假定 s 中各分量的 α 值互不相同。若 s 中含有两个具有同样分布的分量, 例如假设图 2 中的 $\alpha_3 = \alpha_4$, 则上述方法仍适用; 与 α_3, α_4 对应的两个独立分量也是在两次梯度算法中分别估出的。

4 非多项式矩定理及其讨论

前面所采用的非线性函数 $G(\cdot) = \lg p_{s_G}(\cdot)$, 如果 $G(\cdot)$ 不是某一指数型非高斯随机变量的密度函数, 会有什么结果? 有以下的定理。

4.1 非多项式矩定理

定理 1 设参照函数 $G(\cdot) = \lg p_{s_G}(\cdot)$, s_G 是方差

为1, 密度函数为 $p_{s_G}(\cdot)$ 的随机变量, 对应的 $\alpha = \alpha_G$, 则在 $\|\mathbf{w}\| = 1$ 的约束条件下, $E\{G(\mathbf{w}^T \mathbf{z})\}$ 的局部极点 \mathbf{w}^* 对应的 $\mathbf{w}^{*T} \mathbf{z} = s_i$ 应满足

$$E\{s_i g(s_i)\} - E\{g'(s_i)\} > 0, \text{ 或者 } < 0 \quad (20)$$

证明 约束条件 $\|\mathbf{w}\| = 1$ 可以写成 $\mathbf{w}^T \mathbf{w} - 1 = 0$. 建立约束条件极值问题的目标函数

$$F(\mathbf{w}, \beta) = E\{G(\mathbf{w}^T \mathbf{z})\} + \beta(\mathbf{w}^T \mathbf{w} - 1)$$

设 \mathbf{w}^* 为 $F(\mathbf{w}, \beta)$ 的一个极点, ε 是 \mathbf{w} 的小的扰动. 在 \mathbf{w}^* 处将 $F(\mathbf{w}^* + \varepsilon, \beta)$ 展成泰勒级数:

$$F(\mathbf{w}^* + \varepsilon, \beta) = F(\mathbf{w}^*, \beta) + \varepsilon^T \left. \frac{\partial F}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}^*} + \frac{1}{2} \varepsilon^T \left. \frac{\partial^2 F}{\partial \mathbf{w}^2} \right|_{\mathbf{w}=\mathbf{w}^*} \varepsilon + o(\|\varepsilon\|^2)$$

式中依题设

$$\left. \frac{\partial F}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}^*} = E\{\mathbf{z} g(\mathbf{w}^{*T} \mathbf{z})\} + 2\beta \mathbf{w}^* = 0$$

用 \mathbf{w}^{*T} 左乘上式, 并令 $\mathbf{w}^{*T} \mathbf{z} = s_i$, 考虑到 $\mathbf{w}^{*T} \mathbf{w}^* = \|\mathbf{w}^*\|^2 = 1$, 于是得

$$2\beta = -E\{s_i g(s_i)\}$$

又

$$\begin{aligned} \left. \frac{\partial^2 F}{\partial \mathbf{w}^2} \right|_{\mathbf{w}=\mathbf{w}^*} &= E\{\mathbf{z} \mathbf{z}^T g'(\mathbf{w}^{*T} \mathbf{z})\} + 2\beta \mathbf{I} \\ &\approx E\{\mathbf{z} \mathbf{z}^T\} (E\{g'(s_i)\} - E\{s_i g(s_i)\}) \mathbf{I} \\ &= (E\{g'(s_i)\} - E\{s_i g(s_i)\}) \mathbf{I} = -M(s_i) \mathbf{I} \end{aligned}$$

式中 $M(s_i) = E\{s_i g(s_i)\} - E\{g'(s_i)\}$ 称为 s_i 的非多项式矩. 上面推导过程近似认为 $\mathbf{z} \mathbf{z}^T$ 的各元素与 $g'(s_i)$ 是相互独立的, 并且利用了题设条件 $E\{\mathbf{z} \mathbf{z}^T\} = \mathbf{I}$. 可见, 当 $M(s_i) > 0$ 时, \mathbf{w}^* 为极大点; 当 $M(s_i) < 0$ 时, \mathbf{w}^* 为极小点. 即式(20)成立. 证毕

4.2 非多项式矩定理的讨论

(1) 参照函数 $G(\cdot) = \lg p_{s_G}(\cdot) \pm c$, c 为常数, 其作用与 $G(\cdot) = \lg p_{s_G}(\cdot)$ 是一样的.

