

文章编号:1001-8360(2012)04-0069-07

# 非稀疏欠定盲分离及其在语音分离中的应用

陈永强<sup>1,2</sup>, 王宏霞<sup>1</sup>

(1. 西南交通大学 信息科学与技术学院, 四川 成都 610031;  
2. 成都信息工程学院 电子实验中心, 四川 成都 610225)

**摘 要:**本文提出一种基于马尔科夫链蒙特卡洛方法(MCMC)的贝叶斯非稀疏盲源分离算法。用广义高斯分布(GGD)来拟合源信号的分布,通过 MCMC 抽样得到 GGD 参数和隐变量的估计,并由此得到源信号的最小均方误差估计(MMSE),解决了 GGD 参数估计容易陷入局部极值点、鲁棒性差的问题。根据语音信号的局部平稳性,提出基于非稀疏度评判准则的盲分离算法,用 MCMC 方法分离非稀疏区的语音信号,进一步提高了语音信号分离精度。仿真实验证明,本文方法改善了非稀疏信号和语音信号的分离效果,而且具有更好的鲁棒性。

**关键词:** 马尔科夫链蒙特卡洛方法; 欠定盲分离; 贝叶斯方法; 非稀疏度评判准则

**中图分类号:** TN911.7 **文献标志码:** A **doi:**10.3969/j.issn.1001-8360.2012.04.012

## Underdetermined Blind Separation of Non-sparse Signals and Its Application in Speech Separation

CHEN Yong-qiang<sup>1,2</sup>, WANG Hong-xia<sup>1</sup>

(1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China;  
2. Electronic Experiment Center, Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract:** A method applicable to underdetermined blind separation of non-sparse signals is proposed, which is based on the Markov chain Monte Carlo(MCMC)Bayesian framework. The generalized Gaussian distribution (GGD) is used to model the source signals distribution, the model parameters and hidden variables are estimated by MCMC sampling to obtain the least mean square error estimation (MMSE) of source signals, by which the problem is solved that the GGD parameter estimation easily falls into a local extremum point and has poor robustness. Making use of the local stationarity of speech, a blind separation method based on non-sparse judgment criterions is proposed to enhance speech separation accuracy, which separates the speech in the non-sparse zone by MCMC sampling. Computer simulation shows that the proposed method can improve separation performance of non-sparse and speech signals, it also has better robustness.

**Key words:** markov chain monte carlo methods (MCMC); underdetermined blind separation (UBSS); bayesian method; non-sparse judgment criterion

盲源分离广泛应用于语音识别、增强和加密、通信、数字水印等领域。传统的盲源分离方法是独立成分分析<sup>[1-2]</sup>,但是只适用于适定和超定情形。当在欠定情形下,通常的方法是利用音频信号的稀疏特性来分离信号<sup>[3-5]</sup>,该类方法称为稀疏成分分析(SCA)。另一类方法则通过

某一概率模型来拟合源信号在时域上的分布或在某个变换域上的分布,例如:拉普拉斯分布<sup>[6]</sup>、高斯混合模型<sup>[7-8]</sup>、广义高斯分布(GGD)<sup>[9-10]</sup>都可以作为信号的分布模型。与稀疏信号相比,非稀疏信号的欠定盲分离还研究较少。文献[9]通过期望最大化算法(EM 算法)获得 GGD 模型参数的最大似然估计(MLE),同时对稀疏信号和非稀疏信号进行分离。本文则在贝叶斯框架下,通过马尔科夫链蒙特卡洛方法(MCMC)中的 Gibbs 和 Metropolis-Hastings 混合抽样<sup>[11]</sup>获得马尔科夫链的平稳分布。该平稳

收稿日期:2011-07-05;修回日期:2011-12-13  
基金项目:国家自然科学基金(61170226);中央高校基本科研业务费专项资金(SWJTU11CX047)  
作者简介:陈永强(1975—),男,四川达州人,讲师,博士研究生。  
E-mail:chenyongq@cuit.edu.cn

分布就是隐变量和 GGD 模型参数的贝叶斯联合后验分布,由此得到源信号的最小均方差估计(MMSE)。该方法克服了文献[9]算法对初始值依赖大,容易落入局部极值点的缺点,具有很好的鲁棒性。此外,针对文献[3-5]要求局部时间或时频点上的激活源信号个数(即能量不为 0 的源信号)不能超过传感器个数的缺点,将 MCMC 方法与统计稀疏分解原则(SSDP)<sup>[5]</sup>相结合,提出了一种基于非稀疏度评判准则的语音信号分离方法,可以提高语音的分离精度,本文对该方法进行了严格的证明。

## 1 欠定盲分离模型

假设源信号的个数为  $N$ ,传感器的个数为  $M$  ( $M < N$ )。在某个时刻  $t$ ,源信号的瞬时混合模型为

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

式中:  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T$ , 是  $M$  维观测数据向量;  $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ , 是  $N$  维源信号数据向量;  $\mathbf{A}$  是  $M \times N$  维的时不变混合矩阵。源信号经过欠定混合后会丢失一部分信息(即隐变量  $\mathbf{z}$ ),只要估计出隐变量,就可以进一步估计出源信号<sup>[9]</sup>。在本文中,要求  $\mathbf{A}$  行满秩,且各源信号之间相互独立,并假定混合矩阵已经被估计出来<sup>[9,12-13]</sup>。

## 2 基于贝叶斯框架的非稀疏盲信号分离

### 2.1 基于 MCMC 的贝叶斯估计

本文采用广义高斯分布<sup>[9-10]</sup>描述源信号  $s_i(t)$  ( $i = 1, 2, \dots, N$ ) 的分布

$$P(s_i(t) | \beta_i, \sigma_i) = \frac{v(\beta_i)}{\sigma_i} \exp \left\{ -c(\beta_i) \left| \frac{s_i(t)}{\sigma_i} \right|^{\frac{2}{1+\beta_i}} \right\} \quad (2)$$

$$\text{式中} \quad c(\beta_i) = \left\{ \frac{\Gamma[1.5(1+\beta_i)]}{\Gamma[0.5(1+\beta_i)]} \right\}^{\frac{1}{1+\beta_i}} \quad (3)$$

$$v(\beta_i) = \frac{\{\Gamma[1.5(1+\beta_i)]\}^{0.5}}{(1+\beta_i) \{\Gamma[0.5(1+\beta_i)]\}^{1.5}} \quad (4)$$

