论 文

偏度在独立元分析模型中的作用分析及算法设计

刘智勇*, 乔红

中国科学院自动化研究所复杂系统智能控制与管理国家重点实验室(筹), 北京 100190

* 通信作者. E-mail: zhiyong.liu@ia.ac.cn

收稿日期: 2008–10–20; 接受日期: 2010–08–10

国家自然科学基金(批准号:60975002)资助项目

摘要 相对于峭度(kurtosis), 偏度(skewness)历来在独立元分析(ICA)的研究中就没有得到充分重视. 尤其是当关于峭度符号的一比特匹配定理在理论上被证明了以后, 偏度似乎更是变成了 ICA 模型中的一个无用统计量. 但当信号的峭度很小或者其非 Gauss 性主要源自于偏度时, 仅仅利用峭度信息是不足够的. 本文目的就在于分析和讨论在此种情况下独立元分析如何利用偏度信息. 首先从理论上分析了偏度在 ICA 模型中的作用, 结果表明在偏度上并不存在与峭度类似的一比特匹配定理, 也就是说, 算法中模型密度函数的选择无需考虑其偏度与源信号偏度的符号匹配问题. 在此基础上, 本文进一步提出了一套灵活的模型密度函数设计方法, 并提出了一个算法实例, 它可以适用于具有任意偏度和峭度组合的信号.

关键词 独立元分析 偏度统计量 峭度统计量 一比特匹配定理

1 引言及背景介绍

作为一种被广泛用于盲信号分离、特征提取和信号分析的工具, 独立元分析(independent component analysis, ICA)的任务是通过一个线性变换从而使得输出的信号尽可能的相互独立. 在本文中, 我们考虑下述较为简单的 ICA 模型:

- 源信号的数目和观察到的信号数目相等;
- 线性变换是可逆的, 并且变换过程不受噪声的干扰.

在此假设前提下, ICA 可以通过如下线性变换来表示:

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \mathbf{W} \in \mathbb{R}^{n \times n}, \quad (1)$$

这里 $\mathbf{x} \in \mathbb{R}^n$ 表示观察到的 n 个信号, $\mathbf{W}$ 是待求的线性变换矩阵, $\mathbf{y}$ 是模型的输出. 当 $\mathbf{x}$ 是由 n 个相互独立的源信号 $\mathbf{s}$ 经过线性变换 $\mathbf{x} = \mathbf{A}\mathbf{s}$ 而得到, 那么 $\mathbf{y}$ 可以重建 $\mathbf{s}$. 事实上, Tong^[1] 等人在 1993 年就说明了当 $\mathbf{y}$ 的分量皆为非 Gauss 分布(或至多一个为 Gauss)的时候, $\mathbf{y}$ 的相互独立就可以将源信号 $\mathbf{s}$ 分离出来. 此问题在 1994 年的时候由 Comon^[2] 总结为独立元分析问题, 并在近年来无论在算法^[3–9]、理论^[10–15]还是应用(如对脑电信号 EEG^[16] 和金融数据^[17]的分析)等方面都得到了快速的发展.

引用格式: 刘智勇, 乔红. 偏度在独立元分析模型中的作用分析及算法设计. 中国科学: 信息科学, 2011, 41: 998–1012

总体来讲, 根据是否显性需要源信号分布的信息, ICA 算法可以分成两大类. 第一类包括了最小互信息 (minimum mutual information)^[10,11] 和最大似然 (maximum likelihood)^[12] 等方法. 当转换矩阵 $\mathbf{W}$ 是可逆的时候, 它们等价于最小化下面的目标函数,

$$\mathbf{W} = \arg_{\mathbf{W}} \min D_0(\mathbf{W}), \quad D_0(\mathbf{W}) = -H(\mathbf{y}) - \sum_{i=1}^n \int p_{\mathbf{W}}(y_i; \mathbf{W}) \lg p_i(y_i) dy_i, \quad (2)$$

这里

$$H(\mathbf{y}) = - \int p(\mathbf{y}) \lg p(\mathbf{y}) d\mathbf{y}$$

代表 $\mathbf{y}$ 的熵, $p_i(y_i)$ 是算法中的模型密度函数, 理论上应该采用源信号的密度函数形式. 但源信号的密度函数在实际应用中一般是无法预先得知的, 所以如何确定 $p_i(y_i)$ 成为此类方法的一个关键问题. 事实上, Bell^[10] 等 1995 年的论文揭示了 ICA 模型的一个很有意思的现象, 那就是公式 (2) 中的模型密度函数 $p_i(y_i)$ 预先定好为超 Gauss(super-Gaussian, 指归一化四阶统计量峭度大于 0) 就似乎可以重建任意的超 Gauss 信号 (如典型的人的语音信号等), 而无需知道信号的具体分布. 之后, 人们也发现了只要模型密度函数预先定好为亚 Gauss(sub-Gaussian, 指归一化峭度小于 0) 时也可以分离任意的亚 Gauss 信号^[11]. 同时, 一种利用可学习的两个 Gauss 的混合模型来描述信号分布的 LPM-ICA 算法^[4] 被提出, 并且成功地应用于分离超、亚 Gauss 或两者混合的信号. 这些现象使得关于 ICA 的一比特匹配猜想 (one-bit-matching conjecture) 被总结出来, 那就是: “只要模型密度函数和源信号的峭度 (kurtosis) 符号之间存在一一对应的关系, 信号就可以被分离”^[5]. 此猜想也被随后的很多著名 ICA 算法所不断证实, 例如 Flexible ICA Algorithm^[6], Extended Infomax^[7], Constrained EM ICA Algorithm^[8] 等等. 它们的一个共同特点就是利用一比特匹配猜想把模型的密度函数简化到只由一个参数来控制, 而通过学习这个参数可以使得模型的密度函数在超、亚 Gauss 之间切换, 从而使得算法可以分离任意的超、亚 Gauss 信号.

虽然这个猜想在实际中被多数的 ICA 研究者们所承认, 但一开始在理论上它却并没有被证实, 或者说人们对这个猜想为什么成立或在什么条件下成立并不清楚. 文献中, 当只存在两个亚 Gauss 信号时的情况被证明了^[13], 但此结果却不能引申到超 Gauss 或着多于两个信号的情形; 同时, 对于一类特殊的非线性函数, 文献 [14] 也研究了它需要达到稳定和正确 ICA 解的条件, 但它没有讨论如何才能得到正确解. 文献中首次对一般情况下的一比特猜想的研究是在 2004 年, 通过限定模型密度函数的偏度为 0, 文献 [18] 第一次完整证明了任意情况下的一比特定理. 最近, 在文献 [18] 的基础上文献 [15] 进一步证明了满足一比特定理的条件下, 目标函数的任意局部最优值也可以成功分离原信号, 进一步解释了文献中的很多算法虽然无法达到全局最优 (几乎所有常用的 ICA 算法都是一个局部最优的算法), 局部最优也可以达到盲信号分离的目的.