(2) 如果参照函数 $G(\cdot)$ 偏离了 $\lg p_{s_G}(\cdot)$, 例如变成 $\hat{G}(\cdot)$, 相应地 $g(\cdot)$ 和 $g'(\cdot)$ 变成 $\hat{g}(\cdot)$ 和 $\hat{g}'(\cdot)$, $M(s_i)$ 变成 $\hat{M}(s_i)$, 那么, 只要 $\hat{M}(s_i)$, $\hat{g}(s_i)$ 和 $\hat{g}'(s_i)$ 仍满足式(20), $\hat{G}(\cdot)$ 便是可用的参照函数. 研究表明, 有多种非二次型光滑偶函数适合于作为 $\hat{G}(\cdot)$ [5], 以下省去“ \wedge ”.

(3) 定理 1 中给出的非多项式矩是指在 $E\{G(\cdot)\}$ 的极点处对应的随机变量 $\mathbf{w}^{*T} \mathbf{z}$ 的矩. 由于函数 $M(\cdot)$ 一般是连续的, 在极点附近的随机变量 $\mathbf{w}^T \mathbf{z}$ 的非多项式矩 $M(\mathbf{w}^T \mathbf{z})$ 应当与 $M(\mathbf{w}^{*T} \mathbf{z})$ 具有相同的符号, 因此可以采用 $M(\mathbf{w}^T \mathbf{z})$ 的符号作为式(15)中的 γ [5,6], 解决了按式(16)计算 γ 可能遇到的矛盾. 于是式(16)应修正成

$$\gamma = \text{sgn}(E\{\mathbf{w}^T \mathbf{z} g(\mathbf{w}^T \mathbf{z})\} - E\{g'(\mathbf{w}^T \mathbf{z})\}) \quad (21)$$

相应地式(17)应修正成[5]

$$\left. \begin{aligned} \Delta C &\propto \mathbf{w}^T \mathbf{z} g(\mathbf{w}^T \mathbf{z}) - g'(\mathbf{w}^T \mathbf{z}) - C \\ \gamma &= \text{sgn}(C) \end{aligned} \right\} \quad (22)$$

(4) 非线性函数 $G(\cdot)$ 不得采用 2 次型函数. 因为若取 $G(y) = a + by + cy^2$, 其中 a, b 和 c 是常数, 则非多项式矩 $M(y) = E\{yg(y)\} - E\{g'(y)\} = bE\{y\} + E\{2cy^2\} - E\{2c\} = 0$. 可见梯度算法中的 $\gamma \equiv 0$, 算法便无法进行.

5 基于相似度的串行 ICA 算法特点与比较分析

5.1 与采用负熵(近似)的 ICA 方法对比

文献[5]是从负熵的近似——目标函数

$$J(\mathbf{w}^T \mathbf{z}) \propto (E\{G(\mathbf{w}^T \mathbf{z})\} - E\{G(v)\})^2 \quad (23)$$

出发, 文献[6]是从目标函数

$$J(\mathbf{w}) = |E\{F(\mathbf{w}^T \mathbf{x})\} - E\{F(v)\}| \quad (24)$$

出发, 最后都导出本质上与式(15)及式(13)相同的梯度算法公式. 这两种目标函数存在一个共同的问题, 就是不能直接根据它们确定梯度算法中的符号 γ .

本文采用相似度作为目标函数, 导出梯度算法公式(15)、式(13)及算法中的符号 γ 计算公式(21). 从而理顺了目标函数与梯度算法的关系, 避免了采用式(23)或式(24)作为目标函数时可能遇到的矛盾. 另外, 文献[5,6]都提出了非多项式矩定理, 并给出了相同的证明方法. 本文用更为一般化的方法证明了该定理.

