

欠定情形下语音信号盲分离的时域检索平均法

肖 明^{1,2} 谢胜利^{1*} 傅予力

(1. 华南理工大学电子与信息学院, 广州 510640; 2. 茂名学院电子信息工程系, 茂名 525000)

摘要 探讨欠定情形(即观察信号的个数少于源信号个数)下的盲信号分离, 提出一种新的方法, 即时域检索平均法(SAMTD), 该方法可解决目前在频域中难以处理的一类问题, 它利用一类语音信号在时域中的稀疏持续性, 回避像 K 均值聚类或势函数等常用统计聚类方法. 为估计混叠矩阵, 它剔除那些不与基矢量共向或反向的数据样本, 以提高其估计的准确性. 在源信号的恢复上, 提出了一个超完备线性几何 ICA 改进算法. 几个语音信号实验的仿真数据展示了所提方法的性能和实用性.

关键词 欠定盲信号分离 稀疏表示 时域检索平均法 超完备独立分量分析

采用独立分量分析(independent component analysis, ICA)的盲信号分离(blind signal separation, BSS)算法大多要求观测信号的个数 m 不少于源信号个数 n ^[1~3], 而 $m < n$ 的情形通常称为“病态”或“欠定”, 目前对欠定盲信号分离的研究是实时线性混叠的研究难点^[3]. 近年来, 不少学者利用信号的稀疏表示^[4~10], 采用两步法^[9], 先估计混叠矩阵 A , 然后估计源信号, 解决了用 ICA 方法无法解决的问题.

当信号在分析域中是稀疏的, 两个观测信号的散落图就有明显的直线趋势. 以 6 个长笛音符的按空间等分角度线性混叠为例, 两个混叠信号为 $x_1(t)$ 与 $x_2(t)$, 在时域的散落图(图 1(a))中没有直线趋势; 观测信号经过离散 Fourier 变换(DFT)后, 在其实部和虚部的散落图(图 1(b)和图 1(c))中有直线出现, 故信号在频域中表现出稀疏性, 而时域中不存在稀疏性, 所以通常利用频域稀疏表示^[5,6], 采用 K 均值聚类^[8]或势函数^[9]等统计聚类方法估计混叠矩阵.

在欠定的盲信号分离中, 选择的分析域应该是稀疏的. 但有时频域中稀疏性比时域中的要差, 如下面 3 个语音(u2.wav|u3.wav|u5.wav, 两个男音和一个女音的讲话声, 其波形见仿真实验部分)的混叠, 从两个观测信号在时域和频域的散落图知, 应该选择时域作为分析域(参见图 2).

收稿日期: 2005-05-18; 接受日期: 2007-07-11

国家自然科学基金(批准号: U0635001, 60505005, 60674033)、广东省自然科学基金(04205783, 05006508)和科技部重大基础前期研究专项(2005CCA04100)资助项目

* 联系人, E-mail: adshlxie@scut.edu.cn

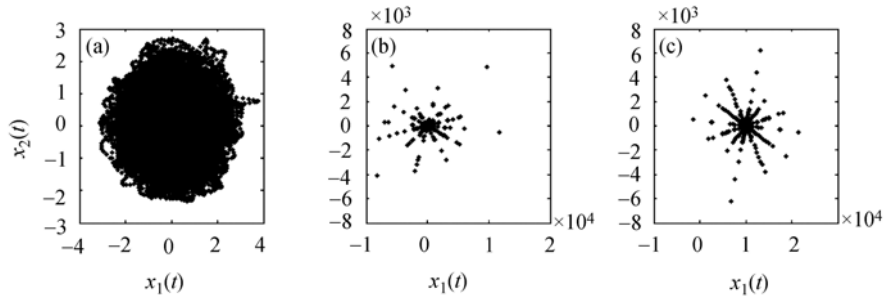


图 1 六个长笛音按空间相等角度混叠成两个混叠信号

(a)在时域; (b)频域实部; (c)虚部的散落图

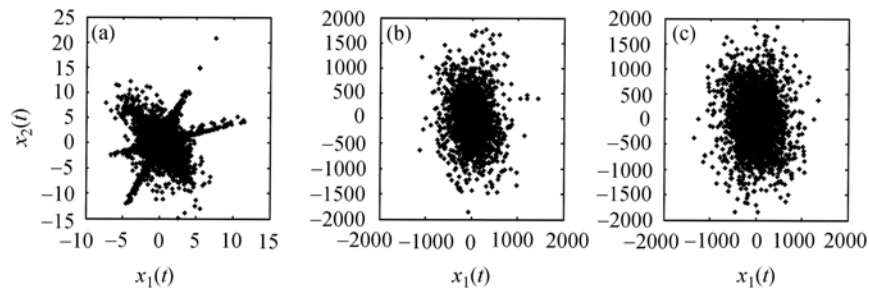


图 2 3 个讲话声(2 男 1 女)混叠成两个混叠信号

(a) 时域; (b) 频域实部; (c) 虚部的散落图

在时域中, 仍可发现其直线的趋势较模糊, 其远离直线周边点将严重干扰统计聚类, 从而使得估计的混叠矩阵出现很大的偏差, 其准确性大大降低. 另外, 在聚类估计混叠矩阵中, 到底应该聚多少类? 即源信号个数的确定也是一大难题, 此外, 对全部样本聚类计算量也太大.

鉴于上述问题, 我们提出了一种新方法, 即时域检索平均法 (searching-and-averaging method in time domain, SAMTD). 该方法利用一类语音信号在时域中的稀疏持续性, 剔除那些不与基矢量共向或反向的样本, 使其计算量大大减少, 回避传统的统计聚类算法.

在源信号的恢复上, Bofill 提出了最短路径的几何算法^[9]; 后来, Theis 证明了最短路径算法等价于 l^1 范数准则^[11], 并提出超完备线性几何 ICA 算法. 本文旨在简化 Theis 的超完备线性几何 ICA 算法, 提出了一个超完备线性几何 ICA 的相应改进算法. 最后, 用几个语音信号实验仿真来证实上述算法的性能.

