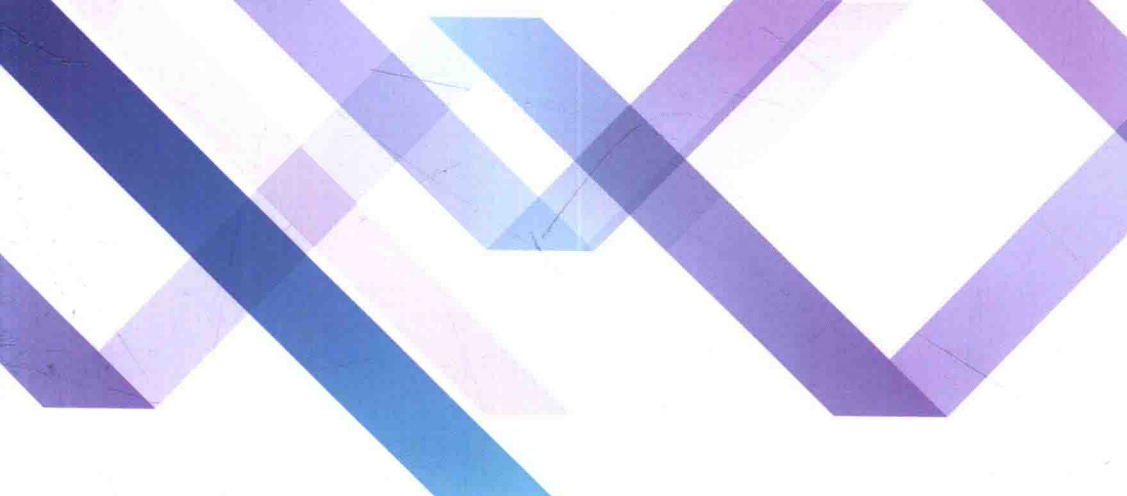




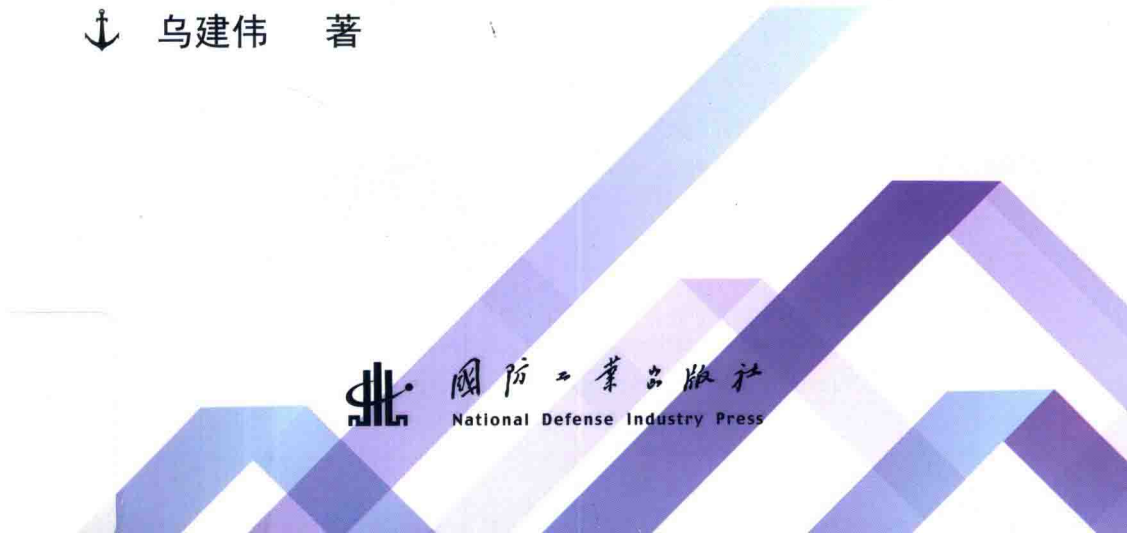
DULI CHENGFEN FENXI ZHONG
DE GAOJIE TONGJILIANG FANGFA

独立成分分析中 的高阶统计量方法

✚ 乌建伟 著



國防工業出版社
National Defense Industry Press



独立成分分析中的 高阶统计量方法

乌建伟 著

国防工业出版社

· 北京 ·

内 容 简 介

独立成分分析是现代数字信号处理理论中的重要研究分支之一,它在许多不同的领域中有着大量的应用。

本书从高阶统计量的角度,以观察数据的瞬时混合模型为主,尽可能地介绍独立成分分析中涉及高阶统计量内容的基本理论和算法,包括一部分传统内容较新的研究进展。

具有一定信号处理、高等数学、线性代数、概率论与数理统计知识的学生和研究者都可以通过本书学习,掌握这一研究方向的基本理论和一些经典的算法,从而可以直接进入这一研究领域。