另一方面, 非 Gauss 最大化方法 (nongaussianity maximization) 是典型的不显性需要源信号分布的 ICA 方法, 它包括了峭度最大/最小化方法^[2,19] 和 FastICA^[20] 等. 和第一种方法比较起来, 此类方法似乎要更为简单, 因为它们算法中不涉及源信号的分布. 但事实上, 此类方法也隐性地对源信号做了某种程度上的假设, 例如峭度最大/最小化方法 (文献 [18] 中已经证明了它其实也是公式 (2) 的一个近似) 至少要求源信号的峭度不能为零.

通过上面关于 ICA 算法发展的背景简述我们可以看出, ICA 研究的重心放在了信号的峭度上. 下面我们进一步考虑信号的偏度. 直观来看, 如果信号的非 Gauss 性主要来自于偏度而非峭度, 也就是说, 信号具有较大的偏度 (绝对值) 和较小的峭度 (绝对值), 那么利用偏度来对其进行处理将可能是一

个较好的选择. 举个简单的例子, 假设源信号 $\mathbf{s}$, 观察信号 $\mathbf{x}$, 和输出信号 $\mathbf{y}$ 皆归一化成均值为零且协方差为单位矩阵, 且假设 $\mathbf{s}$ 的峭度为零, 那么不难得出 $\mathbf{y}$ 的峭度始终为零的结论. 换句话说, 在此种情况下, 基于峭度的 ICA 算法是无效的. 但文献中却鲜有关于偏度在独立元分析中作用的讨论, 文献 [21] 结果显示了基于非线性函数 $g(x) = x^2$ 的 EASI 算法可以适用于重建任意具有零峭度和非零偏度的信号, 而文献 [19, 22] 讨论了如何在目标函数中加入偏度信息. 在此文中, 我们将进一步指出任意具有非零偏度的密度函数均可用于分离具有非零偏度的信号, 而无需考虑偏度的符号问题. 这一点和一比特匹配定理所描述的峭度符号是完全不同的, 它要求模型密度函数和原信号的峭度符号相吻合. 本文的结论结合一比特匹配定理为 ICA 算法中的密度函数设计提供了基本的指导规则.

本文的结构安排如下: 第 2 节从理论上分析偏度在独立元分析模型中的作用; 在此基础上, 在第 3 节我们给出了一套灵活的独立元分析算法设计方法, 及算法设计的一个实例; 第 4 节是实验对比, 然后第 5 节结束本文.

2 偏度在独立元分析模型中的作用分析

我们从公式 (2) 出发来讨论独立元分析模型.

假设相互独立的源信号 $\mathbf{s}$, 观察值 $\mathbf{x}$ 和输出值 $\mathbf{y}$ 都经过预白化具有零均值和单位方差, 那么 $\mathbf{y}$ 和 $\mathbf{s}$ 之间关系可表示为 [18]

$$\mathbf{y} = \mathbf{R}\mathbf{s}, \quad (3)$$

这里 $\mathbf{R} = (r_{ij})_{n \times n}$ 是一个标准正交矩阵. 并且 $\mathbf{y}$ 和 $\mathbf{s}$ 之间的偏度和峭度具有如下关系:

$$\gamma_i^y = \sum_{j=1}^n r_{ij}^3 \gamma_j^s, \quad \nu_i^y = \sum_{j=1}^n r_{ij}^4 \nu_j^s, \quad (4)$$

其中 γ_i^y, ν_i^y 表示 y_i 的偏度和峭度, γ_j^s 和 ν_j^s 表示 s_j 的偏度和峭度.

从公式 (3) 可以看出, 如果 $\mathbf{R}$ 为一置换矩阵, 那么 $\mathbf{y}$ 即可将 $\mathbf{s}$ 重建出来. 所以我们下面要分析的是: 如何选择模型密度函数 $p_i(y_i)$ 的偏度 γ_i^m , 才能保证公式 (2) 最大化可以使得 $\mathbf{R}$ 成为一个 (符号不确定的) 置换矩阵.

首先, $\mathbf{R}$ 是一个标准正交矩阵使得公式 (2) 中的熵 $H(\mathbf{y}) = H(\mathbf{s})$, 也即为与变换无关的一个常数, 所以, 公式 (2) 的最小化等价于下面函数 $\hat{D}$ 的最大化:

$$\mathbf{R} = \arg_{\mathbf{R}} \min D_1(\mathbf{R}), \quad D_1(\mathbf{R}) = - \sum_{i=1}^n \int p_{\mathbf{R}}(y_i; \mathbf{R}) \lg p_i(y_i) dy_i. \quad (5)$$

其中 $p_{\mathbf{R}}(y_i; \mathbf{R})$ 是通过公式 (3) 得到的.

进一步, 借助于 Gram-Charlier 展开式, $p_{\mathbf{R}}(y_i; \mathbf{R})$ 可以被近似为

$$\begin{aligned} p_{\mathbf{R}}(y_i; \mathbf{R}) &\approx g(y_i) \left(1 + \frac{\gamma_i^y}{6} H_3(y_i) + \frac{\nu_i^y}{24} H_4(y_i) \right) \\ &= g(y_i) \left(1 + \frac{\sum_{j=1}^n r_{ij}^3 \gamma_j^s}{6} H_3(y_i) + \frac{\sum_{j=1}^n r_{ij}^4 \nu_j^s}{24} H_4(y_i) \right), \end{aligned} \quad (6)$$

其中 $g(y_i) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{y_i^2}{2})$ 表示一个标准 Gauss 分布密度函数, γ_i^y 和 ν_i^y 由公式 (4) 得到, Chebyshev-Hermite 多项式 $H_3(y_i)$ 和 $H_4(y_i)$ 定义如下:

$$H_3(y) = y^3 - 3y, \quad (7)$$

$$H_4(y) = y^4 - 6y^2 + 3. \quad (8)$$

同时, 模型密度函数也可以近似表示为

$$p_i(y_i) \approx g(y_i) \left(1 + \frac{\gamma_i^m}{6} H_3(y_i) + \frac{\nu_i^m}{24} H_4(y_i) \right), \quad (9)$$

这里 γ_i^m 和 ν_i^m 分别表示模型密度函数 $p_i(y_i)$ 的偏度和峭度.

