

基于奇异值分解的超定盲信号分离¹

朱孝龙 张贤达*

(西安电子科技大学雷达信号处理重点实验室 西安 710071)

*(清华大学自动化系智能技术与系统国家重点实验室 北京 100084)

摘要: 该文研究超定盲信号分离,即观测信号个数不少于源信号个数情况下的盲信号分离问题。作者从分离矩阵的奇异值分解出发,首先提出一种基于独立分量分析的超定盲信号分离代价函数,接着推导了一般梯度学习算法。此后,借助于相对梯度的概念,证明超定盲信号分离与通常的完备盲信号分离具有相同形式的自然梯度算法。仿真试验验证了算法的有效性。

关键词: 盲信号分离,独立分量分析,相对梯度,自然梯度,奇异值分解

中图分类号: TN911.23 **文献标识码:** A **文章编号:** 1009-5896(2004)03-0337-07

Overdetermined Blind Source Separation Based on Singular Value Decomposition

Zhu Xiao-long Zhang Xian-da*

(Key Lab. for Radar Signal Processing, Xidian University, Xi'an 710071, China)

*(State Key Lab of Intelligent Tech. and Sys., Dept. of Automation,
Tsinghua University, Beijing 100084, China)

Abstract The problem of overdetermined Blind Source Separation (BSS) where there are more mixtures than sources is considered. Beginning with the Singular Value Decomposition (SVD) of the separation matrix, a cost function is presented based on Independent Component Analysis (ICA), and then the ordinary gradient learning algorithm is developed. Secondly, resorting to the relative gradient, it is shown that the natural gradient learning algorithm for overdetermined BSS has the same form as that for usual complete BSS, which is verified by simulation results.

Key words Blind source separation, Independent component analysis, Relative gradient, Natural gradient, Singular Value Decomposition (SVD)

1 引言

在源信号和传输信道未知情况下,只利用接收天线输出观测抽取源信号,称为盲信号分离。盲信号分离是当前信号处理研究的热点课题之一,在无线数据通信、雷达、图像、语音、医学以及地震信号处理等领域具有广阔的应用前景。盲信号分离有离线和在线之分。典型的在线技术主要包括熵最大化分析法^[1]、非线性主分量分析法^[2,3]和独立分量分析法^[4-7]三种,而独立分量分析尤其吸引了广泛的研究兴趣,已成为当前在线盲信号分离的主流方法。

盲信号分离通常假定接收天线的个数等于源信号的个数,即现有研究一般只讨论完备的盲信号分离问题。但在大多数实际应用中,源信号的个数是未知的,甚至可能随时间而变化,从而混合矩阵是非方阵和不可逆的。因此,非完备的盲信号分离问题(包括欠完备和超完备或者超定和欠定两种情况)的研究具有重大的现实意义。

¹ 2002-08-30 收到, 2003-03-28 改回
国家自然科学基金资助项目(60072043)

我们注意到, Amari, Zhang 以及 Choi 等人^[8-10]最早开始了这方面的研究, 他们从李群 (Lie group) 和 Stiefel 流形出发, 推导了超定盲信号分离的自然梯度算法. 与他们的方法不同, 本文从分离矩阵的奇异值分解出发, 提出了一种基于独立分量分析的代价函数, 接着推导基于一般梯度的盲信号分离学习算法, 它们是完备的盲信号分离代价函数及对应算法的推广. 然后, 借助于相对梯度^[6], 我们得到了超定盲信号分离的相对梯度算法, 在形式上, 它与通常的完备盲信号分离的自然梯度算法完全相同.

2 超定盲信号分离

2.1 信号模型与假设

考虑不含噪声的线性盲信号分离数学模型^[1-8]:

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k, \quad k = 1, 2, \dots \quad (1)$$

其中 \mathbf{x}_k 和 \mathbf{s}_k 分别表示 m 维的观测信号向量和 n 维的源信号向量, \mathbf{A} 是 $m \times n$ 的混合矩阵.

盲信号分离的基本假设是各个源信号分量相互统计独立, 且最多有一个分量服从高斯分布. 盲信号分离的目的是在混合矩阵未知情况下, 只利用观测数据 \mathbf{x}_k 以及上述假设恢复源信号 \mathbf{s}_k .