图书在版编目(CIP)数据

独立成分分析中的高阶统计量方法/乌建伟著. —北京:国防工业出版社,2016. 12

ISBN 978-7-118-11213-9

I. ①独… II. ①乌… III. ①高阶-统计分析-研究
IV. ①0212. 1

中国版本图书馆 CIP 数据核字(2016)第 325639 号

※

国防工业出版社 出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

三河市众誉天成印务有限公司印刷

新华书店经售

*

开本 710 × 1000 1/16 印张 17½ 字数 332 千字

2016 年 12 月第 1 版第 1 次印刷 印数 1—2000 册 定价 79.00 元

(本书如有印装错误,我社负责调换)

国防书店: (010)88540777

发行邮购: (010)88540776

发行传真: (010)88540755

发行业务: (010)88540717

序

独立成分分析(或盲信号分离)是1990年以来国内外信号处理领域的热点研究方向,它在通信与雷达、图像处理、语音信号处理、医学信号分析、经济及金融数据分析等众多领域有着广泛的应用。

独立成分分析的目的在于恢复被混合的非高斯信号(即源信号),而用于衡量随机变量非高斯性的数量指标是信号的峭度和偏度(即信号样本的四阶累积量和三阶累积量)。这两个数量指标的定义十分简单,同时,也有着很明确的几何解释。因此,在已知的绝大多数算法中,或者直接以观察数据的峭度、偏度为对照函数,或者在算法中直接或间接的估计源信号的峭度或偏度。虽然通过这些方法能够很容易地进入这一领域,但它们(包括某些方法的进一步研究和拓展)都散见于各类文献之中,因此,有必要从这个角度总结分析它们并将其汇集成书。

就目前国内学者出版的同类著作来看,大致可分为两类:第一类以介绍算法为主,但各种方法都混杂在一起,没有一条贯穿始终的主线;第二类以讨论应用为主,所涉及的算法较少。

本书以介绍独立成分分析中与高价统计量有关的算法为主,对于在各个领域中的应用,由于已有国内外大量的文献与专著予以介绍,本书没有将这部分内容包括进来。

对于观察数据的瞬时混合模型,本书直接以实信号和复信号的高阶累积量为主线展开,从对照函数到算法,包括某些算法或算法思想的进一步讨论及扩展,例如算法收敛性的讨论及伪解问题(这部分国内专著很少涉及),以及高阶累积量在其他算法中的应用,都做了较为全面介绍。其中,张量方法是目前国内同类专著所没有涉及的内容。由于篇幅原因,对于噪声估计问题和欠定模型的处理方法,虽然有选择的做了介绍,但内容相对较少,同样的原因,对于盲卷积的内容也没有涉及。

多个学科交叉是独立成分分析这一研究方向的特点,并且每年都有大量新的结果发表出来,由于作者的水平和精力所限,不可能将其全部包括进来,希望读者见谅。同时,对于书中的不足甚至谬误之处,恳请读者批评指正。

作者

2016年12月

目 录

第 1 章 前言	1
1.1 独立成分分析的概念和模型	1
1.1.1 盲信号分离与独立成分分析	1
1.1.2 独立成分分析的概率模型	2
1.2 独立成分分析解的性质	4
1.2.1 ICA 分解的等价性	4
1.2.2 ICA 分解的唯一性	4
1.2.3 ICA 与 PCA 的联系	6
1.3 独立成分分析的发展历史、扩展及应用	6
1.3.1 ICA 的发展历史	6
1.3.2 ICA 的扩展研究	8
1.3.3 ICA 的应用	9
参考文献	9
第 2 章 基本的分离原则、算法和对照函数	14
2.1 几个基本的分离原则	14
2.1.1 最大似然估计	14
2.1.2 互信息最小化	15
2.1.3 信息极大化	16
2.1.4 负熵最大化	17
2.2 其他分离方法	18
2.2.1 消去交叉累积量方法	18
2.2.2 非线性去相关	19
2.2.3 分布比较方法	21
2.2.4 基于几何特征的方法	22
2.3 ICA 中常用的优化方法	23
2.3.1 自然梯度与相对梯度	23

2.3.2	雅可比算法	25
2.3.3	不动点算法	25
2.4	概率密度函数的 Gram - Charlier 和 Edgeworth 展开	26
2.5	目标函数的要求与构造	30
2.6	非对称的对照函数	33
2.7	高阶累积量作为目标函数的一些理论结果	37
2.8	基于交叉累积量的对照函数	44
	参考文献	48
第3章	实信号的基于高阶累积量的分离方法	51
3.1	四阶盲辨识及其扩展方法	51
3.2	基于四阶累积量的快速算法	54
3.2.1	FastICA 算法	54
3.2.2	FastICA 算法收敛性的进一步讨论	57
3.2.3	基于峭度的 P - ICA 算法	61
3.3	峭度之和对照函数及其算法	63
3.4	层级网络方法	67
3.5	高阶统计量目标函数稳定点的讨论与偏度解混算法	71
3.6	有限样本对于高阶统计量对照函数在盲抽取运算中的影响	76
3.7	利用向量峭度的子空间独立成分分析	81
3.8	含噪声数据的高阶统计量盲分离算法	84
3.9	分离源信号某个子集的高阶累积量方法	91
3.10	源信号峭度位于某特定区域的盲抽取算法	93
	参考文献	96
第4章	复值信号的峭度极大化方法	100
4.1	基本的数学知识	100
4.1.1	复数域上的 CR 运算	100
4.1.2	复随机变量及其数字特征	103
4.2	复随机向量及强无关变换	106
4.3	复信号的固定点算法	109
4.4	峭度最大化算法(KMA)	114
4.5	峭度极大化算法的修正算法	120
4.6	基于峭度的梯度算法和固定点算法	123

