

西安电子科技大学
博士学位论文
盲自适应信号分离的并行实现方法研究
姓名：朱孝龙
申请学位级别：博士
专业：信号与信息处理
指导教师：张贤达
20030301

## 摘 要

在源信号和传输信道未知情况下, 仅利用接收天线的观测数据恢复源信号, 称为盲信号分离。盲信号分离因为在无线通信、雷达和声纳、图像、语音以及医学信号处理等诸多领域具有广阔的应用前景, 从而引起了信号处理学界和神经网络学界的共同兴趣。本文围绕这一热点课题展开, 并把研究重点放在盲信号分离的自适应并行实现方面, 主要工作概括如下:

1. 系统阐述了盲信号分离问题。从信号模型出发, 分析了盲信号分离的不确定性和基本假设条件, 研究了盲信号分离的对比函数理论和局部稳定性理论, 总结了不同形式盲信号分离算法的局部稳定性条件, 并给出了算法的性能评价测度。
2. 研究了盲信号分离学习算法的学习速率参数的最佳确定问题。首先定义了描述信号分离状态的相依性测度, 并根据此测度将传统的标量形式的学习速率推广到二维学习速率矩阵, 进而提出了一种分阶段学习的新思想, 它把整个信号分离过程分成三个阶段进行: 初始阶段、捕捉阶段和跟踪阶段。由于学习速率矩阵的元素由信号的分离状态决定, 它既加快了信号分离前半阶段的收敛, 又提高了后半阶段的跟踪和恢复的精度, 因此分阶段学习有效地解决了最小均方 (LMS) 型的盲信号分离算法收敛速度和稳态性能之间的矛盾。
3. 推导了两种基于自然梯度的盲信号分离 RLS 算法。首先推导了一种递归最小二乘 (RLS) 型的预白化算法, 理论证明其输出是希望的白色信号。然后从非线性主分量分析的对比函数出发, 利用分离矩阵的正交约束, 推导了一种基于自然梯度的 RLS 型算法 (RLS-I 算法), 理论证明它能实现预白化的盲信号分离。将 RLS 型的预白化算法和 RLS-I 算法相结合, 经过合理近似又推导出一种新的盲信号分离 RLS 算法 (RLS-II 算法), 它不需要对观测数据进行白化预处理, 并且在理论上是等变化的, 满足分离矩阵的非奇异性条件。因为自然梯度考虑了对比函数的参数结构从而比一般梯度更适合于盲信号分离, 也因为 RLS 型算法比 LMS 型算法具有更快的收敛速度和更强的跟踪能力, 所以提出的两种新算法工作相当出色, 大量的计算机仿真验证了这一点。

4. 研究了混合峰度信号的盲分离问题。从归一化峰度的绝对值之和对比函数出发, 利用正交约束下的自然梯度和现有的串行白化算法, 推导出一种新颖的盲信号分离算法, 它能实现混合亚高斯信号和超高斯信号的盲分离。由于新算法的激励函数不仅是信号自适应的, 而且自动满足局部稳定性条件和稳健性条件, 因此与现有的同类算法相比, 新算法具有运算量小, 收敛性好和稳健性高等优点。
5. 研究了观测信号不少于源信号个数的超定盲信号分离问题。首先从分离矩阵的奇异值分解出发, 提出一种新的基于独立分量分析的盲信号分离对比函数, 它是现有的正定和超定盲信号分离对比函数的推广。然后, 借助于相对梯度严格证明, 完备盲信号分离的自然梯度算法同样适用于超定盲信号分离问题。在混合矩阵列满秩, 源信号是各态历经的平稳随机过程条件下, 结合 Ye 等人提出的源信号个数的确定方法, 我们还成功解决了源信号个数未知甚至动态变化情况下的超定盲信号分离问题。

关键词: 盲信号分离, 独立分量分析, 非线性主分量分析, 自然梯度, 相对梯度, 递归最小二乘, 亚高斯信号, 超高斯信号, 白化, 局部稳定性, 稳健性, 奇异值分解

---

## ABSTRACT

Blind source separation (BSS) consists of recovering mutually independent but otherwise unobserved source signals from their mixtures without any knowledge of the channel. BSS has attracted growing attention in statistical signal processing and unsupervised neural learning, since it is a fundamental problem encountered in various fields, such as wireless telecommunication systems, sonar and radar systems, audio and acoustics, image enhancement, biomedical signal processing, and so on. In this dissertation, we investigate the problem of blind source separation, with great emphasis on its adaptive parallel implementation. The primary contributions and original ideas included in this dissertation are summarized below:

1. The BSS problem is systematically addressed. Based on the signal model, we analyze the indeterminacy and the basic assumptions inherent in the BSS problem. After the contrast function theory and the local stability theory, we also discuss various forms of BSS algorithms and the corresponding local stability conditions. Finally, we present two performance indexes used to measure the BSS algorithms.
2. The optimal learning rate determination problem is studied. To describe the separatedness of source signals, several signal dependence measures are defined, whose recursive update algorithms are developed as well. Based on the signal dependence measures, the traditional scalar learning rate is generalized to a matrix, and then a grading learning algorithm is proposed, whose central idea is to achieve BSS by dividing the whole learning process into three stages: initial stage, capture stage and tracking stage. Since the elements of the learning rate matrix are determined by the separatedness of source signals, the grading learning can provide not only faster convergence, but also better steady-state accuracy. Hence, it is a satisfactory solution to the contradiction between the convergence rate and steady-state performance that lies in the least-mean-square (LMS)-type algorithms for adaptive BSS.
3. Two natural-gradient-based recursive least-squares (RLS) algorithms are proposed. Firstly, we present a new RLS-type prewhitening algorithm, whose output is the expected white vector. Secondly, beginning with the nonlinear principle component analysis contrast function, and exploiting the orthonormal constraint of the separating matrix, we develop a new natural-gradient-based RLS-type algorithm (denoted as RLS-I algorithm), which can be shown theoretically to perform prewhitened BSS.

Combining the RLS-type prewhitening algorithm and RLS-I algorithm, we obtain after reasonable approximations, a novel RLS-type algorithm (denoted as RLS-II algorithm) for BSS, which does not require additional prewhitening process of the observed mixtures. It is proved that the RLS-II algorithm is equivariant and has the property of keeping the separating matrix from becoming singular. On one hand, the natural gradient learning makes full use of the parametric structure of contrast function, and thus is superior to the ordinary gradient learning. On the other hand, the RLS algorithm is better than the LMS algorithm in convergence rate and tracking capability. Therefore, the two newly proposed RLS-type algorithms perform BSS more satisfactorily, which is verified by extensive computer simulations.

4. The problem of blind separation of sources with mixed kurtosis signs is considered. Applying the sum of absolute values of the normalized kurtoses as a contrast function for prewhitened BSS, and exploiting the natural gradient learning with orthogonal constraint together with the existing serial prewhitening algorithm, we develop a new equivariant algorithm, which has moderate computational load, and can perform blind separation of mixed sub-Gaussian and super-Gaussian sources. Since the proposed algorithm uses a signal-adaptive activation function that automatically satisfies the local stability condition and the robustness condition, it works more robustly and efficiently than the existing algorithms.
5. The overdetermined BSS problem is studied. By singular value decomposition of the separating matrix, we present a novel independent-component-analysis (ICA)-based contrast function for BSS, which is a generalization of the existing determined and overdetermined ones. Resorting to the relative gradient, we show that the natural gradient algorithm for overdetermined BSS has the same form as that for usual complete BSS. Finally, if all source signals are ergodic and wide-sense stationary, and the mixing matrix is of full column rank, then we can use the method presented recently by Ye *et al.* to determine the source number, and thus can apply the newly proposed natural gradient algorithm for overdetermined BSS to solve the difficulty problem of BSS with unknown and dynamically changing number of source signals.

**Key words:** blind source separation (BSS), independent component analysis (ICA), nonlinear principle component analysis, natural gradient, relative gradient, recursive least-squares (RLS), sub-Gaussian signal, super-Gaussian signal, prewhitening, local stability, robustness, singular value decomposition (SVD)

## 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了本文中特别加以标注和致谢中所罗列的内容外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中做了明确的说明并表示了谢意。

本人签名: 朱孝松

日期: 2003.3.13

## 关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容；可以允许采用影印、缩印或其它复制手段保存论文。（保密的论文在解密后遵守此规定）

本人签名: 朱孝松

日期: 2003.3.13

导师签名: 张兴达

日期: 2003年3月15日



## 第一章 绪 论

### §1.1 研究历史与现状

1986年4月13日到16日, Herault和Jutten<sup>[76]</sup>在美国举行的以计算神经网络为主题的一个国际会议上, 提出了一种反馈神经网络模型和一种基于Hebb学习规则的学习算法, 它能实现两个独立源信号的盲分离。Herault和Jutten的工作开辟了一个崭新的研究课题——盲信号分离(BSS: Blind Source Separation)。

盲信号分离研究最初的动机是试图解决鸡尾酒会(cocktail party)<sup>[67]</sup>问题, 即在一个同时存在众多说话者的环境中, 听者在眼睛等器官的协助下, 用耳朵来收听他所感兴趣的一个或多个说话者的讲话。显然, 这个问题在通信<sup>[31,126,134,170,171]</sup>、雷达和声纳<sup>[34]</sup>、图像<sup>[74,89]</sup>、语音<sup>[61,162]</sup>、医学<sup>[12,111]</sup>以及地震信号处理等诸多领域都具有广泛应用, 因此对盲信号分离问题的研究具有重大的现实意义。

Herault和Jutten<sup>[76,91]</sup>提出的反馈神经网络只能实现两个源信号的盲分离, 并且用于Hebb训练的奇次非线性函数的选取具有随意性, 缺乏理论解释。此外, Cohen和Andreou<sup>[47]</sup>研究指出, Herault和Jutten的神经网络适合于分离两个亚高斯信号, 当源信号含超高斯信号时, 分离效果将显著下降。为克服Herault和Jutten的盲信号分离算法的不足, 信号处理学界和神经网络学界进行了大量而积极的探索。Tong等人<sup>[142]</sup>分析了盲信号分离问题的不确定性和可分离性, 并给出了一种基于高阶统计量的矩阵代数方法。Cardoso等人<sup>[21]</sup>提出了一类基于高阶统计量的联合对角化盲分离方法, 并用于波束形成。Comon<sup>[48]</sup>详细分析了线性瞬时混合模型下的盲信号分离问题, 明确提出了独立分量分析(ICA: Independent Component Analysis)的概念, 并描述了盲信号分离问题的目标函数即对比函数(contrast function)的性质, 进而证明以Kullback-Leiber散度表征的互信息的相反数是盲信号分离的对比函数。通过对概率密度函数的高阶近似, Comon提出了一种基于特征分解的独立分量分析方法。这种方法需要对观测信号进行白化预处理, 而且是一种批处理算法, 运算量大, 实时性差。为克服这些缺点, Amari等人<sup>[3]</sup>利用截短的Gram-Charlier展开逼近输出信号的概率密度函数, 提出了一种基于前馈神经网络的在线算法, 但仅局限于分离亚高斯信号。Bell等人<sup>[14]</sup>基于信息论, 通过最大化非线性输出节点的熵, 提出了一类熵最大化的在线盲信号分离算法, 但收敛慢, 而且需要矩阵求逆运算, 稳定性也比较差。Yang等人<sup>[153]</sup>证明, 熵最大化分析和独立分量分析是等价的。Cichocki等人<sup>[38,39]</sup>通过大量的仿真提出了一种稳健的

在线盲信号分离算法，形式十分简单，它不需要矩阵求逆运算，即使混合矩阵的条件数很差也能有效地工作，但作者没能从理论上解释算法的工作原理。后来，Amari<sup>[2,6~8]</sup>从信息几何的角度解决了这个问题，并明确了自然梯度的概念，提出了一类盲信号分离的自然梯度算法。此外，Cardoso等人<sup>[22]</sup>从等变化性（equivariance）出发，定义了相对梯度，并提出了盲信号分离的相对梯度算法。业已证明，自然梯度和相对梯度两种算法是等价的。受独立分量分析的启发，Karhunen等人<sup>[92~96,122,124]</sup>在主分量分析<sup>[35,89,121]</sup>（PCA: Principle Component Analysis）方法中引入某种非线性，提出了一类非线性主分量分析的盲信号分离算法。除了同步提取所有源信号的并行算法外，盲信号分离还可以采用一次只提取一个源信号的串行算法<sup>[52]</sup>，比较典型的是Hyvarinen等人<sup>[79,80,86]</sup>的定点（fixed-point）算法和Thawonmas等人<sup>[141]</sup>的级联神经网络提取算法，前者具有二次收敛性，后者克服了串行算法的误差积累问题。在算法研究的同时，盲信号分离研究还诞生了许多理论，如Cao等人<sup>[18]</sup>的行可分解理论、Moreau等人<sup>[116,117]</sup>的对比函数理论、Amari等人<sup>[4]</sup>的半参数统计理论、Cardoso等人<sup>[5,22,27,30,54,78,120]</sup>的局部稳定性理论等。这些理论反过来又促进了许多算法的出现，包括学习速率调整算法<sup>[41,56,90,112,118]</sup>、混合亚高斯和超高斯源信号的盲分离算法<sup>[37,43~45,54,68,101,113,141,148]</sup>以及混合矩阵列亏损病态情况下的盲提取算法<sup>[18,100,104~106]</sup>等。

盲信号分离问题的研究范围不断扩大。Souloumiac<sup>[136]</sup>通过对两个协方差矩阵的联合对角化，提出了一种非平稳源信号的盲分离算法。Matsuoka等人<sup>[115]</sup>基于高斯互信息提出了一种随机梯度算法，Pham等人<sup>[131]</sup>基于极大似然提出了一种具有超效率（super-efficiency）的拟牛顿算法，两者都能实现非平稳源信号的盲分离。Everson等人<sup>[63]</sup>用马尔可夫过程，Zhang等人<sup>[46,149,165,166]</sup>用状态空间模型对信道建模，以解决混合矩阵动态变化情况下的盲信号分离问题。Hyvarinen等人<sup>[87,89]</sup>使用高斯矩（Gaussian moment）函数来应付观测信号中的平稳噪声。冯大政等人<sup>[169]</sup>提出多阶段分解算法，实现空间色噪声背景下的盲信号分离问题。

上面提到的大多数算法考虑的是线性瞬时混合模型。非瞬时或卷积混合模型下的盲信号分离实际上就是盲反卷积<sup>[73,97,158,159]</sup>问题，典型算法有Amari等人<sup>[10]</sup>的自然梯度算法、Douglas等人<sup>[59]</sup>的梯度自适应算法、Smaragdis等人<sup>[135,143]</sup>的傅立叶变换方法、Ehlers等人<sup>[61]</sup>的空时解相关方法、Tugnait等人<sup>[144~146]</sup>的空时高阶统计量方法和Parra等人<sup>[128]</sup>的多重自适应解相关（MAD: Multiple Adaptive Decorrelation）算法等。非线性混合模型下的盲信号分离问题更加复杂，较早涉及这方面研究的是Burel<sup>[17]</sup>，他用一个两层感知器和基于误差后向传输思想的无导师训练算法，通过梯度下降法优化统计独立的测度函数，从而实现非线性混合信号的盲分离。Pajunen等人<sup>[123]</sup>提出用自组织映射（SOM: Self-organization Map）网络从非线性混合信号中恢复源信号，但网络复杂性呈指数增长，并且分离连续性源信号时插值误差严重。Yang等人<sup>[154]</sup>利用两层



感知器网络结构,提出了基于最大熵和最小互信息测度函数的信息后向传输算法,在恰当选择非线性激励函数的前提下能够分离一些特定非线性混合的源信号。Tan等人<sup>[139]</sup>提出用径向基神经网络解决非线性盲分离问题。Taleb等人<sup>[137,138]</sup>就所谓的后非线性混合模型进行了探索。此外,Hyvarinen<sup>[88]</sup>发展了非线性盲分离问题的存在性和惟一性理论,指出非线性盲信号分离问题的解总是存在的,对于两个源信号的情况,在一系列的约束条件下还可获得惟一解。应该说这方面的研究刚刚起步,还有大量的实际问题没有解决。

近年来,在国内也有一些学者从事盲信号处理理论和应用技术的研究。凌燮亭<sup>[172]</sup>用反馈神经网络的Hebb学习算法,实现了近场情况下一般信号的盲分离。何振亚等人<sup>[74,75,170]</sup>在基于特征分析和高阶谱的盲信号分离和盲反卷积研究中,提出了一系列的基于高阶统计量和信息论的测度与算法,并用于解决盲系统参数估计和盲波束形成问题。胡光锐等人<sup>[162]</sup>基于反馈神经网络用混合高斯模型估计概率密度函数,提出一种语音信号的盲分离算法。冯大政等人<sup>[169]</sup>通过系统化的多阶段分解和多阶段重构,提出了一种多阶段分解算法,能实现空间色噪声背景下的盲信号分离问题。焦李成等人<sup>[168,171]</sup>将盲信号分离理论和算法应用于解决盲多用户检测问题。近来,刘琚和何振亚<sup>[173]</sup>,张贤达和保铮<sup>[174]</sup>分别撰写了关于盲信号分离的综述性文献,张贤达和保铮<sup>[177]</sup>在他们的专著《通信信号处理》中对盲信号分离和盲反卷积问题也作了详细介绍,这些工作都极大地推动了国内盲信号处理问题的开展和研究。