由于模型中的乘积因子 \mathbf{A} 和 \mathbf{s}_k 都未知, 因此无法精确辨识源信号 \mathbf{s}_k 各分量的排列顺序、符号和能量, 这就是盲信号分离问题固有的两种不确定性^[4,7]: 信号幅度不确定性和信号排序不确定性. 不失一般性, 假定各个源信号分量具有零均值和单位功率, 从而

$$E\{\mathbf{s}_k \mathbf{s}_k^T\} = \mathbf{I} \quad (2)$$

其中 E 表示数学期望运算, T 表示向量或矩阵的转置, 而 \mathbf{I} 代表单位矩阵.

关于混合矩阵 \mathbf{A} , 除个别文献^[8-12]外, 通常假定它为非奇异方阵. 对于欠定或超完备的盲信号分离^[8,11,12], 即当观测信号个数少于源信号的个数时, 如果对源信号概率分布没有进一步的先验知识可以利用, 即使能正确估计出混合矩阵, 也不可能恢复源信号. 为此, 本文仅限于正定和超定盲信号分离问题的讨论, 此时, \mathbf{A} 为列满秩矩阵 ($m \geq n$).

2.2 基于独立分量分析的代价函数

为实现盲信号分离, 通常用分离矩阵 \mathbf{B}_k 作用于观测数据 \mathbf{x}_k , 使得输出

$$\mathbf{y}_k = \mathbf{B}_k \mathbf{x}_k \quad (3)$$

为源信号向量 \mathbf{s}_k 的拷贝.

独立分量分析^[4,7]用互信息来衡量各输出分量之间的相互独立程度. 利用互信息与微分熵的关系, 我们有下面的代价函数 (为简单起见, 省略时间下标 k):

$$L(\mathbf{y}; \mathbf{B}) = I(y_1, \dots, y_n; \mathbf{B}) = \sum_{i=1}^n H(y_i; \mathbf{B}) - H(\mathbf{y}; \mathbf{B}) \quad (4)$$

其中

$$H(\mathbf{y}; \mathbf{B}) = - \int p(\mathbf{y}; \mathbf{B}) \ln p(\mathbf{y}; \mathbf{B}) d\mathbf{y} \quad (5)$$

$$H(y_i; \mathbf{B}) = - \int p_i(y_i; \mathbf{B}) \ln p_i(y_i; \mathbf{B}) dy_i \quad (6)$$

分别表示 \mathbf{y} 的联合微分熵和边缘微分熵.

因为混合矩阵 A 是 $m \times n$ 的列满秩矩阵, 所以分离矩阵 B 的搜索空间为整个 $n \times m$ 的行满秩矩阵空间. 由矩阵的奇异值分解^[13] 知, 一定存在 $n \times n$ 的正交矩阵 U 和 $m \times m$ 的正交矩阵 V , 使得

$$B = U[\Lambda \ 0]V^T \quad (7)$$

式中 0 表示 $n \times (m - n)$ 的零矩阵, $\Lambda = \text{diag}(\sigma_1, \dots, \sigma_n)$, 而 $\sigma_i (i = 1, \dots, n)$ 为分离矩阵 B 的全部非零奇异值.

将 V 分块, 并记前 n 列构成的子矩阵为 V_1 , 后 $m - n$ 列构成的子矩阵为 V_2 , 则

$$y = Bx = U[\Lambda \ 0]V_1^T x = U\Lambda V_1^T x \quad (8)$$

由上式及微分熵的定义式 (5), 很容易证明 $H(y; B) = \ln |\det(U\Lambda)| + H(V_1^T x)$. 此外, 对于行满秩矩阵 B , 由奇异值分解^[13] 的性质知, $\det(\Lambda) = \sqrt{\det(BB^T)}$. 这样, 代价函数式 (4) 简化为

$$L(y; B) = \sum_{i=1}^n H(y_i; B) - \frac{1}{2} \ln |\det(BB^T)| - H(V_1^T x) \quad (9)$$