1 混叠矩阵的估计

信号在欠定情形下的线性混叠模型

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

其中, \mathbf{A} 和 \mathbf{s}^t 均未知, $\mathbf{x}^t = (x_1^t, \dots, x_m^t)^T$ 是在时间 t 的 m 个观测信号; $\mathbf{s}^t = (s_1^t, \dots, s_n^t)^T$ 是 n 个源信号, 符号 $(\cdot)^T$ 表示矩阵或矢量的转置; $m < n$, 即观测信号少于源信号. 又令 $\mathbf{X} = (\mathbf{x}^1, \dots, \mathbf{x}^T) \in \mathbb{R}^{m \times T}$ 为观测信号矩阵, 混叠矩阵可展开为 $\mathbf{A} = (\mathbf{a}^1, \dots, \mathbf{a}^n) \in \mathbb{R}^{m \times n}$, 其中矢量 $\mathbf{a}^j = (a_{1j}, \dots, a_{mj})^T$ 为基矢量, $\|\mathbf{a}^j\| = (a_{1j}^2 + \dots + a_{mj}^2)^{1/2} = 1$, $\|\cdot\|$ 表示矢量的模. 经过坐标原点且沿矢量 \mathbf{a}^j 方

向的直线称为 \mathbf{a}^j 的基直线.

为了利用信号的稀疏表示, 将(1)式转化为

$$\mathbf{x}^t = \sum_{j=1}^n \mathbf{a}^j s_j^t, \quad t=1, 2, \dots, T. \quad (2)$$

在混叠系统中, 稀疏性一般是指在某些时刻 t_0 上仅仅有某一个源信号的幅值起主要作用, 取非 0 值, 而其他源信号的幅值相对很小或接近于 0, 则有

$$\mathbf{x}^{t_0} \approx \mathbf{a}^j s_j^{t_0} \quad (3)$$

成立. 于是, 矢量 \mathbf{x}^{t_0} 与矢量 \mathbf{a}^j 几乎是在同一直线上.

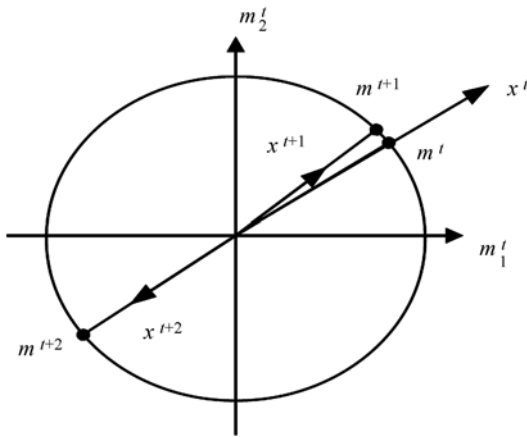


图 3 位于同一基直线附近矢量的 \mathbf{x}^t 和其投影 \mathbf{m}^t

如果源信号在频域中不存在稀疏性, 而在时域中的稀疏性也不强, 则盲分离问题就更加困难, 它不能采用通常的处理方法, 而需另寻途径.

首先, 对矩阵 \mathbf{X} 的每一列矢量 \mathbf{x}^t 进行归一化, 记

$$\mathbf{m}^t = \mathbf{x}^t / \|\mathbf{x}^t\|. \quad (4)$$

令 $\mathbf{M} = (\mathbf{m}^1, \dots, \mathbf{m}^T)$, 则 \mathbf{M} 为 \mathbf{X} 列归一化矩阵, 相当于 \mathbf{X} 的每一列矢量 \mathbf{x}^t 在 m 维的单位超球面上的投影, 当 $m=2$ 时, 是在单位圆上的投影(如图 3).

其次, 对矩阵 \mathbf{M} 的相邻两个列矢量 \mathbf{m}^t

和 \mathbf{m}^{t+1} 作如下运算:

$$\mathbf{d}^t = \begin{cases} \mathbf{m}^t + \mathbf{m}^{t+1}, & \text{当 } \mathbf{m}^t \neq 0 \text{ 时, 若 } m_k^t m_k^{t+1} \leq 0, k = \arg \max_i |m_i^t|; \\ \mathbf{m}^t - \mathbf{m}^{t+1}, & \text{其他,} \end{cases} \quad (5)$$

其中, i 为 $1, \dots, m$ 中的值, 然后设 $\mathbf{D} = (\mathbf{d}^1, \dots, \mathbf{d}^T)$.

当相邻两个列矢量 \mathbf{x}^t 和 \mathbf{x}^{t+1} 几乎位于同一直线时, 矩阵 \mathbf{D} 中将出现近似于 0 的列. 例如, 在图 3 中, 当 \mathbf{m}^t 和 \mathbf{m}^{t+1} 在单位超球面上靠得很近, 而矢量 \mathbf{m}^{t+1} 和 \mathbf{m}^{t+2} 也几乎位于同一直线时, 由(5)式得到 $\mathbf{d}^t \approx 0$ 和 $\mathbf{d}^{t+1} \approx 0$, 所以矩阵 \mathbf{D} 将呈现如下形式:

$$\mathbf{D} = \begin{bmatrix} \cdots & \times & 0 & 0 & \cdots & 0 & \times & \times & \cdots & \times & 0 & \cdots & 0 & \times & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \times & 0 & 0 & \cdots & 0 & \times & \times & \cdots & \times & 0 & \cdots & 0 & \times & \cdots \end{bmatrix}_{m \times T}.$$

如果矩阵 \mathbf{D} 连续出现 5 列近似等于 0, 则在 \mathbf{M} 和 \mathbf{X} 中相应的 5 列和邻近其后的 1 列的矢量几乎位于同一直线上, 而这可能正是欲寻找的基直线.

构造矩阵 \mathbf{D} 的目的: 检索矩阵 \mathbf{X} 中位于同一直线的相邻列矢量, 来估计混叠矩阵.

