

文章编号:1007-5321(2017)06-0024-06

DOI:10.13190/j.jbupt.2016-288

# 基于相似度检测的欠定混合矩阵盲估计

付卫红<sup>1,2</sup>, 周新彪<sup>1</sup>, 农 斌<sup>1</sup>, 王川川<sup>3</sup>

(1. 西安电子科技大学 通信工程学院, 西安 710071; 2. 西安中电科西科大雷达技术协同创新研究院有限公司, 西安 710071;  
3. 电子信息系统复杂电磁环境效应国家重点实验室, 河南 洛阳 471003)

**摘要:** 针对现有欠定盲分离混合矩阵估计方法中存在的估计精度低以及时间复杂度高等缺点, 提出一种基于相似度检测的欠定混合矩阵估计方法. 该方法能够在没有任何先验信息的条件下自适应地估计出源信号数目以及混合矩阵, 而且不需要进行迭代, 时间复杂度低. 仿真结果表明, 与现有的一些混合矩阵估计方法, 如改进  $K$ -均值聚类法和拉普拉斯势函数法相比, 所提出的方法在源信号数目估计准确率、混合矩阵估计精度以及时间复杂度等方面都具有明显优势.

**关键词:** 欠定盲分离; 源数目估计; 混合矩阵盲估计; 相似度检测

中图分类号: TN911.7

文献标志码: A

## Blind Estimation of Underdetermined Mixing Matrix Based on Similarity Measurement

FU Wei-hong<sup>1,2</sup>, ZHOU Xin-biao<sup>1</sup>, NONG Bin<sup>1</sup>, WANG Chuan-chuan<sup>3</sup>

(1. School of Telecommunication Engineering, Xidian University, Xi'an 710071, China;

2. Collaborative Innovation Center of Information Sensing and Understanding, Xi'an 710071, China;

3. State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, Henan Luoyang 471003, China)

**Abstract:** A new algorithm based on similarity measurement was proposed in order to address the issue of low estimation accuracy and high computational complexity in the existing algorithms for underdetermined mixing matrix blind estimation. The proposed algorithm can estimate the number of source signals and the mixing matrix automatically without any prior information or much iteration. Compared with the existing algorithms such as the  $K$ -means clustering and the Laplace's potential function method, the simulations turn out that the proposed has obvious advantages in the estimation accuracy of the number of source signals and the estimation precision of mixing matrix and the computational complexity.

**Key words:** underdetermined blind separation; blind estimation of the source number; estimation of mixing matrix; similarity measurement

盲源分离是指在不知道源信号任何先验信息以及信道混合矩阵的前提下, 仅仅根据接收到的观测信号来估计各源信号的过程. 关于盲源分离技术的研究已成为国际上信号处理领域的一个热点, 并且已经广泛应用于数字通信、图像处理、生物医学、语

音信号处理等领域. 欠定盲分离是盲源分离中的一种病态混叠情况, 由于欠定盲分离中系统是不可逆的, 所以用于求解超定或正定盲分离的传统算法如联合对角化算法<sup>[1-2]</sup>、固定点算法<sup>[3]</sup>、等变自适应分解算法<sup>[4]</sup>等不再适用于欠定盲分离.

收稿日期: 2016-12-15

基金项目: 国家自然科学基金项目(61201134); 高等学校学科创新引智计划项目(D08038)

作者简介: 付卫红(1979—), 女, 副教授, E-mail: whfu@mail.xidian.edu.cn.

对于欠定盲分离问题的求解,“二步法”是一种最常用的分离方法,即首先根据观测信号估计出混合矩阵,然后再进行源信号恢复.对于欠定混合矩阵的估计,常用的方法有 $K$ -均值聚类法<sup>[5]</sup>、势函数法<sup>[6]</sup>、模糊聚类法<sup>[7]</sup>等. Li等<sup>[8]</sup>提出了利用 $K$ -均值算法来估计混合矩阵,但算法中存在初始聚类中心难以选取和普适性不强等缺点. Sun等<sup>[9]</sup>提出了一种基于密度的空间聚类与霍夫变换相结合的混合矩阵估计算法,算法运算复杂度较高,对于存在多路接收信号时的情况实时性不强. 谭等<sup>[10]</sup>提出了一种基于模糊聚类的方法,但算法抗噪能力弱.

