

基于峰态的独立分量分析原理的研究

张旭秀 邱天爽

(大连理工大学电子与信息工程学院 大连 116024)

摘要: 对基于峰态绝对值最大化的 ICA 原理进行了详细分析, 给出了该原理几何解释和适用范围。通过对采用该原理进行 ICA 问题求解过程的分析, 阐明了 ICA 问题解的不确定性的产生原因, 指出并解释了求解过程中表现出的一种概率特性。所得结论对于其它 ICA 方法也具有参考价值。

关键词: 独立分量, ICA, 峰态, 几何解释

中图分类号: TN911.7 **文献标识码:** A **文章编号:** 1009-5896(2005)02-0206-04

The Study on the Principle of Kurtosis Based ICA Method

Zhang Xu-xiu Qiu Tian-shuang

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

Abstract The kurtosis based ICA approach is analyzed particularly and the geometrical explanation of this approach is presented in the paper. Furthermore, we elucidate the reasons of the indeterminacy of ICA solutions and explain the probability property by analyzing the process for finding out the ICA solutions. These analytic results and conclusions are also benefit to the study on other ICA methods.

Key words Independent component, ICA, Kurtosis, Geometry explanation

1 引言

基于峰态的独立分量分析 (Independent Component Analysis, ICA) 方法^[1]是 Hyvariene 提出的一种基本的 ICA 方法。一般来说, 一种 ICA 方法包括两部分: 即 ICA 原理 (目标函数) 和寻优算法 (数值解法)。基于峰态的 ICA 方法, 其目标函数是随机变量峰态的绝对值。这一 ICA 原理的理论根据是中心极限定理和随机变量峰态的性质。根据此原理已经导出了实用的梯度算法和快速迭代算法^[1-3]。有些文献对该原理进行过理论分析^[2], 但是仅仅针对某种特殊情况, 不够全面。本文将对这一 ICA 原理进行系统的理论分析, 并从中得出一些有意义的结论。

2 基本 ICA 问题的模型与假设

基本 ICA 问题不考虑噪声的影响, 其数学模型^[4]如式(1)所示。

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

式(1)中 $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$ 是由 N 个相互独立的随机信号 (这里称为源信号或独立分量) 构成的向量, 简称为源向量, 其中各分量皆为非高斯分布 (或最多有一个为高斯分布), 上标

T 表示转置; $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ 是由上述 N 个源信号按不同比例混合形成的 N 个混合信号构成的向量, 称为混合向量或观测向量; \mathbf{A} 是 $N \times N$ 维常数满秩矩阵, 称为混合矩阵。ICA 问题的一般描述为: 已知 N 个混合信号在自变量 (一般为时间, 但也可以是其它量, 例如位置等) $t = 1, 2, \dots, T$ 各点的观测值, 要求解 N 个源信号在 $t = 1, 2, \dots, T$ 各点的值 (或者变化曲线) 和 (或) \mathbf{A} 阵的各元素。

对于一个已知的向量 \mathbf{x} , 估计出来的向量 \mathbf{s} 及矩阵 \mathbf{A} 不是唯一的, 因而一般不会与原始模型中的向量 \mathbf{s} 及矩阵 \mathbf{A} 相同, 这是由 ICA 问题解的不确定性决定的。第一种不确定性表现为式 (1) 中 \mathbf{s} 的某一个分量缩小 (或增大) k 倍, 同时 \mathbf{A} 阵也增大 (或缩小) k 倍时等式仍成立。为排除这种不确定性的困扰, 在求解时, 一般规定 \mathbf{s} 的各分量的方差皆为 1。另外, 假设 \mathbf{x} 及 \mathbf{s} 各分量的均值皆为零。

3 基于峰态的 ICA 原理

如上所述, ICA 是利用观测向量 \mathbf{x} 求出各个独立分量以及混合矩阵的估计。由中心极限定理可知, 把 \mathbf{x} 的各个分量进行线性组合, 得到一个新的随机变量。一般来说, 这个新随机变量倾向于具有高斯性。但是, 通过适当调整线性组合