在离散时间的语音信号(特别是讲话声)处理中, 经实验分析表明, 稀疏性有一个持续的时段, 而不会仅在孤立的样本上是稀疏的, 可能是连续的 3 个样本, 也可能是连续的 5 个样本,

或者更长的时间段, 这是一种稀疏的持续性.

如果稀疏持续的样本仅有 2 个, 则没有实际意义, 故稀疏持续性可定义如下.

定义 1 在一些连续的 3 个样本以上的小时间段 Δt 上, 仅有一个源信号非 0 而其余的源信号均为 0, 或者一个源信号起主要作用而其他信号值相对很小, 而在这些时间段 Δt 上, (3) 式成立, 即 $\mathbf{x}^t \approx \mathbf{a}^j s_j^t$, 这种性质称为稀疏持续性. 此时, 这些时间段 Δt 为单源区间. 而不具有稀疏持续性的时间段称多源区间.

在单源区间上的列矢量 \mathbf{x}^t 应该近似地位于同一直线上, 该直线通常就是基直线, 在矩阵 \mathbf{D} 中, 由(5)式很容易发现: 在单源区间上, 矩阵 \mathbf{D} 的列矢量都近似为 0; 而在多源区间上, 矩阵 \mathbf{D} 的列矢量为非 0. 根据这一特征, 很容易从矩阵 \mathbf{D} 中分出所有单源区间和多源区间.

给定正整数 Δ , 在矩阵 \mathbf{D} 中, 如果 $t = t_0, t_0 + 1, \dots, t_1$ 且 $t_1 - t_0 + 1 \geq \Delta$, 有 $\mathbf{d}^t = \mathbf{0}$, 当 $t = t_0 - 1, t_1 + 1$, 且 $\mathbf{d}^t \neq \mathbf{0}$ 时, 则时间段 $[t_0, t_1 + 1]$ 视为单源区间; 其他非单源区间的时间段都视为多源区间.

一般情况, Δ 的取值为 3~5 比较合适, Δ 表示单源区间的最小时间长度.

在语音信号的盲分离中, 矩阵 \mathbf{D} 的列矢量 \mathbf{d}^t 完全恒等于 0 几乎没有, 仅仅存在一些接近 0 的 \mathbf{d}^t , 所以单源区间和多源区间应如下检索.

给定的检索误差 $\varepsilon > 0$ 和正整数 Δ , 如果 $t = t_0, t_0 + 1, \dots, t_1$ 且 $t_1 - t_0 + 1 \geq \Delta$, 有 $|\mathbf{d}_1^t| + \dots + |\mathbf{d}_m^t| \leq \varepsilon$, 且当 $t = t_0 - 1, t_1 + 1$ 时, $|\mathbf{d}_1^t| + \dots + |\mathbf{d}_m^t| > \varepsilon$, 则时间段 $[t_0, t_1 + 1]$ 视为单源区间; 其他非单源区间的时间段都视为多源区间.

为了检索所有的单源区间, 设计一个 $1 \times T$ 的行矢量 $\boldsymbol{\Omega}$ 作为确定单源区间的检索矩阵, 设 $\boldsymbol{\Omega} = (\omega^1, \dots, \omega^T)$, $\boldsymbol{\Omega}$ 中的值 ω^t 取值:

$$\omega^t = \begin{cases} 1, & t \text{ 位于单源区间;} \\ 0, & t \text{ 位于多源区间;} \end{cases}$$

即在 $\boldsymbol{\Omega}$ 中与所有单源区间对应的列置 1, 而与多源区间的对应的列置 0.

因为在单源区间上, $\mathbf{x}^t \approx \mathbf{a}^j s_j^t$ 成立, 所以单源区间的数据有益于混叠矩阵的估计, 而多源区间的信号数据干扰混叠矩阵的估计, 造成误差, 甚至淹没基直线而不能聚类, 所以要对其剔除.

当检索矩阵 $\boldsymbol{\Omega}$ 与矩阵 \mathbf{X} 的每一行的对应元素相乘得到矩阵 \mathbf{X}_0 时, 多源区间的数据变为 0, 剔除了所有多源区间的数据, 剩余的仅仅为单源区间的数据, 称矩阵 \mathbf{X}_0 为单源区间数据矩阵. 当检索矩阵 $\boldsymbol{\Omega}$ 与矩阵 \mathbf{M} 的每一行的对应元素相乘得到 $\boldsymbol{\Theta}$, 设 $\boldsymbol{\Theta} = (\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^T)$, 其中 $\boldsymbol{\theta}^t = (\theta_1^t, \dots, \theta_m^t)^T$. 在 $\boldsymbol{\Theta}$ 中, 不在基直线与单位超球面的交点附近的数据已经被置 0.

利用以上方法, 对图 2(a) 中的两个混叠信号进行时域检索, 剔除干扰基直线的多源区间数据后, 3 条模糊直线已经非常清楚地显露出来, 见图 4.

因为矩阵 $\boldsymbol{\Theta}$ 的非 0 列为单源区间样本, 分别位于一些直线附近, 所以必须聚类属于每条直线的样本. 给定偏差 $\delta > 0$, 设其中任意两个时刻 t_i 和 t_j , 则 $\boldsymbol{\theta}^{t_i}$ 和 $\boldsymbol{\theta}^{t_j}$ 为同一直线的相近矢量的判别条件为

$$|\theta_1^{t_i} - \theta_1^{t_j}| + \dots + |\theta_m^{t_i} - \theta_m^{t_j}| < \delta \text{ 或 } |\theta_1^{t_i} + \theta_1^{t_j}| + \dots + |\theta_m^{t_i} + \theta_m^{t_j}| < \delta. \quad (6)$$

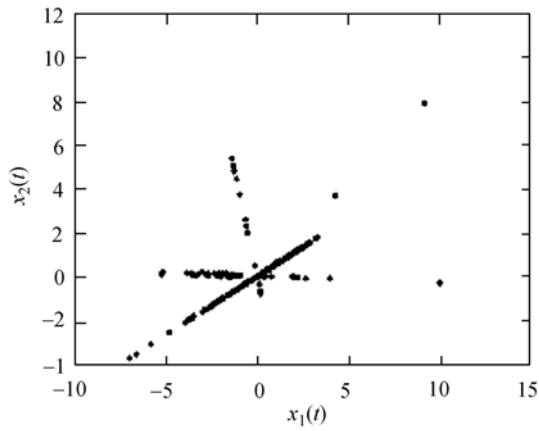


图 4 单源区间数据矩阵 \mathbf{X}_0 中 $x_2(t)$ 对 $x_1(t)$ 的散落图

的可能性非常小, 所以可以设置一个最低样本数 T_0 来删除伪基直线.