4.7	基于峭度的非圆周型信号盲分离算法(K-eBSE)	128
4.8	快速峭度最大化算法与T-快速峭度最大化算法	132
4.9	RobustICA	137
	参考文献	138
第5章	高阶累积量在其他盲分离算法中的应用	141
5.1	双输入双输出问题	141
5.2	JADE 算法	146
5.3	模型匹配算法中峭度的应用:分布的组合	149
5.4	模型匹配算法中峭度的应用:广义 Gaussian 分布	153
5.5	模型匹配算法中峭度的应用:t-分布与广义 Gaussian 分布	160
5.6	一比特匹配猜想的讨论	165
5.7	关于通用匹配函数的存在性	170
5.8	利用互累积量的两个算法	176
	参考文献	181
第6章	张量方法	185
6.1	张量的定义及其基本运算	185
6.2	高阶张量的矩阵表示与秩	187
6.3	超对称张量与张量定义的线性映射	189
6.4	张量的奇异值分解	192
6.5	最优秩-1 与秩- $(R_1, R_1, \dots, R_N)$ 分解	197
6.6	标准分解	203
6.6.1	引言	203
6.6.2	CANDECOP 与联合 EVD	204
6.6.3	联合广义 Schur 分解	206
6.6.4	算法	208
6.7	三阶张量算法标准分解的梯度算法与 ALS 及其改进	210
6.7.1	基于梯度的 Levenberg-Marquardt 算法	210
6.7.2	交替最小二乘算法	211
6.7.3	线搜索与增强的线搜索	212
6.8	基于三阶张量联合对角化盲分离算法	213
6.8.1	三阶张量最大对角化的雅可比方法	214
6.8.2	三阶张量联合对角化(STOTD)的 ICA 算法	216

6.9	欠定情形下的四阶盲辨识方法(FOOBI)	219
6.9.1	FOOBI 算法	219
6.9.2	FOOBI-2 算法	222
6.10	欠定情形下矩阵联合对角化的盲分离算法	224
6.10.1	问题的转化及 PARAFAC 分解的唯一性	224
6.10.2	计算	226
	参考文献	231
第7章	峭度与偏度的直接估计及应用	234
7.1	峭度的直接估计	234
7.1.1	峭度估计算法(KEA)	234
7.1.2	KEA 的两个初步应用	237
7.2	基于峭度估计的 Givens 旋转算法	241
7.2.1	雅可比角的直接估计	242
7.2.2	Givens 旋转矩阵的整体估计	247
7.3	偏度的直接估计	252
7.3.1	偏度估计算法	253
7.3.2	SEA 在选择合适的对照函数或算法方面的应用	256
7.4	分离非对称源信号的 Givens 旋转算法	258
7.5	分离非对称源信号的 Givens 旋转算法(GASS)的一个 理论上的推广	264
7.5.1	基于三阶张量分解的盲分离算法	265
7.5.2	新的混合矩阵估计算法	266
	参考文献	268

第 1 章 前 言

1.1 独立成分分析的概念和模型

1.1.1 盲信号分离与独立成分分析

现实世界里充满了各种各样的信号,它们在客观上表现为各种物理量的变化,能够被人们感知或被仪器纪录。但在许多情况下,人们对于那些感兴趣的信号,需要找到产生这些信号的源信号,因为测得的信号虽然与源信号存在因果关系,但往往因为其他因素(例如干扰)而额外地增加了随机因素。笼统地讲,观察到的信号是源信号的某种变换,并可能混入了不需要的噪声信号。这类信号的变换既包括时间域上的,也包括空间域上的,变换的形式取决于信号源及接收方的空间位置关系、传输介质的物理属性等,可统称为信道特征。

为了获得源信号,或者为了确定某些信道特征,人们需要对这种发送接收的关系建立适当的模型。如果模型的假设条件(限制)很少,其中的参数大多数是不确定的,那么这样的问题就可以称作是盲信号处理(Blind Signal Processing)。“盲”的含义是除了测得的信号之外,其他先验信息很少。盲信号处理主要包括盲信号分离、盲反卷积(盲均衡)、多道的盲反卷积(多道盲均衡)等具有很强实际意义的内容,同时,也恰恰是因为假设条件少,它的应用十分广泛。

盲信号分离(BSS)是盲信号处理当中很重要的一类问题,它针对的是一个多输入多输出(MIMO)系统。考虑系统的输入包含多个源信号以及可能存在的噪声信号,输出即测得的信号是各输入信号的变换组合叠加。盲信号分离的目标是在一定程度上尽可能地把一个个有用的源信号从混杂的信号中分离出来。盲信号分离的一个经典例子是“鸡尾酒会(Cocktail Party)”问题^[1]:鸡尾酒会上的每个人都能从嘈杂的声音中听清自己关注的谈话,而排除其他干扰。假如在现场安置多个麦克风,并用机器来自动处理,能否从采集到的多道混合声音信号中将每个人的说话声分离出来?