针对现有的欠定混合矩阵估计方法中存在的缺点,提出了一种基于相似度检测的欠定混合矩阵估计方法,不仅能够自适应地估计出源信号数目,而且能够以较高的精度计算出混合矩阵,运算复杂度低.

## 1 欠定盲分离数学模型

欠定盲源分离就是当接收信号个数小于源信号个数时,在不知道源信号先验信息以及信道混合矩阵的条件下,仅仅根据观测到的混合信号估计出源信号的过程.对于线性瞬时混合情况下的欠定盲分离,可用数学模型表示为

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

其中: $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$ 为 $n$ 维源信号矢量; $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T$ 为 $m$ 维观测信号矢量, $n > m$ ;  $\mathbf{A}$ 为 $m \times n$ 混合矩阵, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ ,  $\mathbf{a}_i (i=1, 2, \dots, n)$ 为 $\mathbf{A}$ 中的列向量,每个源信号 $s_i$ 对应一个 $\mathbf{a}_i$ ;  $\mathbf{n}(t)$ 为高斯白噪声, $t$ 为观测时刻, $t=1, 2, \dots, T$ .

目前针对欠定盲源分离问题的研究,要求源信号具有一定的稀疏性.只要源信号在某一个域如时域、频域、时频域等内是稀疏的,就能够实现各源信号的分离.以源信号在时域具有稀疏性为例,假设在时刻 $t$ ,只有源信号 $s_i$ 取值占优,其他源信号的幅度取值较小或为零,则可将式(1)写为

$$\mathbf{x}(t) = \mathbf{a}_i s_i(t) + \mathbf{n}(t), \quad t = i_1, i_2, \dots, i_p \quad (2)$$

由式(2)可以看出,在 $t = i_1, i_2, \dots, i_p$ 时刻,由于源信号的稀疏性,观测信号采样点聚集在 $\mathbf{a}_i$ 方向上,呈线聚特性.因此只要对所有时刻的观测信号采样点 $\mathbf{x}(t)$ 进行聚类分析,便可得到混合矩阵 $\mathbf{A}$ 的初步估计.若将观测信号的每一个采样数据点进行归一化处理,由于观测信号的线聚特性,同一个 $\mathbf{a}_i$ 方向上的采样数据点则会聚集在以原点为圆心的单

位球面上的某一个区域上.图1所示为观测通道数目 $m=3$ ,源信号数目 $n=5$ 时,对观测信号采样点进行归一化后的分布情况.其中,为保证一般性,源信号为Matlab随机生成的稀疏信号,混合矩阵 $\mathbf{A}$ 为3行5列的高斯随机矩阵.由图1可知,采样数据点在以原点为圆心的单位超球面上具有相应的聚类特征.

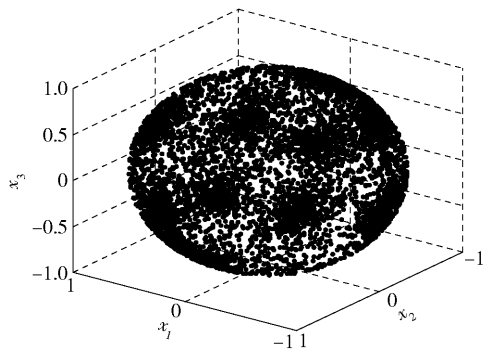


图1  $m=3, n=5$  时的观测信号采样点分布

若源信号在时域上不具有稀疏性,则可以对观测信号进行适当的线性变换,如傅里叶变换、小波变换等,使之变换到相应的稀疏域再进行盲分离.在不考虑噪声的情况下,其变换原理为

$$\mathbf{X}\psi = \mathbf{A}\mathbf{S}\psi \quad (3)$$

其中: $\mathbf{X}$ 为观测信号采样点矩阵,矩阵中的每一列表示一个采样点; $\psi$ 为相应的线性变换; $\mathbf{A}$ 为混合矩阵.根据线性变换原理,可进一步将式(3)表示为

$$\mathbf{C}_x = \mathbf{A}\mathbf{C}_s \quad (4)$$

其中: $\mathbf{C}_x$ 为观测信号在变换域内的表达形式, $\mathbf{C}_s$ 为源信号在变换域内的表达形式.由于线性变换 $\psi$ 是可逆的,所以在相应的变换域内,只要源信号是稀疏的,依然满足式(2)所示的线性瞬时混合模型,并不影响混合矩阵 $\mathbf{A}$ 的估计和最后的源信号恢复.