## §1.2 本论文的主要工作

本论文主要针对线性瞬时混合模型下的盲信号分离问题展开,特别地,我们把工作重点集中在自适应并行实现方面。各章内容具体安排如下:

第二章介绍盲信号分离的基础知识。重点讨论盲信号分离问题的信号模型、对比函数理论、盲信号分离算法的局部稳定性条件以及性能评价测度。

第三章研究盲信号分离学习算法的学习速率参数的最佳确定问题。首先定义描述信号分离状态的相依性测度,并根据此测度将标量形式的学习速率推广至二维学习速率矩阵,进而提出一种分阶段学习的新思想,它把整个信号分离过程分成三个阶段进行:初始阶段、捕捉阶段和跟踪阶段。由于学习速率矩阵的元素由信号的分离状态决定,它既加快了信号分离前半阶段的收敛速度,又提高了后半阶段的跟踪和恢复的精度,因此分阶段学习的盲信号分离算法有效地解决了LMS型的盲分离学习算法收敛速度和稳态性能之间的矛盾。

第四章推导基于自然梯度的盲信号分离RLS算法。首先推导一种RLS型的预白化算法,理论证明其输出是希望的白色信号。接着从非线性主分量分析的对比函数出发,利用分离矩阵的标准正交约束,推导一种基于自然梯度的RLS型算法(RLS-I算法),理论证明它能实现预白化的盲信号分离。将RLS型的预白化算法和RLS-I算法相结合,经合理近似又推导出一种新的盲信号分离RLS算法(RLS-II算法),它不需要对观测数据进行白化预处理就能实现信号的盲分离。因为自然梯度比一般梯度更适合于盲信号分离,也因为RLS型算法比LMS型的算法具有更快的收敛速度和更强的跟踪能力,所以提出的两种新算法工作相当出色。

第五章研究混合亚高斯和超高斯信号的盲分离问题。我们从归一化峰度的绝对值之和对比函数出发,推导一种新颖的盲信号分离算法,以实现混合亚高斯和超高斯信号的自适应并行盲分离。由于新算法的激励函数兼备了信号自适应性以及自动满足局部稳定性条件和稳健性条件这三个显著特点,因此优越于现有的同类算法。

第六章研究观测信号比源信号个数多的超定盲信号分离问题。首先从分离矩阵的奇异值分解出发,提出一种新的基于独立分量分析的盲信号分离对比函数,并严格推导对应的随机梯度学习算法。然后,借助于相对梯度的概念,理论证明完备盲信号分离的自然梯度算法同样适用于超定盲信号分离问题。在混合矩阵列满秩,源信号是各态历经的平稳随机过程条件下,结合Ye等人提出的源信号个数的确定方法,我们还成功解决了源信号个数未知甚至动态变化情况下的超定盲信号分离问题。

## 第二章 盲信号分离的基础知识

### §2.1 引言

盲信号分离问题一诞生, 就很快引起了信号处理学界和神经网络学界的广泛兴趣, 特别是近十年来, 理论研究和实际应用两方面都获得了长足的发展。简而言之, 盲信号分离就是根据观测到的混合数据向量确定一个变换, 以恢复源信号。典型情况下, 观测数据向量是一组传感器的输出, 其中每个传感器接收到的是源信号的不同组合。术语“盲”有两重含义:

- 源信号不能被观测;
- 源信号如何混合是未知的。

显然, 当从源信号到传感器之间的传输信道很难建立数学模型, 或者关于传输的先验知识无法获得时, 盲信号分离是一种很自然的选择。当前, 盲信号分离在无线数据通信<sup>[31,126,134,171]</sup>、雷达和声纳<sup>[34]</sup>、图像<sup>[74,89]</sup>、语音<sup>[61,162]</sup>、医学<sup>[12,111]</sup>以及地震信号处理等领域得到了广泛应用。

为解决盲信号分离问题提出了许多算法。根据信号的提取方式不同, 盲信号分离算法可分为串行算法和并行算法两大类。前者按顺序逐个提取信号, 后者实现所有信号的同步恢复。根据工作方式不同, 盲信号分离算法有离线和在线之分, 前者是批处理算法, 后者是自适应算法。根据适用的范围不同, 盲信号分离算法可分成单一峰度信号的盲分离算法和混合峰度信号的盲分离算法。前者要求所有的源信号要么都是亚高斯信号要么都是超高斯信号, 而后者则没有这种限制条件。

现有的典型算法中, Cardoso等人<sup>[21,24]</sup>的特征矩阵近似联合对角化(JADE: Joint Approximate Diagonalization of Eigenmatrices)算法属于批处理并行算法, Hyvarinen等人<sup>[79,86]</sup>的快速定点算法和冯大政等人<sup>[169]</sup>的多阶段分解算法属于批处理串行算法, 而Amari等人<sup>[3-9,36-39,59,114,153,156]</sup>的自然梯度算法、Cardoso等人<sup>[22]</sup>的等变化(EASI: Equivariant Adaptive Separation via Independence)算法、张贤达等人的分阶段学习算法<sup>[178]</sup>和作者的递归最小二乘算法<sup>[167]</sup>则属于自适应并行算法。典型的混合峰度信号的盲分离算法包括Hyvarinen等人<sup>[79,86]</sup>的快速定点算法、Lee等人<sup>[68,101]</sup>的推广的独立分量分析(ICA: Independent Component Analysis)算法、Choi等人<sup>[37]</sup>的灵活的ICA算法、Vlassis等人<sup>[148]</sup>的源自适应的ICA算法、Mathis等人<sup>[113]</sup>的阈值ICA算法以及Thawonmas等人<sup>[141]</sup>的级联神经网络盲提取算法等。

盲信号分离的算法形式虽多,但研究的信号模型以及涉及到的基本理论却大同小异,本章的目的也就在于介绍这些内容。具体安排如下。首先在第2节介绍盲信号分离的信号模型与基本假设,接着在第3节介绍盲信号分离的对比函数理论,并给出若干常用的对比函数,第4节介绍盲信号分离的局部稳定性条件和算法的性能评价测度,最后在第5节给出本章小结。

## §2.2 信号模型与假设

### 2.2.1 信号模型

在神经网络、信号处理和统计学的研究中,观测数据常常可以用数学模型

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t + \mathbf{v}_t, \quad t=1,2,\dots \quad (2.1)$$

来描述,式中各符号的含义如下:

- $\mathbf{x}_t = [x_{1,t}, \dots, x_{m,t}]^T$   $m$  维的观测信号向量;
- $\mathbf{s}_t = [s_{1,t}, \dots, s_{n,t}]^T$   $n$  维的源信号向量;
- $\mathbf{v}_t = [v_{1,t}, \dots, v_{m,t}]^T$   $m$  维的加性噪声向量;
- $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$   $m \times n$  维的信道响应矩阵,习惯称作混合矩阵;
- $t$  时间下标;
- $T$  向量或矩阵的转置。

盲信号分离的目的在于仅利用观测信号向量  $\mathbf{x}_t$  提取或者恢复源信号向量  $\mathbf{s}_t$ 。因为可利用的信息有限,也因为观测信号含噪声相当于源信号与观测信号之间存在某种非线性关系<sup>[83]</sup>,依据模型(2.1)现实信号的盲分离是困难的。为此,通常假定观测信号向量不含噪声,从而有

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \quad t=1,2,\dots \quad (2.2)$$

这是一种线性瞬时混合模型,盲信号分离研究一般围绕它展开。

事实上,式(2.1)可写作

$$\mathbf{x}_t = [\mathbf{A}, \mathbf{I}] \begin{bmatrix} \mathbf{s}_t \\ \mathbf{v}_t \end{bmatrix} \triangleq \tilde{\mathbf{A}} \tilde{\mathbf{s}}_t$$

其中  $\mathbf{I}$  表示  $m$  维的单位矩阵,因此(2.1)也属于模型(2.2)的范畴。今后如无特别声明,我们只讨论式(2.2)描述的不含噪声的线性盲信号分离模型。



### 2.2.2 基本假设

为了使盲信号分离问题可解,必须对混合矩阵和源信号作某些假设。这些基本的假设条件包括:①源信号向量 $\mathbf{s}_i$ 的各分量相互独立;②最多只有一个源信号分量服从高斯分布;③混合矩阵 $\mathbf{A}$ 列满秩,  $m \geq n$ 。

记第 $i$ 个源信号分量 $s_{i,l}$ 的概率密度函数为 $p_i(s_{i,l})$ ,则独立性假设①意味着源信号向量 $\mathbf{s}_i$ 的联合概率密度函数可表示成各分量的概率密度函数的乘积,即<sup>[125]</sup>

$$p(\mathbf{s}_i) = \prod_{l=1}^n p_i(s_{i,l}) \quad (2.3)$$

式中 $\prod$ 表示连乘运算。独立性假设是盲信号分离的立足点,在很多实际应用中,这也是一条合理的假设。由独立性出发,可推导得到一大类盲信号分离算法:独立分量分析算法,这一部分内容将在第三章介绍。

盲信号分离的基本假设②,一方面是因为多个高斯信号的线性混合仍然服从正态分布,从而是不可惟一再分的;另一方面是因为盲信号分离算法的局部稳定性要求所致(参见本章第4节)。

为确保所有的源信号分量都是可以分离或提取的,盲信号分离约定③混合矩阵 $\mathbf{A}$ 列满秩。换句话说,若混合矩阵 $\mathbf{A}$ 是列亏损的,则只有部分源信号分量能够被提取<sup>[18,105,106]</sup>。显然,对于含噪声的盲信号分离模型,等价的混合矩阵 $\tilde{\mathbf{A}}$ 必定是列亏损的。关于列亏损情况下的病态盲信号分离问题,可参阅文献[18,100,104~106],部分内容将在第六章介绍。

由于模型(2.2)中的乘积因子 $\mathbf{A}$ 和 $\mathbf{s}_i$ 都未知,因此

- (1) 同时交换源信号和混合矩阵与之相对应的列的位置;
- (2) 一个源信号分量和与之相应的混合矩阵 $\mathbf{A}$ 的列之间互换一个固定的比例因子,即

$$\mathbf{x}_i = \mathbf{A}\mathbf{s}_i = \sum_{l=1}^n \frac{\mathbf{a}_l}{\alpha_l} \alpha_l s_{i,l} \quad (2.4)$$

式中 $\alpha_l$ 是任意的复数因子, $\mathbf{a}_l$ 表示由 $\mathbf{A}$ 的第 $l$ 列元素构成的列向量,所得的观测信号向量是相同的。也就是说,源信号向量 $\mathbf{s}_i$ 各分量的排列顺序、符号和能量是无法精确辨识的,这就是盲信号分离问题固有的两种不确定性<sup>[48]</sup>:恢复信号的排列顺序不确定性和幅度不确定性。不失一般性,通常假定④各个源信号分量具有单位功率,从而把源信号振幅的动态变化归并到混合矩阵 $\mathbf{A}$ 的相应列的元素中。

为描述上述两种不确定性,Cardoso等人将两个矩阵的“本质相等”概念引入到



盲信号分离中。

**定义2.1<sup>[22]</sup>** 两个矩阵  $\mathbf{M}$  和  $\mathbf{N}$  称为本质相等，若存在一个矩阵  $\mathbf{G}$  使得  $\mathbf{M} = \mathbf{GN}$ ，其中  $\mathbf{G}$  是一个广义交换矩阵，其任意行任意列有且仅有一个非零元素，且非零元素的模等于1。

借助于定义2.1，盲信号分离问题可叙述为：根据观测信号向量  $\mathbf{x}_l$  以及上述假设①~④辨识混合矩阵  $\mathbf{A}$  的本质相等矩阵与/或恢复源信号向量  $\mathbf{s}_l$ 。

### 2.2.3 算法构造

参见图2.1，为实现盲信号分离，通常用  $n \times m$  维的矩阵  $\mathbf{B}$  作用于观测信号向量  $\mathbf{x}_l$ ，使系统输出向量

$$\mathbf{y}_l = \mathbf{B}\mathbf{x}_l \quad (2.5)$$

是源信号向量  $\mathbf{s}_l$  的某个拷贝。与混合矩阵 (mixing matrix)  $\mathbf{A}$  相对应，在盲信号分离研究中习惯称  $\mathbf{B}$  为分离矩阵 (separating matrix)。

由于存在两种不确定性，实现盲信号分离的最优加权矩阵应满足

$$\mathbf{B}_{opt}\mathbf{A} = \mathbf{DP} \quad (2.6)$$

式中， $\mathbf{D}$  表示任意的非奇异对角矩阵， $\mathbf{P}$  表示任意的交换矩阵。换句话说，合成的混合-分离矩阵

$$\mathbf{C} \triangleq \mathbf{BA} \quad (2.7)$$

是一个广义交换矩阵。

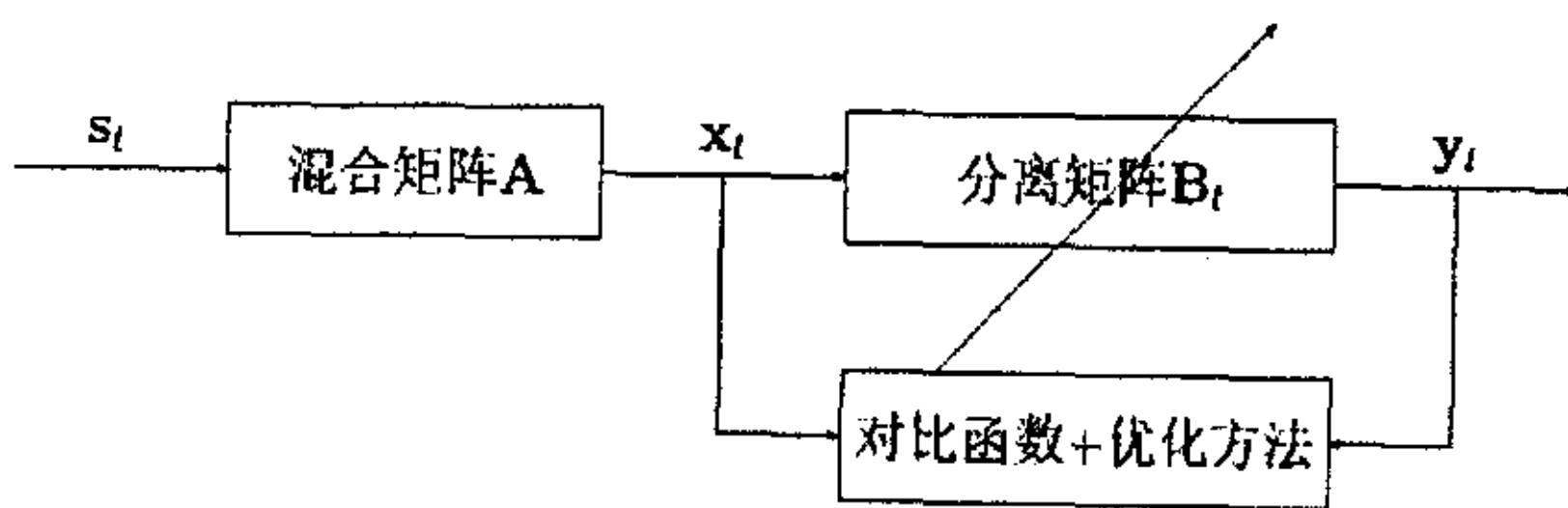


图2.1 盲信号分离原理图

分离矩阵  $\mathbf{B}$  的求解算法多种多样，既有批处理算法也有自适应在线算法；既有基于随机梯度的最小均方 (LMS: Least Mean Square) 算法也有递归最小二乘 (RLS: Recursive Least-Squares) 型算法，等等。算法的形式虽多，推导思路却大同小异。首先选择恰当的对比函数 (contrast function)，然后采用某种优化方法来搜索对比函数

的极值点，即构造过程可以统一表示为<sup>[85]</sup>

$$\text{BSS算法} = \text{对比函数} + \text{优化方法} \quad (2.8)$$

因此，盲信号分离算法的性能取决于两个方面。其中，对比函数决定算法的统计性能，包括一致性、渐近方差和稳健性等；优化方法决定算法的算法性能，包括收敛速度、运算量、需要的内存量和数值稳定性等。

## §2.3 对比函数理论

### 2.3.1 对比函数的定义

盲信号分离的对比函数即目标函数，可以分成两大类。一类是对所有分离状态都等价的对称对比函数，另一类是非对称的对比函数。

在定义对比函数之前，我们先引入下述符号：

- S** 各分量相互独立的源信号向量  $\mathbf{s}$  的全体构成的集合；
- X** 观测信号向量  $\mathbf{x}$  的全体；
- Y** 系统输出向量  $\mathbf{y}$  的全体；
- C** 与混合—分离系统响应矩阵  $\mathbf{C}$  具有相同维数的矩阵集合；
- D** 集合 **C** 的子集，由全体非奇异对角矩阵  $\mathbf{D}$  构成；
- Ω** 集合 **C** 的子集，其元素是任意的广义交换矩阵。