于是, 估计基直线的列矢量越多, 正确的可能性越大, 当 $k > T_0$ 时,

$$\hat{\mathbf{a}}^j = \frac{1}{k} \sum_{i=1}^k (-1)^{f_i} \theta^{t_i}, \quad (7)$$

其中,

$$\begin{cases} f_i = f_1 = 0, & \text{当 } |\theta_1^{t_i} - \theta_1^{t_1}| + \dots + |\theta_m^{t_i} - \theta_m^{t_1}| < \delta \text{ 时;} \\ f_i = 1, & \text{当 } |\theta_1^{t_i} + \theta_1^{t_1}| + \dots + |\theta_m^{t_i} + \theta_m^{t_1}| < \delta \text{ 时,} \end{cases}$$

若 $k \leq T_0$, 则该组数据点所构成的基直线是伪基直线. 依此逐个检索出所有表示基直线的列矢量确定 $\hat{\mathbf{a}}^j$; 上述方法称为时域检索平均法.

为研究源信号在混叠中的稀疏与基矢量估计的关系, 我们作如下定义.

定义 2 设 T_j 为在 Θ 中的与 \mathbf{a}^j 的基直线靠近的样本总数目, T_j 表明了第 j 个源信号的稀疏持续性的样本数目, 故称 T_j 为第 j 个源信号的稀疏持续数; 又设它与样本总数 T 的比值

$P_j = T_j / T$, 称 P_j 为第 j 个源信号的稀疏持续率; 则称 $T_{\text{total}} = \sum_{j=1}^n T_j$ 为总稀疏持续数,

$P_{\text{total}} = \sum_{j=1}^n P_j$ 为总稀疏持续率.

在时域检索平均法(SAMTD)中, 要求 $T_j > T_0 (j=1, 2, \dots, n)$, 才能全部估计基矢量.

下面是时域检索平均法(SAMTD)的算法步骤.

- I. 按照(4)和(5)式计算 \mathbf{M} 和 \mathbf{D} ;
- II. 按定义计算检索矩阵 Ω , 矩阵 Θ , 单源区间数据矩阵 \mathbf{X}_0 ;
- III. 查找检索矩阵 Ω 出现第一个 1 的列 t_1 且 $\theta^{t_1} \neq 0$, 设 $j=1$;
- IV. 按照(6)式检索 Θ 中满足与 θ^{t_1} 在同一条直线附近的时刻 t_1, t_2, \dots, t_k ;
- V. 给定 T_0 和基矢量估计偏差 $\delta > 0$, 如果 $k > T_0$, 按照(7)式计算 \mathbf{a}^j , 且 $j = j+1$;

在(6)式中, 前者表明矢量 θ^{t_i} 和 θ^{t_j} 方向间的夹角接近于 0 , 后者夹角接近 180° . 通过检索, 在 Θ 中得到一组样本时间点为 t_1, t_2, \dots, t_k , 其任意两两矢量满足判别条件, 则易到某个 \mathbf{a}^j .

但是, 在(2)式中, 如果某两个以上的源信号在小时间段 Δt 上保持大小基本恒定而非 0 , 而其余的源信号近似为 0 时, 则 \mathbf{x}^t 在小时间段 Δt 上也近似位于同一直线上, 但它不是基直线, 这样的直线称为伪基直线. 伪基直线出现是可能的, 但出现在同一条伪基直线附近的两个以上小时间段 Δt 的

VI. 将时刻 t_1, t_2, \dots, t_k 上的 ω^t 和 θ^t 置 0, 即 $\omega^t = 0$ 和 $\theta^t = 0$;

VII. 重复第 II 步, 直到 $\Omega = \mathbf{0}_{l \times N}$ 和 $\Theta = \mathbf{0}_{m \times N}$ 停止, 最终得到混叠矩阵的所有列矢量 $\mathbf{a}^1, \dots, \mathbf{a}^n$.

2 时域检索平均法的参数讨论

时域检索平均法的参数有: 检索误差 ε 、最低样本数 T_0 、基矢量估计偏差 δ 、单源区间的最小时间长度 Δ . 单源区间的最小时间长度 Δ 越大, 精度也越高, 而太长会失去很多单源区间, 甚至将基直线漏检, 一般取 3~5 效果已经很好. 检索误差 ε 主要控制基直线的精度, ε 越小直线精度越高, 但太小会失去很多单源区间漏检基直线, ε 太大, 基直线精度低, 出现伪基直线. 基矢量估计偏差 δ 一般较大, 才能保证同一基直线附近的点都能参与基矢量估计. 伪基直线的排除主要取决于最低样本数 T_0 , 一般 T_0 取 2~3 个 Δ 的长度较适宜; 如果 T_0 取得太长, 在删除伪基直线的同时, 也可能删除掉真正的基直线.

时域检索平均法不需预先给定源信号数目, 但它是否能够准确的检索出所有的基直线, 并删除伪基直线呢?

此问题一方面取决于上述参数的取值; 另一方面, 更重要的是, 取决于源信号的稀疏持续数 T_j , 如果 $T_j \leq T_0$, 则基直线被误认为伪基直线而删除. 如果所有源信号的稀疏持续数 $T_j > T_0$, 那么能估计出所有的基矢量; 如果第 j 个源信号不存在单源区间 ($T_j = 0$), 即使我们不排除伪基直线, 也不能检索出 \mathbf{a}^j , 这说明第 j 个源信号全部被其他信号淹没.