## 2 基于相似度检测混合矩阵估计

### 2.1 观测信号预处理

如图1所示,对于时域稀疏的源信号或者经过相应的线性变换后,在相应的变换域内呈现稀疏性的源信号,经过信道后,在得到观测信号中,由于传输环境和噪声的影响,往往存在大量的异常值以及噪声点,以至于观测信号的聚类特征不够明显.因此需要对观测信号进行预处理,包括提取出观测信号采样点中的高能量采样点和高密度采样点.

首先为了更加精确地估计混合矩阵,利用式(5)除去观测信号采样点中的低能量采样点,保留

下高能量采样点,组成一个高能量点矩阵.

$$\mathbf{X}_1 = \{\mathbf{x}(t) \mid \|\mathbf{x}(t)\|_2 \geq \alpha \max \{\|\mathbf{x}(t)\|_2\}, t=1,2,\dots,T\} \quad (5)$$

其中:  $\|\mathbf{x}(t)\|_2$  表示求取矢量  $\mathbf{x}(t)$  的二范数,  $\alpha \in [0.1, 0.5]$ ,  $\mathbf{X}_1$  中的每一列代表一个采样点. 对  $\mathbf{X}_1$  中保留下来的采样点进行归一化:

$$\mathbf{x}(t) = \frac{\mathbf{x}(t)}{\|\mathbf{x}(t)\|_2}, t=1,2,\dots,N \quad (6)$$

其中  $N$  为  $\mathbf{X}_1$  中的总列数,即经过剔除低能量采样点后保留下来的样本点数目.

其次要搜索出高能量采样点矩阵  $\mathbf{X}_1$  中的高密度采样点. 定义样本点之间的夹角为

$$\theta_{i,j} = \arccos \left( \frac{\left| \sum_{k=1}^m \mathbf{x}_i(k) \mathbf{x}_j(k) \right|}{\|\mathbf{x}_i\|_2 \|\mathbf{x}_j\|_2} \right), \quad i=1,2,\dots,N, j=1,2,\dots,N \quad (7)$$

其中:  $\theta_{i,j}$  为  $\mathbf{X}_1$  中第  $i$  个样本点和第  $j$  个样本点之间的角度,  $k$  为每个样本点矢量中的第  $k$  个元素,  $|\cdot|$  表示取绝对值. 样本点经过归一化,因此有  $\|\mathbf{x}_i\|_2 = 1, \|\mathbf{x}_j\|_2 = 1$ , 故式(7)可变为

$$\theta_{i,j} = \arccos \left( \left| \sum_{k=1}^m \mathbf{x}_i(k) \mathbf{x}_j(k) \right| \right) \quad (8)$$

在单位球面上,属于同一聚类的观测信号采样点分布相对紧密,各采样点之间的夹角较小,而不属于同一聚类的观测信号采样点之间的夹角较大,因此各观测信号采样点间的角度可作为各采样点之间相似关系的度量.

对于任意一个观测信号采样点  $p$ ,将距离其最近的  $d$  个采样点组成一个样本集合,该集合称为  $p$  的邻域  $D$ ,  $d$  为邻域  $D$  的邻域容量,其邻域半径  $r_d^p$  即为点  $p$  与这  $d$  个采样点之间距离的最大值,即

$$r_d^p = \max \{ \text{dist}(p, q) \mid q \in D \} \quad (9)$$

笔者利用对象  $p$  与  $q$  之间的角度大小来替代欧氏距离. 如果采样点  $p$  为高密度采样点,则其样本邻域内的采样点分布较为密集,即在一定的邻域容量条件下,与低密度采样点相比,其邻域半径更小. 根据此原理进行高密度采样点的提取.

在提取高密度采样点前,首先根据观测信号采样点数据集合本身的分布特性来设置一个邻域容量  $d$ ,对于  $d$  的取值不宜太大,一般不超过所求聚类中最小聚类的采样点数,  $d$  的取值一般在  $\beta N$  左右,  $N$  为高能量采样点数目,  $\beta \in [0.01, 0.05]$ .