如果假设信号的混合是以最简单的方式,即线性混合且信道没有记忆性(虽然实际情况往往复杂得多),在任何一个时刻 t , (多道)源信号用向量 $\mathbf{s}(t)$ 表

示,那么观察信号可表示为

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) \quad (1.1)$$

式中: $\mathbf{v}(t)$ 为加性噪声; $\mathbf{s}(t)$ 为 m 维; $\mathbf{x}(t)$ 和 $\mathbf{v}(t)$ 为 n 维; $\mathbf{A}$ 为表征这一简单混合信道的 $n \times m$ 未知矩阵,称为混合矩阵。

盲信号分离的任务在已知 $\mathbf{x}(t)$ 的一段纪录后,以尽可能小的误差估计出相应的 $\mathbf{s}(t)$ (称其为保持波形的^[2])和 $\mathbf{A}$ 。从解方程的角度看,虽然这是不可解的问题,但只要加上独立性的假设——在任一时刻 $\mathbf{s}(t)$ 的各分量 $s_1(t), s_2(t), \dots, s_m(t)$ 各自取值是相互独立的,则在一定程度上该问题就变为可解的了,这就产生了所谓的独立成分分析(Independent Component Analysis, ICA),也译为独立分量分析。在实际问题中,只要信号源各自独立地工作,在物理上没有关联,就容易满足这一独立性假设。

1.1.2 独立成分分析的概率模型

首先需要强调的是:独立成分分析虽然最典型的应用就是盲信号分离,但它并不是求解盲信号分离的唯一途径。一般的统计信号处理方法把信号作为随机过程来处理,因此样本的次序是至关重要的,但ICA方法完全忽略了样本的次序。

对任意一个源信号 $s_i(t)$,若不考虑其在不同时刻的取值 $s_i(t_1)$ 和 $s_i(t_2)$ ($t_1 \neq t_2$)的关联性,而仅把这些观察值看作是从某个概率分布里面产生的某些抽样的时候,式(1.1)中的信号都可以用随机变量来代表。此外,噪声项 $\mathbf{v}(t)$ 的存在使得完整地恢复源信号不再可能,所以ICA是忽略掉噪声项的,因此概率模型可表示为

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1.2)$$

式中: $\mathbf{s}$ 为分量相互独立的 m 维随机向量,称作源向量; $\mathbf{x}$ 为 n 维随机向量,且仅是 $\mathbf{s}$ 的一个线性变换,称作观察向量。ICA或BSS的目标是根据已有的 $\mathbf{x}$ 的样本估计混合矩阵 $\mathbf{A}$ 以及每个观察样本所对应的源信号样本。

ICA将源信号的所有采样值都当作是独立同分布的,然而只有白噪声信号才具有这个特征。对大多数信号而言,某时刻的采样值与先前的采样值是有关联的(典型的例子如平稳信号),因此这些样本并不符合独立同分布的假设。图1.1中,(a),(b)分别为产生的两段非周期三角波信号,它们经不同的线性组合产生了(c),(d)所示的混合信号。从ICA的角度,两个离散信号与一张散点图是等价的,因为采样次序已经不起作用了。图1.1中的(e),(f)分别是源信号和混合信号的散点图(分别含1000个样本),从中看出点的分布具有线状轨迹的特征,正是因为这些点由式(1.2)中随机向量的独立抽样产生。不过,只要

样本数量足够多,散点图的总体效果与是否来自独立抽样的关系就不大了。从统计的观点看,三角波信号的瞬时值服从均匀分布,两个混合信号的联合分布是一个平行四边形区域上的均匀分布,图 1.1 (g) 是从这一分布独立抽样得到的散点图,其特征与图 1.1 (f) 是接近的。