3 源信号的恢复

对源信号的恢复, 目前通常根据信息理论的最大可能性原则, 用稀疏表示作为源信号的估计, 得到 l^1 范数准则^[4~8]

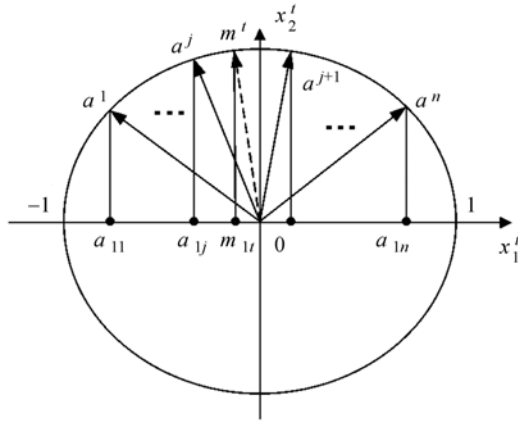
$$\min_{s^t} \sum_{j=1}^n |s_j^t|, \text{ s.t. } \mathbf{x}^t = \sum_{j=1}^n \mathbf{a}^j s_j^t. \quad (8)$$

Bofill 基于 l^1 范数准则提出了最短路径算法^[9], Theis 证明了最短路径算法的理论基础是 l^1 范数准则^[11], 并提出了超完备线性 ICA 几何算法. 在该算法中, 找出与 \mathbf{x}^t 上下最邻近的两条基直线. 下面将提出一种改进的超完备线性 ICA 几何算法, 给出寻找上下最邻近的两条基直线的简易方法, 这里仅考虑两个观测信号的情形, 即 $m = 2$.

矢量 $\mathbf{m}^t = (m_1^t, m_2^t)^T$ 和 $\mathbf{a}^j = (a_{1j}, a_{2j})^T$ 位于坐标平面的单位圆上, 为了便于找出与 \mathbf{x}^t 上下最邻近的两条基直线, 现定义如下.

定义 3 设混叠矩阵 $\mathbf{A} \in \mathbb{R}^{2 \times n}$ 的任意列矢量 \mathbf{a}^j ($\|\mathbf{a}^j\| = 1$) 的方向角为 β_j , 如果 $\pi \leq \beta_j < 2\pi$, 用 $-\mathbf{a}^j$ 替换 \mathbf{a}^j , 则 $0 \leq \beta_j < \pi$ 在 $i = 1, 2, \dots, n$ 均成立, 将 \mathbf{a}^j 在 \mathbf{A} 中按方向角 β_j 从大到小的顺序重新排列, 重新排列后得到的 \mathbf{A} 称为标准混叠矩阵.

定义 4 设矩阵 $\mathbf{M} \in \mathbb{R}^{2 \times N}$ 中的非 0 列矢量 \mathbf{m}^t 的方向角为 β_{m^t} , 当 $\pi \leq \beta_{m^t} < 2\pi$ 时, 则用 $-\mathbf{m}^t$ 替换 \mathbf{m}^t , 使得 $0 \leq \beta_{m^t} < \pi$ 对所有的 $t = 1, 2, \dots, T$ 均成立, 此时称 \mathbf{M} 为标准数据矩阵.

图 5 与 m^t 左、右最邻近基矢量的示意图

为寻找与 m^t 上下最邻近的两条基直线, 先将矩阵 \mathbf{A} 和 \mathbf{M} 标准化. 从图 5 可知, 将 m_1^t 与标准混叠矩阵的第 1 行的值 a_{1j} ($j=1, \dots, n$) 进行比较, 如果 $a_{1j} \leq m_1^t < a_{1,j+1}$, 上下最邻近的基直线的两个列矢量是 a^j 和 a^{j+1} , 并设 $\mathbf{A}_{j,j+1} = [a^j \ a^{j+1}]$, $j=1, \dots, n-1$; 如果 $m_1^t < a_{11}$ 或 $m_1^t \geq a_{1n}$, 上下最邻近的基直线的两个列矢量为 a^1 和 a^n , 设 $\mathbf{A}_{1n} = [a^1 \ a^n]$. 因为 a^j ($j=1, 2, \dots, n$) 两两线性独立, 所以 \mathbf{A}_{1n} 和 $\mathbf{A}_{j,j+1}$ ($j=1, 2, \dots, n-1$) 为 $\mathbb{R}^{2 \times 2}$ 的可逆矩阵.

定理 设 \mathbf{A} 为标准混叠矩阵, \mathbf{M} 为标准数据矩阵, 则

$$\begin{cases} [\hat{s}_1^t \ \hat{s}_n^t]^T = \mathbf{A}_{1n}^{-1} \mathbf{x}^t, \\ \hat{s}_j^t = 0 \ (j \neq 1, n), \end{cases} \quad m_1^t < a_{11} \cup m_1^t \geq a_{1n}; \quad (9)$$

$$\begin{cases} [\hat{s}_j^t \ \hat{s}_{j+1}^t]^T = \mathbf{A}_{j,j+1}^{-1} \mathbf{x}^t, \\ \hat{s}_i^t = 0 \ (i \neq j, j+1), \end{cases} \quad a_{1j} \leq m_1^t < a_{1,j+1}, j=1, \dots, n-1, \quad (10)$$

其中, \hat{s}_j^t 表示第 j 个源信号在时间 t 的估计值.

改进的超完备线性 ICA 几何算法步骤如下:

- I. 将 \mathbf{A} 化为标准混叠矩阵, \mathbf{M} 化为标准数据矩阵;
- II. 构造 \mathbf{A}_{1n} 和 $\mathbf{A}_{j,j+1}$ ($j=1, 2, \dots, n-1$);
- III. 按定理中的(9)和(10)式计算源信号在时刻 t 的估计值 \hat{s}_j^t .

4 仿真实验

实验 1 以 5 个讲话声(4 个男音和 1 个女音)的语音信号作为源信号, 即来自 <http://www.princeton.edu/~srickard/bss.html> 上的 u1.wav|u2.wav|u3.wav|u4.wav|u5.wav. 抽取样本 32000 个, 其波形如图 6.