邻域容量  $d$  确定以后,进一步求出各个采样点

的领域半径,并判断是否为高密度采样点. 通过式(8)计算初始角度矩阵  $\mathbf{G}, \mathbf{G} = \{\theta_{i,j} \mid 1 \leq i \leq N, 1 \leq j \leq N\}$ ,  $\mathbf{G}$  是一个  $N \times N$  的对称阵. 对  $\mathbf{G}$  的每一行分别进行升序排列,得到排序角度矩阵  $\mathbf{W}$ , 这样,  $\mathbf{W}$  的第  $k$  ( $k=1,2,\dots,T$ ) 列中的元素值即表示在邻域容量为  $k$  时各采样点所对应的邻域半径. 对  $\mathbf{W}$  中的第  $d$  列元素进行求均值,得到判别阈值  $\bar{r}_d$ . 由于属于同一个聚类中的采样点集聚密度较大,所以当邻域容量为  $d$  时,聚类中各采样点的邻域半径较小,而对于噪声点和异常值不属于任何一个聚类,其集聚密度一般较小,其邻域半径的值一般较大. 根据此特点,保留下邻域半径小于判别阈值  $\bar{r}_d$  的采样点,组成高密度采样点矩阵  $\mathbf{X}_2$ ,其分布情况如图2所示.

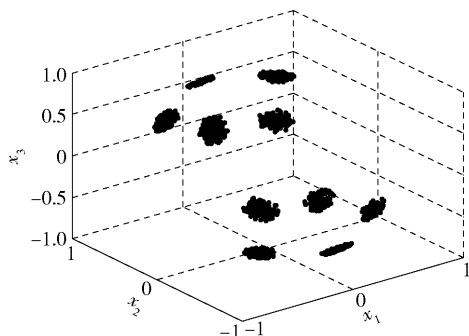


图2 经过预处理得到的高密度采样点

由于源信号  $s_i$  的取值可正可负,由式(2)可知,采样点  $\mathbf{x}(t)$  和  $-\mathbf{x}(t)$  聚集在同一个  $\mathbf{a}_i$  方向上,经归一化后,同一个源信号所对应的采样点在单位超球面上关于原点对称,形成2个聚类,在进行混合矩阵估计时,这样的2个聚类会被合并为一个聚类,即这2个聚类对应的是同一个源信号  $s_i$ . 因此,在图2中采样数据点聚集的类数是源信号个数的2倍.

## 2.2 源信号数目估计和混合矩阵估计

经过预处理后得到的高密度采样点,其聚类特性变得更加明显,而且在高密度采样点矩阵  $\mathbf{X}_2$  中,当邻域容量为  $d$  时,每个高密度采样点的邻域半径  $r_d^k$  ( $k=1,2,\dots,M$ ) 总是小于判别阈值  $\bar{r}_d$ , 其中  $M$  为高密度矩阵  $\mathbf{X}_2$  中观测信号采样点的个数. 而且,  $d$  的取值较小,因此,通常情况下各高密度采样点与其邻域范围内的其他采样点会存在于同一个聚类之中. 利用式(8)重新构建一个用以衡量高密度矩阵  $\mathbf{X}_2$  中各采样点之间相似程度的角度矩阵:

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_{11} & \hat{\theta}_{12} & \cdots & \hat{\theta}_{1M} \\ \hat{\theta}_{21} & \hat{\theta}_{22} & \cdots & \hat{\theta}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\theta}_{M1} & \hat{\theta}_{M2} & \cdots & \hat{\theta}_{MM} \end{pmatrix} \quad (10)$$

其中  $\hat{\theta}_{i,j}$  为高密度矩阵  $\mathbf{X}_2$  中第  $i$  个采样点和第  $j$  个采样点之间的角度. 若  $\hat{\theta}_{i,j}$  小于判别阈值  $\bar{r}_d$ , 则判定第  $i$  个采样点和第  $j$  个采样点归属于同一个聚类.

**定义 1**  $\lambda$ -截矩阵: 对于一个给定的矩阵  $\mathbf{R}$ , 将矩阵中小于  $\lambda$  的元素置为 0, 大于  $\lambda$  的元素置为 1, 得到一个新的矩阵  $\tilde{\mathbf{R}}, \tilde{\mathbf{R}}$  称作矩阵  $\mathbf{R}$  的  $\lambda$ -截矩阵.

根据角度矩阵  $\hat{\boldsymbol{\theta}}$  生成一个  $\lambda$ -截矩阵  $\tilde{\boldsymbol{\theta}}$ , 其中使  $\lambda = \bar{r}_d$ . 在截矩阵中, 矩阵的每一行(列)分别表示在高密度采样点矩阵  $\mathbf{X}_2$  中各个观测信号采样点与其他采样点之间的角度关系. 若在第  $i$  行中,  $\tilde{\theta}_{i,j} = 1$ , 则表示高密度采样点矩阵  $\mathbf{X}_2$  中第  $i$  个采样点和第  $j$  个采样点之间的角度小于判别阈值  $\bar{r}_d$ , 即第  $j$  个采样点在采样点  $i$  的邻域范围内, 此时将第  $i$  个采样点和第  $j$  个采样点划分到同一个聚类之中.