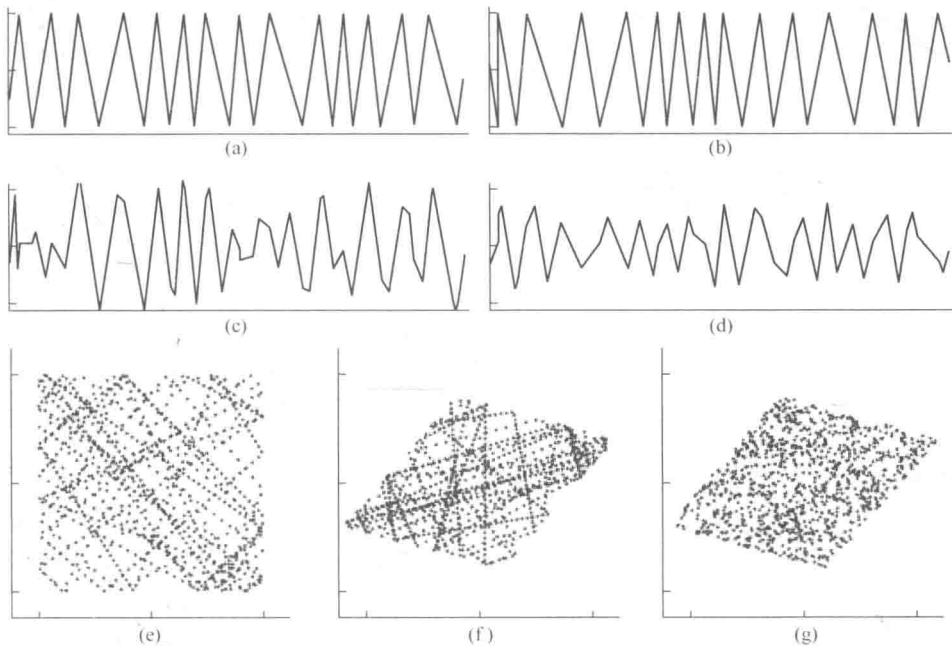


图 1.1 两个三角波信号的混合

- (a) $s_1(t)$; (b) $s_2(t)$; (c) $x_1(t) = 0.5s_1(t) - 0.7s_2(t)$; (d) $x_2(t) = 0.6s_1(t) + 0.2s_2(t)$;
 (e) $s_1(t)$ 与 $s_2(t)$ 的散点图; (f) $x_1(t)$ 与 $x_2(t)$ 的散点图;
 (g) 从联合分布 $(0.5U_1 - 0.7U_2, 0.6U_1 + 0.2U_2)$ 独立抽样 1000 次结果的散点图
 (U_1 与 U_2 独立且都服从 $[-1, 1]$ 上的均匀分布)。

还需要说明的是,虽然 ICA 源于盲信号分离,实际上任给一个随机向量,不论它是否真的是由独立随机变量组合而成,都可以用一套算法得到一种分解。因此,ICA 作为数据处理并不限于用在盲信号的分离。Comon 给出了一个 ICA 的数学定义如下^[3]:

定义 1.1 对于协方差有限的 n 维随机向量 $\mathbf{X}$,其协方差阵记为 $\mathbf{V}$, $\mathbf{X}$ 的独立成分分析是一对矩阵 $\{\mathbf{F}, \mathbf{\Delta}\}$, 满足:

- (1) $\mathbf{X}$ 的协方差阵可以分解为

$$\mathbf{V} = \mathbf{F}\mathbf{\Delta}^2\mathbf{F}^*$$

式中: Δ 为对角元为正的对角矩阵; F 为秩为 ρ 的列满秩阵; $*$ 表示共轭转置。

(2) X 可写为 $X = FY$, 其中 F 为具有协方差阵 Δ^2 的 ρ 维随机向量, 并且其分量(在最大化对照函数的意义下) 是最可能独立的。

在这个定义中, $X = FY$ 称为随机向量 X 的一个 ICA 分解。这个定义是针对一般的随机向量给出的, 并没有假设式(1.2)是满足的, 所以, 作为限定, 有“最可能地”这一措辞。不过, 只要 X 符合式(1.2), 并且其中的 A 是列满秩的, 源向量各分量的方差有限, 那么 $X = AS$ 就是一个 ICA 分解。如果能找到一个矩阵 W , 使得 $WA = I_m$, 则能通过线性变换得到源向量 $S = WX$ 。

反之, 若 A 不是列满秩的, 那么 ICA 得到的矩阵 F 是 $n \times \rho$ 的, 且 $\rho < m$, 这意味着不可能从线性变换反解出 m 个源信号。此时, 式(1.2) 被称为是欠定的 (Underdetermined) 或超完备的 (Overcomplete), 它已经不在 ICA 的基本模型的框架之内。

如果 $m > n$, 则 X 的协方差阵必然是奇异的, 表现为观察信号中存在着冗余, 这种冗余可以通过降维预处理的方法去除。最能反映 ICA 本质的是 $m = n$ 的情形, 即观察信号的维数等于源信号的个数, 且 A 是非奇异的, 这是大多数盲信号分离研究中都采用的假设。

1.2 独立成分分析解的性质

1.2.1 ICA 分解的等价性

不难发现, 定义 1.1 中的 $\{F, \Delta\}$ 并不是唯一确定的, 因为任取一个对角元非零的 $\rho \times \rho$ 对角阵 Λ 和一个置换阵 P , 令 $F' = F\Lambda P$, $\Delta' = P * \Lambda^{-1} \Delta P$, 那么 $\{F', \Delta'\}$ 也满足定义 1.1 中的条件。因此, ICA 的解实际上定义了一个等价类。