Gamma 函数  $\Gamma(\cdot)$  的定义为  $\Gamma(\gamma) = \int_0^\infty u^{\gamma-1} e^{-u} du$ ;  $\beta_i$  的取值范围为  $(-1, +\infty)$ , 控制密度函数衰减速度;  $\sigma_i^2$  为方差,控制密度函数宽度。该模型不但适用于稀疏信号,而且适用于非稀疏信号。

本文将 GGD 模型参数  $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_N]^T$  和  $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_N]^T$  看作随机变量,在贝叶斯框架下估计隐变量  $\mathbf{z} = [z_1, z_2, \dots, z_{N-M}]^T$ 。当得到隐变量和模型参数的联合后验条件概率密度函数  $P(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{x})$ , 通过式(5)可以得到隐变量的 MMSE 估计

$$\hat{\mathbf{z}} = \iiint \mathbf{z} P(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{x}) d\boldsymbol{\beta} d\boldsymbol{\sigma} d\mathbf{z} \quad (5)$$

但这是一个高维多重积分,直接计算几乎是不可能的。本文通过 Gibbs 抽样来解决这个问题。基本思路是:

建立一个以  $P(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{x})$  为平稳分布的马尔科夫链来获得该后验概率密度的若干样本,然后就可以通过这些样本估计隐变量  $\mathbf{z}$ 。具体步骤如下:

步骤 1 设定初始值  $\boldsymbol{\beta}^{(0)}, \boldsymbol{\sigma}^{(0)}$ , 令  $l=0$ ;

步骤 2 根据全条件概率密度  $P(\mathbf{z}^{(l+1)} | \boldsymbol{\beta}^{(l)}, \boldsymbol{\sigma}^{(l)}, \mathbf{x})$  对  $\mathbf{z}$  抽样,得到  $\mathbf{z}$  在第  $l+1$  步的样本  $\mathbf{z}^{(l+1)}$ 。

这又分为以下子步骤:

对  $P(z_1^{(l+1)} | z_2^{(l)}, z_3^{(l)}, \dots, z_{N-M}^{(l)}, \boldsymbol{\beta}^{(l)}, \boldsymbol{\sigma}^{(l)}, \mathbf{x})$  抽样,得到  $z_1^{(l+1)}$ ;

对  $P(z_2^{(l+1)} | z_1^{(l+1)}, z_3^{(l)}, \dots, z_{N-M}^{(l)}, \boldsymbol{\beta}^{(l)}, \boldsymbol{\sigma}^{(l)}, \mathbf{x})$  抽样,得到  $z_2^{(l+1)}$ ;

⋮

对  $P(z_{N-M}^{(l+1)} | z_1^{(l+1)}, z_2^{(l+1)}, \dots, z_{N-M-1}^{(l+1)}, \boldsymbol{\beta}^{(l)}, \boldsymbol{\sigma}^{(l)}, \mathbf{x})$  抽样,得到  $z_{N-M}^{(l+1)}$ 。

步骤 3 根据全条件概率密度  $P(\boldsymbol{\beta}^{(l+1)} | \mathbf{z}^{(l+1)}, \boldsymbol{\sigma}^{(l)}, \mathbf{x})$  对  $\boldsymbol{\beta}$  抽样,得到第  $l+1$  步  $\boldsymbol{\beta}$  的样本  $\boldsymbol{\beta}^{(l+1)}$ 。由于各  $\beta_i$  之间相互独立,只需对  $P(\beta_i^{(l+1)} | \mathbf{z}^{(l+1)}, \boldsymbol{\sigma}^{(l)}, \mathbf{x})$  抽样得到  $\beta_i^{(l+1)}$ ,  $i=1, 2, \dots, N$ 。

步骤 4 根据全条件概率密度  $P(\boldsymbol{\sigma}^{(l+1)} | \mathbf{z}^{(l+1)}, \boldsymbol{\beta}^{(l+1)}, \mathbf{x})$  对  $\boldsymbol{\sigma}$  抽样,得到第  $l+1$  步  $\boldsymbol{\sigma}$  的样本  $\boldsymbol{\sigma}^{(l+1)}$ 。各  $\sigma_i$  之间也相互独立,只需对  $P(\sigma_i^{(l+1)} | \mathbf{z}^{(l+1)}, \boldsymbol{\beta}^{(l+1)}, \mathbf{x})$  抽样得到  $\sigma_i^{(l+1)}$ ,  $i=1, 2, \dots, N$ ; 令  $l=l+1$ , 返回步骤 2。

以上步骤 2~步骤 4 迭代进行,马尔科夫链最终达到平稳分布。假设此时是第  $L$  步,此后的采样值  $(\mathbf{z}^{(L)}, \boldsymbol{\beta}^{(L)}, \boldsymbol{\sigma}^{(L)})$  ( $l > L$ ) 就是  $P(\mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{x})$  的抽样,根据这些抽样就可以获得  $\mathbf{z}$  的 MMSE 估计

$$\hat{\mathbf{z}} = \frac{1}{P} \sum_{l=L+1}^{L+P} \mathbf{z}^{(l)} \quad (6)$$

式中:  $P$  为达到平稳分布后的采样个数。

然后利用式(7)就可以获得源信号  $\mathbf{s}$  的估计<sup>[9-10]</sup>

$$\mathbf{s} = \mathbf{A}^+ \mathbf{x} + \mathbf{V}\mathbf{z} \quad (7)$$

式中:  $\mathbf{A}^+$  是 Moore-Penrose 伪逆;  $\mathbf{V}$  的各列是构成  $\mathbf{A}$  零空间的一组基。