利用得到的  $\lambda$ -截矩阵, 对高密度采样点矩阵  $\mathbf{X}_2$  中的采样点进行聚类, 得到初步的分类  $\mathbf{Y} = \{\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_K\}$ , 其中  $\tilde{\mathbf{X}}_i (i=1, 2, \dots, K)$  表示分类后得到的第  $i$  个聚类,  $K$  为聚类数目,  $K$  即为初步估计得到的源信号数目.

考虑到实际传输环境中信道参数和噪声的影响, 可能会存在混合矩阵中某 2 列相似度较高的情况, 致使这 2 列对应的 2 个聚类比较接近, 甚至存在交叠的部分, 从而错误地形成一个大的聚类. 因此在初步得到的聚类  $\mathbf{Y} = \{\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_K\}$  中, 依然有可能存在聚类数目偏少的情况. 基于此, 方法中对  $\mathbf{Y}$  中的每一个聚类做了进一步的评估和校正, 评估校正方法如下.

对于得到的分类  $\mathbf{Y} = \{\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_K\}$ , 计算各分类  $\tilde{\mathbf{X}}_i (i=1, 2, \dots, K)$  中样本点间的最大夹角  $\theta_{\max}$ , 对于  $\theta_{\max} \leq \bar{r}_d + \delta$  ( $\delta$  为常数) 的聚类保持不变, 而对于  $\theta_{\max} > \bar{r}_d + \delta$  的聚类  $\tilde{\mathbf{X}}_i$ , 利用模糊  $C$ -均值聚类 (FCM, fuzzy  $C$ -means) 方法重新分为 2 类, 其中  $\bar{r}_d$  为判别阈值. 由于判别阈值  $\bar{r}_d$  是在 2.1 节中根据观测信号采样点自身的分布特点得到的, 能够随着信噪比 (SNR, signal noise ratio) 的变化而自适应地改变, 能够很好地表征观测信号采样点的分布离散程度. 而对于参数  $\delta$  的选取, 为保证普适性, 在源信号由 Matlab 随机生成, 混合矩阵  $\mathbf{A}$  为高斯随机矩阵的情况下, 在不同信噪比下进行多次实验, 验证不同的  $\delta$  值对源信号数目估计准确率的影响, 从而选取最

佳的  $\delta$  值. 其实验结果如表 1 所示, 为兼顾不同信噪比的情况,  $\delta$  的取值在  $[12, 18]$  之间最为合适.

表 1  $\delta$  的取值与源信号数目估计准确率的关系

SNR/ dB	$\delta / (^{\circ})$							
	3	6	9	12	15	18	21	24
10	18	18.5	22.5	50	79	85.5	79.5	73.5
12	89	87.5	89.5	92	92.5	88	75.5	73
14	92.5	93.5	95.5	97	96.5	92	82.5	70
16	95.5	97.5	96.5	98	94.5	90.5	87	85
18	97.5	98	99.5	96.5	96.5	93.5	93.5	94
20	96	99.5	99.5	99	99.5	98.5	98	98.5

经过校正后, 得到一组更新后的分类  $\check{\mathbf{Y}} = \{\check{\mathbf{X}}_1, \check{\mathbf{X}}_2, \dots, \check{\mathbf{X}}_{K'}\}$ , 其中  $K'$  为进行重新分类后得到的聚类个数,  $\check{\mathbf{X}}_j (j=1, 2, \dots, K')$  为更新分类后得到的第  $j$  个聚类. 各聚类的聚类中心表示为

$$\mathbf{a}_j = \frac{1}{B_j} \sum \check{\mathbf{X}}_j, j=1, 2, \dots, K' \quad (11)$$

其中  $B_j$  为分类  $\check{\mathbf{X}}_j$  中采样点的个数. 最后得到混合矩阵的估计  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{K'}]$ .

综上所述, 基于相似度检测的欠定混合矩阵估计算法的主要实现流程如下.