根据熵与互信息的关系以及微分熵对正交变换的不变性, 我们有

$$H(V_1^T x) = H(x) + I(V_1^T x; V_2^T x) - H(V_2^T x) \quad (10)$$

因为分离矩阵 $B = U\Lambda V_1^T$ 不依赖于子矩阵 V_2 , 所以可任意选择子矩阵 V_2 , 使得 $H(V_1^T x)$ 不依赖于分离矩阵 B . 具体地, 对于无噪声的线性混合模型式 (1), 我们选择 V_2 为 $R_x = AE\{s_k s_k^T\}A^T = AA^T$ 的零空间, 则 $AA^T V_2 = 0$, 两端同时左乘以列满秩矩阵 A 的广义逆矩阵 $A^+ = (A^T A)^{-1} A^T$ 有 $A^T V_2 = 0$, 进而 $V_2^T x = V_2^T A s = 0$, 显然 $H(V_1^T x) = H(x)$ 不依赖于 B . 对于观测数据含噪声的情况, 如果噪声功率足够低, 能确保 V_1 张成观测数据的信号子空间, V_2 张成噪声子空间, 从而 $I(V_1^T x; V_2^T x) = I(V_1^T s; V_2^T n) = 0$, 则 $H(V_1^T x) = H(x) - H(V_2^T x)$ 也不依赖于 B .

因为 $H(V_1^T x)$ 不是分离矩阵 B 的函数, 它可以从式 (9) 中省去, 因此有下面的超定盲信号分离的代价函数:

$$L(y; B) = \sum_{i=1}^n H(y_i; B) - \frac{1}{2} \ln |\det(BB^T)| \quad (11)$$

显然, 通常的完备盲信号分离的代价函数:

$$L(y; B) = \sum_{i=1}^n H(y_i; B) - \ln |\det(B)| \quad (12)$$

是式 (11) 在 B 取可逆方阵情况下的一个特例. 此外, Zhang 等人^[9] 提出了代价函数:

$$L(y; B) = \sum_{i=1}^n H(y_i; B) - \ln |\det(BE^T)| \quad (13)$$

其中 E 表示李群 $Gl(n, m)$ 下的一个恒等算子. 因为 $\ln |\det(BE^T)| = \frac{1}{2} \ln |\det(BB^T)|$, 所以实现超定盲信号分离的代价函数式 (11) 和式 (13) 是等价的.

2.3 梯度学习算法

本节推导代价函数式(11)最小化的分离矩阵 B 的自适应更新, 即梯度学习算法. 代价函数 $L(\mathbf{y}; B)$ 的微分:

$$dL(\mathbf{y}; B) = L(B + \epsilon) - L(B) = \sum_{i,j} \frac{\partial L(\mathbf{y}; B)}{\partial B_{ij}} dB_{ij} \quad (14)$$

其中 ϵ 表示分离矩阵 B 的微小扰动矩阵, 而 B_{ij} 表示矩阵 B 的第 i 行第 j 列元素.

利用性质 $\partial \ln |\det(\mathbf{W})| / \partial W_{ij} = (\mathbf{W}^{-1})_{ji}$, 可以证明

$$\begin{aligned} \frac{\partial \ln |\det(\mathbf{B}\mathbf{B}^T)|}{\partial B_{ij}} &= \sum_{p,q=1}^n \frac{\partial \ln |\det(\mathbf{B}\mathbf{B}^T)|}{\partial (\mathbf{B}\mathbf{B}^T)_{pq}} \cdot \frac{\partial (\mathbf{B}\mathbf{B}^T)_{pq}}{\partial B_{ij}} \\ &= \sum_{p,q=1}^n [(\mathbf{B}\mathbf{B}^T)^{-1}]_{pq} \cdot \frac{\partial \sum_{l=1}^m B_{pl} B_{ql}}{\partial B_{ij}} \\ &= \sum_{p,q=1}^n [(\mathbf{B}\mathbf{B}^T)^{-1}]_{pq} \cdot (B_{qj} \delta_{pi} + B_{pj} \delta_{qi}) = 2[(\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B}]_{ij} \end{aligned} \quad (15)$$

此外,

$$\frac{\partial \sum_{l=1}^n H(y_l; B)}{\partial B_{ij}} = -\frac{\partial \sum_{l=1}^n E\{\ln p_l(y_l; B)\}}{\partial B_{ij}} = -E\left\{\frac{p'_i(y_i; B)}{p_i(y_i; B)} x_j\right\} \quad (16)$$