利用公式 (6) 和 (9), 公式 (5) 变成了如下形式:

$$\begin{aligned} \hat{D} &= \sum_{i=1}^n \int g(y_i) \left(1 + \frac{\gamma_j^y}{6} H_3(y_i) + \frac{\nu_i^y}{24} H_4(y_i) \right) \\ &\quad \times \lg(g(y_i) \left(1 + \frac{\gamma_i^m}{6} H_3(y_i) + \frac{\nu_i^m}{24} H_4(y_i) \right)) dy_i \\ &= \hat{D}_1 + \hat{D}_2, \end{aligned} \quad (10)$$

这里

$$\begin{aligned} \hat{D}_1 &\triangleq \sum_{i=1}^n \int g(y_i) \left(1 + \frac{\gamma_j^y}{6} H_3(y_i) + \frac{\nu_i^y}{24} H_4(y_i) \right) \\ &\quad \times \lg \left(1 + \frac{\gamma_i^m}{6} H_3(y_i) + \frac{\nu_i^m}{24} H_4(y_i) \right) dy_i, \\ \hat{D}_2 &\triangleq \sum_{i=1}^n \int g(y_i) \left(1 + \frac{\gamma_j^y}{6} H_3(y_i) + \frac{\nu_i^y}{24} H_4(y_i) \right) \\ &\quad \times \lg g(y_i) dy_i, \end{aligned}$$

我们首先考虑 $\hat{D}_2$. 根据 Chebyshev-Hermite 多项式的如下正交特性^[23]:

$$\int H_n(x) H_m(x) e^{-\frac{x^2}{2}} dx = n! \delta_{mn}, \quad (11)$$

这里 $\delta_{mn} = 1$ 当 $m = n$, 否则 $\delta_{mn} = 0$, 不难求得

$$\begin{aligned} \int g(y_i) \frac{\gamma_j^y}{6} H_3(y_i) \lg g(y_i) dy_i &= 0, \\ \int g(y_i) \frac{\nu_i^y}{24} H_4(y_i) \lg g(y_i) dy_i &= 0, \end{aligned}$$

所以得到

$$\hat{D}_2 = \sum_{i=1}^n \int g(y_i) \lg g(y_i) dy_i = -\frac{n}{2} (1 + \lg 2\pi). \quad (12)$$

因此, 最大化 $\hat{D}$ 等价于最大化 $\hat{D}_1$.

下面进一步考虑 $\hat{D}_1(\mathbf{R})$. $\hat{D}_1(\mathbf{R})$ 中的

$$\sum_{i=1}^n \int g(y_i) \lg(g(y_i) \left(1 + \frac{\gamma_i^m}{6} H_3(y_i) + \frac{\nu_i^m}{24} H_4(y_i) \right)) dy_i$$

是与 $\mathbf{R}$ 无关的量, 所以 $\hat{D}_1$ 可以进一步等价于如下:

$$\begin{aligned}\tilde{D}_1 &= \sum_{i=1}^n \int g(y_i) \left(\frac{\gamma_j^y}{6} H_3(y_i) + \frac{\nu_i^y}{24} H_4(y_i) \right) \\ &\quad \times \lg \left(1 + \frac{\gamma_i^m}{6} H_3(y_i) + \frac{\nu_i^m}{24} H_4(y_i) \right) dy_i.\end{aligned}\quad (13)$$

因为我们着重分析偏度在上述目标函数中的作用, 为了简单起见, 我们假设原始信号的峭度为零, 即 $\nu_i^s = 0$, 那么通过公式 (4) 可以得出 $\nu_i^y = 0$. 同时, 我们限制模型密度函数的峭度也为零, 这样, 上述的目标函数 $\tilde{D}_1$ 就可以进一步简化为

$$\begin{aligned}\tilde{D}_1 &= \sum_{i=1}^n \int g(y_i) \frac{\gamma_j^y}{6} H_3(y_i) \lg \left(1 + \frac{\gamma_i^m}{6} H_3(y_i) \right) dy_i \\ &= \gamma_j^y \sum_{i=1}^n \int g(y_i) \frac{H_3(y_i)}{6} \lg \left(1 + \frac{H_3(y_i)}{6} \gamma_i^m \right) dy_i.\end{aligned}\quad (14)$$

我们分析 $\sum_{i=1}^n \int g(y_i) \frac{H_3(y_i)}{6} \lg(1 + \frac{H_3(y_i)}{6} \gamma_i^m) dy_i$. 令 $f(x) = h \lg(1 + xh)$, 那么 $f'(x) = \frac{h^2}{1+xh} \geq 0$ 且当且仅当 $h = 0$ 时 $f'(x) = 0$. 同时有 $f(0) = 0$, 所以当 $h \neq 0$ 时, $f(x)$ 和 x 是保持同号的. 重新回头来看 $\tilde{D}_1$, 因为我们考虑偏度的影响, 所以当然假设 $H_3(y_i) \neq 0$, 那么 $\sum_{i=1}^n \int g(y_i) \frac{H_3(y_i)}{6} \lg(1 + \frac{H_3(y_i)}{6} \gamma_i^m) dy_i$ 是一个和 γ_i^m 同号的常数, 我们记

$$\lambda_i^m \triangleq \sum_{i=1}^n \int g(y_i) \frac{H_3(y_i)}{6} \lg \left(1 + \frac{H_3(y_i)}{6} \gamma_i^m \right) dy_i.\quad (15)$$

这样, $\tilde{D}_1$ 就可以最终简化成以下的形式:

$$\begin{aligned}\tilde{D}_1 &= \gamma_j^y \sum_{i=1}^n \lambda_i^m = \sum_{j=1}^n r_{ij}^3 \gamma_j^s \sum_{i=1}^n \lambda_i^m \\ &= \sum_{j=1}^n \sum_{i=1}^n r_{ij}^3 \gamma_j^s \lambda_i^m.\end{aligned}\quad (16)$$

于是, ICA 的目标函数 (2) 可以最终简化成下面的形式:

$$\mathbf{R} = \arg_{\mathbf{R}} \max \sum_{i=1}^n \sum_{j=1}^n r_{ij}^3 \gamma_j^s \lambda_i^m.\quad (17)$$

对于公式 (17), 我们给出如下定理:

定理 1 只要 γ_j^s 和 λ_i^m 皆不为零, 那么公式 (17) 的最大化将由一个符号不确定的置换矩阵 $\mathbf{R} = \mathbf{P}$ 得到, 而与 γ_i^s 和 λ_i^m 的大小和符号无关, 也即与 γ_i^m 的大小和符号无关.

证明 见附录 A.

从上述定理可以得出如下的独立元分析算法中关于偏度的规则:

规则 1 当信号的峭度趋于零并且限制模型密度函数的峭度趋于零的情况下, 任意具有非零偏度的模型密度函数皆可用于 ICA 算法.