混叠矩阵由计算机随机产生为

$$\mathbf{A} = \begin{bmatrix} 0.9995 & 0.8823 & -0.2491 \\ -0.0316 & 0.4708 & 0.9685 \end{bmatrix}.$$

然后, 取 u2.wav|u3.wav|u5.wav, u1.wav|u2.wav|u3.wav 和 u3.wav|u4.wav|u5.wav 作为 3 组输入的源信号, 进行 3 组实验, 在 3 组实验中随机产生的混叠矩阵不变.

在混叠信号波形上, 仅给出以 u2.wav|u3.wav|u5.wav 为源信号的混叠结果, 其波形如图 7.

在 3 组源信号的时域检索平均法的算法仿真中, 先设置检索误差 ε , 单源区间长度 Δ , 基矢量估计偏差 δ , 基直线最低样本数 T_0 , 见表 1.

从表 1 可知, 检索误差 ε 是时域检索平均法的最重要的参数, 当检索误差 ε 较大时, 需适

当调节基直线最低样本数 T_0 , 而其他参数基本可以不需调整.

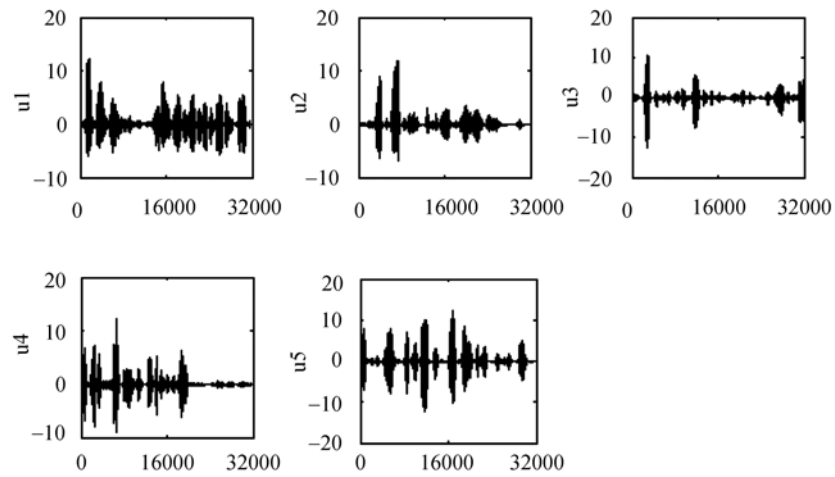


图 6 源信号波形图

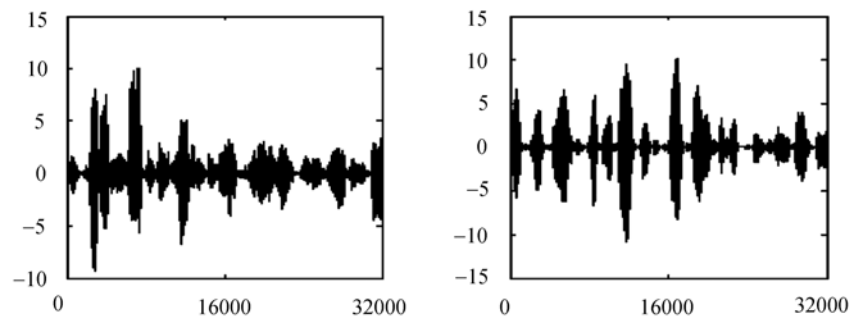


图 7 混叠信号波形图(实验 1)

表 1 时域检索平均法的参数设置

输入源信号	设置参数			
	Δ	ε	δ	T_0
u2.wav u3.wav u5.wav	4	0.008	0.2	10
u1.wav u2.wav u3.wav	4	0.015	0.2	10
u3.wav u4.wav u5.wav	4	0.012	0.2	10
u6.wav u7.wav u8.wav	4	0.030	0.2	20

经仿真, 每组实验的 3 个源信号的稀疏持续数 T_1 , T_2 , T_3 和总稀疏持续数 T_{total} , 以及稀疏持续率分别为 P_1 , P_2 , P_3 和总稀疏持续率 P_{total} , 参见表 2.

表 2 时域检索平均法的计算结果

输入源信号	计算结果							
	T_1	T_2	T_3	T_{total}	$P_1/\%$	$P_2/\%$	$P_3/\%$	$P_{\text{total}}/\%$
u2.wav u3.wav u5.wav	58	13	353	424	0.18	0.04	1.10	1.33
u1.wav u2.wav u3.wav	255	27	324	606	0.80	0.08	1.01	1.89
u3.wav u4.wav u5.wav	259	194	18	471	0.81	0.61	0.06	1.47
u6.wav u7.wav u8.wav	404	677	34	1115	1.26	2.12	0.11	3.48

从表 2 可知, 经时域的检索后, 干扰基矢量估计的数据点已被剔除, 剩下的数据很少, 占整个数据点的 1%~4%, 所以, 计算量减小, 精度提高.

在混叠矩阵的估计中, 仅给出以 u2.wav|u3.wav|u5.wav 为源信号的实验结果, 则

$$\hat{\mathbf{A}} = \begin{bmatrix} -0.99952 & -0.24867 & 0.88209 \\ 0.030901 & 0.96859 & 0.47108 \end{bmatrix}.$$

估计混叠矩阵 $\hat{\mathbf{A}}$ 的性能用 $\hat{\mathbf{A}}$ 与 \mathbf{A} 的对应列矢量的角度偏差来度量, 参见表 3.