从盲信号分离的角度看, 因为源信号在混合的过程中有幅度的缩放, 就无从通过观察信号的强度去推断源信号的强度(平均能量, 统计意义上即为方差)。即使各信号源有一个排列次序, 它们一经混合, 次序信息也丢失了, 也就是说, 除非有对源信号某些统计性质的先验知识, 在最好的情况下某个恢复信号对应于哪一个源信号, 以及与源信号的尺度关系, 都是无法确定的^[4]。若 M 是可能的混合矩阵, 那么 $M\Lambda P$ 也是可能的混合矩阵(Λ, P 如上定义)。

设 $X = F_1 Z_1$, $X = F_2 Z_2$ 是两个不同的 ICA 分解, 只要存在 $\rho \times \rho$ 的对角矩阵 Λ 以及置换矩阵 P , 使得 $F_1 = F_2 \Lambda P$ 成立, 则这两个分解就是等价的。

1.2.2 ICA 分解的唯一性

等价性表明 ICA 分解有无穷多种可能, 但是否任意两个解都等价? 就盲信

号分离问题来说,只有与 $X=AS$ 等价的 ICA 分解才能分离出源信号。如果任意的分解都与此等价,就称此时的 ICA 分解是唯一的(在等价类的意义下)。满足这种唯一性的线性混合模型称为可辨识的(identifiable)^[5]。

考虑混合矩阵为非奇异矩阵的情形,为了寻找 ICA 分解唯一性的条件,可以借助如下的 Darmois - Skitovich 定理^[6]:

定理 1.1 (Darmois - Skitovich 定理) 设 $X_1, X_2, \dots, X_n$ 是相互独立的一组随机变量, $a_1, \dots, a_n, b_1, \dots, b_n$ 是一组常数,

$$Y_1 = a_1 X_1 + \dots + a_n X_n$$

$$Y_2 = b_1 X_1 + \dots + b_n X_n$$

又假设 Y_1 与 Y_2 独立,若有 $a_i \neq 0, b_i \neq 0$, 则 X_i 必定是高斯(正态)分布。

这个定理说明,一组独立的随机变量经过不同的线性组合之后,所得到的“混合随机变量”要独立所需要满足的条件。在 ICA 中,若通过分离变换 $Y = WX = WAs$ 从观察向量 X 获得一个分量相互独立的重构向量 Y , 那么矩阵 $C \triangleq WA$ 应该具有的性质可以由这个定理推出。

定理 1.2 (文献[3]中的定理 11) 设 X 是 n 维随机向量,其每个分量的分布都是独立且非退化的,至多有一个是高斯(正态)分布。令 C 是 n 阶正交矩阵, $Z = CX$, 则以下 3 个性质是等价的:

- (1) Z 的分量 z_i 两两独立;
- (2) Z 的分量 z_i 相互独立;
- (3) $C = AP$, 其中 A 是对角阵, P 是置换阵。

这个定理表明,只要源信号向量的各分量相互独立且高斯分量至多只有一个,那么依据独立性是可以唯一地(在等价类的意义下)确定混合矩阵的。这就意味着盲信号分离可以通过样本分布的特征来实现。若源向量中高斯分量超过一个,则这些高斯成分是无法分开的。下面的定理也可以由定理 1.1 推出,是定理 1.2 的推广。

定理 1.3 (Cao 和 Liu, 文献[7]中的定理 3) 设 $Y = AS, S = (s_1, \dots, s_n)^T$ 的分量相互独立, $Y = (y_1, \dots, y_n)^T$ 的分量两两独立, A 是可逆矩阵。设 s_i 和 y_i 的方差都是 1, 那么改变 S 和 Y 的分量的排列次序, 可使 A 具有分块对角阵的形式 $A = \text{diag}(A_1, \dots, A_k; I)$, 其中 I 为单位矩阵, A_i 都是正交阵。并且那些对应于 A_i 所在行的 S 的分量都是高斯分布, 而对应于 I 所在行的 S 的分量可具有任意分布。

定理 1.2 和定理 1.3 都说明,要得到源向量,只需恢复分量之间的两两独立性就够了。在 ICA 的模型下重构向量分量的两两独立性与相互独立性是等价的,但验证起来显然是容易了许多,这便为算法的设计提供了方便。

1.2.3 ICA 与 PCA 的联系

按照定义 1.1, 随机向量 $\mathbf{X}$ 的 ICA 分解 $\mathbf{X} = \mathbf{FY}$ 中, $\mathbf{Y}$ 的协方差阵是对角阵。这一性质说明 ICA 分解去除了各分量间的统计相关性^①, 即 $\mathbf{Y}$ 的任意两个不同分量 y_p 和 y_q 都满足:

$$E[(y_p - E(y_p))(y_q - E(y_q))] = 0 \quad (p \neq q) \quad (1.3)$$

这些性质也是主成分分析(Principle Component Analysis, PCA)所具有的, 只不过 PCA 不要求 $\mathbf{Y}$ 的分量独立, 而是要求 $\mathbf{F}$ 的各列正交。