利用独立元分析的一比特匹配定理, 可以得到以下的独立元分析算法中关于峭度的规则:

规则 2 当限制模型密度函数的偏度趋于零的情况下, 独立元分析算法中模型密度函数的峭度必须与源信号的峭度符号保持一致.

下面我们讨论如何在上述两个规则指导下设计合理的模型密度函数来处理具有不同偏度和峭度组合的信号.

3 模型密度函数设计及对应算法

我们设计一种灵活的密度函数, 它只需要调整一个参数就可以具有不同的偏度和峭度组合. 由于理论上混合 Gauss 模型 (mixture of Gaussian, MOG) 可以模拟任意的非 Gauss 分布, 在本文中我们选取作它为我们所要设计模型密度函数的原型函数. 首先我们需要求取混合 Gauss 模型的偏度和峭度的表达式.

3.1 混合 Gauss 模型偏度和峭度

k 个 Gauss 的混合模型可以表示为

$$p(y) = \sum_{j=1}^k \alpha_j G(y|m_j, \sigma_j^2), \quad (18)$$

其中 $G(y|m_j, \sigma_j^2)$ 表示的是以 m_j 为均值, 以 σ_j^2 为方差的 Gauss 分布, α_j 是指相应的 Gauss 分布的先验概率. 它的偏度和峭度可以通过下面的表达式来求解 [19]:

$$\begin{aligned} \varphi(\tau) &= \int_{-\infty}^{+\infty} e^{i\tau y} p(y) dy \\ &= \int_{-\infty}^{+\infty} e^{i\tau y} \sum_{j=1}^k \alpha_j G(y|m_j, \sigma_j^2) dy \\ &= \sum_{j=1}^k \alpha_j \exp \left\{ \tau m_j i - \frac{\tau^2 \sigma_j^2}{2} \right\}, \end{aligned} \quad (19)$$

这里 $i \triangleq \sqrt{-1}$.

基于累积量母函数 (cumulant generating function)

$$\phi(\tau) = \ln(\varphi(\tau)),$$

我们可以通过下面的公式来计算累积量 c_n :

$$c_n = (-i)^n \frac{d^n \phi(\tau)}{d\tau^n}. \quad (20)$$

最终得到如下结果:

$$\begin{aligned} c_1 &= m_1 = \mu_{10}, \\ c_2 &= m_2 = \mu_{20} + \mu_{02} - \mu_{10}^2, \\ c_3 &= m_3 = \mu_{30} + 3\mu_{12} + 2\mu_{10}^3 - 3\mu_{10}\mu_{02} - 3\mu_{10}\mu_{20}, \\ c_4 &= m_4 - 3m_2^2 \end{aligned}$$

$$= \mu_{40} + 6\mu_{22} + 3\mu_{04} + 12\mu_{10}^2\mu_{02} + 12\mu_{10}^2\mu_{20} - 12\mu_{10}\mu_{12} - 4\mu_{10}\mu_{30} - 3\mu_{02}^2 - 3\mu_{20}^2 - 6\mu_{02}\mu_{20} - 6\mu_{10}^4, \quad (21)$$

这里

$$\mu_{pq} \triangleq \sum_{j=1}^k \alpha_j m_j^p \sigma_j^q,$$

m_n 表示 n 阶矩, $c_1, \dots, c_4$ 分别表示混合 Gauss 模型的均值、方差、偏度和峭度.

3.2 模型密度函数设计

基于前面关于密度函数偏度和峭度的规则 1 和 2, 我们设计如下的模型密度函数:

$$p(y|\theta) = 0.5G(y|\theta, (2-\theta)^2) + 0.5G(y|-\theta, 1), \quad 0 \leq \theta \leq 1. \quad (22)$$

根据公式 (21), 得到上述函数的偏度和峭度如下:

$$c_3 = 1.5\theta^3 - 6\theta^2 + 4.5\theta, \quad (23)$$

$$c_4 = -1.25\theta^4 - 6\theta^3 + 16.5\theta^2 - 18\theta + 6.75. \quad (24)$$

函数偏度和峭度随 θ 变化的曲线如图 1 所示, 同时, 图 2 是密度函数形状随 θ 变化的直观示意图. 我们可以注意到函数的以下两点特性:

- 当 $\theta = 0$ 和 $\theta = 1$ 时, 密度函数分别为超、亚 Gauss 函数, 且偏度为零;
- 当 $\theta = 0.69$ 时, 函数的峭度为零, 且对应一个较大的偏度为 0.76.

上述两点特性是符合规则 1 和 2 的. 这样, 通过调整参数 θ 从 0 到 1 的变化函数可以平滑地在超 Gauss 和亚 Gauss 之间进行切换, 并可以实现零偏度, 同时, 函数也可以处理当信号的非 Gauss 性主要来自于偏度的情形.

3.3 基于上述模型密度函数的 ICA 算法

基于上述的模型密度函数, 我们提出如下只需估计一个参数就可以调节模型密度函数来适用于具有不同偏度和峭度组合的信号 ICA 算法.

步骤 1 固定 $\Theta = [\theta_1, \theta_2, \dots, \theta_m]^T$, 按照如下 natural gradient 规则更新 $\mathbf{W}^{[4]}$:

$$\mathbf{W}^{\text{new}} = \mathbf{W}^{\text{old}} + \eta(\mathbf{I} + \phi(\mathbf{y})\mathbf{y}^T)\mathbf{W}, \quad (25)$$

这里

$$\phi_i(y_i) = r_{i1} \frac{\theta_i - y_i}{(2 - \theta_i)^2} - r_{i2}(y_i + \theta_i),$$

并且

$$r_{i1} = \frac{0.5G(y_i|\theta_i, (2-\theta_i)^2)}{p(y_i|\theta_i)},$$

图 1 模型密度函数的偏度和峭度随 θ 变化的曲线
Figure 1 Changes of the skewness and kurtosis of the model pdf with θ

图 2 模型密度函数随 θ 变化的三维示意图
Figure 2 Changes of the shape of the designed model pdf with θ

$$r_{i2} = \frac{0.5G(y_i | -\theta_i, 1)}{p(y_i | \theta_i)}.$$

步骤 2 固定 $\mathbf{W}$, 更新 Θ :

$$\Theta^{\text{new}} = \Theta^{\text{old}} + \tau \Delta \Theta, \quad (26)$$

这里 $\Delta \Theta = [\Delta \theta_1, \Delta \theta_2, \dots, \Delta \theta_m]^T$, 并且

$$\Delta \theta_i = r_{i1} \frac{\theta_i^2 + (y_i - 6)\theta_i - y_i^2 + 2y_i + 4}{(2 - \theta_i)^3} - r_{i2}(y_i + \theta_i).$$

算法中 $\theta_i = \frac{1}{1 + \exp(\zeta_i)}$ 可以引入来限制 $0 \leq \theta \leq 1$.