5.2 与采用极大似然估计的 ICA 方法对比

为求解 ICA 问题导出的极大似然函数[5]为

$$\frac{1}{M} \lg L(\mathbf{B}) = E \left\{ \sum_{i=1}^n \lg p_i(b_i^T \mathbf{X}) \right\} + \lg |\det \mathbf{B}| \quad (25)$$

式中 \mathbf{B} 是分离矩阵, $p_i(\cdot)$ 是假想的第 i 个源变量的分布密度函数, M 是观测向量 \mathbf{X} 的样本个数. 这里分离阵改用正交阵 \mathbf{W} , 于是式(25)应改写成

$$\frac{1}{M} \log L(\mathbf{W}) = E \left\{ \sum_{i=1}^n \log p_i(\mathbf{W}_i^T \mathbf{X}) \right\} + \log |\det \mathbf{W}| \quad (26)$$

由于 \mathbf{W} 阵是正交阵, $|\det \mathbf{W}| = 1$, 因此, 求 $(1/M) \cdot \log L(\mathbf{W})$ 极大值问题变成在 \mathbf{W} 为正交阵的约束条件下, 求式(26)右端第 1 项的极大值问题. 如果选取的假想密度函数 $p_i(\cdot)$ 使得每一个 $E\{\lg p_i(\mathbf{W}_i^T \mathbf{z})\}$ 都大于零, 则上式第一项可改写成

$$E\{\lg p_i(\mathbf{w}_i^T \mathbf{z})\} = \sum_{i=1}^n E\{\lg p_i(\mathbf{W}_i^T \mathbf{z})\} \quad (27)$$

与式(11)对照可知, $E\{\lg p_i(\mathbf{W}_i^T \mathbf{z})\}$ 就是为第 i 个随机源变量选取的相似度函数.

在极大似然法中, $G_i(\cdot)$ 必须选取得与对应源变量 s_i 的分布类型一致, 以保证式(26)中第 1 项极大化。而本文中的 $G(\cdot)$ 则不受此限制, 同一个 $G(\cdot)$ 可以用来估计所有的源变量。

6 结论

本文定义了两个随机变量之间的相似度。详细讨论了利用求相似度极点的方法实现 ICA 模型中观测数据(预白化后)线性组合 $\mathbf{w}^T \mathbf{z}$ 的非高斯性最大化, 从而串行估计各个独立分量的原理和算法。从更一般的角度证明了独立分量的非多项式定理, 以此定理为依据, 阐明了以一般的非 2 次型光滑偶函数 $G(\cdot)$ 的期望近似代替相似度的可行性。本文给出的式(15)、式(13)及式(21)组成的梯度算法, 虽然与现有文献[5,6]等给出的相应公式具有相同的形式, 但导出过程的出发点与文献[5,6]是不同的, 从而避免了该文献中的目标函数与梯度算法中符号 γ 计算公式之间的矛盾。进一步分析表明, 本文给出的相似度函数相当于在预白化条件下, 单个源变量的极大似然函数。

参 考 文 献

- [1] Comon P. Independent component analysis : A new concept [J]? *Signal Processing*, 1994, 36(3): 287-314.
- [2] Bell A J and Sejnowski T J. An information-maximization approach to blind separation and blind deconvolution[J]. *Neural Computation*, 1995, 7: 1129-1159.
- [3] Cardoso J F and Adali T. The maximum likelihood approach to complex ICA[C]. The 2006 IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP), Toulouse, France, May 14-19 2006, Vol.V: 673-676.
- [4] Hiroe A. Solution of permutation problem in frequency domain ica, using multivariate probability density functions[C]. 6th International Conference on Independent Component Analysis and Blind Source Separation, Carlestone, SC, USA, March 5-8, 2006: 601-608.
- [5] Hyvariene A, Karhunen J, and Oja E. Independent Component Analysis[M]. New York, John Wiley & Sons Inc., 2001: 165-181.
- [6] Hyvarinen A and Oja E. Independent component analysis by general nonlinear Hebbian-like learning rules[J]. *Signal Processing*, 1998, 64(3): 301-313.
- [7] Hyvariene A and Oja E. Fast and robust fixed-point algorithms for independent component analysis[J]. *IEEE Transactions on Neural Networks*, 1999, 10(3): 626-634.

张旭秀: 女, 1968年生, 副教授, 博士后, 研究领域为信号处理、智能控制等。

邱天爽: 男, 1954年生, 教授, 博士生导师, 博士, 研究领域为数字信号处理理论、生物医学信号处理、非平稳与非高斯信号处理等。