由以上步骤可以看到,要得到  $\hat{\mathbf{z}}$ , 关键是要获得  $\mathbf{z}$ ,  $\boldsymbol{\beta}, \boldsymbol{\sigma}$  各自的全条件概率密度函数。下面将推导出所有全条件概率密度函数的表达式。

### 2.2 基于 GGD 模型的全条件概率密度

由式(7)可知隐变量和观测数据的联合条件概率密度正比于源信号的概率密度<sup>[9]</sup>, 即

$$P(\mathbf{z}, \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\sigma}) \propto P(\mathbf{s} | \boldsymbol{\beta}, \boldsymbol{\sigma}) \quad (8)$$

本文假设各源信号相互独立,由式(2)、式(7)可得

$$P[s(t) | \boldsymbol{\beta}, \boldsymbol{\sigma}] \propto \prod_{i=1}^N \frac{v(\beta_i)}{\sigma_i} \exp \left\{ -c(\beta_i) \left| \frac{\mathbf{a}_i^+ \mathbf{x}(t) + \mathbf{v}_i \mathbf{z}(t)}{\sigma_i} \right|^{\frac{2}{1+\beta_i}} \right\} \quad (9)$$

式中:  $\mathbf{a}_i^+$ 、 $\mathbf{v}_i$  分别是  $\mathbf{A}^+$ 、 $\mathbf{V}$  的第  $i$  行构成的行向量。

根据贝叶斯定理和式(8)、式(9),可以导出  $z_i(t)$  ( $i=1,2,\cdots,N-M$ ) 的全条件密度函数为

$$\begin{aligned} P[z_i(t) | z_{-i}(t), \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{x}(t)] &= \frac{P[z_i(t) | \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{x}(t)]}{P[z_{-i}(t) | \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{x}(t)]} = \\ &= \frac{P[z_i(t), \mathbf{x}(t) | \boldsymbol{\beta}, \boldsymbol{\sigma}]}{P[\mathbf{x}(t) | \boldsymbol{\beta}, \boldsymbol{\sigma}] P[z_{-i}(t) | \boldsymbol{\beta}, \boldsymbol{\sigma}, \mathbf{x}(t)]} \propto \\ &P[z_i(t), \mathbf{x}(t) | \boldsymbol{\beta}, \boldsymbol{\sigma}] \propto \exp\left(-\sum_{i=1}^N c(\beta_i) \left| \frac{\mathbf{a}_i^+ \mathbf{x}(t) + \mathbf{v}_i z_i(t)}{\sigma_i} \right|^{\frac{2}{1+\beta_i}}\right) \end{aligned} \tag{10}$$

式中:  $z_{-i}(t)$  表示  $z(t)$  除去  $z_i(t)$  后,由剩余元素构成的向量。

根据贝叶斯定理和式(8)、式(9),并利用  $\beta_i, \sigma_i$  ( $i=1,2,\cdots,N$ ) 之间都相互独立,可以导出  $\beta_i, \sigma_i$  的全条件密度函数为

$$\begin{aligned} P(\beta_i | \mathbf{z}, \boldsymbol{\sigma}, \mathbf{x}) &= P(\beta_i | \boldsymbol{\beta}_{-i}, \mathbf{z}, \boldsymbol{\sigma}, \mathbf{x}) = \\ &= \frac{P(\boldsymbol{\beta} | \mathbf{z}, \boldsymbol{\sigma}, \mathbf{x})}{P(\boldsymbol{\beta}_{-i} | \mathbf{z}, \boldsymbol{\sigma}, \mathbf{x})} = \frac{P(\mathbf{z}, \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\sigma}) P(\boldsymbol{\beta} | \boldsymbol{\sigma})}{P(\mathbf{z}, \mathbf{x} | \boldsymbol{\sigma}) P(\boldsymbol{\beta}_{-i} | \mathbf{z}, \boldsymbol{\sigma}, \mathbf{x})} \propto \\ &P(\mathbf{z}, \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\sigma}) P(\boldsymbol{\beta}) \propto v(\beta_i)^Q \cdot \\ &\exp\left(-\sum_{i=1}^Q c(\beta_i) \left| \frac{\mathbf{a}_i^+ \mathbf{x}(t) + \mathbf{v}_i z_i(t)}{\sigma_i} \right|^{\frac{2}{1+\beta_i}}\right) P(\beta_i) \end{aligned} \tag{11}$$

$$\begin{aligned} P(\sigma_i | \mathbf{z}, \boldsymbol{\beta}, \mathbf{x}) &= P(\sigma_i | \boldsymbol{\sigma}_{-i}, \mathbf{z}, \boldsymbol{\beta}, \mathbf{x}) = \\ &= \frac{P(\boldsymbol{\sigma} | \mathbf{z}, \boldsymbol{\beta}, \mathbf{x})}{P(\boldsymbol{\sigma}_{-i} | \mathbf{z}, \boldsymbol{\beta}, \mathbf{x})} = \frac{P(\mathbf{z}, \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\sigma}) P(\boldsymbol{\sigma} | \boldsymbol{\beta})}{P(\mathbf{z}, \mathbf{x} | \boldsymbol{\beta}) P(\boldsymbol{\sigma}_{-i} | \mathbf{z}, \boldsymbol{\beta}, \mathbf{x})} \propto \\ &P(\mathbf{z}, \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\sigma}) P(\boldsymbol{\sigma}) \propto \\ &\frac{1}{\sigma_i^Q} \exp\left(-\sum_{i=1}^Q c(\beta_i) \left| \frac{\mathbf{a}_i^+ \mathbf{x}(t) + \mathbf{v}_i z_i(t)}{\sigma_i} \right|^{\frac{2}{1+\beta_i}}\right) P(\sigma_i) \end{aligned} \tag{12}$$

式中:  $\boldsymbol{\beta}_{-i}$  表示  $\boldsymbol{\beta}$  除去  $\beta_i$  后,由剩余的元素构成的向量;  $\boldsymbol{\sigma}_{-i}$  表示  $\boldsymbol{\sigma}$  除去  $\sigma_i$  后,由剩余的元素构成的向量;  $Q$  是观测数据个数。