的系数, 可以使这个新随机变量的高斯性变弱, 非高斯性变强; 当该随机变量的非高斯性变得最强的时候, 便可以认为它就是一个独立分量的估计。

峰态是随机变量非高斯性的一种重要的度量方法。随机变量 y 的峰态定义为^[1]

$$\text{Kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \quad (2)$$

根据峰态的数值, 可以将随机变量划分为高斯型 ($\text{Kurt}(y) = 0$) 和非高斯型 ($\text{Kurt}(y) \neq 0$)。其中, 峰态值大于零的称为超高斯型变量, 而峰态值小于零的称为亚高斯型变量。在方差相等的条件下, 峰态的绝对值越大, 随机变量的非高斯性越强。

峰态具有以下性质: (1) 峰态是偶函数; (2) 对常数 α , 有 $\text{Kurt}(\alpha y) = \alpha^4 \text{Kurt}(y)$; (3) 对于两个独立的随机变量 x_1, x_2 , 有 $\text{Kurt}(x_1 + x_2) = \text{Kurt}(x_1) + \text{Kurt}(x_2)$ 。

现在以二维为例, 考察观测数据的线性组合

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{b}^T \mathbf{A} \mathbf{s} = \mathbf{q}^T \mathbf{s} = q_1 s_1 + q_2 s_2 \quad (3)$$

其中 $\mathbf{x} = [x_1, x_2]^T$ 是观测数据向量, \mathbf{A} 是混合矩阵, $\mathbf{s} = [s_1, s_2]^T$ 是源向量, $\mathbf{b} = [b_1, b_2]^T$ 是线性组合系数向量, $\mathbf{q}^T = \mathbf{b}^T \mathbf{A} = [q_1, q_2]$ 。

把峰态性质(2)和(3)用于式(3), 得

$$\text{Kurt}(y) = \text{Kurt}(\mathbf{b}^T \mathbf{x}) = q_1^4 \text{Kurt}(s_1) + q_2^4 \text{Kurt}(s_2) \quad (4)$$

式中 \mathbf{x} 是给定的。为了方便, 使 $E\{y^2\} = 1$, 并将式(3)代入, 整理得

$$E\{y^2\} = E\{(\mathbf{q}^T \mathbf{s})^2\} = q_1^2 E\{s_1^2\} + q_2^2 E\{s_2^2\} = 1$$

因为前面已假定 s_1 和 s_2 的方差皆为 1, 所以有

$$q_1^2 + q_2^2 = 1 \quad (5)$$

定义目标函数为

$$\begin{aligned} F(\mathbf{b}) = F(\mathbf{q}) &= |\text{Kurt}(y)| = |\text{Kurt}(\mathbf{q}^T \mathbf{s})| \\ &= |q_1^4 \text{Kurt}(s_1) + q_2^4 \text{Kurt}(s_2)| \end{aligned} \quad (6)$$

在式(5)的约束条件下, 通过调整 b_1 和 b_2 来改变 q_1 和 q_2 , 使上述目标函数达到最大值, 从而使 y 的非高斯性达到最强, 则此时的 y 便是独立分量 s_1 或者 s_2 (关于这一点, 后面将给出详细证明)。这是一个求条件极值的问题。由于目标函数不能写成 b_1 和 b_2 的解析函数, 只能采用搜索寻优方法求解。

寻优过程中, 要满足条件式(5)。由式(3)还可以写出与之等价的约束条件

$$E\{y^2\} = E\{(\mathbf{b}^T \mathbf{x})^2\} = \mathbf{b}^T E\{\mathbf{x} \mathbf{x}^T\} \mathbf{b} = 1 \quad (7)$$