**定义2.2**（对称对比函数<sup>[48,116]</sup>） 由集合 **Y** 到实数集  $\mathbb{R}$  的一个多变量映射  $J$  称为盲信号分离的对比函数，若满足条件

- T1)  $\forall \mathbf{y} \in \mathbf{Y}, \forall \mathbf{W} \in \Omega, \text{有 } J(\mathbf{W}\mathbf{y}) = J(\mathbf{y})$ ;
- T2)  $\forall \mathbf{s} \in \mathbf{S}, \forall \mathbf{W} \in \mathbf{C}, \text{有 } J(\mathbf{W}\mathbf{s}) \leq J(\mathbf{s})$ ;
- T3)  $\forall \mathbf{s} \in \mathbf{S}, \forall \mathbf{W} \in \mathbf{C}, J(\mathbf{W}\mathbf{s}) = J(\mathbf{s}) \Leftrightarrow \mathbf{W} \in \Omega$ 。

条件T1说明对比函数对变量具有幅度和顺序不变性，从而所有的分离状态是等价的；条件T2说明对比函数是有上界函数；而条件T3确保，当且仅当合成系统的传输矩阵是广义交换矩阵，即输出信号的分量相互独立时，对比函数取得全局极大值。对变量的顺序不变性又称对称性，因此它是一种对称的对比函数。

盲信号分离经常使用的是对称对比函数，它取全局极大值是实现盲信号分离的充分必要条件。实际上，可对上述定义作推广，保证对比函数取全局极大值是实现盲信号分离的充分条件即可。

**定义2.3** (非对称对比函数<sup>[117]</sup>) 设  $\Omega_d$  是广义交换矩阵集合  $\Omega$  的一个非空子集, 则由集合  $\mathbf{Y}$  到实数集  $\mathbb{R}$  的一个多变量映射  $J$  称为盲信号分离的对比函数, 若满足条件

$$\text{GT1)} \quad \forall \mathbf{y} \in \mathbf{Y}, \quad \forall \mathbf{D} \in \mathbf{D}, \quad \text{有 } J(\mathbf{D}\mathbf{y}) = J(\mathbf{y});$$

$$\text{GT2)} \quad \forall \mathbf{s} \in \mathbf{S}, \quad \forall \mathbf{W} \in \mathbf{C}, \quad \text{有 } J(\mathbf{W}\mathbf{s}) \leq J(\mathbf{s});$$

$$\text{GT3)} \quad \forall \mathbf{s} \in \mathbf{S}, \quad \forall \mathbf{W} \in \mathbf{C}, \quad \exists \Omega_d \subset \Omega, J(\mathbf{W}\mathbf{s}) = J(\mathbf{s}) \Leftrightarrow \mathbf{W} \in \Omega_d.$$

与定义2.2作比较不难发现, 非对称的对比函数只满足对变量的幅度不变性, 而对顺序不变性不作要求。另一方面, 与T3要求对任意广义交换矩阵, 对比函数都取全局极大值不同, 条件GT3仅要求存在某个广义交换矩阵, 使对比函数取得全局极大值。显然, GT1~GT3定义的对比函数取全局极大值一定导致盲信号分离, 但反之不成立, 因此所有的分离状态不再等价, 这也是术语“非对称”的由来。

### 2.3.2 常用对比函数

对比函数是盲信号分离研究的出发点。上一小节定义了两种类型的对比函数, 为便于查阅, 下面列出其中常用的典型对比函数。我们指出, 部分参考文献给出的是代价函数, 其最小化实现信号的盲分离, 因此下述函数与参考文献给出的函数可能存在一个符号差异。此外, 为简单起见, 省略时间下标  $t$ 。

最大熵对比函数<sup>[14]</sup>:

$$J_{ME}(\mathbf{y}; \mathbf{B}) = H(g(\mathbf{y})) = H(g(\mathbf{B}\mathbf{x})) \quad (2.9)$$

式中  $H(x) = -\int p(x) \ln p(x) dx$  表示微分熵<sup>[49]</sup>,  $g(\mathbf{y}) = [g_1(y_1), \dots, g_n(y_n)]^T$  是一非线性的列向量。非线性函数  $g_i(y_i)$  的形式取决于算法的局部稳定性条件。

极大似然对比函数<sup>[25,27]</sup>:

$$J_{ML}(\mathbf{y}; \mathbf{B}) = \ln |\det(\mathbf{B})| + \sum_{i=1}^n E \{ \log p_i(y_i; \mathbf{B}) \} \quad (2.10)$$

其中  $p_i(\cdot)$  表示第  $i$  个源信号的概率密度函数。

最小互信息对比函数<sup>[3,153]</sup>:

$$J_{MI}(\mathbf{y}; \mathbf{B}) = \ln |\det(\mathbf{B})| + H(\mathbf{x}) - \sum_{i=1}^n H(y_i) \quad (2.11)$$

业已证明<sup>[25,119,153]</sup>, 若非线性函数取作源信号的概率密度函数, 即  $g_i(\cdot) = p_i(\cdot)$ ,  $i = 1, \dots, n$ , 则上述三种对比函数是等价的。

如果所有源信号具有相同的模值, 如通信系统中的BPSK和QPSK调制信号, 则盲信号分离可使用下面的对比函数, 即

恒模对比函数<sup>[31]</sup>:

$$J_{CMA}(\mathbf{y}; \mathbf{B}) = -\sum_{i=1}^n E\left\{\left[|y_i|^2 - 1\right]^2\right\} - \rho \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left|E\{y_i y_j^*\}\right|^2 \quad (2.12)$$

式中  $\rho$  是一个正的实常数, 以调整第一项恒模代价函数与第二项去相关代价函数之间的比重。上标\*表示复共轭运算, 而  $E\{\cdot\}$  表示对变量求数学期望。

对于白化预处理的观测信号, 除上述四种对比函数外, 还经常用到下面的对比函数 (为区别起见, 对比函数加上标  $w$ )。

非线性主分量分析对比函数<sup>[96,124,167]</sup>:

$$J_{NPCA}^w(\mathbf{y}; \mathbf{B}) = -E\left\{\left|\tilde{\mathbf{x}} - \mathbf{B}^T g(\mathbf{B}\tilde{\mathbf{x}})\right|^2\right\} \quad (2.13)$$

其中  $\tilde{\mathbf{x}}$  表示观测信号  $\mathbf{x}$  经白化预处理获得的信号向量, 满足  $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I}$ 。

高阶累积量对比函数1<sup>[48]</sup>:

$$J_{HOC1}^w(\mathbf{y}; \mathbf{B}) = \sum_{i=1}^n \left[\text{cum}_4(y_i)\right]^2 \quad (2.14)$$

其中  $\text{cum}_4(y_i)$  表示  $y_i$  的四阶累积量, 参阅附录A。Moreau等人<sup>[116]</sup>证明, (2.14) 式中的平方运算可以简化用绝对值运算代替, 从而有

高阶累积量对比函数2<sup>[116,127]</sup>:

$$J_{HOC2}^w(\mathbf{y}; \mathbf{B}) = \sum_{i=1}^n |\text{cum}_4(y_i)| \quad (2.15)$$

如果所有源信号的四阶累积量具有相同的符号  $\varepsilon = \text{sgn}[\text{cum}_4(s_i)]$ , 则对比函数 (2.15) 可进一步简化, 得

高阶矩对比函数<sup>[19,22,116]</sup>:

$$J_{HOM}^w(\mathbf{y}; \mathbf{B}) = \varepsilon \sum_{i=1}^n E\{y_i^4\} \quad (2.16)$$

累积量匹配对比函数<sup>[23,27,117]</sup>:

$$J_{CM}^w(\mathbf{y}; \mathbf{B}) = \sum_{r \geq 2} \sum_{i_1, \dots, i_r=1}^n \left[ \text{cum}(y_{i_1}, \dots, y_{i_r}) - \text{cum}(s_{i_1}, \dots, s_{i_r}) \right]^2 \quad (2.17)$$

负熵对比函数<sup>[48,86]</sup>:

$$J_{NE}^w(\mathbf{y}; \mathbf{B}) = \sum_{i=1}^n Q(y_i) \quad (2.18)$$

其中  $Q(\mathbf{y}) = H(\mathbf{y}_{gau}) - H(\mathbf{y})$  表示向量  $\mathbf{y}$  的负熵，而  $\mathbf{y}_{gau}$  则表示与  $\mathbf{y}$  具有相同协方差矩阵的高斯随机向量。Hyvarinen<sup>[82]</sup>推导了负熵的最大熵逼近公式，并提出了最大熵逼近的负熵对比函数<sup>[82,86]</sup>：

$$J_{ANE}^w(\mathbf{y}; \mathbf{B}) = \sum_{i=1}^n \left[ E\{L(y_i)\} - E\{L(v)\} \right]^2 \quad (2.19)$$

式中  $v$  表示标准的高斯随机变量， $L(\cdot)$  表示任意的非二次函数，常用的函数有

$$L_1(u) = \frac{1}{a} \ln \cosh(au), \quad 1 \leq a \leq 2 \quad (2.20a)$$

$$L_2(u) = -\frac{1}{b} \exp\left(-\frac{b}{2}u^2\right), \quad b \approx 1 \quad (2.20b)$$

$$L_3(u) = \frac{1}{4}u^4 \quad (2.20c)$$

很容易证明，当取非线性函数  $L_3(\cdot)$  时，对比函数 (2.19) 与 (2.14) 是等价的。因此，可将 (2.19) 看成是一种推广的高阶累积量对比函数。

关于高阶累积量对比函数，还有下面的定理。

**定理2.1<sup>[117]</sup>** 设  $p$  个  $y_i$  和  $q$  个  $y_i^*$  (\*代表复数共轭) 的高阶累积量为  $C_p^q[y_i]$ ， $p$  与  $q$  之和不小于3并且取偶数， $\lambda_1 \geq \dots \geq \lambda_n > 0$ ，函数  $f(\cdot)$  是一个严格单调增的凸函数，则对于白化预处理的观测信号，

$$J_{p,q}^w(\mathbf{y}; \mathbf{B}) = \sum_{i=1}^n \lambda_i f(|C_p^q[y_i]|) \quad (2.21)$$

是盲信号分离的对比函数。

所谓凸函数，指  $\forall x_1, x_2 \in \mathbb{R}$ ， $\forall \alpha \in [0, 1]$  有  $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$  成立。定理2.1的证明参见文献[117]。若参数  $\lambda_i$  取相同的值，则式 (2.21) 属于对称对比函数的范畴。反之，它给出的是一种非对称的对比函数。



## §2.4 局部稳定性条件

### 2.4.1 算法框架

有了对比函数, 再选择合适的优化方法<sup>[71,176]</sup>, 比如最速下降法、共轭梯度法、自然梯度法、拟牛顿法、卡尔曼 (Kalman) 滤波方法以及RLS方法等, 即可推导出各种各样的盲信号分离算法。其中, 自适应并行算法具有如下形式<sup>[27,69,72,89,99]</sup>:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i \mathbf{F}(\mathbf{y}_i, \mathbf{B}_i) \quad (2.22)$$

式中  $\eta_i > 0$  为学习速率或学习步长参数,  $\mathbf{F}(\mathbf{y}_i, \mathbf{B}_i)$  是以  $\mathbf{y}_i$  和  $\mathbf{B}_i$  为自变量的矩阵函数, 在盲信号分离研究中, 习惯称作估计函数 (estimating function) <sup>[4,26,27]</sup>。显然, 在算法的平衡点处, 估计函数的数学期望值等于零矩阵, 即

$$E\{\mathbf{F}(\mathbf{y}_i, \mathbf{B}_i)\} = \mathbf{O} \quad (2.23)$$

从盲信号分离的半参数统计模型 (semiparametric statistical model) 出发, Amari 和 Cardoso<sup>[4]</sup>证明最优估计函数具有形式

$$\mathbf{F}(\mathbf{y}_i, \mathbf{B}_i) = [\mathbf{I} - \Phi(\mathbf{y}_i) \mathbf{y}_i^T] \mathbf{B}_i \quad (2.24)$$

式中  $\Phi(\mathbf{y}_i) = [\varphi_1(y_{1,i}), \dots, \varphi_n(y_{n,i})]^T$  是一个非线性的列向量, 其分量

$$\varphi_i(u) = -\frac{p'_i(u)}{p_i(u)} \quad (2.25)$$

称为盲信号分离的分值函数 (score function), 而  $p'_i(\cdot)$  表示概率密度函数  $p_i(\cdot)$  的一阶导数。最优估计函数 (2.24) 也可从极大似然对比函数 (2.10) 推导得到。

最优估计函数取决于源信号的概率分布, 但是在盲信号分离问题中, 它总是未知的。为此, 通常用一个确定性的非线性函数或者含自适应参数的非线性函数  $g_i$  替代原来的分值函数  $\varphi_i$ , 从而有下面的算法<sup>[4,7,22,27,38,39]</sup>:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i [\mathbf{I} - \mathbf{g}(\mathbf{y}_i) \mathbf{y}_i^T] \mathbf{B}_i \quad (2.26)$$

式中  $g_i$  习惯称作盲信号分离的激励函数 (activation function)。

通过转置 (2.26) 中的估计函数, Hoff 等人<sup>[78]</sup>提出了下面的盲信号分离算法:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i [\mathbf{I} - \mathbf{y}_i \mathbf{g}^T(\mathbf{y}_i)] \mathbf{B}_i \quad (2.27)$$

对算法 (2.26) 的平衡点运用矩阵求逆引理<sup>[176]</sup>, 也可以推导得到算法 (2.27)。

此外, 盲信号分离还可以同时使用两组非线性函数  $f_i$  和  $g_i$ , 得到算法<sup>[43,51,95]</sup>

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i [\mathbf{I} - \mathbf{f}(\mathbf{y}_i) \mathbf{g}^T(\mathbf{y}_i)] \mathbf{B}_i \quad (2.28)$$

显然, 算法 (2.26) 和 (2.27) 都是算法 (2.28) 的特例。

仿真分析表明<sup>[43,51,95]</sup>, 当非线性函数选择恰当, 算法 (2.28) 可获得快速收敛和高鲁棒性等优良性能, 但理论研究表明<sup>[4]</sup>, 它不属于渐进最优类的盲信号分离算法, 因此稳态性能要低于算法 (2.26) 和 (2.27)。

除了 (2.26)、(2.27) 和 (2.28) 三种形式外, Cardoso 等人<sup>[22]</sup>借助于相对梯度提出了下面的EASI算法:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i [\mathbf{I} - \mathbf{y}_i \mathbf{y}_i^T + \mathbf{g}(\mathbf{y}_i) \mathbf{y}_i^T - \mathbf{y}_i \mathbf{g}^T(\mathbf{y}_i)] \mathbf{B}_i \quad (2.29)$$

式中的激励函数  $g_i$  根据局部稳定性条件可以取不同的形式, 从而得到一类算法。

如果观测信号经过白化预处理, 或者混合矩阵是正交矩阵, 则分离矩阵必定也是一个正交矩阵, 那么它可以用非线性主分量分析 (PCA: Principle Component Analysis) 方法

$$\mathbf{B}_{i+1}^{\text{ort}} = \mathbf{B}_i^{\text{ort}} + \eta_i \mathbf{g}(\mathbf{y}_i) [\tilde{\mathbf{x}}_i^T - \mathbf{g}^T(\mathbf{y}_i) \mathbf{B}_i^{\text{ort}}] \quad (2.30)$$

求解得出, 式中  $\tilde{\mathbf{x}}_i = \mathbf{U} \mathbf{x}_i$  表示观测信号  $\mathbf{x}_i$  经白化预处理获得的信号向量,  $\mathbf{U}$  代表白化矩阵, 上标 ort 是正交矩阵的缩写。因此这种情况下, 总的分离矩阵  $\mathbf{B} = \mathbf{B}^{\text{ort}} \mathbf{U}$ 。

上述几种算法基于随机梯度, 从而可归纳为LMS型算法。除此之外, 还有RLS型算法, 这一部分内容将在第四章介绍。

#### 2.4.2 局部稳定性条件

盲信号分离的分值函数取决于源信号的概率分布, 因此难以获得最优估计函数, 解决的途径是用激励函数代替未知的分值函数。一个很自然的问题是: 选用的激励函数在什么样的条件下能够实现盲信号分离?