4 实验分析

本节将对几种典型的 ICA 算法进行对比实验分析, 包括 Extended Infomax^[7], FastICA^[20](其中非线性函数采用 $\tanh(u)$), LPM-ICA^[4], 以及本文提出的方法. 实验环境是一台 P4 3.6 GHz 台式电脑, 所有程序均在 MatLab 6.5 平台下完成.

上述 4 种算法将在分别在人工数据和实际数据上进行比较测试. 实验结果由两个指标来评价, 第一是效率, 即实验所花费的时间, 第二是效果. 我们设计下面的指标来对信号重建的效果进行衡量,

$$E = \frac{1}{d} \max_i \left(\sum_{j=1}^d \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \frac{1}{d} \max_j \left(\sum_{i=1}^d \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right), \quad (27)$$

这里 d 表示信号的个数, $\mathbf{P} \triangleq \mathbf{W}\mathbf{A}$. 它利用了重建信号中最差的一个来衡量算法的效果, 某种程度上代表了算法的最差表现, 即其性能下限.

4.1 人工数据上的实验

实验中共采用了 4 个样本数目均为 10000 个的归一化的人工信号 (即均值为 0 和方差为 1), 它们

表 1 实验中 4 个原始信号的偏度和峭度

Table 1 The statistics of the source signals used in the experiment

	Signal 1	Signal 2	Signal 3	Signal4
Skewness	-0.288	0.317	0.284	0.014
Kurtosis	-0.060	0.048	0.64	-0.551

相应的偏度和峭度列于表 1 中, 可由链接 <http://www.ia.ac.cn/zhiyong.liu/icadata.zip> 下载. 从表 1 可以看出, 信号 1 和 2 的峭度基本为零, 所以它们的非 Gauss 性主要来自于偏度; 而信号 3 和 4 分别是超 Gauss 和亚 Gauss 信号.

基于上述 4 个信号我们进行了两组实验. 实验 1 是对信号 1 和 2 进行处理, 以对比测试 4 种算法在处理非 Gauss 性主要来自于偏度的信号时的性能; 而实验 2 将对整个 4 个信号进行处理, 以测试它们的综合性能.

实验 1 的结果列于表 2, 从表中可以看出, LPM-ICA 和本文提出的方法成功地重建了原始信号, 而 Extended-Infomax 和 FastICA 却失败了. FastICA 和本文方法所得到的 $\mathbf{P}$ 矩阵见公式 (28) 和 (29). 原始信号 1 和 2 的重建必须利用它们的偏度信息, 而 LPM-ICA 成功是因为它采用了 3 个 Gauss 的混合模型作为密度函数, 通过适当的学习过程, 它显然可以使得密度函数具有较大偏度且较小的峭度, 只是它每个信号都需要估计 8 个参数而使得它的速度较慢 (其实它更为严重的问题不只是速度慢, 而是它的参数在学习过程中可能陷入局部最优而导致密度函数不能匹配原信号, 这一点在后面的实验中还会讨论). 本文所提出算法通过自动学习参数 θ 而使得密度函数具有较大偏度 (可以参见图 3 中的 θ_1 和 θ_2 学习曲线最后都大概收敛到 0.62 左右) 从而成功地分离了原信号. 还有一点可以看出的是, 虽然信号 1 的偏度是负数, 而本文算法中的密度函数偏度始终大于 0, 但还是成功地分离了它, 只不过重建符号相反 (对应的 $\mathbf{P}$ 的元素为 -1.30), 这一点和定理 1 的结论是相符合的. 另一方面, Extended-Infomax 算法是严格遵守一比特匹配定理的, 它利用两个偏度为零的超、亚 Gauss 密度函数之间的切换来达到峭度符合匹配, 而 FastICA 也只是利用了峭度信息, 所以它们两者都失败了. 这个实验也说明了当信号的非 Gauss 性主要来自于偏度时, 关于峭度符号的一比特匹配定理可能是无效的, 信号的成功重建必须利用偏度信息.

$$\mathbf{P}_{\text{FICA}} = \begin{pmatrix} 0.399 & -0.923 \\ -0.917 & -0.385 \end{pmatrix}, \quad (28)$$

$$\mathbf{P}_{\text{ours}} = \begin{pmatrix} -1.300 & -0.054 \\ 0.056 & 1.492 \end{pmatrix}, \quad (29)$$

实验 2 的结果也列于表 2. 通过公式 (30) 可以看出 FastICA 成功重建了信号 3 和 4, 即两个超亚 Gauss 信号, 但仍然无法重建信号 1 和 2. Extended-Infomax 的效果甚至要劣于 FastICA, 这一点可以从它们两者得到的 E 指标得到反映. LPM-ICA 的效果要优于 FastICA, 但由于它庞大的参数空间 (32 个参数) 使得它极易陷入局部最优, 所以决定了它无论是在效率还是效果上都不如本文所提出的方法. 公式 (31) 表明本文的算法成功地重建了原始的 4 个信号, 同时通过图 3 中的参数学习曲线可以看出 4 个模型密度函数是怎样逐步适应 4 个源信号的.

图 3 人工数据的实验 2 中本文提出方法中的 4 个算法在一个模型密度函数的参数 θ 的学习曲线, 循环以后就已经收敛, 我们每隔 10 个样本记录一次, 即 1000 次
Figure 3 Learning process of θ , with two converged to skewed pdf, one to sub-Gaussian and one to super-Gaussian, where the algorithm converged after only one iteration and we took a record every 10 samples for the total 10000 samples

图 4 实际数据上的实验: 本文提出方法中的 10 个模型密度函数的参数 θ 的学习曲线收敛为 7 个超 Gauss 和 3 个亚 Gauss
Figure 4 Convergence process of our algorithm on real data-set, which shows that the model pdf's converged to seven super-Gaussian and three sub-Gaussian

$$P_{\text{FICA}} = \begin{pmatrix} -0.064 & -0.049 & 0.016 & \mathbf{0.997} \\ -0.000 & 0.091 & \mathbf{-0.996} & 0.013 \\ 0.954 & -0.310 & -0.031 & 0.041 \\ -0.294 & -0.945 & -0.085 & -0.063 \end{pmatrix}, \quad (30)$$

$$P_{\text{ours}} = \begin{pmatrix} -0.128 & 0.146 & -0.062 & \mathbf{-1.3188} \\ -0.002 & 0.021 & \mathbf{1.409} & -0.123 \\ 0.097 & \mathbf{1.471} & 0.001 & 0.121 \\ \mathbf{-1.291} & -0.033 & -0.187 & 0.028 \end{pmatrix}. \quad (31)$$

最后讨论一下 4 种算法的效率问题. 通过上面的实验数据可以看出 FastICA 毫无疑问是最快的, Extended-Infomax 算法要快于我们的算法在于它无需估计模型参数, 而 LPM-ICA 较慢的原因是它需要估计过于庞大的参数集. 这其实启发我们是否可以将本文得到的有关偏度的结论综合到 FastICA 框架, 以得到快速且有效的 ICA 算法, 而这当然是下一步的研究内容.