以上推导均是在时域里进行的,在变换域中也是类似的。只需要将时域中各时刻的采样值换成变换域中的各系数即可,而且也用 GGD 来描述这些系数的分布。

这里需要说明的是,推导式(10)~式(12)时,认为相邻时刻的样本(或者变换域相邻的系数)之间互相独立。虽然语音等信号在相邻时刻(或变换域的相邻系数)是相关的,但这并不与实际情况矛盾,因为这里只是没有利用时频结构而已。

$\boldsymbol{\beta}$  和  $\boldsymbol{\sigma}$  既然被看作随机变量,因此存在一个如何选取  $\boldsymbol{\beta}$  和  $\boldsymbol{\sigma}$  的先验概率分布的问题。本文将  $\boldsymbol{\beta}$  和  $\boldsymbol{\sigma}$  的先验分布都取为均匀分布。事实上,当观测数据较多时,先验分布的影响会变得微乎其微。

可以看到,以上各条件概率密度都不是常见的概率分布形式。因此,对式(10)~式(12)抽样时,均采用 Metropolis-Hastings 抽样。所以,本文的抽样方法其实是

一种 Gibbs 和 Metropolis-Hastings 混合抽样。下面将本文第 2 节所描述的盲分离方法称为 GMC 算法。

3 基于非稀疏度评判准则的语音信号盲分离

文献[4]通过观测信号的空间时频分布(STFD)矩阵的对角化来分离信号。文献[5]则根据语音信号在时间域的局部平稳性,通过观测信号协方差矩阵的对角化来分离语音信号。但由于语音在时间上不够稀疏,所以分离效果不好。而且,这两种方法都要求激活源信号个数不超过传感器个数。本文将语音信号做修改离散余弦变换(MDCT),式(1)在 MDCT 域中可写成

$$\bar{\mathbf{x}}(t, k) = \mathbf{A} \bar{\mathbf{s}}(t, k) \tag{13}$$

式中:  $\bar{\mathbf{x}}(t, k)$ ,  $\bar{\mathbf{s}}(t, k)$  分别为观测信号和源信号在 MDCT 域中的系数向量。这里  $t$  是时间帧的序号,  $k$  是每帧系数的序号。本文将  $(t, k)$  构成的平面称为 MDCT 平面,语音信号在整个 MDCT 平面上的能量分布呈现出稀疏性,但从某些局部来看却不具有稀疏性,而呈现出非稀疏的特点。根据语音信号的局部平稳性,MDCT 平面上的局部区域内处于激活状态的源信号始终是相同的。根据这一特点,本文将得到两个重要定理。为了引出这两个定理,这里先给出 3 个定义:

定义 1 MDCT 平面上的某个小的矩形区域  $\Omega_{t,k}$  称为局部区域,即

$$\Omega_{t,k} = \{ (t + \Delta t, k + \Delta k) | 0 \leq \Delta t \leq t_0, 0 \leq \Delta k \leq k_0 \} \tag{14}$$

定义 2 若某个局部区域上激活源信号的个数超过传感器的个数,称该局部区域为非稀疏区。

定义 3 令

$$\begin{aligned} \mathbf{B}_j &= \mathbf{A}_j^{-1} \bar{\mathbf{A}}_j = \begin{bmatrix} b_{1,1}^{(j)} & b_{1,2}^{(j)} & \cdots & b_{1,N-M}^{(j)} \\ b_{2,1}^{(j)} & b_{2,2}^{(j)} & \cdots & b_{2,N-M}^{(j)} \\ \vdots & \vdots & & \vdots \\ b_{M,1}^{(j)} & b_{M,2}^{(j)} & \cdots & b_{M,N-M}^{(j)} \end{bmatrix} \\ &k=1, \cdots, C_N^M \end{aligned} \tag{15}$$

$$\begin{aligned} \mathbf{H}_j &= [h_1^{(j)} \quad h_2^{(j)} \quad \cdots \quad h_{N-M}^{(j)}] \\ &k=1, \cdots, C_N^M \end{aligned} \tag{16}$$

式中:

$$\mathbf{h}_i^{(j)} = [b_{1,i}^{(j)} \quad b_{2,i}^{(j)} \quad \cdots \quad b_{1,i}^{(j)} \quad b_{M,i}^{(j)} \quad b_{2,i}^{(j)} \quad b_{3,i}^{(j)} \quad \cdots, \\ b_{2,i}^{(j)} \quad b_{M,i}^{(j)} \quad \cdots, b_{M-1,i}^{(j)} \quad b_{M,i}^{(j)}]^\text{T} \quad i=1,2,\cdots,N-M$$

$\mathbf{A}_j$  是  $M \times M$  维矩阵,由混合矩阵  $\mathbf{A}$  的任意  $M$  个列向量构成,一共有  $C_N^M$  种组合;  $\bar{\mathbf{A}}_j$  表示  $\mathbf{A}$  除去  $\mathbf{A}_j$  后,由剩余列构成的矩阵。

由以上的定义,可以得到如下定理:

定理 1 若在某个局部区域,找不到一个  $\mathbf{A}_j$  ( $j=1, \cdots, C_N^M$ ),能够使得  $\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-\text{T}}$  是对角矩阵,那么该

局部区域一定是非稀疏区。

证明: $\mathbf{R}_{ss}, \mathbf{R}_{xx}$  分别为源信号的 MDCT 系数在该局部区域的协方差矩阵,根据  $\mathbf{A}$  行满秩和源信号相互独立的前提条件以及定义 3,对于任意的  $\mathbf{A}_j$  都有

$$\begin{aligned} \mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T} &= \mathbf{A}_j^{-1} \mathbf{A} \mathbf{R}_{ss} \mathbf{A}^T \mathbf{A}_j^{-T} = \\ \mathbf{A}_j^{-1} \begin{bmatrix} \mathbf{A}_j & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{A}}_j \end{bmatrix} \begin{bmatrix} \mathbf{A}_j^T \\ \bar{\mathbf{A}}_j^T \end{bmatrix} \mathbf{A}_j^{-T} &= \\ \begin{bmatrix} \mathbf{I} & \mathbf{B}_j \end{bmatrix} \begin{bmatrix} \mathbf{A}_j & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{A}}_j \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{B}_j^T \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_j & \mathbf{B}_j \bar{\mathbf{A}}_j \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{B}_j^T \end{bmatrix} = \\ \mathbf{A}_j + \mathbf{B}_j \bar{\mathbf{A}}_j \mathbf{B}_j^T \end{aligned} \tag{17}$$