在前面提到的几种盲信号分离算法中, 形式 (2.26) 最典型也最常用, 因此下面就以它为例进行具体讨论。

在 (2.26) 的两端同时右乘以  $\mathbf{A}$ , 并记合成系统的响应矩阵为  $\mathbf{C}$ , 则有

$$\mathbf{C}_{i+1} = \mathbf{C}_i + \eta_i [\mathbf{I} - \mathbf{g}(\mathbf{y}_i) \mathbf{y}_i^T] \mathbf{C}_i \quad (2.31)$$

该算法能够实现盲信号分离, 要求

- (1)  $\mathbf{C}$  取广义交换矩阵是算法 (2.31) 的稳定平衡点;
- (2) 除  $\mathbf{C}$  取广义交换矩阵外, 算法 (2.31) 再没有其它的稳定平衡点。

不难看出, 第二个条件要求算法是全局收敛的。遗憾的是, 并行盲分离算法的全局收敛性是目前尚未解决的难题之一, 因此这里只讨论局部收敛性。具体地讲, 激励函数  $g_i$  和源信号在满足何种条件下, 合成矩阵  $\mathbf{C}$  取广义交换矩阵是算法 (2.31) 的稳定平衡点? 下面的定理回答了这个问题。为书写方便, 省略时间下标  $t$ 。

**定理 2.2**<sup>[5,30,78,120]</sup> 设  $a_i = E[g'_i(s_i)s_i^2]$ ,  $b_i = E[g_i(s_i)s_i]$ ,  $\sigma_i^2 = E[s_i^2]$  和  $m_i = E[g'_i(s_i)]$ , 其中  $g'_i$  表示激励函数  $g_i$  的一阶导数。合成系统的响应矩阵  $\mathbf{C}$  取广义交换矩阵是算法 (2.31) 的稳定平衡点, 若对任意的  $i, j$  ( $i \neq j$ ), 条件

$$\begin{aligned} \text{I: } & a_i + b_i > 0, \quad m_i > 0 \\ \text{II: } & \frac{m_i \sigma_i^2}{b_i} \cdot \frac{m_j \sigma_j^2}{b_j} > 1 \end{aligned} \quad (2.32)$$

都成立。

证明: 简单起见, 考虑在平衡点  $\mathbf{C}_t = \mathbf{I}$  处算法的局部稳定性。设  $\mathbf{C}_t$  的扰动矩阵为  $\Delta_t$ , 则  $\mathbf{C}_t = \mathbf{I} + \Delta_t$ , 从而输出信号  $\mathbf{y}_t = \mathbf{s}_t + \Delta_t \mathbf{s}_t$ 。非线性列向量  $\mathbf{g}(\mathbf{y}_t)$  在  $\mathbf{y}_t = \mathbf{s}_t$  处的一阶 Taylor 级数展开为

$$\mathbf{g}(\mathbf{y}_t) = \mathbf{g}(\mathbf{s}_t) + \mathbf{D}_{g'} \Delta_t \mathbf{s}_t \quad (2.33)$$

其中  $\mathbf{D}_{g'} = \text{diag}[g'_1(s_{1,t}), \dots, g'_n(s_{n,t})]$  是一个对角矩阵。

将式 (2.33) 代入到算法 (2.31) 中, 两边同取数学期望, 利用源信号的相互独立性并忽略扰动矩阵的二次项, 有

$$E\{\Delta_{t+1}\} = E\{\Delta_t\} - \eta_t E\{\mathbf{g}(\mathbf{s}_t) \mathbf{s}_t^T \Delta_t^T + \mathbf{D}_{g'} \Delta_t \mathbf{s}_t \mathbf{s}_t^T\} \quad (2.34)$$

记  $\mathbf{D}_{gs} = E\{\mathbf{g}(\mathbf{s}_t) \mathbf{s}_t^T\}$ ,  $\mathbf{D}_{ss} = \text{diag}[s_1^2, \dots, s_n^2]$ , 并假设扰动  $\Delta_t$  与源信号  $\mathbf{s}_t$  不相关, 则

$$E\{\Delta_{t+1}\} = E\{\Delta_t\} - \eta_t [\mathbf{D}_{gs} E\{\Delta_t^T\} + E\{\mathbf{D}_{g'} \Delta_t \mathbf{D}_{ss}\}] \quad (2.35)$$

考虑式 (2.35) 中的对角元素, 有

$$E\{\Delta_{ii,t+1}\} = [1 - \eta_t (a_i + b_i)] E\{\Delta_{ii,t}\} \quad (2.36)$$

若学习速率参数  $\eta_i$  足够小, 则对角元素收敛要求  $a_i + b_i > 0$ 。

考虑式 (2.35) 中的非对角元素, 有

$$E \begin{Bmatrix} \Delta_{ij,l+1} \\ \Delta_{ji,l+1} \end{Bmatrix} = (\mathbf{I} - \eta_i \Omega) E \begin{Bmatrix} \Delta_{ij,l} \\ \Delta_{ji,l} \end{Bmatrix} \quad (2.37)$$

其中,

$$\Omega = \begin{bmatrix} m_i \sigma_j^2 & b_i \\ b_j & m_j \sigma_i^2 \end{bmatrix} \quad (2.38)$$

显然, 非对角元素收敛要求矩阵  $\Omega$  的特征值具有正实部, 从而要求  $m_i > 0$  和式 (2.32) 中的条件II成立。证明完毕。■

由定理2.2可以看出, 盲信号分离的局部稳定性取决于两个方面: 选用的激励函数  $g_i$  和源信号的概率分布。若源信号服从对称分布,  $g_i$  取单调增的奇函数, 则条件I必定成立。这样, 定理2.2描述的局部稳定性条件可简化为<sup>[5,30,59,78,120]</sup>

$$E[g'_i(s_i)] \cdot E[s_i^2] - E[g_i(s_i)s_i] > 0 \quad (2.39)$$

由定理2.2还可以看出, 对于两个高斯信号, 稳定性条件不可能成立, 因为

$$\begin{aligned} m_i \sigma_i^2 &= \sigma_i^2 \cdot \int_{-\infty}^{\infty} g'_i(x) \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma_i^2}\right] dx \\ &= \int_{-\infty}^{\infty} g_i(x) x \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma_i^2}\right] dx = b_i \end{aligned}$$

即条件II的左边恒等于1。因此, 盲信号分离假定最多只有一个源信号服从正态分布, 也就是基本假设②。

此外, 若激励函数取三次函数  $g_i(y_i) = y_i^3$ , 则条件II等价于

$$3E^2\{s_i^2\} - E\{s_i^4\} > 0$$

从而要求所有的源信号都服从亚高斯分布。准确地说, 是要求源信号是负峰度信号。关于亚/超高斯信号, 正/负峰度信号的定义以及两者的联系和区别, 参见第五章。

若激励函数取符号函数  $g_i(y_i) = \text{sgn}(y_i)$ , 并为源信号建立广义高斯分布模型,

$$p(s, \alpha) = \frac{\alpha}{2\lambda\Gamma(\alpha^{-1})} \exp\left[-|\lambda^{-1}s|^\alpha\right]$$

其中参数  $\alpha$  称作广义高斯分布的高斯指数 (Gaussian exponent),  $\lambda$  与信号的偶数阶矩

存在如下关系:

$$E\{s_i^{2k}\} = \lambda^{2k} \Gamma[(2k+1)\alpha^{-1}] / \Gamma(\alpha^{-1})$$

Gamma函数的定义为

$$\Gamma(x) = \int_0^{\infty} t^{x-1} \exp[-t] dt$$

利用Gamma函数的性质, 条件II的左边等于

$$E\{s_i^2\} E\{\text{sgn}'(s_i)\} / E\{|s_i|\} = \frac{\alpha}{2} \cdot \Gamma(3\alpha^{-1}) / [\Gamma(\alpha^{-1}) \Gamma(2\alpha^{-1})] \triangleq f(\alpha)$$

可以证明  $f(\alpha)$  是单调减函数, 且  $f(2)=1$ 。因此, 算法的局部稳定性条件成立要求  $\alpha < 2$ , 换句话说, 除一个源信号可以是高斯信号外, 其它的源信号都必须是超高斯信号。

类似地, 还可以证明双曲正切函数  $g_i(y_i) = \tanh(y_i)$  可以实现超高斯源信号的稳定分离。

以上讨论的是形式 (2.26) 的局部稳定性条件, 对于形式 (2.27), 经过类似的推导, 不难证明其稳定性条件与 (2.26) 的区别仅在于将条件II中的大于号改成小于号即可。也就是说, 形式 (2.27) 的简化的局部稳定性条件是<sup>[78]</sup>

$$E[g'_i(s_i)] \cdot E[s_i^2] - E[g_i(s_i)s_i] < 0 \quad (2.40)$$

关于形式 (2.28), 由于使用了两个非线性的函数, 其局部稳定性分析要复杂得多, 也因为这种形式并不常用, 故目前尚没有文献对此进行讨论。

关于EASI类算法 (2.29), Cardoso等人<sup>[22,30]</sup>证明它的局部稳定性条件要比形式 (2.26) 稍广一些, 但具体操作时, 一般仍以定理2.2为依据选择激励函数。

Karhunen等人在文献[96]中讨论了非线性PCA类算法 (2.30) 与EASI算法的联系, 最近, 作者借助于自然梯度更加明确了这种联系, 详情请参阅文献[167]。根据两者之间的联系不难得出, 算法 (2.30) 的局部稳定性条件与式 (2.40) 相同。

最后, 需要强调的是本节讨论的局部稳定性条件, 都以源信号是线性过程为前提。当源信号是非线性过程时, 观测样本不再满足独立同分布 (i.i.d: independent identical distribution) 的性质, 算法局部稳定性条件的推导将更加复杂而繁琐, 这一部分内容可参阅文献[120]。



### 2.4.3 算法的性能评价

下面讨论盲信号分离算法的性能评价测度。

由于盲信号分离问题存在两种固有的不确定性<sup>[48]</sup>：信号排列顺序不确定性和幅度不确定性，因此计算恢复信号与源信号之差或者与之相关的信噪比没有意义。盲信号分离的目的是从观测信号中恢复源信号的波形，因此可以用合成矩阵  $\mathbf{C}_l = [c_{ij,l}] = \mathbf{B}_l \mathbf{A}$  偏离广义交换矩阵的程度来衡量盲信号分离算法的性能。它有两种测度，一种是性能指数（PI: Performance Index）

$$PI_{cl,l} = \sum_{i=1}^n \left( \sum_{j=1}^n \frac{|c_{ij,l}|}{\max_l |c_{il,l}|} - 1 \right) + \sum_{j=1}^n \left( \sum_{i=1}^n \frac{|c_{ij,l}|}{\max_l |c_{jl,l}|} - 1 \right) \quad (2.41)$$

另一种是性能指数

$$PI_l = \sum_{i=1}^n \left( \sum_{j=1}^n \frac{c_{ij,l}^2}{\max_l c_{il,l}^2} - 1 \right) + \sum_{j=1}^n \left( \sum_{i=1}^n \frac{c_{ij,l}^2}{\max_l c_{jl,l}^2} - 1 \right) \quad (2.42)$$

显然，合成矩阵  $\mathbf{C}_l$  偏离广义交换矩阵的程度越大，上述两种性能指数的取值就越大，盲信号分离的效果也越差。反之，性能指数越小，盲信号分离的效果就越好。当  $\mathbf{C}_l$  等于广义交换矩阵时，性能指数等于零，它对应于理想情况下的盲信号分离。

设合成矩阵  $\mathbf{C}_l$  在平衡点处的微小扰动矩阵为  $\Delta$ ，则性能指数（2.42）的稳态值等于<sup>[150]</sup>

$$PI_{ss} = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n [E\{\Delta_{ij}^2\} + E\{\Delta_{ji}^2\}] \quad (2.43)$$

Hoff等人<sup>[78]</sup>建立了扰动矩阵的稳态均方误差与激励函数  $g_i$  以及源信号概率分布之间的关系，证明对于形式（2.26）和（2.27），有

$$E\{\Delta_{ij}^2\} = \frac{\eta}{2K} \frac{\gamma_+/\chi_+ + \gamma_-/\chi_-}{2} \quad (2.44)$$

式中  $\eta$  是学习速率参数的稳态值， $K$  表示批处理盲信号分离算法块的长度，对于在线算法  $K=1$ ，而参数

$$\begin{aligned} \gamma_{\pm} &= E\{g_i^2(s_i)\} E\{s_i^2\} \pm E^2\{g_i(s_i)s_i\} \\ \chi_{\pm} &= E\{g_i'(s_i)\} E\{s_i^2\} \pm E\{g_i(s_i)s_i\} \end{aligned} \quad (2.45)$$

业已证明<sup>[4,25,78,114]</sup>，在所有的非线性函数中，分值函数(2.25)是最优的，它工作最稳健，对应的稳态均方误差值也最小。

## §2.5 小 结

盲信号分离问题因为在无线通信、雷达和声纳、图像、语言、医学以及地震信号处理等领域具有广阔的应用前景，因此一诞生就迅速引起了信号处理学界和神经网络学界的广泛兴趣。为了对盲信号分离有一个比较全面的了解，也为后续章节奠定基础，本章介绍了盲信号分离的一些基础知识，主要内容有：

(1) 介绍了盲信号分离研究的信号模型，讨论了模型的两种不确定性，以及对模型所作的一些基本假设；

(2) 介绍了盲信号分离算法的构造过程，研究了盲信号分离的对比函数理论，并列举了若干常用的对比函数；

(3) 讨论了盲信号分离算法的基本框架，并针对不同框架研究了算法的局部稳定性条件，为激励函数的实际选择提供理论依据；

(4) 介绍了盲信号分离算法的性能评价测度。

## 第三章 分阶段学习的盲信号分离

### § 3.1 引言

在第二章中我们看到,盲信号分离的目的是根据观测数据以及对混合矩阵和源信号的一些基本假设来恢复源信号的波形。盲信号分离算法的构造过程是首先选择恰当的对比函数,然后采用某种优化方法来搜索对比函数的极大值。

对比函数有很多种可能的选择,而最直接的就是依据源信号之间的独立性假设来构造,并由此产生了三类对比函数,即最大熵<sup>[14]</sup>、极大似然<sup>[25,27]</sup>和最小互信息<sup>[3,153]</sup>对比函数,业已证明<sup>[25,119,153]</sup>在一定条件下,它们是等价的。

从最小互信息对比函数出发,可以得到一大类盲信号分离算法——独立分量分析法。独立分量分析法是Comon<sup>[48]</sup>在1994年首先提出的,它已经成为当前盲信号分离研究的主流方法。独立分量分析有很多有效的算法,如Amari等人<sup>[3,153]</sup>的自然梯度算法、Cardoso等人<sup>[7]</sup>的EASI算法、推广的ICA算法<sup>[101]</sup>、灵活的ICA算法<sup>[37]</sup>,迭代求逆ICA算法<sup>[51]</sup>等等。虽然具体形式不同,但这些算法都可归类为LMS型的算法,因此存在一个学习速率参数的优选问题,这也是本章研究的核心问题。

当学习速率或步长参数取常数时,LMS型的盲信号分离算法存在收敛速度和稳态性能之间的矛盾:步长小,算法的稳态性能好,源信号的恢复质量高,但算法收敛慢;反之,收敛快但稳态性能差。解决上述矛盾的最简单做法是令学习速率随时间递减<sup>[3,7,156]</sup>,但这又会产生新的问题:如果信号没能分离而学习速率又降至很小,则未分离的信号会因学习速率太小而长时间甚至永远得不到分离。解决上述矛盾的另一类做法是采用自适应的步长参数,如基于辅助变量的自适应步长算法<sup>[41,56,90,118]</sup>和自适应梯度步长算法<sup>[112]</sup>等。这些算法的自适应步长的选择取决于辅助变量,而与神经网络输出信号之间的分离状态无关,并且对所有输出分量使用相同的学习速率。据我们所知,还没有一种自适应步长学习规则是与信号分离状态有关的。

在本章中,我们提出一种具有信号自适应学习速率的分阶段学习算法:在不同的阶段,根据各个输出分量与其它输出分量之间的分离状态的不同,对它们赋以不同的学习速率,以自动控制它们的收敛与跟踪。

本章的结构安排如下。在第2节介绍独立分量分析法,并重点讨论自然梯度和等变化性。第3节定义描述信号分离状态的相依性测度。第4节是本章的主体部分,首先提出一种基于分阶段学习的盲信号分离算法,然后将它与现有的相关算法作比较。第

5节用计算机仿真验证新算法的有效性，最后在第6节给出本章小结。

## § 3.2 独立分量分析法

### 3.2.1 对比函数

源信号向量  $\mathbf{s}_t$  的各分量相互独立意味着  $\mathbf{s}_t$  的联合概率密度函数等于其边缘概率密度函数的乘积。独立分量分析方法的中心思想是：调整分离矩阵  $\mathbf{B}_t$ ，如果使得网络输出  $\mathbf{y}_t = \mathbf{B}_t \mathbf{x}_t$  的各分量相互独立，则  $\mathbf{y}_t$  就是源信号向量  $\mathbf{s}_t$  的某个拷贝。

记  $\mathbf{y}_t = [y_{1,t}, \dots, y_{n,t}]^T$  的联合概率密度函数为  $p(\mathbf{y}_t)$ ，那么分量  $y_{i,t}$  的概率密度函数

$$p_i(y_{i,t}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\mathbf{y}_t) dy_{1,t} \cdots dy_{i-1,t} dy_{i+1,t} \cdots dy_{n,t} \quad (3.1)$$

由概率论的基本知识，若  $\mathbf{y}_t$  的分量相互统计独立，则<sup>[125]</sup>

$$p(\mathbf{y}_t) = \prod_{i=1}^n p_i(y_{i,t}) \quad (3.2)$$

如果用互信息 (MI: Mutual Information) 来衡量各分量之间的独立性程度，式 (3.2) 意味着输出向量  $\mathbf{y}_t$  的互信息等于零。

互信息有两种定义形式。一种是以 K-L 散度 (Kullback-Leibler Divergence) 为测度，即<sup>[49]</sup>

$$I_{KL}(\mathbf{y}_t; \mathbf{B}_t) = \text{KL}[p(\mathbf{y}_t) \| \tilde{p}(\mathbf{y}_t)] \triangleq \int_{-\infty}^{\infty} p(\mathbf{y}_t) \ln \frac{p(\mathbf{y}_t)}{\tilde{p}(\mathbf{y}_t)} d\mathbf{y}_t \quad (3.3)$$