表 3 两种方法的 $\hat{\mathbf{a}}^j$ 与 \mathbf{a}^i 的角度偏差比较表

源 信 号	方法	$\hat{\mathbf{a}}^j$ 与 \mathbf{a}^i 的角度偏差		
		$\hat{\mathbf{a}}^1$ 与 $\mathbf{a}^1/(^\circ)$	$\hat{\mathbf{a}}^2$ 与 $\mathbf{a}^3/(^\circ)$	$\hat{\mathbf{a}}^3$ 与 $\mathbf{a}^2/(^\circ)$
u2.wav u3.wav u5.wav	时域检索平均法	0.041172	0.025239	0.025714
	Theis 方法	20.5763	0.0540	5.9728
u1.wav u2.wav u3.wav	时域检索平均法	1.2304	3.2014	0.1197
	Theis 方法	8.2421	7.2489	0.2414
u3.wav u4.wav u5.wav	时域检索平均法	0.15575	0.17344	0.41032
	Theis 方法	0.0529	4.8243	5.1273
u6.wav u7.wav u8.wav	时域检索平均法	0.1647	1.9217	0.49129
	Theis 方法	15.7112	2.1465	0.4982

在相同源信号和混叠矩阵的情形下, 用 Theis^[11]的算法对 3 组信号进行仿真, 得到角度偏差, 并与本文的方法进行比较, 见表 3.

从表 3 中可看出, 本文的方法在混叠矩阵的估计中, 在每组实验中, 较 Theis^[11]在 2004 年发表的方法有更高的精确度, 而且, Theis 的方法要求源信号数目已知.

在源信号恢复中, 常用估计信号与源信号的相关系数来衡量其算法的性能, 并可观察其波形. 在此, 仅给出以 u2.wav|u3.wav|u5.wav 为源信号的实验结果, 同时与采用 Theis^[11]的方法仿真实验进行比较, 见表 4.

表 4 两种方法的源信号与估计信号之间的相关系数比较表

源信号 (语音信号)	时域检索平均法的估计信号			Theis 方法的估计信号		
	\hat{s}_1	\hat{s}_2	\hat{s}_3	\hat{s}_1	\hat{s}_2	\hat{s}_3
s_1	-0.85338	-0.025098	0.13997	-0.7638	-0.1689	0.3670
s_2	-0.16545	0.093504	0.92766	0.0120	0.1189	0.9767
s_3	0.17137	0.988010	0.18811	0.3541	0.9043	0.1185

从表 4 也可以明显地看出, 本文方法所估计的信号与源信号比 Theis 方法有更高的相似度. 在估计信号的波形图上, 仅给出以 u2.wav|u3.wav|u5.wav 为源信号的实验结果, 见图 8.

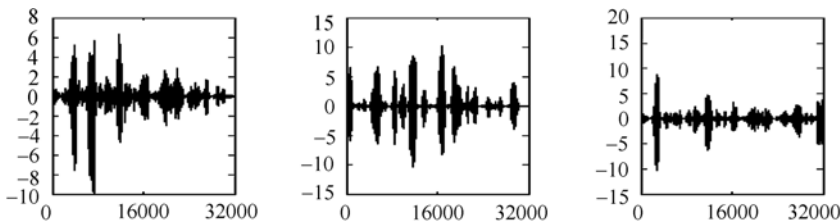


图 8 估计信号的波形图(实验 1)

与图 6 相比, 图 8 左边的是 $u2.wav$ 的估计波形, 有小部分的失真, 中间的是 $u5.wav$ 的估计波形, 右边的是 $u3.wav$ 的估计波形, 后面两个波形与源信号也极为相似.

实验 2 为了进一步验证我们的方法, 再选择 3 个讲话声(2 个女音和 1 个男音)的语音信号作为源信号, 其波形图参见图 9, 混叠矩阵与实验 1 相同, 3 个源信号来自 <http://www.princeton.edu/~srickard/bss.html> 中的 $u6.wav | u7.wav | u8.wav$, 抽取样本 32000 个, 其混叠信号参见图 10.

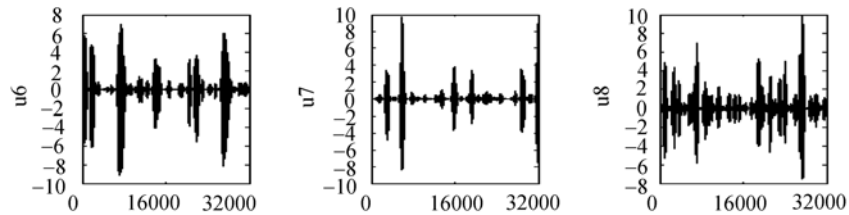


图 9 源信号波形图(实验 2)

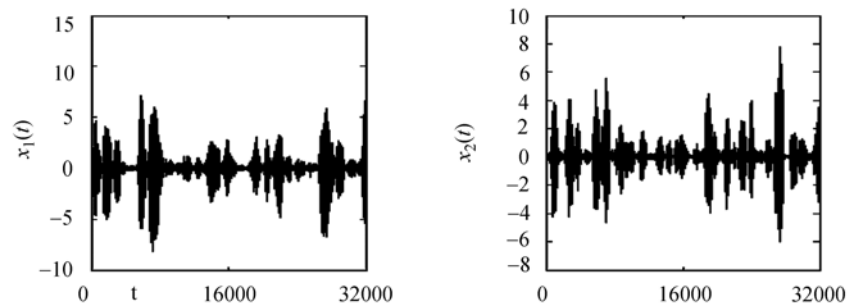


图 10 混叠信号波形图(实验 2)

时域检索平均法的算法的检索误差 ε 、单源区间长度 Δ 和基矢量估计偏差 δ 的设定参见表 1; 经检索计算, 得到关于源信号的稀疏持续数和稀疏持续率的实验结果, 参见表 2.

单源区间数据矩阵 \mathbf{X}_0 中 $x_2(t)$ 对 $x_1(t)$ 的散落图如图 11. 在图 11(a)中几乎完全不存在基直线趋势的条件下, 只要加大检索误差 ε , 就能检索出基直线. 很明显, 图 11(b)易估计基矢量, 不需知道源信号的个数, 而图 11(a)几乎无法聚类估计混叠矩阵.