式中: $\mathbf{A}_j$  为  $M \times M$  维对角矩阵,其对角线元素为  $\mathbf{A}_j$  各列对应的源信号方差; $\bar{\mathbf{A}}_j$  为  $(N-M) \times (N-M)$  维对角矩阵,其对角线元素为  $\bar{\mathbf{A}}_j$  各列对应的源信号方差; $\mathbf{I}$  是单位矩阵。

假设不存在  $\mathbf{A}_j$  能够使得  $\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T}$  是对角矩阵,且该局部区域激活源信号个数小于或等于传感器个数(等于  $M$ )。那么总存在一个  $\mathbf{A}_j$ ,可以使  $\bar{\mathbf{A}}_j$  对角线上的元素全为 0,从而  $\mathbf{B}_j \bar{\mathbf{A}}_j \mathbf{B}_j^T = 0, \mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T}$  为对角矩阵。这就与假设相矛盾,从而证明定理 1 成立。

定理 2 若  $\mathbf{H}_j (j=1, \dots, C_N^M)$  都列满秩,当某个局部区域是非稀疏区,那么一定找不到一个  $\mathbf{A}_j (j=1, \dots, C_N^M)$ ,能够使得  $\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T}$  是对角矩阵。

证明:由式(17),若  $\lambda_1^{(j)}, \lambda_2^{(j)}, \dots, \lambda_{N-M}^{(j)}$  是  $\bar{\mathbf{A}}_j$  对角线上的元素,那么要使  $\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T}$  为对角阵,  $\mathbf{B}_j \bar{\mathbf{A}}_j \mathbf{B}_j^T$  的非对角线上的元素必须全为 0。根据定义 3,这就要使

$$\mathbf{H}_j [\lambda_1^{(j)}, \lambda_2^{(j)}, \dots, \lambda_{N-M}^{(j)}]^T = 0 \tag{18}$$

但当该局部区域是非稀疏区时,对于任意的  $\mathbf{A}_j, \bar{\mathbf{A}}_j$  对角线上的元素至少有一个不为 0。所以当  $\mathbf{H}_j$  列满秩时,式(18)一定不成立,  $\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T}$  不可能是对角矩阵,从而证明定理 2 成立。

定理 1 和定理 2 说明,若  $\mathbf{H}_j (j=1, \dots, C_N^M)$  都列满秩,不但可以找到非稀疏区,而且可以找到全部的非稀疏区。事实上,即使  $\mathbf{H}_j$  不列满秩,式(18)能够成立的可能性也是很小的,依然可以保证找到几乎所有的非稀疏区。

根据定理 1 和定理 2,本文得到一种语音信号的盲分离方法,本文称之为 NSJC 算法,可描述为:首先将整个 MDCT 平面划分为若干个互不相交的局部区域,然后对各个局部进行非稀疏度评价,评判准则如式(19)所示

$$\xi = \min \left\{ \sum_{u>v} |(\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T})_{u,v}| / \sum_{u,v} |(\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T})_{u,v}| \right\} \tag{19}$$

$j = 1, \dots, K$

式中: $|(\cdot)_{u,v}|$  表示矩阵元素的模值;  $K=C_N^M$ ;  $\min\{\cdot\}$  表示取集合中的最小值。

式(19)可以衡量  $\mathbf{A}_j^{-1} \mathbf{R}_{xx} \mathbf{A}_j^{-T}$  非对角化的程度。当  $\xi$  大于  $\xi_0$  就认为该区域是非稀疏区,将所有非稀疏区上的观测数据作为本文 GMC 算法的输入,可以估计源信号在这些区域上的系数;当  $\xi$  小于  $\xi_0$ ,则通过文献[5]的 SSDP 算法由式(20)估计各激活源信号在该区域的系数,并将非激活源信号的系数置 0。不过这里 SSDP 算法是在 MDCT 域进行的。

$$\bar{\mathbf{s}}_{\text{act}} = \mathbf{A}_{\text{opt}}^{-1} \bar{\mathbf{x}} \tag{20}$$

式中: $\bar{\mathbf{s}}_{\text{act}}, \bar{\mathbf{x}}$  分别是激活源信号和观测信号 MDCT 系数构成的向量; $\mathbf{A}_{\text{opt}}$  是式(19)集合中最小元素所对应的  $\mathbf{A}_j$ 。

最后,通过逆 MDCT 变换得到源信号的时域估计。

4 计算机仿真

为了比较本文提出的算法和文献[5,9]算法的性能,三类源信号被用于仿真。第一类源信号全部为次高斯分布信号(即非稀疏信号);第二类源信号由次高斯分布信号和超高斯分布信号(稀疏信号)共同组成;第三类源信号取自语音信号。若  $s_i(t)$  和  $\hat{s}_i(t)$  分别为源信号和分离信号,  $Q$  为采样个数,则每个信号的分离效果可以用信干比(SIR)衡量

$$(SIR)_{dB} = 10 \lg \left\{ \sum_{i=1}^Q s_i(t)^2 / \sum_{i=1}^Q [(s_i(t) - \hat{s}_i(t))^2] \right\} \tag{21}$$

$i = 1, \dots, N$

以下仿真都是在混合矩阵已知的情况下进行的。本文将文献[9]的算法称为 MLEM 算法。在本文中,文献[5]的 SSDP 算法在 MDCT 域中进行。为了区分,这里称之为 MDCT-SSDP 算法。