式中  $\tilde{p}(\mathbf{y}_t) = \prod_{i=1}^n p_i(y_{i,t})$ 。另一种是以 Renyi 互信息为测度<sup>[77]</sup>,

$$I_{R_\alpha}(\mathbf{y}_t; \mathbf{B}_t) = \frac{1}{\alpha-1} \ln \int_{-\infty}^{\infty} p(\mathbf{y}_t) \ln \frac{[p(\mathbf{y}_t)]^\alpha}{[\tilde{p}(\mathbf{y}_t)]^{\alpha-1}} d\mathbf{y}_t \quad (3.4)$$

式中  $\alpha > 1$  是一个常数。

相对而言，形式 (3.3) 比 (3.4) 简单，因此更常用，本章的讨论也围绕它展开。关于形式 (3.4)，可参阅文献<sup>[77]</sup>。

K-L 散度是 Bayesian 统计意义下的风险函数，它具有以下几个重要性质<sup>[2,22]</sup>：

P1)  $\text{KL}[p(y) \| q(y)] \geq 0$ ，当且仅当  $p(y) = q(y)$  时等号成立；

P2) K-L散度对数据样本的可逆非线性变换是不变的, 即

$$\text{KL}[p(f(y))\|q(f(y))]=\text{KL}[p(y)\|q(y)] \quad (3.5)$$

P3) 当三个点  $p$ ,  $q$  和  $r$  在某些条件下构成“广义右三元组”时, 广义Pythagorean定理成立, 即

$$\text{KL}[p(y)\|q(y)]=\text{KL}[p(y)\|r(y)]+\text{KL}[r(y)\|q(y)] \quad (3.6)$$

根据上述性质, Comon<sup>[48]</sup>证明

$$J_{MI}(\mathbf{y}_i; \mathbf{B}_i) = -I_{KL}(\mathbf{y}_i; \mathbf{B}_i) \quad (3.7)$$

是盲信号分离的对比函数, 当且仅当合成系统的响应矩阵  $\mathbf{C}_i = \mathbf{B}_i \mathbf{A}$  是一个广义交换矩阵时, 式 (3.7) 取全局极大值零。

为简单起见, 假设源信号个数与传感器的个数相同, 即  $m = n$ 。关于  $m > n$  情况下的超定盲信号分离将在第六章介绍。这样, 利用信息论中微分熵的定义<sup>[49]</sup>

$$H(\mathbf{y}_i; \mathbf{B}_i) = -\int_{-\infty}^{\infty} p(\mathbf{y}_i) \ln p(\mathbf{y}_i) d\mathbf{y}_i \quad (3.8)$$

$$H(y_{i,j}; \mathbf{B}_i) = -\int_{-\infty}^{\infty} p(y_{i,j}) \ln p(y_{i,j}) dy_{i,j} \quad (3.9)$$

我们可以得到下面的对比函数<sup>[3,153]</sup>:

$$J_{MI}(\mathbf{y}_i; \mathbf{B}_i) = H(\mathbf{y}_i; \mathbf{B}_i) - \sum_{j=1}^n H(y_{i,j}; \mathbf{B}_i) \quad (3.10)$$

此外, 因为线性盲信号分离系统的Jacobian矩阵

$$\mathbf{M}_{Jac} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \mathbf{B}_i \quad (3.11)$$

所以输出  $\mathbf{y}_i$  的概率密度函数  $p(\mathbf{y}_i; \mathbf{B}_i)$  与观测信号  $\mathbf{x}_i$  的概率密度函数  $p(\mathbf{x}_i)$  之间存在以下函数关系:

$$p(\mathbf{y}_i; \mathbf{B}_i) = \frac{p(\mathbf{x}_i)}{|\det(\mathbf{B}_i)|} \quad (3.12)$$

将 (3.12) 代入微分熵的定义式 (3.8), 有



$$H(y_i; \mathbf{B}_i) = \ln |\det(\mathbf{B}_i)| + H(\mathbf{x}_i) \quad (3.13)$$

进而有

$$J_{MI}(y_i; \mathbf{B}_i) = \ln |\det(\mathbf{B}_i)| - \sum_{i=1}^n H(y_{i,i}; \mathbf{B}_i) + H(\mathbf{x}_i) \quad (3.14)$$

显然，式 (3.14) 就是第二章中提到的最小互信息对比函数。

### 3.2.2 自然梯度

上一小节简单推导了独立分量分析法的对比函数 (3.14)，下面采用随机梯度算法，更确切地，采用随机自然梯度 (natural gradient) 算法来学习分离矩阵  $\mathbf{B}_i$ 。

对比函数 (3.14) 共有三项组成，其中观测信号的微分熵  $H(\mathbf{x}_i)$  不依赖于分离矩阵  $\mathbf{B}_i$ ，因此推导  $\mathbf{B}_i$  的自适应更新算法时可以不预考虑。

记  $\mathbf{B}_i$  的第  $i$  行第  $j$  列元素  $\mathbf{B}_{ij}$  的代数余子式为  $\text{cof}(\mathbf{B}_{ij})$ ，由基本的线性代数知识有  $\det(\mathbf{B}_i) = \sum_j \mathbf{B}_{ij} \text{cof}(\mathbf{B}_{ij})$ ，因此

$$\frac{\partial}{\partial \mathbf{B}_{ij}} \ln |\det(\mathbf{B}_i)| = \frac{\text{cof}(\mathbf{B}_{ij})}{\det(\mathbf{B}_i)} \quad (3.15)$$

根据逆矩阵的定义  $\mathbf{B}_i^{-1} = \text{Adj}(\mathbf{B}_i) / \det(\mathbf{B}_i)$ ， $\text{Adj}(\mathbf{B}_i)$  表示  $\mathbf{B}_i$  的伴随矩阵，很容易证明

$$\frac{\partial}{\partial \mathbf{B}_i} \ln |\det(\mathbf{B}_i)| = \frac{[\text{Adj}(\mathbf{B}_i)]^T}{\det(\mathbf{B}_i)} = \mathbf{B}_i^{-T} \quad (3.16)$$

式中  $\mathbf{B}_i^{-T} = [\mathbf{B}_i^{-1}]^T$ 。

由单变量微分熵的定义式 (3.9)， $-\sum_i H(y_{i,i}; \mathbf{B}_i) = \sum_i E\{\log p_i(y_{i,i})\}$ ，得

$$\frac{\partial}{\partial \mathbf{B}_{ij}} \sum_{i=1}^n E\{\ln p_i(y_{i,i})\} = E\left\{\frac{p'_i(y_{i,i})}{p_i(y_{i,i})} x_{j,i}\right\} \quad (3.17)$$

令分值函数  $\varphi_i(y_{i,i}) = -p'_i(y_{i,i})/p_i(y_{i,i})$ ，则可将 (3.17) 写成如下的矩阵形式：

$$\frac{\partial}{\partial \mathbf{B}_i} \sum_{i=1}^n E\{\ln p_i(y_{i,i})\} = -E\{\Phi(\mathbf{y}_i) \mathbf{x}_i^T\} \quad (3.18)$$

综合 (3.16) 和 (3.18)，对比函数 (3.14) 的梯度等于

$$\nabla J_{MI}(y_i; \mathbf{B}_i) = \frac{\partial}{\partial \mathbf{B}_i} J_{MI}(y_i; \mathbf{B}_i) = \mathbf{B}_i^{-T} - E\{\Phi(\mathbf{y}_i) \mathbf{x}_i^T\} \quad (3.19)$$

从而有下面的随机（一般）梯度学习算法

$$\frac{d\mathbf{B}_i}{dt} = \eta_i [\mathbf{B}_i^{-T} - \Phi(\mathbf{y}_i) \mathbf{x}_i^T] \quad (3.20)$$

式中  $\eta_i > 0$  是学习速率参数。

事实上，式（3.20）就是Bell等人<sup>[14]</sup>提出的信息最大化（infomax）算法。该算法形式非常简单，但矩阵求逆运算导致数值稳定性差。为克服这个缺点，Amari等人<sup>[6,7]</sup>提出用自然梯度代替一般梯度来学习分离矩阵  $\mathbf{B}_i$ ，简单介绍如下。

记盲信号分离的代价函数  $J$  对分离矩阵  $\mathbf{B}$  的一般梯度为  $\nabla J$ 。当参数空间是均匀的Euclidean空间时， $\nabla J$  给出对比函数的最陡上升方向。由于两个非奇异矩阵的乘积仍然是非奇异矩阵，因此全体  $n \times n$  的可逆矩阵  $\mathbf{B}$  构成的矩阵空间  $Gl(n)$  组成李群（Lie group），它是一个Riemannian空间，其测度矩阵  $\mathbf{G}(\mathbf{B})$  是一个  $n^2 \times n^2$  的矩阵。定义一个将  $m \times m$  的矩阵映射到另一个  $n \times n$  的矩阵的矩阵算子“ $\odot$ ”，则对比函数  $J$  对分离矩阵  $\mathbf{B}$  的自然梯度为

$$\tilde{\nabla} J = \mathbf{G}^{-1}(\mathbf{B}) \odot \nabla J \quad (3.21)$$

Amari<sup>[6,7]</sup>证明自然梯度（3.21）不仅给出对比函数在Riemannian参数空间的最陡上升方向，而且是Fisher有效的。

设李群  $Gl(n)$  在  $\mathbf{B}$  点的正切空间（tangent space）为  $T_{\mathbf{B}}$ ， $\mathbf{B}$  的扰动为  $d\mathbf{B}$ ，则

$$d\mathbf{B} = d\mathbf{B}\mathbf{B}^{-1} \quad (3.22)$$

是正切空间  $T_{\mathbf{B}}$  在  $\mathbf{B}$  点的一个基，它将函数在  $\mathbf{B}$  点的扰动  $d\mathbf{B}$  映射到函数在单位矩阵  $\mathbf{I}$  点的扰动  $d\mathbf{B}$ ，从而考虑了参数空间的局部结构信息。

由导数的链式法则，对比函数  $J$  对  $\mathbf{B}_{ij}$  的偏导数

$$\frac{\partial J}{\partial \mathbf{B}_{ij}} = \sum_{p,q=1}^n \frac{\partial J}{\partial \mathbf{B}_{pq}} \cdot \frac{\partial \mathbf{B}_{pq}}{\partial \mathbf{B}_{ij}} \quad (3.23)$$

此外，根据定义（3.22）不难证明

$$\frac{\partial \mathbf{B}_{pq}}{\partial \mathbf{B}_{ij}} = \delta_{pi} \mathbf{B}_{jq} \quad (3.24)$$

其中  $\delta_{pi}$  是Kronecker函数，当  $p=i$  时  $\delta_{pi}=1$ ；否则， $\delta_{pi}=0$ 。

将（3.24）代入到（3.23）并写成矩阵形式，有

$$\frac{\partial J}{\partial \mathbf{B}} = \frac{\partial J}{\partial \mathbf{B}} \mathbf{B}^T \quad (3.25)$$

因此,

$$\frac{d\mathbf{B}}{dt} = \frac{d\mathcal{B}}{dt} \cdot \mathbf{B} = \frac{\partial J}{\partial \mathcal{B}} \cdot \mathbf{B} = \frac{\partial J}{\partial \mathbf{B}} \mathbf{B}^T \cdot \mathbf{B} \quad (3.26)$$

从而证明非奇异矩阵空间下的自然梯度

$$\tilde{\nabla} J = \nabla J \cdot \mathbf{B}^T \mathbf{B} \quad (3.27)$$

结合对比函数的一般梯度 (3.19), 可以得到盲信号分离的随机自然梯度算法<sup>[6,7]</sup>

$$\frac{d\mathbf{B}_t}{dt} = \eta_t [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T] \mathbf{B}_t \quad (3.28)$$

或者写成离散时间形式

$$\mathbf{B}_{t+1} = \mathbf{B}_t + \eta_t [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T] \mathbf{B}_t \quad (3.29)$$

关于自然梯度算法, 我们说明如下<sup>[177]</sup>。

**说明1:** 式 (3.29) 所示算法是Cichocki等人<sup>[38,39]</sup>在1994年首先提出的, 后来Amari等人<sup>[6,7]</sup>从理论上证明了其有效性, 指出它实际上是一种随机自然梯度学习算法。

**说明2:** 与一般梯度算法 (3.20) 不同, 自然梯度算法不需要矩阵求逆运算, 因此算法的运算量降低, 数值稳定性也大大增强。事实上可以证明<sup>[153,156]</sup>, 若初始矩阵  $\mathbf{B}_0$  非奇异, 那么当前时刻的分离矩阵  $\mathbf{B}_t$  也一定是非奇异的。

**说明3:** 除了满足分离矩阵的非奇异性条件外, 自然梯度算法还是等变化<sup>[22]</sup>的。换句话说, 算法的收敛性不依赖于具体的混合矩阵。

**说明4:** 用相对扰动量  $d\epsilon \cdot \mathbf{B}$  代替绝对扰动  $d\mathbf{B}$ , Cardoso和Laheld<sup>[22]</sup>定义了对比函数的相对梯度  $\hat{\nabla} J$ , 并证明  $\hat{\nabla} J = \nabla J \cdot \mathbf{B}^T$ 。因为最终获得的相对梯度算法与 (3.29) 具有相同的形式, 因此相对梯度算法和自然梯度算法二者是等价的。

### 3.3.3 等变化性

在统计检验中, 经常要求对数据所作的某种变换等价于对参数的同一变换。这一等价性称为等变化性 (equivariance), 它是统计学中的一个重要性质<sup>[103]</sup>。

考虑批处理方式的盲信号分离, 并根据  $L$  个观测样本  $\mathbf{x}_1, \dots, \mathbf{x}_L$  来估计混合矩阵  $\mathbf{A}$ , 记

$$\hat{\mathbf{A}} = \mathcal{A}(\mathbf{X}_L) \quad (3.30)$$

其中  $\mathcal{A}(\cdot)$  表示估计器, 而  $\mathbf{X}_L = [\mathbf{x}_1, \dots, \mathbf{x}_L] = \mathbf{A}\mathbf{S}_L$ 。

定义3.1<sup>[22]</sup> 对于任意可逆矩阵 $\mathbf{M}$ ，若 $\mathbf{A}$ 的估计器 $\mathcal{A}$ 均满足条件

$$\mathcal{A}(\mathbf{M}\mathbf{X}_L) = \mathbf{M}\mathcal{A}(\mathbf{X}_L) \quad (3.31)$$

则称估计器 $\mathcal{A}$ 是等变化的。

盲信号分离的等变化估计器有一个关键性质：提供盲信号分离的均匀性能。对均匀性能，可作如下理解<sup>[22]</sup>。假定源信号的估计 $\hat{\mathbf{s}}_i = \hat{\mathbf{A}}^{-1}\mathbf{x}_i$ ，其中 $\hat{\mathbf{A}}$ 由一个等变化的估计器获得，利用等变化公式(3.31)，我们有

$$\begin{aligned} \hat{\mathbf{s}}_i &= [\mathcal{A}(\mathbf{X}_L)]^{-1} \mathbf{x}_i = [\mathcal{A}(\mathbf{A}\mathbf{S}_L)]^{-1} \mathbf{A}\mathbf{s}_i \\ &= [\mathbf{A}\mathcal{A}(\mathbf{S}_L)]^{-1} \mathbf{A}\mathbf{s}_i = [\mathcal{A}(\mathbf{S}_L)]^{-1} \mathbf{s}_i \end{aligned} \quad (3.32)$$

因此，源信号的估计 $\hat{\mathbf{s}}_i$ 不依赖于混合矩阵 $\mathbf{A}$ ，而只与源信号 $\mathbf{S}_L = [\mathbf{s}_1, \dots, \mathbf{s}_L]$ 有关。

综合以上分析，所谓“均匀性能”指“盲信号分离算法的性能与源信号如何混合无关”。显然，这是盲信号分离算法期望具有的一个基本性能。

以批处理算法为例，上述研究了等变化性及其提供的均匀性能的涵义，那么对于自适应的盲信号分离算法，在什么样的条件下具有等变化性呢？Cardoso和Laheld<sup>[22]</sup>回答了这个问题。

设 $\mathbf{B}_i$ 是某个自适应算法的分离矩阵，若合成混合一分离系统的传递矩阵 $\mathbf{C}_i = \mathbf{B}_i\mathbf{A}$ 的更新具有如下形式：

$$\mathbf{C}_{i+1} = \mathbf{C}_i + \eta_i \mathbf{F}(\mathbf{C}_i \mathbf{s}_i) \mathbf{C}_i \quad (3.33)$$

其中 $\mathbf{F}(\mathbf{C}_i \mathbf{s}_i)$ 与单独的混合矩阵 $\mathbf{A}$ 和单独的分离矩阵 $\mathbf{B}_i$ 无关，则这种盲信号分离算法是等变化的。

式(3.33)意味着混合矩阵 $\mathbf{A}$ 只影响初始值 $\mathbf{C}_0$ ，而对 $\mathbf{C}_i$ 的收敛性不起决定性的作用。这正是盲信号分离算法所希望的均匀性能。根据条件(3.33)不难看出，自然梯度算法具有等变化性，而infomax算法(3.20)则不具备这种良好性质。

### § 3.3 信号相依性测度

盲信号分离的目的在于搜索最优加权矩阵 $\mathbf{B}_i$ ，使得网络输出 $\mathbf{y}_i$ 的各分量相互统计独立。根据概率论的基本知识，信号分量 $y_{i,i}$ 和 $y_{j,i}$ 相互独立意味着对任意可度量的函数 $g_1$ 和 $g_2$ ，有<sup>[125]</sup>

$$E\{g_1(y_{i,i})g_2(y_{j,i})\} - E\{g_1(y_{i,i})\}E\{g_2(y_{j,i})\} = 0, \quad \forall i \neq j \quad (3.34)$$