其中 $E\{\mathbf{x} \mathbf{x}^T\}$ 是观测数据向量的自相关矩阵, 计算很麻烦。

因此, 在调整 b_1 和 b_2 的同时, 根据式(7)来保持 $E\{y^2\} = 1$, 从而满足条件式(5)是有困难的。为此, 预先对观测向量进行白化处理^[2]: 用白化矩阵 $\mathbf{V} = \mathbf{D}^{-1/2} \mathbf{E}^T$, 左乘式(1)两端, 其中, $\mathbf{E} = (e_1, e_2, \dots, e_N)$, e_i 是观测向量 \mathbf{x} 的自相关阵的特征向量, \mathbf{D} 是由该阵的特征值构成的对角阵。令 $\mathbf{z} = \mathbf{V} \mathbf{x}$, 则有

$$\mathbf{z} = \mathbf{V} \mathbf{A} \mathbf{s} \quad (8)$$

其中 \mathbf{z} 的两个分量互不相关, 且方差皆为 1。令 $\mathbf{b}^T = \mathbf{w}^T \mathbf{V}$, 代入式(3), 得

$$y = \mathbf{b}^T \mathbf{x} = \mathbf{w}^T \mathbf{V} \mathbf{x} = \mathbf{w}^T \mathbf{V} \mathbf{A} \mathbf{s} = \mathbf{q}^T \mathbf{s} = \mathbf{w}^T \mathbf{z}$$

于是有

$$\mathbf{q}^T = \mathbf{w}^T \mathbf{V} \mathbf{A} \quad (9)$$

由于 $E\{y^2\} = \mathbf{w}^T E\{\mathbf{z} \mathbf{z}^T\} \mathbf{w} = \mathbf{w}^T \mathbf{w} = 1$, 所以约束条件式(5)现在可以写成

$$\mathbf{w}^T \mathbf{w} = \|\mathbf{w}\|^2 = 1 \quad (10)$$

于是问题变成: 在保持向量 \mathbf{w} 长度不变的条件下, 调整 \mathbf{w} 的方向, 使得目标函数

$$F(\mathbf{q}) = F(\mathbf{w}) = |\text{Kurt}(y)| = |\text{Kurt}(\mathbf{w}^T \mathbf{z})| = |\text{Kurt}(\mathbf{q}^T \mathbf{s})| \quad (11)$$

达到极大值。

4 寻优过程的几何解释

根据式(4), 将式(11)进一步写成

$$\begin{aligned} F(\mathbf{q}) = F(\mathbf{w}) &= |\text{Kurt}(y)| = |\text{Kurt}(\mathbf{w}^T \mathbf{z})| \\ &= |\text{Kurt}(\mathbf{q}^T \mathbf{s})| = |q_1^4 \text{Kurt}(s_1) + q_2^4 \text{Kurt}(s_2)| \end{aligned} \quad (12)$$

4.1 $\text{Kurt}(s_1) > 0, \text{Kurt}(s_2) > 0$ 的情形

当满足 $\text{Kurt}(s_1) > 0$ 和 $\text{Kurt}(s_2) > 0$ 时, 式(12)可以写成

$$F(\mathbf{q}) = F(\mathbf{w}) = \text{Kurt}(y) = q_1^4 \text{Kurt}(s_1) + q_2^4 \text{Kurt}(s_2) \quad (13)$$

约束条件为式(5)。为作图方便, 作变量代换, 令标量

$$t_1 = q_1^2, \quad t_2 = q_2^2 \quad (14)$$

和向量

$$\mathbf{t} = [t_1, t_2]^T \quad (15)$$

于是式(5)和式(13)可改写成

$$t_1 + t_2 = 1 \quad (16)$$

和

$$F(\mathbf{q}) = F(\mathbf{w}) = F(\mathbf{t}) = t_1^2 \text{Kurt}(s_1) + t_2^2 \text{Kurt}(s_2) \quad (17)$$

根据式(16)和式(17)可以绘出图 1。图中, 式(16)表现为直线段 AC ; 式(17)根据 $F(\mathbf{q})$ 即 $F(\mathbf{t})$ 的不同取值, 表现为一些 1/4 椭圆代表的等高线, 这里假设 $\text{Kurt}(s_2) > \text{Kurt}(s_1)$, B 点是线段 AC 与 1/4 椭圆的交点。