独立性要求联合概率密度是可分解的, 这意味着对任意 $\mathbb{R}$ 上的函数 $f(\cdot)$ 和 $g(\cdot)$ 都有

$$E[f(y_p)g(y_q)] = E[f(y_p)]E[g(y_q)] \quad (p \neq q) \quad (1.4)$$

成立(假定所涉及的期望都存在), 显然, 式(1.4)成立的条件远比式(1.3)强。仅使用去相关处理, 即主成分分解是不能实现盲源分离的: 不妨考虑 $\mathbf{S}$ 的每个分量都具有零均值、单位方差, 且相互独立; 再设 $\mathbf{U}$ 是任意正交矩阵, 那么 $\mathbf{Z} = \mathbf{US}$ 的协方差阵是 $E[\mathbf{USS}^T \mathbf{U}^T] = \mathbf{I}$, 即 $\mathbf{Z}$ 的分量总是不相关的, $\mathbf{Z} = \mathbf{IZ}$ 就是一个 PCA 分解, 但显然其任一分量都是 $\mathbf{S}$ 各分量的线性组合。

从另一角度看, 若混合矩阵是 n 阶可逆矩阵, “不相关”要求仅包含了 $n(n-1)/2$ 个约束条件, 离 n^2 个未知参数还差了很多^[9]。PCA 可以作为 ICA 过程的第一步: 标准化^[3], 即减少待定变元的个数。用 PCA 处理完二阶统计信息之后, 余下的过程是确定出一个高维空间的旋转变换, 这才是 ICA 最为本质之处。

如果源向量中含有两个或更多个高斯分量, 对任意两个高斯分量来说, 不相关就意味着独立, 这也从另一角度解释了为什么 ICA 的解要唯一就不能存在两个及两个以上的高斯分布。

1.3 独立成分分析的发展历史、扩展及应用

1.3.1 ICA 的发展历史

根据 Jutten 和 Taeb 的综述性文章^[10], 独立成分分析问题的提出可以追溯

① 沿用文献[3]的定义, 但也有其他学者持不同观点, 例如 Cardoso 认为“最可能地独立”并不蕴含不相关性(见文献[8]的 1.2 节)。但这种分歧只发生在源向量分量不具有独立性的情况下, 在此不作进一步的讨论。

到 1982 年,当时在神经生理学的研究中关于肌肉收缩运动的编码有一个简化的模型:

$$\begin{aligned}f_1(t) &= a_{11}v(t) + a_{12}p(t) \\f_2(t) &= a_{21}v(t) + a_{22}p(t)\end{aligned}$$

在这个模型中,位置(角度) $p(t)$ 和角速度 $v(t) = dp(t)/dt$ 的信息是混合在一起传导给中枢神经系统的,可以肯定的是中枢神经系统能分辨出这两种信息。这一模型正是无噪声的线性瞬时混合信号模型。此外,文献[3]和文献[10]也都提及了鲜为人知的 Bar - Ness 在通信领域对于干扰消除的研究,其方法和 ICA 相似。

第一个成功求解了 ICA 的方法是由 Héroult 和 Jutten 等人以反馈式人工神经网络实现的,简称 H - J 网络^[11]。关于这一方法的第一篇文献以法文发表于 1984 年,随后在 1985 年法国的 GRETSI 会议,1985 年在巴黎召开的 Cognitive 会议以及 1986 年在 Snowbird 召开的神经网络会议上,相关结果都有所发表(不过 1987 年以前都还没有使用 ICA 或 BSS 这个称谓),引起了一些学者的关注。但当时人们看到这一结果时,都认为是不可思议的。

稍晚些时候,Lacoume 尝试使用四阶统计量来表达独立性,也成功地实现了盲信号分离^[12]。在 80 年代,高阶统计量方法开始逐渐兴起。Donoho 的盲反卷积方法^[13]所采用的技术已经和 ICA 很接近。1989 年,第一届高阶谱分析会议召开,Comon^[4]和 Cardoso^[14]都分别提供了使用四阶累积量进行盲源分离的方法和结果。

虽然在 20 世纪 80 年代用 ICA 技术已经成功地实现了盲源分离,但它的关注度被同时期的 BP 算法、Hopfield 网络、SOM 等神经网络的热点所掩盖。20 世纪 90 年代初,关于盲信号分离和 ICA 的理论体系才建立起来,以可辨识性或可分离性的探讨为标志^[2,15]。Comon 提出了对照函数的概念,给出了 ICA 的一个数学界定。当然,那个时候关于盲源分离问题的研究已不完全是在 ICA 理论框架之内。

Bell 和 Sejnowski^[42]将信息极大化原理用于盲信号分离取得了成功之后,引起了广泛的关注。他们的算法以及之后的一些改进算法在众多的领域中被使用,使得 ICA 这一技术的影响力大为拓宽,更多的学者加入到这一领域。芬兰学者 Hyvärinen 和 Oja^[17]提出的 FastICA 算法由于其计算效率高,对 ICA 的应用和推广做出了重要贡献。1999 年,第一届 ICA 专题国际会议召开,迄今已举办了十几届。