换句话说,  $g_1(y_{i,t})$  和  $g_2(y_{j,t})$  彼此不相关。因此, 为判断信号  $y_{i,t}$  和  $y_{j,t}$  是否相互独立, 理论上需要遍历  $g_1(y_{i,t})$  和  $g_2(y_{j,t})$  的相关性, 但这在实践中无法实现, 也没有必要。

为衡量信号  $y_{i,t}$  和  $y_{j,t}$  的相依性程度, 我们定义

$$sc_{ij,t} = \text{cov}[y_{i,t}, y_{j,t}] / \sqrt{\text{cov}[y_{i,t}] \text{cov}[y_{j,t}]} \quad (3.35)$$

$$hc_{ij,t} = \text{cov}[\varphi_i(y_{i,t}), y_{j,t}] / \sqrt{\text{cov}[\varphi_i(y_{i,t})] \text{cov}[y_{j,t}]} \quad (3.36)$$

其中  $\varphi_i(\cdot)$  是盲信号分离的分值函数, 也可以是其它形式的激励函数, 而

$$\text{cov}[x_t, y_t] = E\{[x_t - \bar{x}][y_t - \bar{y}]\} \quad (3.37)$$

$$\text{cov}[x_t] = E\{[x_t - \bar{x}]^2\} \quad (3.38)$$

分别表示  $x_t$  与  $y_t$  的协方差函数和  $x_t$  的方差, 式中  $\bar{x} = E\{x_t\}$ 。

显然,  $sc_{ij,t}$  表示信号  $y_{i,t}$  与  $y_{j,t}$  的二阶相关系数, 且  $sc_{ij,t} = sc_{ji,t}$ 。由于  $\varphi_i(y_{i,t})$  是  $y_{i,t}$  的非线性函数, 我们称  $hc_{ij,t}$  是信号  $y_{i,t}$  与  $y_{j,t}$  的高阶相关系数。需要说明的是  $hc_{ij,t} \neq hc_{ji,t}$ , 因为  $hc_{ij,t}$  描述的是  $\varphi_i(y_{i,t})$  和  $y_{j,t}$  之间的相关性, 而  $hc_{ji,t}$  描述的则是  $\varphi_j(y_{j,t})$  和  $y_{i,t}$  之间的相关性。

在盲信号分离中, 两个信号的分离状态可以用它们之间相互依赖的程度来度量, 并有下面的基本事实:

- 信号  $y_{i,t}$  和  $y_{j,t}$  强相依, 如果二阶相关系数  $sc_{ij,t}$  以及高阶相关系数  $hc_{ij,t}$  和  $hc_{ji,t}$  三者中有一个取值很大;
- 信号  $y_{i,t}$  和  $y_{j,t}$  弱相依, 如果二阶相关系数  $sc_{ij,t}$  以及高阶相关系数  $hc_{ij,t}$  和  $hc_{ji,t}$  三者取值都比较小;
- 信号  $y_{i,t}$  和  $y_{j,t}$  近似相互独立, 如果二阶相关系数  $sc_{ij,t}$  以及高阶相关系数  $hc_{ij,t}$  和  $hc_{ji,t}$  三者取值都足够小。

**定义3.2** 相关系数  $sc_{ij,t}$ ,  $hc_{ij,t}$  及  $hc_{ji,t}$  的最大绝对值定义为信号  $y_{i,t}$  与  $y_{j,t}$  之间的相依性测度, 并记为  $D_{ij,t}$ , 即

$$D_{ij,t} = \max\{|sc_{ij,t}|, |hc_{ij,t}|, |hc_{ji,t}|\} \quad (3.39)$$

显然, 定义3.2与上述基本事实相符。换句话说, 若  $D_{ij,t}$  很大 (例如  $0.25 < D_{ij,t} < 1$ ), 则信号  $y_{i,t}$  和  $y_{j,t}$  强相依; 若  $D_{ij,t}$  比较小 (如  $0.05 < D_{ij,t} < 0.25$ ); 则信号  $y_{i,t}$  和  $y_{j,t}$  弱相依; 当  $D_{ij,t}$  足够小 (如  $0 < D_{ij,t} < 0.05$ ) 时, 可近似认为信号  $y_{i,t}$  和  $y_{j,t}$  相互独立。特别



地, 对于相干信号  $y_{i,t}$  和  $y_{j,t}$  有  $D_{ij,t} = 1$ ; 对于相互独立信号  $y_{i,t}$  和  $y_{j,t}$  有  $D_{ij,t} = 0$ 。

定义3.2对于预白化的信号同样成立, 在这种情况下, 二阶相关系数  $sc_{ij,t} = 0$ , 相依性测度简化为  $D_{ij,t} = \max\{|hc_{ij,t}|, |hc_{ji,t}|\}$ 。

此外, 定义3.2确保  $D_{ij,t} = D_{ji,t}$ , 这与信号之间的相依性是对称的这一事实相符。

为了描述一个信号与其它信号相互依赖的程度, 以及所有信号的分离状况, 我们对定义3.2作如下推广。

**定义3.3** 信号  $y_{i,t}$  与其它所有信号之间的相依性测度定义为

$$D_{i,t} = \max_{\forall j, j \neq i} D_{ij,t} = \max_{\forall j, j \neq i} \{|sc_{ij,t}|, |hc_{ij,t}|, |hc_{ji,t}|\} \quad (3.40)$$

**定义3.4** 所有信号之间的相依性测度描述了所有信号的总体分离状态, 定义为

$$D_t = \max_i \{D_{i,t}\} = \max_{\forall i, j, j \neq i} \{|sc_{ij,t}|, |hc_{ij,t}|, |hc_{ji,t}|\} \quad (3.41)$$

为在线跟踪信号相依性测度, 我们需要推导相关系数的自适应更新。假定网络输出  $y_t$  是各态历经的平稳随机过程, 其统计平均可用时间平均来替代。那么, 对任意的平稳离散时间序列  $x_t$ , 其  $t$  时刻的均值可以写成

$$\bar{x}_t = \frac{1}{t} \sum_{i=1}^t \lambda^{t-i} x_i = \lambda \frac{t-1}{t} \bar{x}_{t-1} + \frac{1}{t} x_t$$

其中  $0 < \lambda < 1$  为遗忘因子。记

$$R_{ij,t} = \text{cov}[y_{i,t}, y_{j,t}] \quad (3.42)$$

$$P_{ij,t} = \text{cov}[\varphi_i(y_{i,t}), y_{j,t}] \quad (3.43)$$

$$Q_{ii,t} = \text{cov}[\varphi_i(y_{i,t})] \quad (3.44)$$

经类似的推导, 可得

$$\bar{y}_{i,t} = \lambda \frac{t-1}{t} \bar{y}_{i,t-1} + \frac{1}{t} y_{i,t} \quad (3.45)$$

$$\Delta_{i,t} = \bar{y}_{i,t} - \bar{y}_{i,t-1} \quad (3.46)$$

$$R_{ij,t} = \lambda \frac{t-1}{t} [R_{ij,t-1} + \Delta_{i,t} \Delta_{j,t}] + \frac{1}{t} [y_{i,t} - \bar{y}_{i,t}] [y_{j,t} - \bar{y}_{j,t}] \quad (3.47)$$

$$\bar{\varphi}_{i,t} = \lambda \frac{t-1}{t} \bar{\varphi}_{i,t-1} + \frac{1}{t} \varphi_i(y_{i,t}) \quad (3.48)$$

$$\nabla_{i,j} = \bar{\varphi}_{i,j} - \bar{\varphi}_{i,j-1} \quad (3.49)$$

$$P_{ij,t} = \lambda \frac{t-1}{t} [P_{ij,t-1} + \nabla_{i,j} \Delta_{j,t}] + \frac{1}{t} [\varphi_i(y_{i,t}) - \bar{\varphi}_{i,j}] [y_{j,t} - \bar{y}_{j,t}] \quad (3.50)$$

$$Q_{ii,t} = \lambda \frac{t-1}{t} [Q_{ii,t-1} + \nabla_{i,i}^2] + \frac{1}{t} [\varphi_i(y_{i,t}) - \bar{\varphi}_{i,i}]^2 \quad (3.51)$$

其中  $i, j = 1, \dots, n$ 。这样，一旦递推计算出  $R_{ij,t}$ ,  $P_{ij,t}$  和  $Q_{ii,t}$ ，即可利用式 (3.35) 和 (3.36) 计算  $sc_{ij,t}$ ,  $hc_{ij,t}$  和  $hc_{ji,t}$ ，进而求得信号相依性测度  $D_{ij,t}$ ,  $D_{i,t}$  和  $D_t$ 。

### § 3.4 分阶段学习的盲信号分离

由式 (3.29) 可以看出，自然梯度算法取决于两个方面：学习速率参数  $\eta_i$  和非线性激励函数  $\varphi_i(\cdot)$ 。下面讨论学习速率  $\eta_i$  的确定问题。因为自然梯度算法可以归类为 LMS 型算法，因此当学习速率  $\eta_i$  取固定的常数时，要兼顾收敛速度和稳态性能（它决定源信号的恢复质量）是困难的。为此， $\eta_i$  通常取时变函数。最简单的是采用时间递减函数<sup>[3,7,156]</sup>，更好的选择是使用自适应的步长参数<sup>[41,56,112,118]</sup>，但它们都没有和信号的分离状态或者相依性直接挂钩，因此效果有限。此外，采用一维参数  $\eta_i$  意味着对所有信号使用相同的步长进行学习，也不尽合理，因为先分离出来的信号显然应使用小的步长以保持跟踪，而与其它信号强相依的信号则需要大步长以加速其分离。基于这一考虑，我们提出分阶段学习的盲信号分离，其核心思想是用学习速率矩阵  $\Lambda_t = [\eta_{ij,t}]$  代替原有算法中的一维参数  $\eta_i$ 。以自然梯度算法 (3.29) 为例，分阶段学习算法的形式如下：

$$\mathbf{B}_{t+1} = \mathbf{B}_t + \left\{ \Lambda_t \circ [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T] \right\} \mathbf{B}_t \quad (3.52)$$

式中“ $\circ$ ”表示矩阵的 Hadamard 乘积，即  $\mathbf{A} \circ \mathbf{B} = [a_{ij} b_{ij}]$ ，而  $\Lambda_t$  取决于相依性测度  $D_{ij,t}$ ,  $D_{i,t}$  和  $D_t$ 。

借助于  $D_t$ ，我们将整个信号分离过程分成三个阶段：

- 初始阶段：  $0.25 \leq D_t \leq 1$ ；
- 捕捉阶段：  $0.05 \leq D_t < 0.25$ ；
- 跟踪阶段：  $0 \leq D_t < 0.05$ 。

初始阶段可视为信号的预分离阶段，实现信号的初步分离；捕捉阶段实现信号的粗分离，以捕捉各个源信号；最后，跟踪阶段实现信号的精细分离，尽可能地提高源信号的恢复质量。

对于算法 (3.52), 我们有下面的定理。

**定理3.1** 分阶段学习的盲信号分离算法 (3.52) 是等变化的, 并满足分离矩阵的非奇异性条件。

证明 在算法 (3.52) 的两边同时右乘以分离矩阵  $\mathbf{A}$ , 并令  $\mathbf{C}_t = \mathbf{B}_t \mathbf{A}$ , 则

$$\mathbf{C}_{t+1} = \mathbf{C}_t + \left\{ \Lambda_t \circ [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T] \right\} \mathbf{C}_t$$

因为  $\mathbf{F}(\mathbf{y}_t) = \mathbf{F}(\mathbf{C}_t \mathbf{s}_t) = \Lambda_t \circ [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T]$  与单独的  $\mathbf{A}$  和  $\mathbf{B}_t$  均无关, 根据式 (3.33), 算法 (3.52) 是等变化的。

记  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^T \mathbf{Y}) = \sum_{ij} \mathbf{X}_{ij} \mathbf{Y}_{ij}$ ,  $|\mathbf{B}_t| = \det(\mathbf{B}_t)$ , 类似于文献[153,156], 有

$$\begin{aligned} \frac{d|\mathbf{B}_t|}{dt} &= \left\langle \frac{\partial |\mathbf{B}_t|}{\partial \mathbf{B}_t}, \frac{d\mathbf{B}_t}{dt} \right\rangle \\ &= \left\langle |\mathbf{B}_t| \mathbf{B}_t^{-T}, \left\{ \Lambda_t \circ [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T] \right\} \mathbf{B}_t \right\rangle \\ &= |\mathbf{B}_t| \cdot \text{trace} \left\{ \Lambda_t \circ [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T] \right\} \\ &= |\mathbf{B}_t| \sum_{i=1}^n \eta_{ii,t} [1 - \varphi_i(y_{i,t}) y_{i,t}] \end{aligned}$$

因此,

$$\det(\mathbf{B}_t) = \det(\mathbf{B}_0) \cdot \exp \left[ \int_0^t \sum_{i=1}^n \eta_{ii}(\tau) \{1 - \varphi_i(y_{i,\tau}) y_{i,\tau}\} d\tau \right]$$

显然, 若初始分离矩阵  $\mathbf{B}_0$  非奇异, 则当前时刻的分离矩阵  $\mathbf{B}_t$  也必定是非奇异的, 从而分阶段学习算法 (3.52) 满足分离矩阵的非奇异性条件。证明完毕。 ■

下面我们讨论学习速率矩阵  $\Lambda_t$  的设计, 并对分阶段学习算法 (3.52) 与现有的几种著名的变步长算法作比较。

#### 3.4.1 初始阶段

在盲信号分离的初始阶段, 各信号通常是强相依的, 因此需要使用较大的学习速率以加速信号的分离。显然, 在这个阶段对所有信号使用同一学习速率是合理的, 即学习速率矩阵取常数矩阵:  $\eta_{ij,t} = \eta_t$ , 从而分阶段学习的盲信号分离算法 (3.52) 退化为一般的自然梯度算法 (3.29)。

参数  $\eta_t$  有多种选择, 最简单的是选择常数。选择之二是采用时间递减函数, 如退

火规则 (cooling scheme) 步长<sup>[7]</sup>, 即学习速率从某个初始值开始, 然后按规则  $\frac{1}{2^t}$  递减。最近, Yang 提出使用指数衰减函数<sup>[156]</sup>

$$\eta_t = \begin{cases} \eta_0, & t \leq T_0 \\ \eta_0 \exp[-T_d(t - T_0)], & t > T_0 \end{cases} \quad (3.53)$$

式中  $\eta_0$ ,  $T_0$  及  $T_d$  是适当选择的常数。

时间递减步长存在的问题是: 若在  $\eta_t$  已变得很小时, 还有信号未被分离, 那么这些信号将难于被分离。为克服这一缺点, 又提出了第三种选择, 即自适应步长。最早的方法是所谓的自适应梯度步长算法<sup>[56,90,112]</sup>, 即

$$\eta_{t+1} = \eta_t + \rho_t \cdot \frac{\partial J(\mathbf{B}_t)}{\partial \eta_t} \quad (3.54)$$

式中  $J(\cdot)$  为算法的对比函数,  $\rho_t$  是“步长之步长”参数。Douglas 等人<sup>[56]</sup>指出, 除了鲁棒性较差外, 梯度步长的最大困难在于参数  $\rho_t$  的选择。事实上, 这一困难导致该方法并不实用。

利用分离矩阵在每次迭代中的调整量这一信息, Murata 等人<sup>[118]</sup>提出了下面的自适应步长算法:

$$\Gamma_{t+1} = (1 - \xi)\Gamma_t - \xi[\mathbf{I} - \Phi(\mathbf{y}_t)\mathbf{y}_t^T]\mathbf{B}_t \quad (3.55)$$

$$\eta_{t+1} = \eta_t + \alpha\eta_t[\beta\|\Gamma_t\|_F - \eta_t] \quad (3.56)$$

式中  $\alpha$ ,  $\beta$  和  $0 < \xi < 1$  为待定参数, 而  $\|\Gamma_t\|_F = \sqrt{\sum_{ij} \Gamma_{ij}^2}$  表示矩阵  $\Gamma_t$  的 Frobenius 范数。若考虑稳定性, 还需对学习速率加约束条件  $\eta_{\min} \leq \eta_t \leq \eta_{\max}$ 。

与上述方法都不同, 分阶段学习算法在初始阶段根据测度  $D_t$  选择自适应步长  $\eta_t$ , 并使用下面的简单线性函数

$$\eta_t = \begin{cases} \eta_0, & t \leq T_0 \\ a + (\eta_0 - a)D_t, & t > T_0 \text{ 且 } D_t \geq 0.25 \end{cases} \quad (3.57)$$

式中  $a$ ,  $\eta_0$  及  $T_0$  是待定常数。当  $t \leq T_0$  时, 我们选择  $\eta_t = \eta_0$  是因为观测样本少时, 相关系数  $sc_{ij,t}$ ,  $hc_{ij,t}$  和  $hc_{ji,t}$  的估计误差较大。此外, 在盲信号分离的最初阶段, 信号一般强相依也部分说明了这种选择的合理性。