**步骤 1** 对接收到的观测信号利用式 (5) 提取出高能量采样点, 并归一化, 参数  $\alpha = 0.1$ .

**步骤 2** 提取高密度采样点.

1) 计算初始角度矩阵  $\mathbf{G} = \arccos(|\mathbf{X}_1^T \mathbf{X}_1|)$ ;

2) 计算排序角度矩阵  $\mathbf{W} = \text{sort}(\mathbf{G})$ ,  $\text{sort}(\cdot)$  表示对矩阵的每一行进行升序排序;

3) 设置邻域容量  $d = \beta N$ , 参数  $\beta = 0.03$ ;

4) 计算判别阈值  $\bar{r}_d = E(\mathbf{W}(:, d))$ ,  $\mathbf{W}(:, d)$  表示矩阵  $\mathbf{W}$  的第  $d$  列元素,  $E(\cdot)$  表示求均值;

5) 保留下邻域半径小于判别阈值  $\bar{r}_d$  的采样点, 组成高密度采样点矩阵  $\mathbf{X}_2$ .

**步骤 3** 进行初步的聚类划分.

1) 构建角度矩阵  $\hat{\boldsymbol{\theta}} = \arccos(|\mathbf{X}_2^T \mathbf{X}_2|)$ ;

2) 根据  $\hat{\boldsymbol{\theta}}$  得到  $\lambda$ -截矩阵,  $\lambda = \bar{r}_d$ ;

3) 根据  $\lambda$ -截矩阵和高密度采样点矩阵  $\mathbf{X}_2$ , 得到初步的聚类划分  $\mathbf{Y} = \{\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \dots, \tilde{\mathbf{X}}_K\}$ .

**步骤 4** 对各聚类进行评估和校正.

1) 根据角度矩阵, 找到各分类  $\tilde{\mathbf{X}}_i (i=1, 2, \dots, K)$  中样本点间的最大夹角  $\theta_{\max}$ ;



2) 若  $\theta_{\max} \leq \bar{r}_d + \delta$ , 聚类保持不变, 否则利用 FCM 方法将  $\tilde{X}_i$  重新划分为 2 类, 参数  $\delta = 15$ .

**步骤 5** 根据校正后的聚类  $\tilde{Y} = \{\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_{K'}\}$  计算聚类中心  $a_j = \frac{1}{B_j} \sum \tilde{X}_j (j = 1, 2, \dots, K')$ , 得到混合矩阵的估计  $A = [a_1, a_2, \dots, a_{K'}]$ .

### 3 算法性能分析和时间复杂度分析

在分析所提算法的时间复杂度时, 以迭代次数和每次迭代中的乘法次数作为计算复杂度评价准则, 并与改进  $K$ -均值聚类法<sup>[5]</sup>和拉普拉斯势函数法 (LMMPF, Lapulace mixed model potential function)<sup>[6]</sup>进行了比较.

在文献[5]的改进  $K$ -均值方法中, 需要事先进行采样点预处理, 并给定一组聚类中心数目  $[\varepsilon_{\min}, \varepsilon_{\max}]$ , 在不同聚类数目的情况下, 计算初始聚类中心, 然后再根据初始聚类中心进行  $K$ -均值聚类, 并计算出聚类结果的戴维森堡丁指数 (DBI, Davies-Bouldin index), 最后通过不同聚类数目下 DBI 指标值的大小来确定最优的聚类数目. 其总的计算复杂度为

$$P_{K\text{-means}} = O((\varepsilon_{\max} - \varepsilon_{\min} + 1)(\alpha m^2 T + (\tau_0 + 1)IT_0)) \quad (12)$$

其中:  $\alpha$  为常数,  $m$  为接收传感器个数,  $T$  为采样点总个数,  $I$  为设定的聚类数目,  $\tau_0$  为  $K$ -均值迭代次数,  $T_0$  为预处理后的采样点个数. 文献[6]中的 LMMPF 法需要预先估计 2 个参数, 并将所有的观测数据作为初始聚类中心处理, 其计算复杂度为

$$P_{\text{LMMPF}} = O(T + a\tau_1 T^2 + \tau_2 r T^2) \quad (13)$$

其中:  $a$  为常数,  $\tau_1$  为参数估计时需要的迭代次数,  $\tau_2$  为计算聚类中心时需要的迭代次数,  $r$  为算法中计算出的参数值.