其中 p'_i 表示 p_i 的一阶导数. 记 $\phi_i = -p'_i/p_i$, $i = 1, \dots, n$, 则有

$$dL(\mathbf{y}; B) = \sum_{i,j} (E\{\phi_i x_j\} - [(\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B}]_{ij}) dB_{ij} \quad (17)$$

从而代价函数式(11)的一般梯度为

$$\nabla L(\mathbf{y}; B) = E\{\Phi(\mathbf{y})\mathbf{x}^T\} - (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} \quad (18)$$

式中列向量 $\Phi(\mathbf{y}) = [\phi_1, \dots, \phi_n]^T$. 分离矩阵 B 的一般随机梯度更新算法为

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \eta_k \cdot [(\mathbf{B}_k \mathbf{B}_k^T)^{-1} \mathbf{B}_k - \Phi(\mathbf{y}_k) \mathbf{x}_k^T] \quad (19)$$

其中 η_k 为学习步长参数. 不难看出, 通常的一般梯度盲信号分离算法^[1,5,7]是式(19)在行满秩矩阵 B_k 取非奇异方阵情况下的特例.

如果在式(14)中, 分离矩阵取相对扰动 $\epsilon \cdot B$ 而非绝对扰动 ϵ , 则可获得代价函数的相对梯度 $\tilde{\nabla} L(\mathbf{y}; B)$, Cardoso 和 Laheld^[6]证明 $\tilde{\nabla} L(\mathbf{y}; B) = \nabla L(\mathbf{y}; B) \cdot B^T$. 因此有下面的相对梯度学习算法:

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \eta_k \tilde{\nabla} L(\mathbf{y}; \mathbf{B}_k) \mathbf{B}_k = \mathbf{B}_k + \eta_k \cdot [\mathbf{I} - \Phi(\mathbf{y}_k) \mathbf{y}_k^T] \mathbf{B}_k \quad (20)$$

显然, 它与完备盲信号分离的自然梯度算法^[4-7,14]具有相同的形式. 但区别在于, 式(20)中的分离矩阵 B_k 是行满秩的, 而不必要求是非奇异方阵.

我们注意到, 式(20)是 Zhang 等人^[9]从李群出发首先得到的. 与他们的方法不同, 本文由分离矩阵的奇异值分解出发, 提出了代价函数式(11), 并借助于相对梯度得到算法式(20). 两种方法的代价函数虽然等价, 却不尽相同. 相比之下, 本文方法也更容易理解.

与一般梯度学习算法式 (19) 相比, 算法式 (20) 具有如下显著特点^[6,7,14]:

- (1) 不需要矩阵求逆运算, 从而运算速度大大提高, 同时节省内存空间;
- (2) 若初始分离矩阵行满秩, 则分离矩阵行满秩, 因此确保了算法的稳定性;
- (3) 满足串行更新 (serial updating) 规则, 算法的收敛性能不依赖于具体的混合矩阵, 从而是一种等变化^[6]的盲分离算法, 收敛速度大大加快.

特点 (2) 的证明如下. 设 $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^T \mathbf{Y}) = \sum_{ij} \mathbf{X}_{ij} \mathbf{Y}_{ij}$, $|\mathbf{B}\mathbf{B}^T| = \det(\mathbf{B}\mathbf{B}^T)$, 类似于文献 [7], 有

$$\begin{aligned} \frac{d|\mathbf{B}\mathbf{B}^T|}{dt} &= \left\langle \frac{\partial |\mathbf{B}\mathbf{B}^T|}{\partial (\mathbf{B}\mathbf{B}^T)}, \frac{\partial (\mathbf{B}\mathbf{B}^T)}{dt} \right\rangle \\ &= \langle |\mathbf{B}\mathbf{B}^T| (\mathbf{B}\mathbf{B}^T)^{-1}, \eta \mathbf{F}(\mathbf{y}) \mathbf{B} \cdot \mathbf{B}^T + \mathbf{B} \cdot [\eta \mathbf{F}(\mathbf{y}) \mathbf{B}]^T \rangle \\ &= 2\eta |\mathbf{B}\mathbf{B}^T| \text{trace}(\mathbf{F}(\mathbf{y})) \end{aligned}$$