### 3.4.2 捕捉阶段

本阶段盲信号分离的主要目的是捕捉到所有的信号分量。考虑到部分信号可能在

初始阶段已被分离或者部分分离, 因此对所有信号使用相同步长进行学习不再最优。为了在跟踪已分离信号的同时, 加速捕捉未分离信号, 我们对不同信号分量采用不同的学习速率, 即学习速率矩阵是每行元素取相同值的特殊矩阵。

设矩阵  $\Lambda_i$  的第  $i$  行元素为  $\eta_{ij,i} = \eta_{i,i}$ , 并记对角阵  $\mathcal{D}_i = \text{diag}[\eta_{1,i}, \dots, \eta_{n,i}]$ , 则基于特殊结构  $\Lambda_i \circ \mathbf{B}_i = \mathcal{D}_i \mathbf{B}_i$ , 分阶段学习算法 (3.52) 可简化为

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \mathcal{D}_i [\mathbf{I} - \Phi(\mathbf{y}_i) \mathbf{y}_i^T] \mathbf{B}_i \quad (3.58)$$

式中  $\mathcal{D}_i$  的对角元素

$$\eta_{i,i} = \mathcal{J}(D_{i,i}), \quad i = 1, \dots, n \quad (3.59)$$

而  $\mathcal{J}(\cdot)$  是一个适当选择的非线性函数。

最近, Yang 由 Fisher 信息矩阵出发, 提出了一种串行更新算法<sup>[156]</sup>,

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i \mathcal{D}_\mu^{-1} [\mathbf{I} - \Phi(\mathbf{y}_i) \mathbf{y}_i^T] \mathbf{B}_i \quad (3.60)$$

其中  $\mathcal{D}_\mu = \text{diag}[\mu_1, \dots, \mu_n]$ 。算法 (3.60) 有以下约束条件:

C1) 所有源信号的概率密度函数是偶函数;

C2)  $\mathcal{H} = E \left\{ \text{vec} [\mathbf{I} - \Phi(\mathbf{y}_i) \mathbf{y}_i^T] \text{vec}^T [\mathbf{I} - \Phi(\mathbf{y}_i) \mathbf{y}_i^T] \right\}$  的三元组  $(\mu_i, \lambda_i, \nu_i)$  满足关系式  $\nu_i = \mu_i \lambda_i$ , 其中矩阵算子  $\text{vec}[\mathbf{M}]$  表示将矩阵  $\mathbf{M}$  从第一列开始一直到最后一列, 列与列相连得到的列向量;

C3) 混合矩阵  $\mathbf{A}$  是方阵。

我们强调, 算法 (3.58) 和算法 (3.60) 在形式上类似, 但存在本质上的区别。首先算法 (3.60) 中的对角矩阵  $\mathcal{D}_\mu^{-1}$  的元素与信号的分离状态无关, 而算法 (3.58) 中的对角矩阵  $\mathcal{D}_i$  则取决于相依性测度  $D_{i,i}$ 。其次分阶段学习的盲分离算法 (3.58) 没有条件 C1)~C3) 的限制。

### 3.4.3 跟踪阶段

一旦捕捉到所有的源信号分量, 信号分离便进入到跟踪阶段。在此阶段, 学习速率矩阵的各个元素取比较小的值, 以尽可能减小各分量之间的影响, 并提高信号的恢复质量, 实现源信号的精细分离。从算法收敛看, 本阶段要求网络收敛到平衡点, 使权矩阵  $\mathbf{B}_i$  达最优解  $\mathbf{B}_{opt}$ 。为此, 我们取



$$\eta_{ij,t} = \begin{cases} \mathcal{J}(D_{i,t}) & , \quad i = j \\ \mathcal{J}(D_{ij,t}) & , \quad i \neq j \end{cases} \quad (3.61)$$

其中函数  $\mathcal{J}(\cdot)$  与式 (3.59) 取同一非线性函数。

我们注意到Cichocki和Amari等人在文献[41]中最早提出了一种对神经网络各突触权采用不同学习速率的盲信号分离算法：

$$\frac{d\mathbf{B}_{ij,t}}{dt} = \eta_{ij,t} \left[ \mathbf{B}_{ij,t} - \varphi_i(y_{i,t}) \sum_{p=1}^n \mathbf{B}_{pj,t} y_{p,t} \right] \quad (3.62)$$

式中

$$\mathbf{g}_{ij,t} = \mathbf{B}_{ij,t} - \varphi_i(y_{i,t}) \sum_{p=1}^n \mathbf{B}_{pj,t} y_{p,t} \quad (3.63)$$

$$\tau_1 \frac{dv_{ij,t}}{dt} = -v_{ij,t} + |\mathbf{g}_{ij,t}| \quad (3.64)$$

$$\tau_2 \frac{d\eta_{ij,t}}{dt} = -\eta_{ij,t} + \gamma |v_{ij,t}| \quad (3.65)$$

而  $\tau_1, \tau_2$  及  $\gamma$  是恰当选择的正常数。将 (3.62) 写成矩阵形式，有

$$\mathbf{B}_{t+1} = \mathbf{B}_t + \Lambda_t \circ \left\{ \left[ \mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T \right] \mathbf{B}_t \right\} \quad (3.66)$$

比较算法 (3.66) 和分阶段学习算法 (3.52)，尽管它们在形式上相似，但有根本的区别，因为 (3.66) 所示的盲信号分离算法不具有等变化性。证明是简单的。式 (3.66) 两边同时右乘以混合矩阵  $\mathbf{A}$ ，得

$$\mathbf{C}_{t+1} = \mathbf{C}_t + \left\{ \Lambda_t \circ \left[ \left( \mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T \right) \mathbf{B}_t \right] \right\} \mathbf{A}$$

因此，除  $\Lambda_t$  取常数矩阵或者各行元素取相同值这两种特殊情况外，合成系统  $\mathbf{C}_t$  的更新依赖于混合矩阵  $\mathbf{A}$  和分离矩阵  $\mathbf{B}_t$ ，从而 (3.66) 不是一种等变化的算法。此外，算法 (3.66) 也不满足分离矩阵的非奇异性条件。

如前所述，等变化性提供盲信号分离的均匀性能，确保算法收敛与具体的混合矩阵无关，而分离矩阵的非奇异性是算法稳定工作的前提条件，它们是任何一种有效的盲信号分离算法应该具备的两个关键性能。可以预见，算法 (3.66) 难以获得满意的收敛性。

## § 3.4 性能仿真

为了验证所提出的分阶段学习盲信号分离算法的有效性,我们以自然梯度算法为例,对以下信号进行了仿真实验:

- S1: 符号信号  $\text{sign}(\cos(2\pi 155t))$ ;
- S2: 高频正弦信号  $\sin(2\pi 800t)$ ;
- S3: 低频正弦信号  $\sin(2\pi 90t)$ ;
- S4: 相位调制信号  $\sin(2\pi 300t - 6\cos(2\pi 60t))$ ;
- S5: 幅度调制信号  $\sin(2\pi 9t)\sin(2\pi 300t)$ ;
- S6: 在 $[-1,1]$ 均匀分布的随机噪声信号;
- S7: 参数为1的瑞利分布随机信号的平方根信号。

仿真中,传输信道即混合矩阵 $\mathbf{A}$ 的元素为在 $[-1,1]$ 区间均匀分布或者均值为零、方差为0.5的高斯分布的随机数。为书写方便,前者记为 $\mathbf{A} \sim \mathcal{U}(-1,1)$ ,后者记为 $\mathbf{A} \sim \mathcal{N}(0,0.5)$ 。以速率10kHz对传感器接收信号采样以产生离散观测信号 $\mathbf{x}_l$ ,且 $L=4000$ ,即用了4000个观测样本。

盲信号分离算法的性能用“串音”误差<sup>[3,78,153,156]</sup>来评价,即

$$E_{cl} = \sum_{k=1}^n \left( \sum_{l=1}^n \frac{|c_{kl}|}{\max_p |c_{kp}|} - 1 \right) + \sum_{l=1}^n \left( \sum_{k=1}^n \frac{|c_{kl}|}{\max_p |c_{pl}|} - 1 \right) \quad (3.67)$$

式中 $\mathbf{C}_l = [c_{kl}] = \mathbf{B}_l \mathbf{A}$ 。

除了学习速率参数外,决定盲信号分离学习算法的另一个因素是非线性激励函数 $\Phi(\mathbf{y}_l) = [\varphi_1(y_{1,l}), \dots, \varphi_n(y_{n,l})]^T$ 。由第二章的基本知识以及本章第2节对比函数的推导过程可以看出,激励函数的最优选择是分值函数。但分值函数取决于未知的源信号的概率分布。为解决这个问题,通常采用两种方法。一种是所谓的“自底而顶”的方法,它根据算法的局部稳定性条件选择激励函数,这部分内容在第二章已经作了介绍。另一种是所谓的“自顶而底”的方法,它首先为源信号的概率分布建立模型,然后根据观测信号样本估计模型参数,并由此确定激励函数。作为第二种方法的一个典型, Yang和Amari<sup>[153]</sup>利用截短的Gram-Charlier展开为源信号的分布建模,经过推导,最终获得的激励函数具有如下形式:

$$\varphi_l(y_{l,l}) = f_l(\kappa_{3l}, \kappa_{4l}) y_{l,l}^2 + g_l(\kappa_{3l}, \kappa_{4l}) y_{l,l}^3 \quad (3.68)$$

式中 $\kappa_{3l} = E\{y_{l,l}^3\}$ 表示信号 $y_{l,l}$ 的偏度(skewness),  $\kappa_{4l} = E\{y_{l,l}^4\} - 3$ 表示信号的(未归一

化的)峰度(kurtosis),而

$$f_i(\kappa_{3i}, \kappa_{4i}) = -\frac{1}{2}\kappa_{3i} + \frac{9}{4}\kappa_{3i}\kappa_{4i} \quad (3.69)$$

$$g_i(\kappa_{3i}, \kappa_{4i}) = -\frac{1}{6}\kappa_{4i} + \frac{3}{2}\kappa_{3i}^2 + \frac{3}{4}\kappa_{4i}^2 \quad (3.70)$$

偏度和峰度用下面的公式更新:

$$\kappa_{3i,t+1} = \kappa_{3i,t} - \mu_t \cdot (\kappa_{3i,t} - y_{i,k}^3) \quad (3.71)$$

$$\kappa_{4i,t+1} = \kappa_{4i,t} - \mu_t \cdot (\kappa_{4i,t} - y_{i,k}^4 + 3) \quad (3.72)$$

式中 $\mu_t$ 为偏度和峰度更新的步长参数。仿真中取 $\mu_t = 60 \times T$ ,而 $T=0.0001$ 表示对观测信号的采样周期。

作为比较,我们同时运行学习速率参数取常数、时间递减函数以及自适应步长三种算法,为实现快速收敛并获得良好的稳态性能(提高源信号的恢复质量),对应的近似最优参数选择如下:

- (1) Yang-Amari 算法(简称 Y-A 算法<sup>[153]</sup>)取常数步长,  $\eta_k = 65T$ ;
- (2) Yang算法<sup>[156]</sup>取指数衰减步长(3.53),且 $\eta_0 = 140T$ ,  $T_0 = 500$ ,  $T_d = 15T$ ;
- (3) MMZA 算法<sup>[118]</sup>取自适应步长(3.55)~(3.56),参数 $\eta_{\max} = 140T$ ,  $\eta_{\min} = 5T$ ,  $\xi = 5T$ ,  $\alpha = 0.02$ ,  $\beta = 150 \|\Gamma_t\|^{0.8} / \max_{p \leq t} \|\Gamma_p\|^{1.8}$ ;

(4) 分阶段学习算法:式(3.45)~(3.51)取遗忘因子 $\lambda = 0.999$ ;式(3.57)取 $\eta_0 = 140T$ ,  $T_0 = 500$ ,  $a = 40T$ ;式(3.59)和(3.61)中的非线性函数 $\mathcal{J}(\cdot)$ 必须是一个单调不增的有界函数,它可以有多种选择(例如指数函数、S形函数和分段线性函数等)。为了提供更好的捕捉和跟踪能力,我们选择捕捉阶段的 $\mathcal{J}(\cdot)$ 为凸函数,跟踪阶段的 $\mathcal{J}(\cdot)$ 为凹函数,合起来取S形函数:

$$\mathcal{J}(x) = T \cdot \begin{cases} 1350x^{1.3} & , 0 \leq x < 0.05 \\ 27.5 + 500(x - 0.05)^{0.7} & , 0.05 \leq x < 0.2 \\ 160 & , 0.2 \leq x \end{cases} \quad (3.73)$$

为进一步了解上述算法在信道与/或信号不同情况下的性能,实验仿真了以下四种情况:

- 源信号由确定性信号S1~S5及随机信号S6组成,信道 $A \sim \mathcal{U}(-1,1)$ ;
- 源信号由确定性信号S1~S5及随机信号S6组成,信道 $A \sim \mathcal{N}(0,0.5)$ ;
- 源信号由确定性信号S1~S5及随机信号S6~S7组成,信道 $A \sim \mathcal{U}(-1,1)$ ;

- 源信号由确定性信号S1~S5及随机信号S6~S7组成, 信道  $A \sim \mathcal{N}(0, 0.5)$ 。

初始分离矩阵取  $B_0 = 0.5I$ 。每种情况独立运行200次, 每次运行中的随机信号S6~S7以及混合矩阵  $A$  都是随机产生的。表3.1是四种算法不同情况下不成功分离(指到算法结束仍有部分信号未能成功抽取)次数的统计结果。图3.1和图3.2分别给出了四种算法针对6个源信号S1~S6的400次独立运行(其中200次混合矩阵  $A \sim \mathcal{U}(-1, 1)$ , 另200次混合矩阵  $A \sim \mathcal{N}(0, 0.5)$ ) 串音误差的均值和标准离差; 而图3.3和图3.4则分别是四种算法对7个源信号S1~S7的400独立运行的串音误差的均值与标准离差。

从这些表与图可以看出, 本文提出的分阶段学习算法明显优于其它三种算法, 它既能以最快的速度收敛(串音误差下降最快), 而且稳态性能也最好(串音误差均值最小)。当信号与/或信道变化时, 它工作也最稳定(标准离差最小, 不成功分离的次数最少)。

表3.1 不成功分离次数统计

信号与信道		学习速率算法			
源信号个数	信道矩阵	Y-A算法	Yang算法	MMZA算法	分阶段算法
$n = 6$	$A \sim \mathcal{U}(-1, 1)$	3 (1.5%)	7 (3.5%)	5 (2.5%)	1 (0.5%)
$n = 6$	$A \sim \mathcal{N}(0, 0.5)$	5 (2.5%)	9 (4.5%)	7 (3.5%)	1 (0.5%)
$n = 7$	$A \sim \mathcal{U}(-1, 1)$	15 (7.5%)	40 (20%)	20 (10%)	3 (1.5%)
$n = 7$	$A \sim \mathcal{N}(0, 0.5)$	15 (7.5%)	38 (19%)	22 (11%)	3 (1.5%)