4.2 基于实际数据的算法比较

本实验将利用实际数据对上面的 4 种算法进行比较测试. 实验中采用的数据是 50000 个十维观察信号, 它们是由 7 个超 Gauss 和 3 个亚 Gauss 信号经过线形变换而得到, 实验中的数据可以从链接 <http://sweat.cs.unm.edu/~bap/domos.html> 下载得到.

表 2 人工数据上的实验结果

Table 2 Experimental results on artificial data-set

Experiment 1			Experiment 2	
	Cost time	E	Cost time	E
E-Infomax	1.34 s	0.412	2.78 s	0.261
FastICA	0.39 s	0.437	1.45 s	0.236
LPM-ICA	3.67 s	0.052	6.22 s	0.187
Our algorithm	2.01 s	0.043	3.42 s	0.115

表 3 实际数据上的实验结果

Table 3 Experimental results on real data-set

	Cost time	E	Number of recovered signals
E-Infomax	8.75 s	2.71	6
FastICA	2.36 s	0.15	10
LPM-ICA	22.42 s	3.43	3
Our algorithm	12.23 s	0.24	10

实验结果列于表 3 中, 其中, FastICA 和本文中的方法都成功地分离了 10 个原始信号, 而 LPM-ICA 和 Extended-Infomax 只成功分离了 3 个和 6 个. 它们没有成功分离的信号导致其性能指标 E 偏大, 而它们的失败可能是由于算法陷入了局部最优而使得密度函数没有成功地匹配原始信号 (这里局部最优指的是模型密度函数的峭度符号没有收敛到与源信号的峭度符号一致). 本文提出算法的参数收敛过程见图 4. 另外在效率上, 同上面人工数据上的实验类似, FastICA 也是远快于其他 3 种方法, 包括本文提出的算法.

5 讨论及结论

本文首先从理论上分析了偏度在独立元分析模型中的作用, 同时根据理论分析得到了实际算法设计中模型密度函数的偏度设置规则. 结合关于峭度符号的一比特匹配定理, 我们给出了两个规则来设计适用于任意偏度和峭度组合信号的独立元分析算法. 在此基础上我们提出了利用混合 Gauss 模型作为模型密度函数的灵活的 ICA 算法设计方法, 并给出了一个实例. 实验结果表明本文提出的算法可以有效处理信号的非 Gauss 性来自偏度的情形, 而这是传统的基于峭度的算法, 如典型的 Extended-Infomax 和 FastICA 所无法处理的. 另外通过实验也启发我们可能为偏度和峭度的组合寻求更快速的优化框架, 如 FastICA, 这为我们下一步的研究提供了启示.

参考文献

- 1 Tong L, Inouye Y, Liu R. Waveform-preserving blind estimation of multiple independent sources. *IEEE Trans Signal Process*, 1993, 41: 2461–2470
- 2 Comon P. Independent component analysis—a new concept? *Signal Process*, 1994, 36: 287–314
- 3 Bell A, Sejnowski T. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput*, 1995, 7: 1129–1159

- 4 Amari S I, Cichocki A, Yang H. A new learning algorithm for blind separation of sources. *Adv Neural Inf Process Syst*, 1996, 8: 757–763
- 5 Cardoso J F. Infomax and maximum likelihood for source separation. *IEEE Signal Process Lett*, 1997, 4: 112–114
- 6 Xu L, Cheung C C, Yang H H, et al. Independent component analysis by the information-theoretic approach with mixture of density. In: *Proc of 1997 IEEE Intl Conf on Neural Networks (IEEE-INNS IJCNN97) III*. Houston, 1997. 1821–1826
- 7 Xu L, Cheung C C, Amari S I. Learned parametric mixture based ICA algorithm. *Neurocomput*, 1998, 22: 69–80
- 8 Xu L, Cheung C C, Amari S I. Further results on nonlinearity and separation capability of a liner mixture ICA method and learned LPM. In: *Fyfe C, ed. Proceedings of the I&ANN'98, La Laguna, 1998*. 39–45
- 9 Everson R, Roberts S. Independent component analysis: a flexible nonlinearity and decorrelating manifold approach. *Neural Comput*, 1999, 11: 1957–1983
- 10 ELee T W, Girolami M, Sejnowski T J. Independent component analysis using an Extended Infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. *Neural Comput*, 1999, 11: 417–441
- 11 Welling M, Weber M. A constrained EM algorithm for independent component analysis. *Neural Comput*, 2001, 13: 677–689
- 12 Cheung C C, Xu L. Some global and local convergence analysis on the information-theoretic independent component analysis approach. *Neurocomput*, 2000, 30: 79–102
- 13 Amari S I, Chen T P. Stability analysis of adaptive blind source separation. *Neural Netw*, 1997, 10: 1345–1351
- 14 Liu Z Y, Chiu K C, Xu L. One-bit-matching conjecture for independent component analysis. *Neural Comput*, 2004, 16: 383–399
- 15 Xu L. One-bit-matching theorem for ICA, convex-concave programming on polyhedral set, and distribution approximation for combinations. *Neural Comput*, 2007, 19: 964–973
- 16 Hulle M M V. Constrained subspace ICA based on mutual information optimization directly. *Neural Comput*, 2007, 19: 546–569
- 17 Dyrholm M, Makeig S, Hansen L K. Model selection for convolutive ICA with an application to spatiotemporal analysis of EEG. *Neural Comput*, 2007, 19: 934–955
- 18 Cha S M, Chan L W. Applying independent component analysis to factor model in finance. In: *Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning, Data Mining, Financial Engineering, and Intelligent Agents*. Hong Kong, 2000. 538–544
- 19 Stuart A, Ord J. *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory*. London: Edward Arnold, 1994
- 20 Hyvainen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw*, 1999, 10: 626–634
- 21 Choi S J, Liu R W, Cichocki A. A spurious equilibria-free learning algorithm for the blind separation of non-zero skewness signals. *Neural Process Lett*, 1998, 7: 61–68
- 22 Blaschke T, Wiskott L. Cubica: Independent component analysis by simultaneous third-and fourth-order cumulant diagonalization. *IEEE Trans Signal Process*, 2004, 52: 1250–1256
- 23 Zarzoso V, Phlypo R, Comon P. A contrast for independent component analysis with priors on the source kurtosis signs. *IEEE Signal Process Lett*, 2008, 15: 501–504

附录 A 定理 1 的证明

为了证明定理 1, 我们首先引入如下引理.