其中 $\mathbf{F}(\mathbf{y}) = \mathbf{I} - \Phi(\mathbf{y})\mathbf{y}^T$. 因为 $|\mathbf{B}_t \mathbf{B}_t^T| = |\mathbf{B}_0 \mathbf{B}_0^T| \exp[2 \int_0^t \eta(\xi) \text{trace}(\mathbf{F}(\mathbf{y}(\xi))) d\xi]$, 所以若初始分离矩阵行满秩, 则当前时刻的分离矩阵也一定是行满秩的. 证毕

3 计算机仿真与分析

为验证超定盲信号分离算法式 (20) 的有效性, 考虑以下信号^[5,7] 的分离: (1) 符号信号 $\text{sgn}(\cos(2\pi 155t))$; (2) 高频正弦信号 $\sin(2\pi 800t)$; (3) 低频正弦信号 $\sin(2\pi 90t)$; (4) 相位调制信号 $\sin(2\pi 300t - 6 \cos(2\pi 60t))$; (5) 幅度调制信号 $\sin(2\pi 9t) \sin(2\pi 300t)$; (6) 在 $[-1, 1]$ 均匀分布的随机噪声信号. 以速率 10kHz 对源信号采样以产生离散源信号 s_k . 图 1 是某一次运行中的源信号 (为清晰起见, 仅给出在样本 5000~6000 之间的源信号).

仿真采用 9 个传感器, 信道即混合矩阵 \mathbf{A} 的各个元素为在 $[-1, 1]$ 区间均匀分布的随机数. 算法的性能用“串音”误差^[5,7]

$$E_{ct} = \sum_{p=1}^n \left[\sum_{q=1}^n \frac{|c_{pq}|}{\max_l |c_{pl}|} - 1 \right] + \sum_{q=1}^n \left[\sum_{p=1}^n \frac{|c_{pq}|}{\max_l |c_{lq}|} - 1 \right] \quad (21)$$

来衡量, 其中 $\mathbf{C}_k = [c_{pq}] = \mathbf{B}_k \mathbf{A}$ 表示合成系统的传递矩阵.

在算法式 (20) 中取激励函数 $\phi_i = y_i^3$, 固定步长 $\eta_k = 40 \times T (T = 10^{-4}$ 为采样周期), 初始分离矩阵 \mathbf{B}_0 为 6×9 的随机矩阵, 其元素在 $[-1, 1]$ 区间均匀分布. 图 2 给出了 200 次独立运行 (每次运行中的随机信号 (6) 以及混合矩阵 \mathbf{A} 都是随机产生的) 平均“串音”误差曲线图. 为说明噪声对算法式 (20) 的影响, 我们设观测信号含零均值的高斯白噪声, 其方差分别为 0.05 和 0.1 (对应的信噪比为 13dB 和 10dB), 为便于比较, 对应的 200 次平均“串音”误差曲线也在图 2 中给出. 图 3 至图 6 给出了其中某一次运行的“串音”误差及各自的分离结果. 不难看出: 在无噪声或者噪声足够低的环境下, 算法式 (20) 实现超定盲信号分离的结果令人满意. 但随着噪声功率的提高, 分离结果逐步变差, 甚至无法达到提取源信号的目的. 这与理论分析是一致的, 因为随着噪声功率的提高, \mathbf{V}_1 不再张成观测数据的信号子空间, \mathbf{V}_2 也不再张成噪声子空间, 从而代价函数的最小化会导致提取噪声和部分信号被抑制.