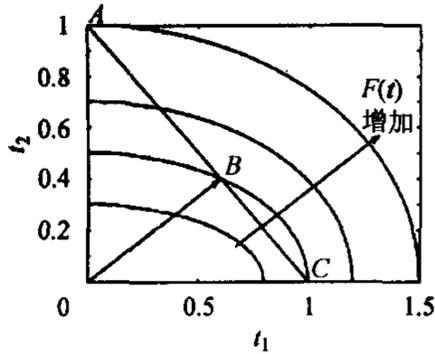


图1 寻优过程几何解释($\text{Kurt}(s_2) > \text{Kurt}(s_1) > 0$)

寻优过程中,通过改变 w 的方向,寻找 $F(w)$,即 $F(q)$,亦即 $F(t)$ 的极大值。在此过程中, t 始终在直线 AC 上移动。随着 $F(t)$ 值的增大,等高线向外扩展。显然,在约束条件式(16)之下, $F(t)$ 的极大点 t^* 有两个:一个是 $t^* = (1,0)^T$,另一个是 $t^* = (0,1)^T$ 。第一个 t^* 对应的是 $q^* = (1,0)^T$,代入式(13),得 $\text{Kurt}(y) = \text{Kurt}(s_1)$,于是有 $s_1 = y = w^T z$ 。第二个 t^* 对应的是 $q^* = (0,1)^T$,代入式(13),得 $\text{Kurt}(y) = \text{Kurt}(s_2)$,于是有 $s_2 = y = w^T z$ 。每一次寻优只能寻找到一个极大点,至于能寻到哪一个极大点,是无法预测的。因为每一次寻优开始时,要对向量 w 进行初始化,因而向量 q 和 t 也被初始化。在有关算法中,一般是用随机数作为初始化向量 w 的元素。以第一次寻优为例,参看图1,如果与初始化向量 w 对应的向量 t 位于向量 OB 的上方,则寻到的极大点将是 $t^* = (1,0)^T$;如果在下方,则寻到的将是 $t^* = (0,1)^T$;如果初始化向量 w 对应的向量 t 恰好与 OB 重合,则寻到的可能是 $t^* = (1,0)^T$,也可能是 $t^* = (0,1)^T$ 。

由于 w 的初始化状态是随机的,而 t 对 q 和 w 的正负是不敏感的。因此,作为对同一个极大点的寻优结果 w^* 有两种可能,二者相差一个负号。相应地,同一个独立分量,估计的结果也有两种可能,二者相差一个负号。因而,估计出的独立分量也可能与原始的源信号相差一个负号。实际上,这是前面提到的第一种不确定性的特例—— k 取负值。另外,由图1可知, $\text{Kurt}(s_2)$ 越大,它与直线 ABC 的交点 B 越向下移, ΔAOB 的面积越大,初始化的向量 t 进入该区域的概率越大, s_2 被先估计出的概率也越大。这表明,在同一种类型的独立分量中,峰态值较大的独立分量被优先分离出来的概率较大。

根据上面的分析还可以解释以下现象:(1)用 ICA 方法估计出的独立分量,它们的排列顺序可能与它们在 ICA 模型中的顺序不同,这是 ICA 问题解的第二种不确定性;(2)同一个 ICA 模型,用同一种方法,例如用本文所讨论的方法,求解几遍所估计出的独立分量,其排列顺序可能不同;(3)估计出的独立分量可能与原始模型中对应的独立分量相差

一个负号,如前所述,这是 ICA 问题解的第一种不确定性的—种特殊情况。

若在第1次寻优中,寻到一个极大点,记作 t_1^* ,譬如 $t_1^* = (0,1)^T$,与 t_1^* 对应的 q 和 w 分别记作 q_1^* 和 w_1^* ,则在第2次寻优中,为了避免再一次寻到 t_1^* ,应该增加一个约束,即在寻优过程中还要保持 t 与 t_1^* 正交,这样,对于图1所示情况,便可以寻到第2个极大点 $t_2^* = (1,0)^T$ 。实际上,直接的寻优变量为 w 。定理1确定了变量 w 与变量 t 的在寻优过程中的关系。

定理1 在寻优过程中,若保持 w 与 w_1^* 正交,则 t 与 t_1^* 正交。

证明 设第 i 个 w 向量 w_i 与第 j 个 w 向量 w_j 是正交的,即满足 $w_i^T w_j = 0$,则根据式(9)有 $q_i^T = w_i^T VA$ 及 $q_j^T = w_j^T VA$ 。