引理 1 当 $1 > \alpha_2 > \alpha_3 > \cdots > \alpha_n, 1 > \beta_2 > \beta_3 > \cdots > \beta_n$, 那么对于矩阵 C

$$C = \begin{pmatrix} 1 & \beta_2 & \cdots & \beta_n \\ \alpha_2 & \alpha_2\beta_2 & \cdots & \alpha_2\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_n & \alpha_n\beta_2 & \cdots & \alpha_n\beta_n \end{pmatrix} \quad (A1)$$

对任意的 $i > l$ 且 $j > k$, 有下面的结论:

$$c_{iljk} \triangleq \alpha_i\beta_j + \alpha_l\beta_k - \alpha_i\beta_k - \alpha_l\beta_j = (\alpha_i - \alpha_l)(\beta_j - \beta_k) > 0. \quad (A2)$$

下面我们首先考虑公式 (17) 中所有的 γ 和 λ 同号的情形, 也即模型密度函数的偏度符号和信号的偏度符号相匹配的情况. 对此, 我们提出下面的引理 2.

引理 2 当信号和模型的偏度同号, 且信号和模型具有不同的偏度值, 那么公式 (17) 的最大值将由 $R = P$ 得到, 这里 P 是一个置换矩阵.

证明 首先引入 $\gamma_2^s = \alpha_2\gamma_1^s, \dots, \gamma_n^s = \alpha_n\gamma_1^s, \lambda_2^m = \beta_2\lambda_1^m, \dots, \lambda_n^m = \beta_n\lambda_1^m$,

$$\sum_{j=1}^n \sum_{i=1}^n r_{ij}^3 \gamma_i^s \lambda_j^m = - \sum_{i=1}^n \sum_{j=1}^n r_{ij}^3 \gamma_i^s \lambda_j^m = -\gamma_1^s \lambda_1^m \sum_{i=1}^n \sum_{j=1}^n r_{ij}^3 \alpha_i \beta_j \geq -\gamma_1^s \lambda_1^m \sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 \alpha_i \beta_j, \quad (A3)$$

上式中的 $\alpha_1 = \beta_1 = 1$. 因为假设了所有密度函数偏度以及信号偏度都不相同, 不失一般性我们假设 $1 > \alpha_2 > \cdots > \alpha_n > 0, 1 > \beta_2 > \cdots > \beta_n > 0$. 那么根据引理 1 得到如下推导:

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 \alpha_i \beta_j \\ & \leq \left(\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 \alpha_i \beta_j + r_{21}^2 c_{2121} \right) \\ & \quad (\text{假设 } r_{12}^2 \geq r_{21}^2, \text{ 反之先消除 } r_{12}^2) \\ & = S \begin{pmatrix} r_{11}^2 + r_{21}^2 & (r_{12}^2 - r_{21}^2)\beta_2 & \cdots & r_{1n}^2\beta_n \\ 0 & (r_{22}^2 + r_{21}^2)\alpha_2\beta_2 & \cdots & r_{2n}^2\alpha_2\beta_n \\ r_{31}^2\alpha_3 & r_{32}^2\alpha_3\beta_2 & \cdots & r_{3n}^2\alpha_3\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1}^2\alpha_n & r_{n2}^2\alpha_n\beta_2 & \cdots & r_{nn}^2\alpha_n\beta_n \end{pmatrix} \\ & \leq S \begin{pmatrix} r_{11}^2 + r_{21}^2 & (r_{12}^2 - r_{21}^2)\beta_2 & \cdots & r_{1n}^2\beta_n \\ 0 & (r_{22}^2 + r_{21}^2)\alpha_2\beta_2 & \cdots & r_{2n}^2\alpha_2\beta_n \\ r_{31}^2\alpha_3 & r_{32}^2\alpha_3\beta_2 & \cdots & r_{3n}^2\alpha_3\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1}^2\alpha_n & r_{n2}^2\alpha_n\beta_2 & \cdots & r_{nn}^2\alpha_n\beta_n \end{pmatrix} \\ & \quad + (r_{12}^2 - r_{21}^2)c_{3121} \\ & = S \begin{pmatrix} r_{11}^2 + r_{12}^2 & 0 & \cdots & r_{1n}^2\beta_n \\ 0 & z_1 & \cdots & r_{2n}^2\alpha_2\beta_n \\ z_2 & z_3 & \cdots & r_{3n}^2\alpha_3\beta_n \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1}^2\alpha_n & r_{n2}^2\alpha_n\beta_2 & \cdots & r_{nn}^2\alpha_n\beta_n \end{pmatrix} \\ & \quad \vdots \end{aligned}$$

$$\leq S \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \hat{r}_{22}^2 \alpha_2 \beta_2 & \hat{r}_{23}^2 \alpha_2 \beta_3 & \cdots & \hat{r}_{2n}^2 \alpha_2 \beta_n \\ 0 & \hat{r}_{32}^2 \alpha_3 \beta_2 & \hat{r}_{33}^2 \alpha_3 \beta_3 & \cdots & \hat{r}_{3n}^2 \alpha_3 \beta_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{r}_{n2}^2 \alpha_n \beta_2 & \hat{r}_{n3}^2 \alpha_n \beta_3 & \cdots & \hat{r}_{nn}^2 \alpha_n \beta_n \end{pmatrix}$$

其中运算符 S 定义为 $S[a_{ij}]_{i,j=1}^n \triangleq \sum_{i=1}^n \sum_{j=1}^n a_{ij}$, $z_1 \triangleq (r_{22}^2 + r_{21}^2) \alpha_2 \beta_2$, $z_2 \triangleq (r_{31}^2 + r_{21}^2 - r_{12}^2) \alpha_3$, $z_3 \triangleq (r_{32}^2 + r_{12}^2 - r_{21}^2) \alpha_3 \beta_2$. 最终第一行和第一列可以同时消除为 0 是因为 $\mathbf{R}$ 是一个标准的正交矩阵, 它满足 $\sum_{i=2}^1 r_{ij}^2 = \sum_{j=1}^n r_{ij}^2 = 1$, 同时上述结果满足 $\sum_{i=2}^n \hat{r}_{ij}^2 = \sum_{j=2}^n \hat{r}_{ij}^2 = 1$.