(1) 非稀疏信号

非稀疏信号按文献[14]中的方法产生,即由反双曲正弦函数  $y = \arcsinh(x)$  产生。当  $x$  是服从正态分布的随机数时,  $y$  服从次高斯分布。用这种方法产生 4 个源信号,每个信号有 20 000 个样本,并得到 3 个混合信号。算法在时域内进行,共进行 20 次仿真实验,每次仿真实验均采用不同的混合矩阵。将这 20 组数据平均后,得到各源信号的信干比,如表 1 所示。对各源信号的平均信干比再次取平均,得到总体的分离效果(如表 1 最后一列所示)。可以看到,GMC 算法比 MLEM 算法平均提高了 2.09 dB。

表 1 各算法(SIR)<sub>dB</sub>比较

类型	算 法	$S_1$	$S_2$	$S_3$	$S_4$	平均值
第 I 类	MLEM	9.11	6.90	5.04	4.36	6.35
	GMC	10.13	8.72	8.85	6.07	8.44
第 II 类	MLEM	10.22	14.94	4.61	9.60	9.84
	GMC	13.37	17.82	6.85	10.32	12.09

(2) 非稀疏和稀疏混合

4 个源信号如图 1(a) 所示,前两个源信号为非稀疏信号,后两个源信号为稀疏信号。产生 20 次不同的混合矩阵,源信号经混合后得到 3 个混合信号。图 2 是混合矩阵

$$\begin{bmatrix} -0.009\ 7 & 0.925\ 3 & 0.758\ 1 & 0.454\ 7 \\ 0.894\ 6 & 0.370\ 0 & -0.372\ 5 & -0.359\ 7 \\ -0.446\ 8 & 0.078\ 9 & 0.535\ 3 & -0.598\ 4 \end{bmatrix}$$

时的 3 个混合信号。在时域内对混合信号进行分离,从图 1(b)和图 1(c)可看出,非稀疏信号的分离效果比稀疏信号还好。同时看到,GMC 算法的分离效果要好于 MLEM 算法。在表 1 中,给出了两种算法对各源信号分离的平均信干比,GMC 比 MLEM 平均提高了 2.25 dB。

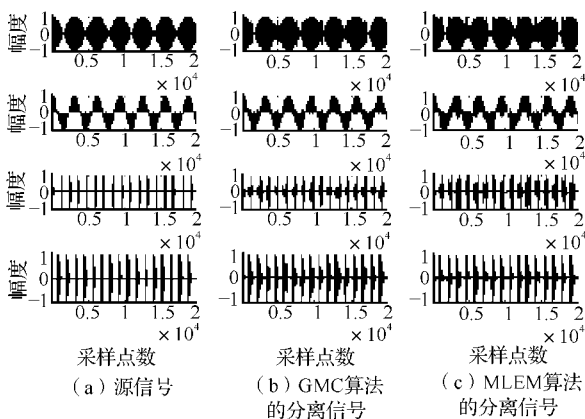


图 1 分离信号比较

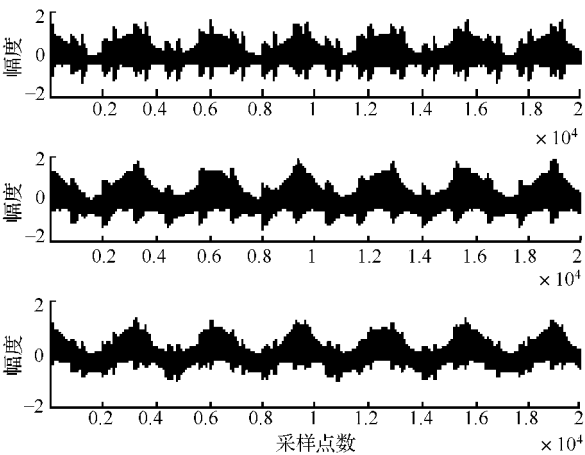


图 2 混合信号

(3) 语音信号

4 个语音信号的采样频率为 16 kHz,时长为 2 s,如图 3(a)所示。通过混合矩阵得到 3 个混合信号,分离算法在 MDCT 域进行。MDCT 采用重叠 50%、宽度为 512 的正弦窗。对于 GMC 算法,仿真时将式(6)中的  $L$ 、 $P$  都取为 20。对于 MDCT-SSDP 和 NSJC 算法,MDCT 平面的局部区域由一个矩形窗定义。仿真发现:当式(14)中的  $t_0=4$ , $k_0=8$  时分离效

果最佳。NSJC 算法的具体实现方式是:MDCT-SSDP 算法在非稀疏区的估计数据被 GMC 算法在该区域的估计数据替换后,再进行逆 MDCT 变换得到各源的估计,这里的  $\xi_0$  取为 0.15。图 3(b)和图 3(c)分别是 MLEM 算法和 NSJC 算法的分离结果。可以看到,NSJC 算法的分离效果更好。图 4 是混合信号,混合矩阵与图 2 的混合矩阵相同。