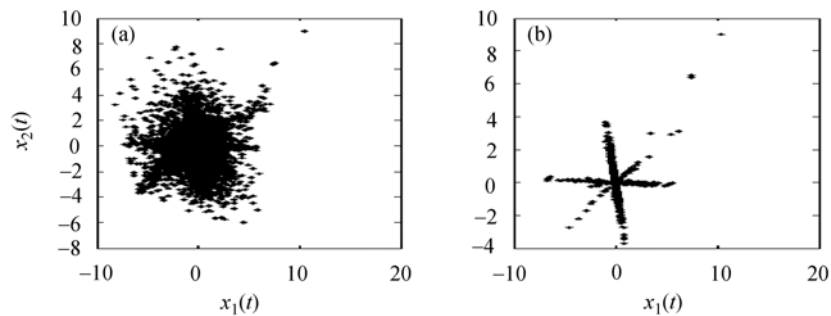


图 11 矩阵 \mathbf{X} 和矩阵 \mathbf{X}_0 中 $x_2(t)$ 对 $x_1(t)$ 的散落图

(a) 矩阵 \mathbf{X} ; (b) 矩阵 \mathbf{X}_0

根据矩阵 Θ , 得到混叠矩阵的估计

$$\hat{\mathbf{A}} = \begin{bmatrix} -0.99959 & -0.24079 & 0.89759 \\ 0.028746 & 0.97058 & 0.44084 \end{bmatrix}.$$

计算 $\hat{\mathbf{A}}$ 与 \mathbf{A} 的对应列矢量的角度偏差并与 Theis^[11]的方法进行比较, 见表 3. 从表 3 可以看出, 本文的方法也优于 Theis^[11]的方法. 由于提高了检索误差 ε , 故基直线的角度偏差高于实验 1, 说明检索误差 ε 和角度偏差的关系. 同时, 可以得出两种方法的源信号与估计信号之间的相关系数比较表, 见表 5.

从表 5 可以进一步知道本文的方法优于 Theis^[11]方法. 估计信号的波形图见图 12.

表 5 两种方法的源信号与估计信号之间的相关系数比较表

源信号 (语音信号)	时域检索平均法的估计信号			Theis 方法的估计信号(同源信号)		
	\hat{s}_1	\hat{s}_2	\hat{s}_3	\hat{s}_1	\hat{s}_2	\hat{s}_3
s_1	-0.8900	-0.1099	0.2156	-0.8699	-0.1372	0.2693
s_2	-0.0983	0.0939	0.9711	-0.0289	0.0709	0.9559
s_3	0.2554	0.8426	0.2564	0.4758	0.8011	0.2454

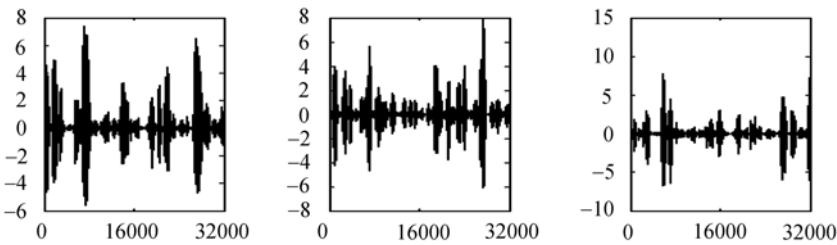


图 12 估计信号的波形图(实验 2)

语音信号实验结果表明, 检索平均法省略了干扰基矢量估计的样本, 实验中仅用几百个样本用于基矢量估计, 整个样本的 1%~4%, 也可不需 K 均值聚类, 而仅将在同一直线附近的列归一化后进行平均, 得到混叠矩阵, 计算量远远小于对所有样本进行 K 均值聚类; 特别地, 检索平均法能够处理在稀疏持续率低至 1%~4%的混叠语音信号. 最后, 时域检索平均法的参数设置也较容易.

5 结论

本文分析了一类语音信号的稀疏表示在欠定的盲信号分离中的应用, 在两步法的基础上, 给出了一种新的估计混叠矩阵的方法——时域检索平均法. 时域检索平均法具有三大优点: 计算量小; 精度高; 在所有源信号具有一定稀疏持续数时, 源信号数目可以未知. 同时它是处理语音信号混叠的好方法. 超完备线性 ICA 几何改进算法与 Theis^[11]几何算法一致, 在寻找两个相邻矢量的方法上进行了小的改进. 几个语音信号的实时的线性混叠实验证实了所提算法的性能.

参 考 文 献

1 Cardoso J F. Blind signal separation: Statistical principles. Proc IEEE (Special Issue on Blind Identification and Estimation), 1998, 90(10): 2009—2026

- 2 Cichocki A, Amari S. Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. New York: Wiley, 2002
- 3 Li Y, Wang J. Sequential blind extraction of instantaneously mixed sources. IEEE Trans Signal Process, 2002, 50(5): 997—1006
- 4 Chen S, Donoho D L, Saunders M A. Atomic decomposition by basis pursuit. SIAM J Sci Comput, 1998, 20(1): 33—61
- 5 Olshausen B A, Sallee P, Lewicki M S. Learning sparse image codes using a wavelet pyramid architecture. Adv Neural Inform Process Syst, 2001, 13: 887—893
- 6 Lee T-W, Lewicki M, Girolami M, et al. Blind source separation of more sources than mixtures using overcomplete representations. IEEE Signal Process Lett, 1999, 6(4): 187—90
- 7 Zibulevsky M, Pearlmutter B A. Blind source separation by sparse decomposition in signal dictionary. Neural Comput, 2001, 13(4): 863—882
- 8 Li Y, Andrzej C, Amari S. Analysis of sparse representation and blind source separation. Neural Comput, 2004, 16(6): 1193—1234
- 9 Bofill P, Zibulevsky M. Underdetermined blind source separation using sparse representations. Signal Process, 2001, 81(11): 2353—2362
- 10 Bofill P. Underdetermined blind separation of delayed sound sources in the frequency domain, Neurocomputing, 2003, 55(3): 627—644
- 11 Theis F J, Lang W E, Puntonet C G. A geometric algorithm for overcomplete linear ICA. Neurocomputing, 2004, 56: 381—398