由式(8)可推得, VA 是正交矩阵,故 $q_i^T q_j = w_i^T VA(VA)^T w_j = w_i^T I w_j = w_i^T w_j = 0$ 取 $w = w_i^T$, $w_1^* = w_j$,及 $q = q_i$, $q_1^* = q_j$ 则有 $w^T w_1^* = 0$ 及 $q^T q_1^* = (q_1, q_2) \begin{pmatrix} 0 \\ q_2^* \end{pmatrix} = q_2 q_2^* = 0$ 。这样, $t^T t_1^* = (q_1^2, q_2^2) \begin{pmatrix} 0 \\ q_2^* \end{pmatrix} = q_2^2 q_2^{*2} = 0$ 即 t 与 t_1^* 正交。

证毕

由定理1可知,在寻优过程中,只要保持 w 与 w_1^* 正交,就可以确保 t 与 t_1^* 正交。

4.2 Kurt(s_1) 和 Kurt(s_2) 均小于零的情形

式(12)可以写成

$$F(q) = F(w) = |\text{Kurt}(y)| = q_1^4 |\text{Kurt}(s_1)| + q_2^4 |\text{Kurt}(s_2)| \quad (18)$$

于是,只要把情形1中的 $\text{Kurt}(y)$, $\text{Kurt}(s_1)$ 和 $\text{Kurt}(s_2)$ 分别用 $|\text{Kurt}(y)|$, $|\text{Kurt}(s_1)|$ 和 $|\text{Kurt}(s_2)|$ 取代即可。图1仍然适用。

4.3 Kurt(s_1) 和 Kurt(s_2) 异号的情形

假设 $\text{Kurt}(s_1) > 0$, $\text{Kurt}(s_2) < 0$ 。对 $\text{Kurt}(y)$ 暂不取绝对值,令

$$\begin{aligned} \bar{F}(q) &= \bar{F}(w) = \text{Kurt}(y) = q_1^4 \text{Kurt}(s_1) + q_2^4 \text{Kurt}(s_2) \\ &= q_1^4 \text{Kurt}(s_1) - q_2^4 |\text{Kurt}(s_2)| \end{aligned} \quad (19)$$

或

$$\bar{F}(t) = \text{Kurt}(y) = t_1^2 \text{Kurt}(s_1) - t_2^2 |\text{Kurt}(s_2)| \quad (20)$$

根据式(16)和式(20)可以绘出图2。其中式(16)表现为直线段 CE ; 式(20)表现为两族双曲线。以 $\bar{F}(t) = 0$ 时的直线 $t_2 = \sqrt{[\text{Kurt}(s_1)/|\text{Kurt}(s_2)|]} t_1$ 为分界,在该曲线右下方的是 $\bar{F}(t) > 0$ 对应的双曲线族,在左上方的是 $\bar{F}(t) < 0$ 的双曲线

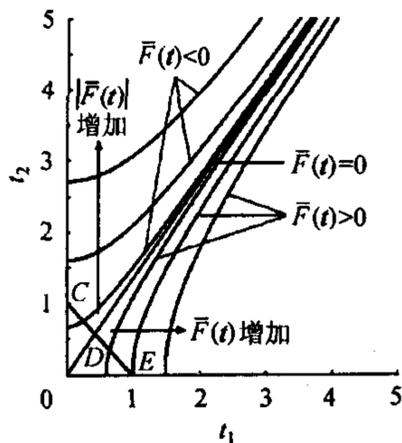


图 2 寻优过程的几何解释 ($Kurt(s_1) > 0, Kurt(s_2) < 0$)