虽然对于基本的 ICA 模型的求解到目前为止仍在不断地出现新算法,从 20 世纪 90 年代开始,ICA 的研究就早已不局限于最基本的模型了,因为它远不能

满足应用的要求。

1.3.2 ICA 的扩展研究

基本的 ICA 模型是从实际问题简化而来的,即信源数等于观测数,且没有噪声的理想状态下的线性、瞬时混合。求解 ICA 方法也就是如何确定混合矩阵。有的 ICA 解法是以逐次提取方式进行的,如文献[18,19],可以看作是最简单的扩展形式,因为实际问题可能有从高维数据中挖掘出一部分有用分量的需要。介于逐次提取和同步求解之间的则是部分提取的技术^[20]。

复值 ICA 在 BSS 研究的早期研究中就被纳入考虑了,因为窄带信号是用复值信号表示的。早期的基于累积量的算法,如 JADE^[21]已考虑了信源为复值,但许多算法最初提出都是基于实变量的,后来才被推广到处理复信号的形式,如文献[22,23]等。另外,还有所谓多维 ICA (MICA)^[24],它假设源信号向量的分量分为若干组,组与组之间是相互独立的,但是每组包含的几个变量不要求相互独立。

如果考虑到观测数据中存在噪声,基本的 ICA 模型是无法将噪声去除的,因为这个模型是不可逆的。即使在一定条件下能够求出混合矩阵,也无法采用线性的方法将测得的信号还原为源信号,所以需加降噪处理^[25]。在完全没有先验信息的情况下,要区分信号和噪声根本上是不现实的。考虑有噪声的 ICA 模型需要同时信源和噪声建模,使用最大后验估计等技术减少噪声的影响,如文献[26]。独立因子分析^[27]也属于这一类推广。

关于欠定模型的盲分离,由于信源数比观测数多,依靠基本的 ICA 手段是无法实现分离的,这种模型称为欠定的或超完备的,它有着很强的实际应用意义。如果信源分布具有某些特征,比如稀疏性,还是有方法对其进行一定程度上的分离,如文献[28,29],这种分离方法当然也不可能是线性的,这与有噪 ICA 的处理相似。更多的方法可参考综述性文献^[30]。

ICA 假设的瞬时混合这一条件在现实中常常不能满足,比较接近真实的模型是信号经过未知的线性滤波器之后再作线性混合。于是,自然地,恢复这些信号也就不能简单地用一个线性变换实现了。卷积型混合信号的盲分离以及多道盲反卷积^[31]需要将 ICA 学习的解混矩阵扩展为滤波器矩阵,仍可以使用 ICA 的独立性准则进行类似方式的学习。还可以使用瞬时傅里叶变换的方法将问题转换为多个频带上的近似瞬时混合,再用 ICA 方法处理,这方面的方法可参见文献[32]的综述。

ICA 最困难的推广是所谓“非线性 ICA”,也就是假设混合过程是源向量的非线性变换。但是如果没有先验知识或者不加特定的约束条件,这个概念太宽,

无法使用。比较多的研究是后非线性(Postnonlinear)混合模型,即混合过程仍然是线性的,仅是观测经过了非线性映射。沿着这一方向的研究相对于ICA基本模型的研究要少得多,可参见文献[10]的综述。

1.3.3 ICA 的应用

ICA 的早期研究和天线阵列信号的处理联系很紧密,特别是应用与波束形成^[21]。因此,ICA 可应用于雷达信号处理中,如文献[33]。在目前的无线网络中也有用武之地,如文献[34]。

在生物医学和脑科学的研究中,ICA 具有极为重要的应用,包括脑电信号(Electro Encephalo Graphic, EEG)^[35,36]、脑磁信号(Magneto Encephalo Graphic, MEG)^[37]、心电信号(Electro Cardio Gram, ECG)^[38],以及功能核磁共振成像(fMRI)^[39]的时间序列分析等。

ICA 也可用于金融时间序列的分析中,探索出数据背后隐藏着的某些因素^[40,41]。

在传统使用PCA等多元数据分析的场合,ICA有可能作为替代或补充,用来进行特征提取。例如用于图像的特征提取^[42],图像去噪^[43],图像的分类、分割和增强^[44]。又例如用于人脸识别^[45]、语音信号的特征提取^[46]、基因微阵列数据的分析^[47]等。

在混合声音的分离^[32]、工业生成中监控^[48]、合成孔径雷达(SAR)成像的处理^[49],数字图像水印^[50]等很多领域中都有广泛的应用(见文献[51]),此处就不再赘述了。

参考文献

- [1] Haykin S, Chen Z. The cocktail party problem[J]. Neural Computation, 2005, 17(9), 1875 - 1902.
- [2] Tong L, Inouye y, Liu R. Waveform - preserving blind estimation of multiple independent sources[J]. IEEE Trans. on Signal Processing, 1993, 41(7): 2461 - 2470.
- [3] Comon P. Independent component analysis, a new concept[J]. Signal Process, 1994, 36: 287 - 314.
- [4] Comon P. Separation of stochastic processes[C]. Proc. Workshop on Higher - Order Spectral Analysis, 1989: 174 - 179.
- [5] Eriksson J, Karvanen J. Identifiability, separability, and uniqueness of linear