4 结论

本文从分离矩阵的奇异值分解出发, 提出了一种基于独立分量分析的超定盲信号分离代价函数, 并推导了一般梯度的盲信号分离算法. 它们是通常的完备盲信号分离代价函数及对应算法的推广. 借助于相对梯度的概念, 本文还证明实现完备盲信号分离的自然梯度算法同样适用

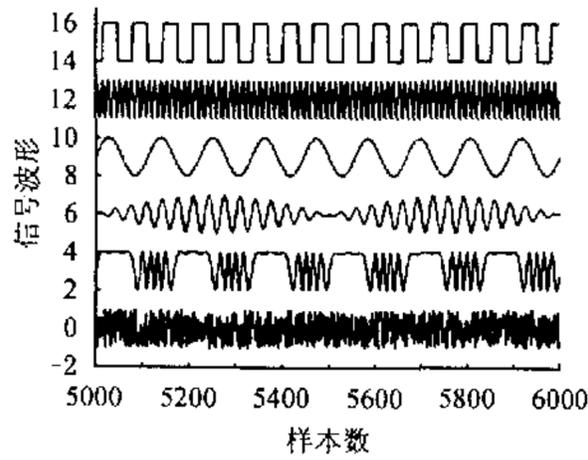


图 1 六个源信号的波形图

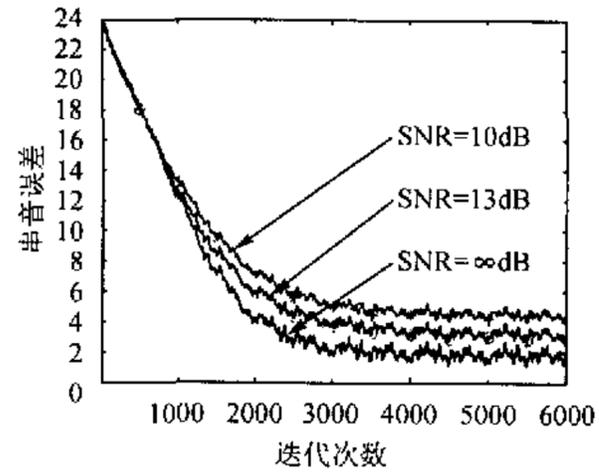


图 2 200 次独立运行串音误差均值曲线图

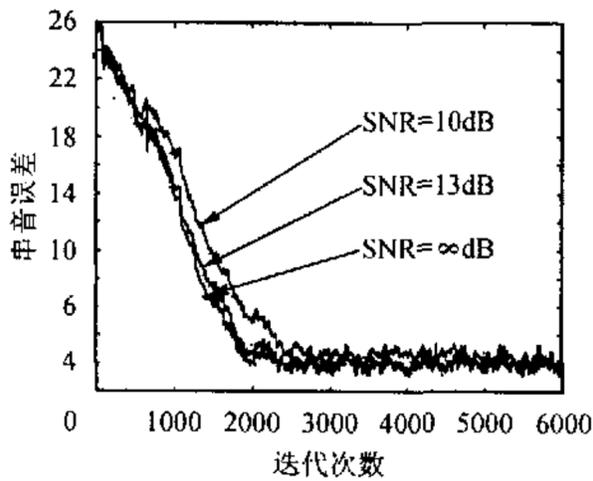


图 3 某一次运行串音误差曲线图

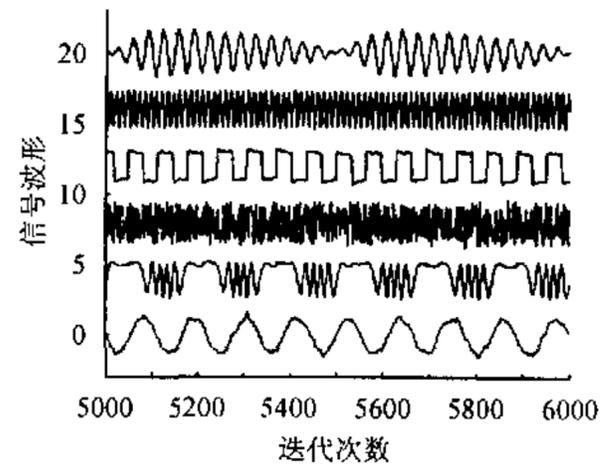


图 4 无噪声分离结果, SNR=∞

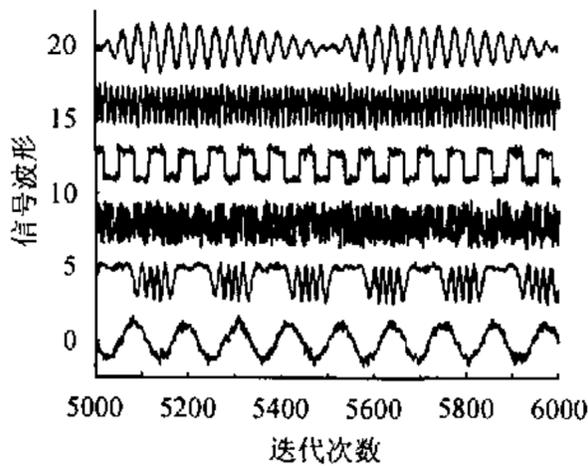


图 5 分离结果, SNR=13dB

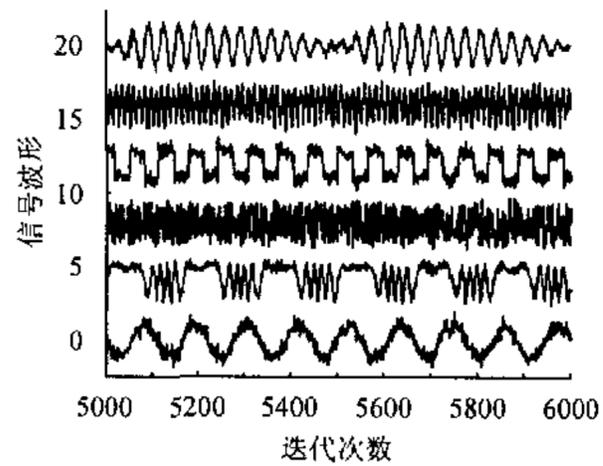


图 6 分离结果, SNR=10dB

于超定盲信号分离问题。如果噪声功率足够低, 则算法能够在抑制噪声的同时提取信号。但随着噪声功率的提高, 算法性能逐步恶化, 直至失效。仿真实验验证了算法的有效性。

参 考 文 献

[1] Bell A J, Sejnowski T J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 1995, 7(6): 1129-1159.

- [2] Karhunen J, Joutsensalo J. Representation and separation of signals using nonlinear pca type learning. *Neural Networks*, 1994, 7(1): 113-127.
- [3] Karhunen J, Pajunen J, Oja E. The nonlinear PCA criterion in blind source separation: Relations with other approaches. *Neurocomputing*, 1998, 22(1): 5-20.
- [4] Comon P. Independent component analysis, a new concept? *Signal Processing*, 1994, 36(3): 287-314.
- [5] Amari S I, Cichocki A, Yang H H. A new learning algorithm for blind signal separation. In D. S. Touretzky, M. C. Mozer & M. E. Hasselmo (Eds.), *Advance in Neural Information Processing Systems*, Cambridge, MA: MIT Press, 1996: 757-763.