族。D 是直线段 CE 与 $\bar{F}(t) = 0$ 的交点。

在式(10)的约束条件下, 优化算法以 $\bar{F}(t) = Kurt(y)$ 的绝对值作为目标函数进行寻优。参看图 2, 寻优过程中, 向量 t 的端点始终在直线 CE 上。第 1 次寻优时, 如果初始化向量 w 对应的向量 t 在直线 OD 的左上方, 则寻到的极大点 w_1^* 对应的 $t_1^* = (0, 1)^T$; 如果在右下方, 则 $t_1^* = (1, 0)^T$; 如果恰好在直线 OD 上, 则可能是前者, 也可能是后者。像前面的情况一样, 第 2 次寻优时, 要增加一个约束, 即保持 w 与 w_1^* 正交。

由上述分析可见, 在一个 ICA 模型中, 对于独立分量峰态取值有正有负的情况, 基于峰态的方法也适用。

另外, 从图 2 还可以看出, $|Kurt(s_2)|$ 越大, s_2 被先估计出的概率也越大。

4.4 独立分量中有一个为高斯型的情形

假设 s_2 为高斯型变量, 此时 $Kurt(s_2) = 0$, 式(12)变成

$$F(q) = F(w) = |Kurt(y)| \\ = |Kurt(w^T z)| = |Kurt(q^T s)| = q_1^4 |Kurt(s_1)|$$

即

$$F(t) = |Kurt(y)| = t_1^2 |Kurt(s_1)| \quad (21)$$

于是可以绘出图 3。显然, 寻优将寻到 $t^* = (1, 0)^T$, 从而得 $s_1 = y = w^* z$ 。 s_2 不能被估计了。

以上的讨论可以推广到多维, 譬如, 假设已经估计出两个独立分量, 得到的极大点分别是 w_1^* 和 w_2^* , 那么第 3 次寻优时, 应保持 w 与 w_1^* 及 w_2^* 正交; 再譬如, 假定共有 3 个独立分量, 其中包含 1 个(甚至 2 个)高斯型的, 则除了高斯型分量以外, 其它独立分量都可以估计, 等等。

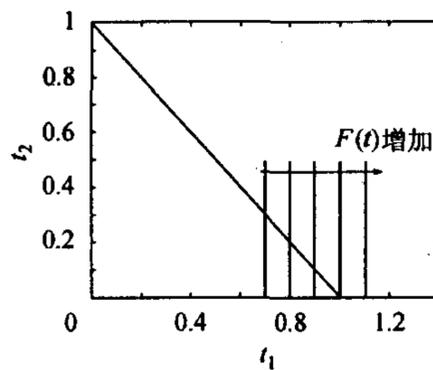


图 3 寻优过程的几何解释 (s_2 为高斯型)

5 结论

通过以上分析表明: 如果不考虑算法对观测数据中远离中心的无关数据过分敏感^[2]的话, 采用基于峰态绝对值最大化、对各独立分量逐个估计的方法解 ICA 问题, 从理论上讲, 可以给出每一个非高斯型独立分量的精确估计(精确到峰态相等), 只是每一个估计的方差均为 1(可能与原始源信号的方差不同), 并且可能与原始源信号相差一个负号。另外, 用上述方法估计出的独立分量, 其排列顺序(按估出的先后排列)不一定与其在原始模型中的顺序相同。而且, 对同一个模型估计几遍, 各遍估计出的独立分量, 其排列顺序也不一定相同。最后, 采用上述方法解 ICA 问题时, 峰态绝对值越大的独立分量被优先分离出来的概率越大。

参 考 文 献

- [1] Hyvariene A, Oja E. Fast and robust fixed-point algorithms for independent component analysis[J]. *IEEE Trans. on Neural Networks*, 1999, 10(3): 626 - 634.
- [2] Hyvariene A, Karhunen J, Oja E. Independent Component Analysis [M]. New York: John Wiley & Sons Inc., 2001: 165 - 181.
- [3] Hyvarinen A. Survey on independent component analysis[J]. *Neural Computing Surveys*, 1999, 2: 94 - 128.
- [4] Comon P. Independent component analysis: A new concept?[J]. *Signal Processing*, 1994, 34(4): 287 - 314.

张旭秀: 女, 1968 年生, 博士生, 副教授, 从事信号处理方面的研究工作。

邱天爽: 男, 1954 年生, 教授, 博士生导师, 从事信号与信息处理的研究工作。