对于所提出的方法, 其中观测信号预处理和提取高密度采样点的复杂度为

$$P_1 = O(mT + m^2 T_2) \quad (14)$$

其中:  $m$  为接收传感器个数,  $T_2$  为高能量采点的总数目. 而在源信号数目及混合矩阵估计步骤中, 复杂度为

$$P_2 = O(m^2 M + bM + \tau_3 L) \quad (15)$$

其中:  $M$  为高密度采样点的总数目,  $b$  为常数,  $L$  为需要校正的聚类中的采样点数目,  $\tau_3$  为 FCM 重新划分聚类时需要的迭代次数. 若聚类无需进行校正时,

则  $\tau_3 L = 0$ . 由于  $L < M < T_2 < T$ , 因此所提方法的总计算复杂度为

$$P_3 = O(mT + m^2 T_2 + m^2 M + bM + \tau_3 L) < O((2m^2 + m + \beta + \tau_3)T) \quad (16)$$

由上述分析可知, LMMPF 法的复杂度要远远高于其他 2 种方法, 其复杂度的量级为  $O(T^2)$ , 而改进  $K$ -均值方法与所提方法复杂度的量级均为  $O(T)$ .

### 4 算法仿真和结果分析

下面将通过仿真来测试所提出算法在不同实验条件下的混合矩阵估计精度和运行时间, 并与现有的改进  $K$ -均值算法和 LMMPF 法进行比较, 说明该算法的优势. 采用估计误差衡量各个算法的混合矩阵估计精度, 估计误差的定义为

$$e = \frac{\|A - \hat{A}\|_F}{\|A\|_F} \quad (17)$$

其中:  $A$  为真实的混合矩阵,  $\hat{A}$  为已经消除模糊性估计得到的混合矩阵,  $\|\cdot\|_F$  表示求 Frobenius 范数. 若矩阵  $\hat{A}$  中的列向量个数多于或少于真实混合矩阵中的列向量个数时, 则仅对混合矩阵  $\hat{A}$  中估计正确的列向量进行误差评估.

在仿真实验中利用 Matlab 产生随机的 5 路时域稀疏信号作为源信号, 接收传感器数目  $m = 3$ , 源信号数学模型为

$$S = \text{randn}(n, T) \odot \max(0, \text{sign}(\text{rand}(n, T) - v)) \quad (18)$$

其中:  $S$  为源信号矩阵, 矩阵中的每一行对应一路源信号;  $\text{randn}(\cdot)$  用以生成服从标准正态分布的随机数; “ $\odot$ ”表示求点积;  $\text{rand}(\cdot)$  用以生成服从均匀分布的随机数;  $\text{sign}(\cdot)$  为符号函数;  $T$  为采样数据长度; 参数  $v$  的取值在  $[0, 1]$  之间,  $v$  的取值越大, 信号的稀疏度越高, 反之越小, 实验中取  $v = 0.9$ . 源数目  $n = 5$ , 信号采样点数  $T = 5\,000$ . 信噪比在  $6 \sim 22$  dB 之间, 每个信噪比下仿真 400 次. 为验证各个算法的普适性, 每次仿真中所使用的高斯混合矩阵都重新随机生成, 对应于不同的信道状态.

图 3 所示为不同 SNR 下对各算法的源信号数目估计准确率的比较, 与改进  $K$ -均值算法和 LMMPF 相比, 所提出的算法具有更高的源信号估计准确率.

图 4 所示为各算法的混合矩阵估计偏差比较, 可见, 在不同 SNR 下, 使用所提出的方法得到的混

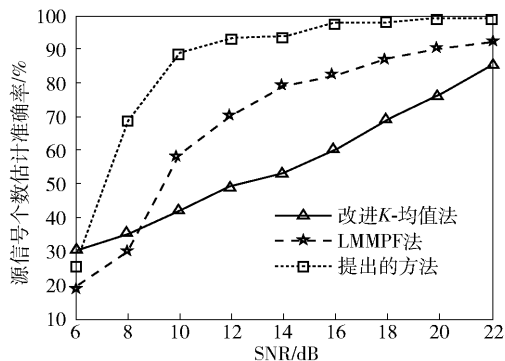


图 3 各算法源信号数估计准确度比较

合矩阵的估计误差最小。综合图 3 和图 4 的结果可知,所提出的欠定混合矩阵估计方法在源信号数目估计以及矩阵估计精度方面明显优于其他 2 种方法,在对应不同的信道状态时,能够保持较好的普适性。