重复上述的过程,

$$\begin{aligned} \gamma_1^s \lambda_1^m S \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \hat{r}_{22}^2 \alpha_2 \beta_2 & \hat{r}_{23}^2 \alpha_2 \beta_3 & \cdots & \hat{r}_{2n}^2 \alpha_2 \beta_n \\ 0 & \hat{r}_{32}^2 \alpha_3 \beta_2 & \hat{r}_{33}^2 \alpha_3 \beta_3 & \cdots & \hat{r}_{3n}^2 \alpha_3 \beta_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \hat{r}_{n2}^2 \alpha_n \beta_2 & \hat{r}_{n3}^2 \alpha_n \beta_3 & \cdots & \hat{r}_{nn}^2 \alpha_n \beta_n \end{pmatrix} &\leq \gamma_1^s \lambda_1^m S \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \alpha_2 \beta_2 & 0 & \cdots & 0 \\ 0 & 0 & \tilde{r}_{33}^2 \alpha_3 \beta_3 & \cdots & \tilde{r}_{3n}^2 \alpha_3 \beta_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \tilde{r}_{n3}^2 \alpha_n \beta_3 & \cdots & \tilde{r}_{nn}^2 \alpha_n \beta_n \end{pmatrix} \vdots \\ &\leq \gamma_1^s \lambda_1^m S \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & \alpha_2 \beta_2 & 0 & \cdots & 0 \\ 0 & 0 & \alpha_3 \beta_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_n \beta_n \end{pmatrix} = \gamma_1^s \lambda_1^m \sum_{i=1}^n \alpha_i \beta_i. \end{aligned}$$

上述等号成立的条件是当且仅当 $r_{11} = r_{22} = \cdots = r_{nn} = 1$. 所以公式 (17) 的最大值为 $\sum_{i=1}^n \alpha_i \beta_i$ 当且仅当 $\mathbf{R} = \mathbf{I}$ 得到. 一般情况下, 当假设 $1 > \alpha_2 > \alpha_3 > \cdots > \alpha_n > 0$ 和 $1 > \beta_2 > \beta_3 > \cdots > \beta_n > 0$ 不成立 (但它们仍然不相等), 那么所以公式 (17) 的最大值将由 $\mathbf{R} = \mathbf{P}$ 得到, $\mathbf{P}$ 为一置换矩阵.

在引理 1 和 2 的基础上, 得到以下引理 3.

引理 3 当信号和模型的偏度同号, 那么公式 (17) 的最大值将由 $\mathbf{R} = \mathbf{P}$ 得到, 这里 $\mathbf{P}$ 是一个置换矩阵.

证明 类似引理 2 的证明引入 $\alpha_2, \dots, \alpha_n, \beta_2, \dots, \beta_n$, 并且不失一般性假设 $1 > \alpha_2 > \cdots > \alpha_k = \alpha_{k+1} > \cdots > \alpha_n > 0, 1 > \beta_2 > \beta_3 > \cdots > \beta_n > 0$. 引入常数 $\epsilon > 0$, 满足

$$\alpha_{k-1} > \alpha'_k = \alpha_k + \epsilon > \alpha_{k+1}. \quad (\text{A4})$$

假设 α'_k 是新的 α_k , 那么根据引理 2, 公式 (17) 的最大值得到如下:

$$\max \left(\sum_{j=1}^n \sum_{i=1}^n r_{ij}^3 \gamma_i^s \lambda_j^m \right) = \lim_{\epsilon \rightarrow 0} \gamma_1^s \lambda_1 \sum_{i=1}^n (\alpha_i \beta_i + \epsilon \beta_k) = \gamma_1^s \lambda_1 \sum_{i=1}^n \alpha_i \beta_i, \quad (\text{A5})$$

且当 $\mathbf{R}$ 为单位矩阵得到. 其实, 上述的结果也可以通过假设 $\alpha_{k-1} > \alpha'_{k+1} = \alpha_{k+1} + \epsilon > \alpha_k$ 得到, 只不过这时候的 $\mathbf{R}$ 变成了置换矩阵, 即 $r_{(k+1)k} = r_{k(k+1)} = 1$ 而非 $r_{kk} = r_{(k+1)(k+1)} = 1$. 所以, 当等号出现的时候, 相应的置换矩阵和单位矩阵都可以使得公式 (17) 达到最大值. 当等号出现在 β 之中或者 α, β 两者皆有等号的证明同上类似, 不再赘述.

最后, 在引理 3 的基础上, 定理 1 的证明就变得非常直接, 因为定理 1 和引理 3 的差别就在于引理 3 没有考虑信号偏度和模型密度函数偏度异号时的情况.

证明 (定理 1) 当信号偏度和模型密度函数偏度存在异号时, 即 γ_i^s 和 λ_j^m 存在正负数, 我们首先取它们的绝对值, 即考虑 $|\gamma_i^s|$ 和 $|\lambda_j^m|$, 那么显然有

$$\max \left(\sum_{j=1}^n \sum_{i=1}^n r_{ij}^3 \gamma_i^s \lambda_j^m \right) \leq \max \sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 |\gamma_i^s| |\lambda_j^m|. \quad (\text{A6})$$

根据引理 3, $\sum_{i=1}^n \sum_{j=1}^n r_{ij}^2 |\gamma_i^s| |\lambda_j^m|$ 的最大值 (用 D 表示) 将由 $\mathbf{R} = \mathbf{P}$ (置换矩阵) 得到. 根据公式 (40),

$$\max \left(\sum_{j=1}^n \sum_{i=1}^n r_{ij}^3 \gamma_i^s \lambda_j^m \right) \leq D, \quad (\text{A7})$$

并且, 上述等号成立的条件为 $\mathbf{R} = \mathbf{P}'$, 这里 $\mathbf{P}'$ 为一个同 $\mathbf{P}$ 有同样结构, 但是当对应的 $\gamma_i^s \lambda_j^m$ 为负数时, $p_{ij} = -1$, 也即从一般意义上来讲为一个符号不确定的置换矩阵.

Investigation on the skewness for independent component analysis

LIU ZhiYong* & QIAO Hong

Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

*E-mail: zhiyong.liu@ia.ac.cn

Abstract Skewness has received much less attention than kurtosis in the independent component analysis (ICA). In particular, the skewness seems to become a useless statistics after the kurtosis related one-bit-matching theorem was proven. However, as the non-Gaussianity of one signal comes mainly from skewness, it is intuitively understandable that its recovery should not rely on kurtosis. In this paper we discuss the skewness based ICA, and show that any probability density function (pdf) with non-zero skewness can be employed by ICA for the recovery of the source with non-zero skewness, without needing to consider the skewness sign. The observation together with the one-bit-matching theorem provides a basic guideline for the model pdf design in ICA algorithm.

Keywords independent component analysis, skewness, kurtosis, one-bit-matching theorem