- [6] Cardoso J F, Laheld B. Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 1996, 44(12): 3017-3030.
- [7] Yang H H, Amari S I. Adaptive on-line learning algorithms for blind separation—maximum entropy and minimum mutual information. *Neural Computation*, 1997, 9(5): 1457-1482.
- [8] Amari S I. Natural gradient learning for over- and under-complete bases in ICA. *Neural Computation*, 1999, 11(8): 1875-1883.
- [9] Zhang L Q, Cichocki A, Amari S I. Natural gradient algorithm for blind separation of overdetermined mixture with additive noise. *IEEE Signal Processing Letters*, 1999, 6(11): 293-295.
- [10] Choi S, Cichocki A, Zhang L Q, Amari S I. Approximate maximum likelihood source separation using natural gradient. 3rd IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, Taoyuan, Taiwan, 2001: 20-23.
- [11] Lee T W, Lewicki M S, Girolami M, Sejnowski T J. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing Letters*, 1999, 6(4): 87-90.
- [12] Lewicki M S, Sejnowski T J. Learning overcomplete representation. *Neural Computation*, 2000, 12(2): 337-365.
- [13] 张贤达. 信号处理中的线性代数. 北京: 科学出版社, 1997, 第6章.
- [14] Amari S I. Natural gradient works efficiently in learning. *Neural Computation*, 1998, 10(2): 251-276.

朱孝龙: 男, 1976年生, 博士生, 研究方向为盲信号处理、阵列信号处理等.

张贤达: 男, 1946年生, 教授, 博士生导师, 研究方向为信号处理及其在雷达和通信中的应用.