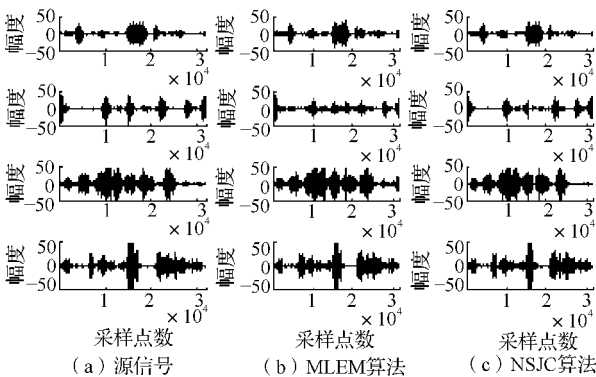


图 3 语音信号的分离结果比较

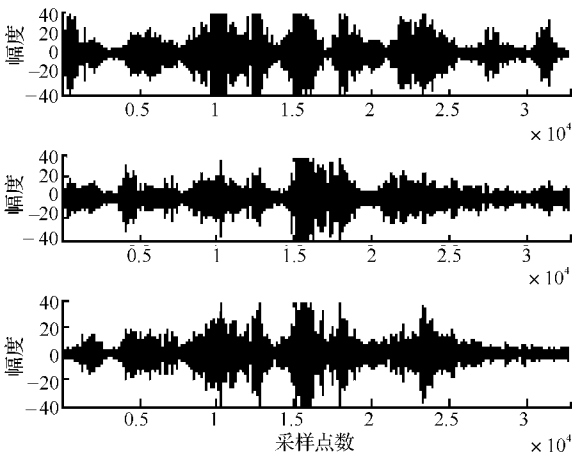


图 4 语音混合信号

进行 20 次实验,每次产生不同的混合矩阵。图 5(a)和图 5(b)是各算法的平均信干比 SIR 和信噪比 SNR 的关系曲线图。这里考虑了两种噪声源,图 5(a)是高斯白噪声,图 5(b)是高斯有色噪声。仿真时采用文献[9]的方法估计混合矩阵,仿真中考虑了混合矩阵的估计误差。从图 5(a)可以看到,GMC 算法的分离效果明显好于 MLEM 算法。虽然在信噪比较高时 GMC 算法比 MDCT-SSDP 算法稍差,但在低信噪比时明显好于 MDCT-SSDP 算法。NSJC 算法在信噪比较高时,具有最好的分离性能。这是因为 GMC 算法在非稀疏区的分离性能更好,而 MDCT-SSDP 算法则在稀疏区更具优势,而 NSJC 算法利用非稀疏评判准则,在不同区域采用不同算法,综合了 GMC 算法和 MDCT-SSDP 算法的优点,因此性能得到了进一步提高。从图 5(a)还看到,随着信噪

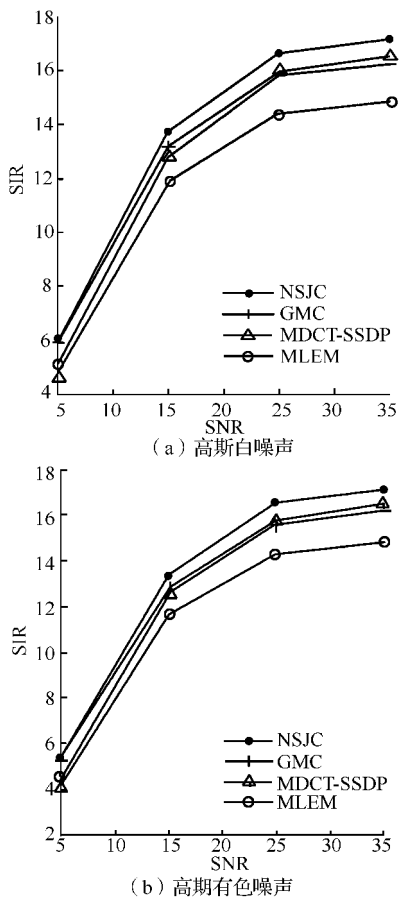


图 5 各算法分离性能比较

比的降低,MDCT-SSDP 算法性能下降最快,这是因为当噪声增加以及混合矩阵估计误差增大时,MDCT-SSDP 算法判断激活信源的准确度会下降。当激活信源的判断出现错误时,会带来更大的估计误差,因此分离性能下降最快。基于类似的原因,NSJC 算法寻找非稀疏区的准确度也会下降,因此分离性能下降也较快。但是由于 NSJC 算法结合了 GMC 算法的优点,所以分离性能的下降速度相比 NSJC 算法要小一些。可以看到,在信噪比为 5 dB 时,GMC 算法的分离性能最好,而 MDCT-SSDP 算法的分离性能则变得最差。

从图 5(b)看到,各算法的性能比较与高斯白噪声的情况类似,但同时看到,相比白色噪声,有色噪声对欠定盲分离算法的分离性能影响更大。当信噪比为 5 dB 时,各算法的 SIR 比图 5(a)平均下降了 0.8 dB。