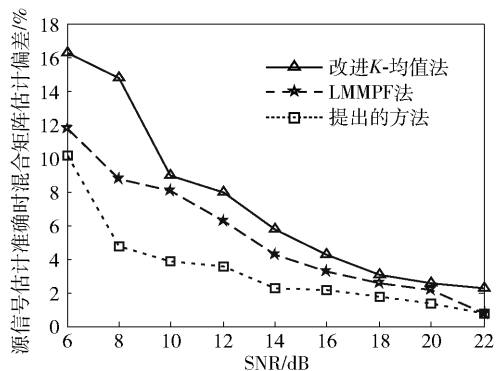


图 4 各算法的混合矩阵估计偏差比较

图 5 给出了在不同信噪比下各算法运行一次所需时间的比较结果,图中纵坐标采用以 10 为底的对数坐标来表示。由图 5 可知,在不同信噪比下,LMMPF 运行一次所需时间明显要长得多,时间复杂度最高,其次是改进 K-均值聚类方法,所提出的方法的时间复杂度最低。

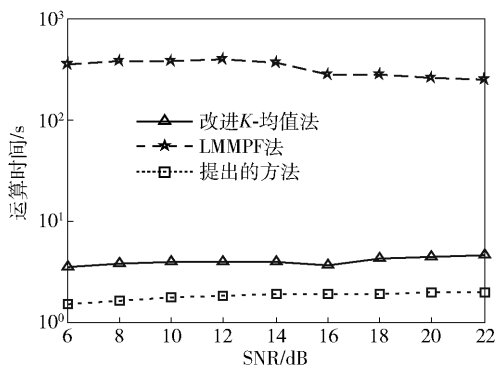


图 5 各算法运行一次所需要的时间比较

## 5 结束语

针对现有的欠定混合矩阵估计方法中存在的估计精度低以及时间复杂度高缺点,提出了一种基于相似度检测的欠定混合矩阵估计方法。与现有的欠定混合矩阵估计算法相比,所提方法在矩阵估计精度和时间复杂度方面都具有明显的优势,实验仿真结果也进一步验证了所提算法具有较好的普适性,在非理想的通信环境中依然适用。

## 参考文献:

- [1] Cardoso J F. Source separation using higher order moments[C]//1989 International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE, 1989: 2109-2112.
- [2] 冯大政, 保铮, 张贤达. 信号盲分离问题多阶段分解算法[J]. 自然科学进展, 2002, 12(3): 324-328.  
Feng Dazheng, Bao Zheng, Zhang Xianda, et al. Multi-stage decomposition algorithm for blind source separation [J]. Progress in Natural Science: Materials International, 2002, 12(5): 378-382.
- [3] Learned-Miller E G, Iii J W F. ICA using spacings estimates of entropy [J]. Journal of Machine Learning Research, 2003, 4(4): 1271-1295.
- [4] Cardoso J F, Laheld B H. Equivariant adaptive source separation[J]. IEEE Transactions on Signal Processing, 1996, 44(12): 3017-3030.
- [5] Fu Weihong, Ma Lifan, Li Aili. Blind estimation of underdetermined mixing matrix based on improved K-means clustering[J]. Journal of Systems Engineering and Electronics, 2014, 36(11): 2143-2148.
- [6] Zhang Ye, Fang Yong. A new method to estimate the number of the sources for underdetermined blind separation based on Lapulacial potential function [J]. Signal Processing, 2009, 25(11): 1719-1725.
- [7] Li Yibing, Nie Wei, Ye Fang, et al. A mixing matrix estimation algorithm for underdetermined blind source separation [J]. Circuits, Systems, and Signal Processing, 2016, 35(9): 3367-3379.
- [8] Li Yuanqing, Cichocki A, Amari S. Analysis of sparse representation and blind source separation [J]. Neural Computation, 2004, 16(6): 1193-1234.
- [9] Sun Jiedi, Li Yuxia, Wen Jiangtao, et al. Novel mixing matrix estimation approach in underdetermined blind source separation [J]. Neurocomputing, 2015, 173 (P3): 623-632.
- [10] 谭北海, 杨祖元, 周郭许, 等. 欠定盲分离中源的个数估计和分离算法[J]. 中国科学: 信息科学, 2009, 39(3): 349-356.  
Tan Beihai, Yang Zuyuan, Zhou Guoxu, et al. Source number estimation and separation algorithms of underdetermined blind separation [J]. Scientia Sinica Informationis, 2009, 39(3): 349-356.