图 6 是 GMC 算法的各源信号 GGD 参数  $\beta$ 、 $\sigma$  的迭代曲线。可以看到,经过 15 步迭代后,两个参数都开始收敛。

(4) 鲁棒性

图 7 是参数取不同的初始值时,GMC 算法和 MLEM 算法在分离非稀疏信号时的参数迭代曲线。从图 7(a)看到,对于不同的初始值,GMC 算法的收敛

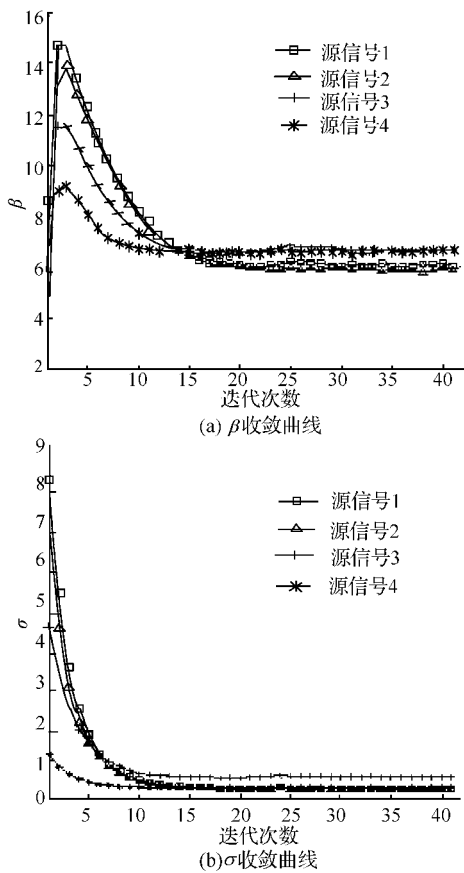


图 6 语音信号 GGD 参数收敛曲线

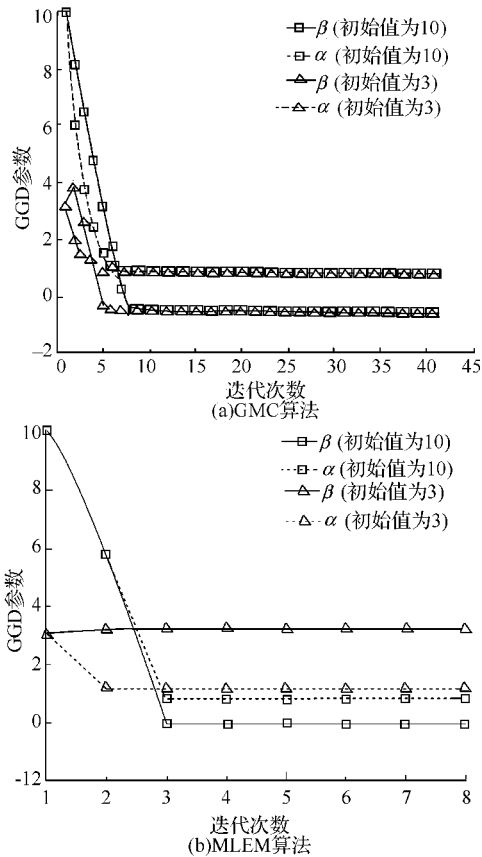


图 7 鲁棒性比较

结果完全一样,表明 GMC 算法的性能稳定,鲁棒性

好。但从图 7(b)看到,MLEM 算法在不同的初始值下, $\beta, \sigma$  的收敛结果明显不同。由于源信号是非稀疏信号, $\beta$  的收敛值应该小于 0。但从图 7(b)看到,当初始值取为 5 时, $\beta$  的收敛值明显大于 0,表明算法陷入了局部极值点。

( 5 )算法运算时间

经过多次仿真,GMC 算法均在 5~20 步开始收敛。为了取平均值,GMC 算法运行 40 步。经过仿真发现,MLEM 算法并不总能如文献[7]所述,在第 2 步就可以收敛,有时要在第 6、第 7 步才能收敛,本文仿真时运行 7 步。虽然 MLEM 算法迭代次数少一些,但由于 MLEM 每一步都需要进行一次辅助函数的优化,因此总的运算时间反而更长。在处理器为 Pentium(R)Dual-Core T4500 2.3 GHz,内存为 2G 的硬件配置条件下,对 32 000 个采样数据进行盲分离,GMC 算法的运算时间为 9 min 左右,而 MLEM 算法需要运行 16 min,表明本文方法计算复杂度更低。

5 结 束 语

本文在贝叶斯框架下提出了可以在欠定情形下分离非稀疏信号的 GMC 算法,通过 MCMC 方法估计各源信号的广义高斯模型参数和隐变量,进而获得源信号的估计。这个算法在低信噪比时的分离性能明显好于 MDCT-SSDP 算法。更为重要的是,这个算法允许局部区域的激活信源个数超过传感器个数,因此可以对非稀疏信号进行有效分离。GMC 算法比 MLEM 算法的鲁棒性更好,不易陷入局部极值点,对初始值无依赖,而且具有更低的计算复杂度。本文在 GMC 算法的基础上,进一步提出基于非稀疏度评判准则的 NSJC 算法,利用语音信号的局部平稳性提高了语音分离性能。

参考文献:

[1] Choi S, Cichocki A, Park H M, et al. Blind Source Separation and Independent Component Analysis: A Review [J]. Neural Information Processing, 2005, 6(1):1-57.  
[2] Li Y Q, Wang J. Sequential Blind Extraction of Instantaneously Mixed Sources [J]. IEEE Trans. on Signal Processing, 2002, 50(5): 997-1006.  
[3] Aissa-El-Bey A, Linh-Trung N, Abed-Meraïmet K, et al. Underdetermined Blind Separation of Nondisjoint Sources in the Time-frequency Domain [J]. IEEE Trans. on Signal

Processing, 2007, 55(3): 897-907.  
[4] Peng D Z, Xiang Y. Underdetermined Blind Source Separation Based on Relaxed Sparsity Condition of Sources [J]. IEEE Trans. on Signal Processing, 2009, 57(2):809-814.  
[5] Xiao M, Xie S L, Fu Y L. A Statistically Sparse Decomposition Principle for Underdetermined Blind Source Separation[C] // Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems. New York:IEEE, 2005:165-168.  
[6] Lewicki M S, Sejnowski T J. Learning Overcomplete Representations[J]. Neural Computation,2000,12(2): 337-365.  
[7] Davies M E, Mitianoudis N. Simple Mixture Model for Sparse Overcomplete ICA[C] // IEE Proceedings on Vision, Image and Signal Processing. 2004, 151(1):35-43.  
[8] Ozerov A, Philippe P, Bimbot F, et al. Adaptation of Bayesian Models for Single-channel Source Separation and Its Application to Voice/Music Separation in Popular Songs [J]. IEEE Trans. on Audio, Speech and Language Processing, 2007, 15(5): 1564-1578.  
[9] Kim S G, Yoo C D. Underdetermined Blind Source Separation Based on Subspace Representation [J]. IEEE Trans. on Signal Processing, 2009, 57(7):2604-2614.  
[10] Vincent E. Complex Nonconvex l-p Norm Minimization for Underdetermined Source Separation[C] // Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation. Berlin: Springer-Verlag, 2007: 430-437.  
[11] Andrieu C, DE Freitas N, Doucet A, Jordan M I. An Introduction to MCMC for Machine Learning [J]. Machine Learning, 2003, 50(1/2):5-43.  
[12] Reju V G, Koh S N, Soon I Y. An algorithm for Mixing Matrix Estimation in Instantaneous Blind Source Separation [J]. Signal Processing, 2009, 89(9):1762-1773.  
[13] Arberet S, Gribonval R, Bimbot F. A Robust Method to Count and Locate Audio Sources in a Multichannel Underdetermined Mixture [J]. IEEE Trans. on Signal Processing, 2010, 58(1):1-13.  
[14] Kim S G, Yoo C D. Underdetermined Blind Source Separation Based on Generalized Gaussian Distribution[C] // Proceedings of 2006 IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing. New York:IEEE, 2006: 103-108.

(责任编辑 魏京燕)