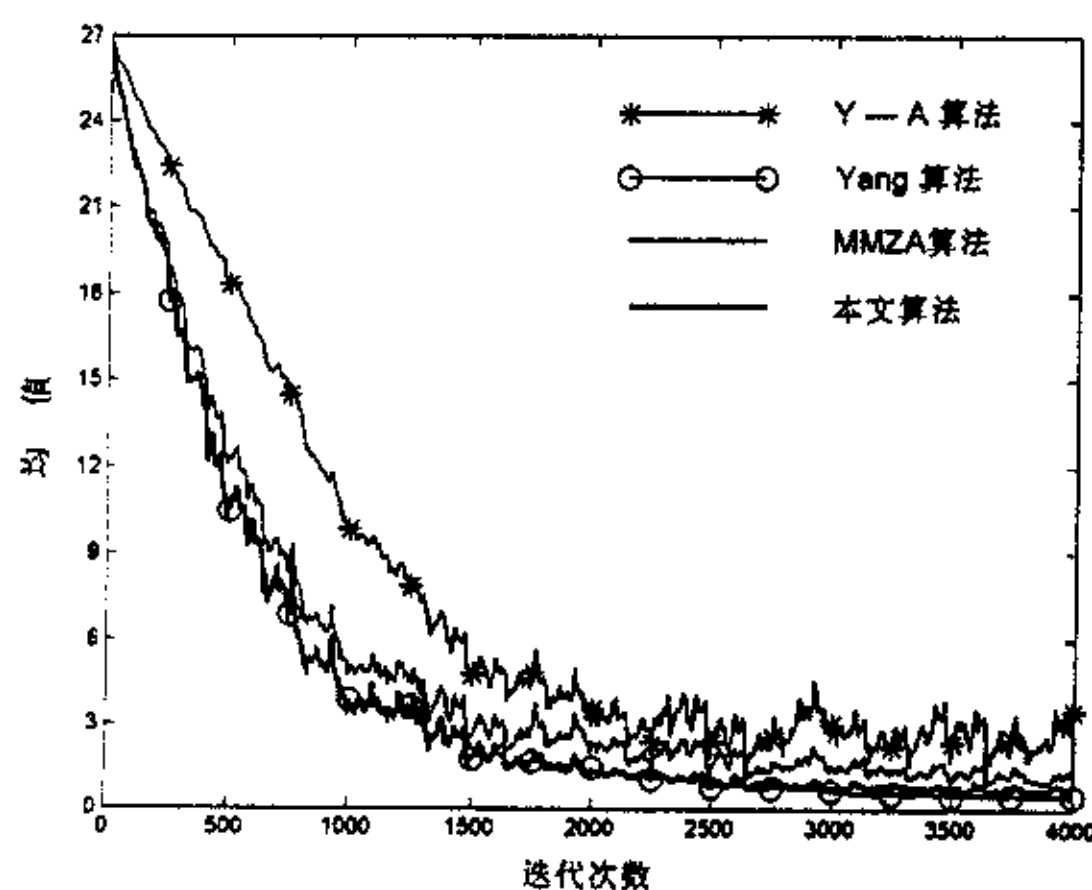


图3.1 串音误差均值图,  $n = 6$

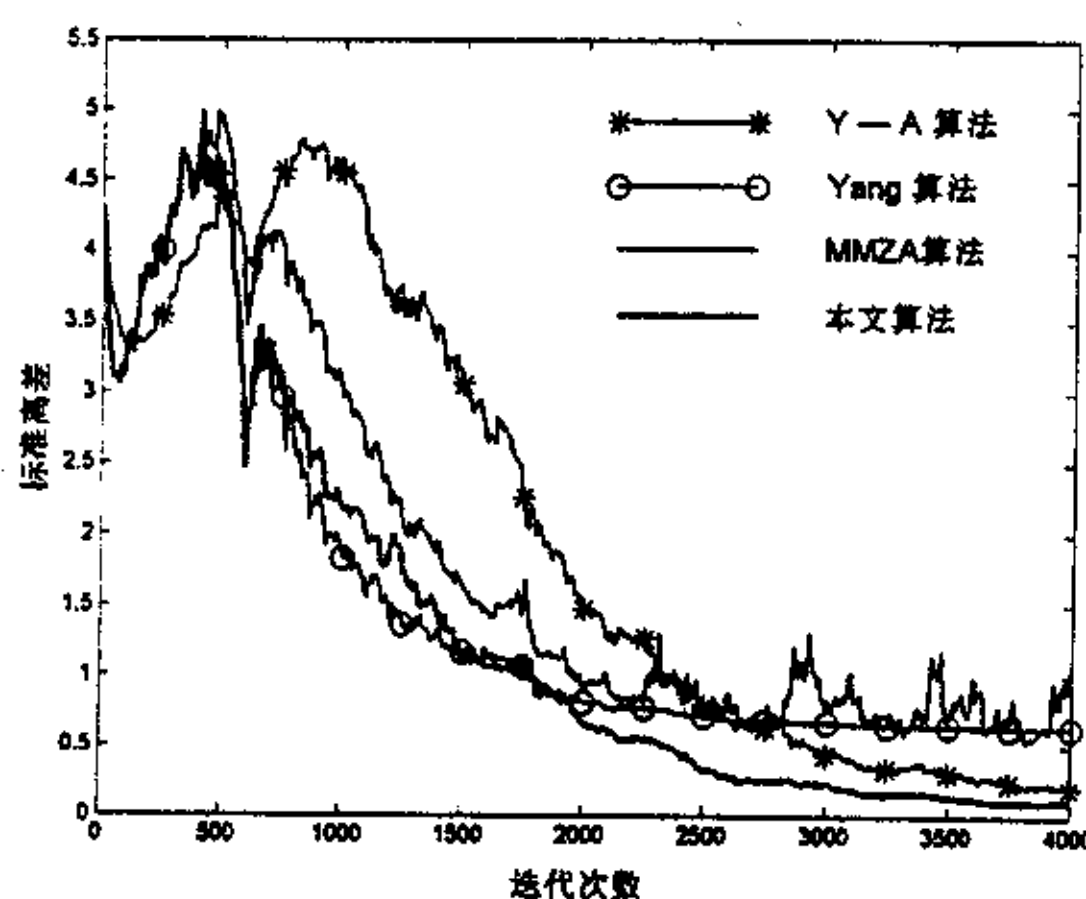
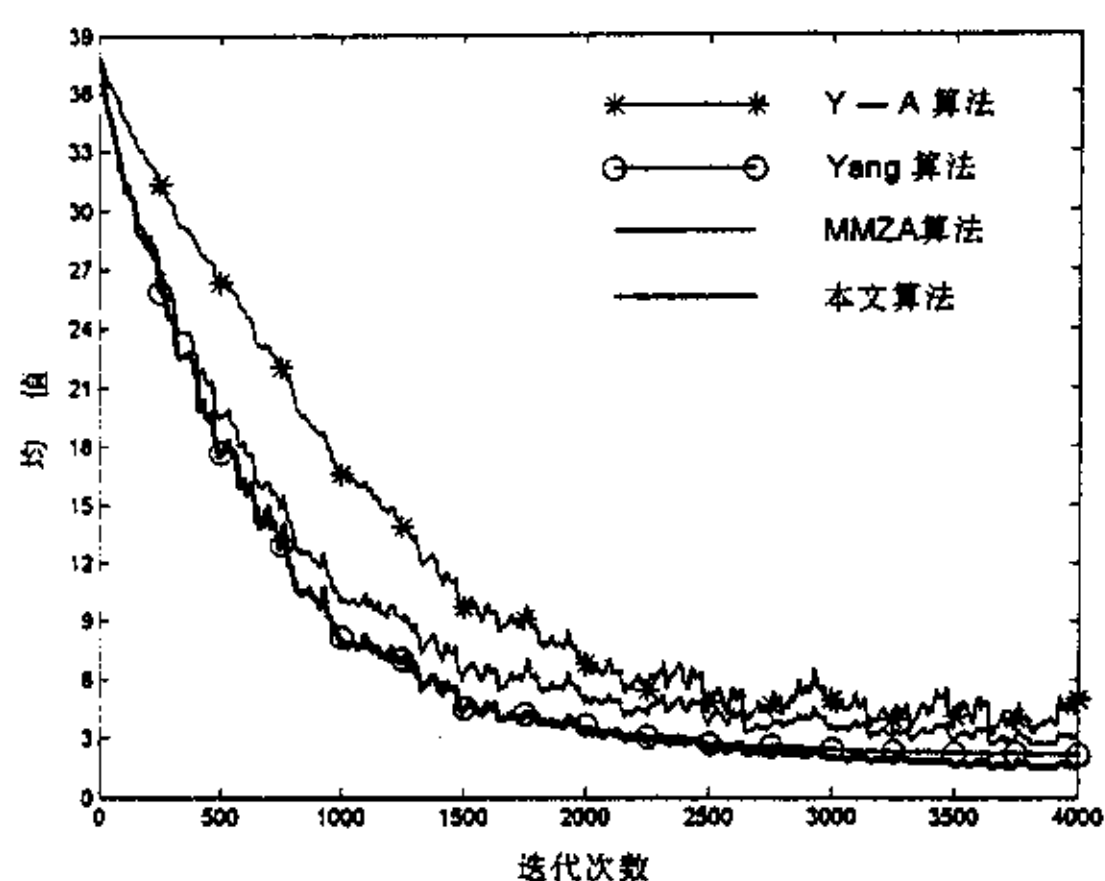
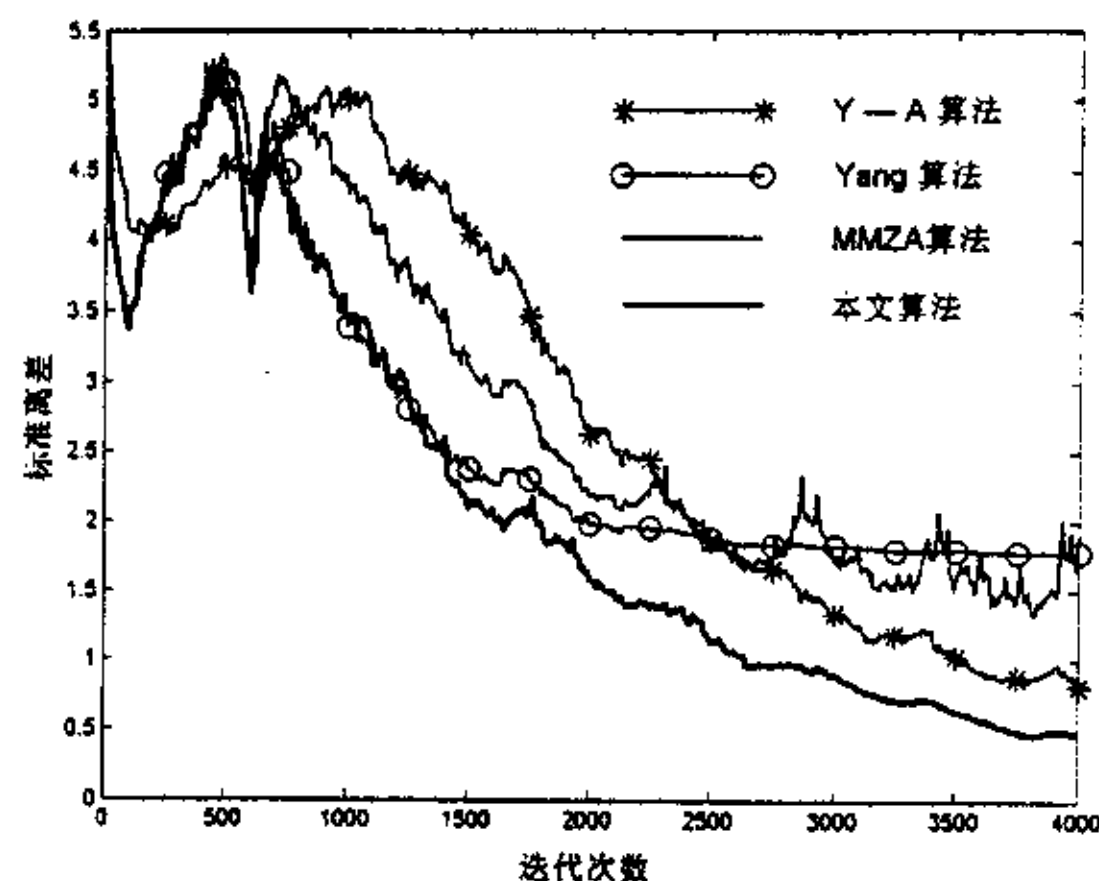


图3.2 串音误差标准离差图,  $n = 6$

图3.3 串音误差均值图,  $n=7$ 图3.4 串音误差标准离差图,  $n=7$ 

我们指出, 学习速率参数是一个经验值, 除了函数(3.73)外, 还可以有其它的形式。仿真表明在一个合理的范围内, 它虽然对算法单次运行的性能有影响, 但统计结果是相当的。此外, 本章提出的分阶段学习的思想除了应用于自然梯度算法外, 还可以广泛应用到其它的盲信号分离算法, 如EASI算法和非线性PCA算法等。

### § 3.6 小 结

如何提高算法的收敛速度和改进算法的稳态性能一直是盲信号分离的一个热点课题。为解决这个问题, 本章提出了一种分阶段学习的新思想。本章主要内容包括:

(1) 推导了独立分量分析的对比函数, 并介绍了盲信号分离中经常用到的自然梯度和等变化性这两个概念。

(2) 定义了两个信号之间的相依性测度, 并推广到一个信号与其它信号之间以及所有信号之间的相依性测度。

(3) 利用信号相依性测度, 提出了一种分阶段学习的新算法, 它将学习速率参数由一维标量推广到二维学习速率矩阵, 并把整个信号分离过程分成三个阶段进行。学习速率矩阵在不同阶段有不同的形式: 在初始阶段取常数矩阵, 在捕捉阶段取对角矩阵, 在跟踪阶段则为一般矩阵。由于学习速率矩阵的元素由信号的分离状态决定, 因此新算法既能加快信号分离前半阶段的收敛, 又能提高后半阶段的跟踪和恢复的精度, 从而很好地克服了收敛速度和信号恢复质量之间的矛盾。

(4) 比较分析了分阶段学习的盲信号分离算法与现有的固定或可变步长盲信号分离算法的联系和区别, 并通过大量的仿真试验验证了新算法的有效性。



## 第四章 基于自然梯度的RLS盲信号分离

### §4.1 引言

盲信号分离算法的构造首先是选择合适的对比函数,然后采用某种优化方法来搜索对比函数的极大值,以确定最优的分离矩阵。现有的绝大多数自适应盲信号分离算法采用随机梯度也就是最陡下降法来学习分离矩阵,比较典型的是Bell等人<sup>[96]</sup>的信息最大化算法、Amari等人<sup>[3~9,69,72,89,99,153]</sup>的自然梯度算法、Cardoso等人<sup>[22]</sup>的EASI算法以及Karhunen等人<sup>[92~96,122]</sup>的非线性PCA算法等。这些算法均属最小均方(LMS)算法,因此存在学习速率参数的最优确定这一难题。

为解决LMS型的盲信号分离算法收敛速度和稳态性能之间的矛盾,上一章中我们定义了描述信号分离状态的相依性测度,并由此提出了分阶段学习的盲信号分离算法,收到了很好的效果。虽然分阶段学习算法对信号与/或信道的变化具有较高的稳定性,但总得来说,学习速率参数仍然是一个经验值。由信号相依性测度确定学习速率参数的函数需要小心选择。此外,信号相依性测度的自适应学习也必定增加了算法的运算量。

信号处理的滤波器理论<sup>[71]</sup>表明,与LMS型算法相比较,递归最小二乘(RLS)型算法因为每次更新中使用了更多的观测样本,从而收敛更快,跟踪能力更强。就盲信号分离问题而言,最早的RLS型算法是Pajunen等人<sup>[96,124]</sup>提出的,它是Yang<sup>[151,152]</sup>的投影逼近子空间跟踪(PAST: Project Approximation Subspace Tracking)算法的一个简单的非线性推广。

在上一章中我们看到,Amari等人<sup>[6,7]</sup>提出的自然梯度是一般梯度的推广,它考虑了对比函数的Riemannian参数结构,因此代表真正意义上的最陡上升方向。由于Pajunen等人<sup>[96,124]</sup>的RLS型盲信号分离算法是根据一般梯度推导出的,因此一个很自然的问题是:是否存在一种盲信号分离算法,它既是RLS型的算法,同时又基于自然梯度?答案是肯定的。

本章主要推导两种基于自然梯度的盲信号分离RLS算法,具体内容组织如下。第2节推导一种RLS型的预白化算法。第3节从非线性PCA的对比函数出发,利用分离矩阵的标准正交约束,推导一种基于自然梯度的RLS型盲信号分离算法(称作RLS-I算法),理论证明它能实现预白化的盲信号分离。第4节将RLS型的预白化算法与RLS-I算法相结合,经过合理的近似,推导出一种新的盲信号分离RLS算法(称作RLS-II算

法), 理论上它是等变化的, 并满足分离矩阵的非奇异性条件, 大量的计算机仿真则表明它无需对观测信号进行白化预处理, 就能实现信号的盲分离。最后第5节对本章进行了小结。

## §4.2 RLS型的预白化算法

分离矩阵  $\mathbf{B}$  的确定有两种方法。一种是直接法, 它直接调整  $\mathbf{B}$  以实现信号的盲分离; 另一种是间接法: 先求矩阵  $\mathbf{U}$ , 用它对观测数据进行预白化, 然后更新获得正交矩阵  $\mathbf{W}$ , 并作用于白化后的数据向量  $\mathbf{v}_i = \mathbf{U}\mathbf{x}_i$  以提取源信号。现有算法中, Bell等人<sup>[14]</sup>的信息最大化算法、Amari等人<sup>[3~9, 69, 72, 89, 99, 153]</sup>的自然梯度算法、Cardoso等人<sup>[22]</sup>的EASI算法以及上一章介绍的分阶段学习算法<sup>[178]</sup>都属于直接法, 而Karhunen等人的非线性PCA算法<sup>[92~95, 122]</sup>和RLS算法<sup>[96, 124]</sup>则属于间接法。本章主要研究间接法, 在此类方法中, 总的分离矩阵  $\mathbf{B} = \mathbf{W}\mathbf{U}$ 。

对于盲信号分离的观测信号向量  $\mathbf{x}_i$ , 白化预处理的目的是寻找一个线性变换  $\mathbf{U}$ , 使得

$$\mathbf{v}_i = \mathbf{U}\mathbf{x}_i \quad (4.1)$$

为白色信号向量。换句话说,  $\mathbf{v}_i$  的自相关函数矩阵

$$\mathbf{R}_v = E\{\mathbf{v}_i \mathbf{v}_i^T\} \quad (4.2)$$

等于单位矩阵  $\mathbf{I}$ 。

设  $\mathbf{x}_i$  的自相关函数矩阵  $\mathbf{R}_x = E\{\mathbf{x}_i \mathbf{x}_i^T\}$  的特征值分解  $\mathbf{R}_x = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ , 则  $\mathbf{U} = \mathbf{\Lambda}^{-0.5}\mathbf{Q}^T$  是白化预处理问题的一个解, 因为

$$E\{\mathbf{v}_i \mathbf{v}_i^T\} = \mathbf{U}E\{\mathbf{x}_i \mathbf{x}_i^T\}\mathbf{U}^T = \mathbf{\Lambda}^{-0.5}\mathbf{Q}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{Q}\mathbf{\Lambda}^{-0.5} = \mathbf{I}.$$

很遗憾, 这种方法需要作特征值分解, 因此是一种离线算法, 运算量大, 不适应实时应用的需要。

为了实现在线白化, Cardoso等人<sup>[22]</sup>提出了代价函数

$$J_1(\mathbf{U}) = \text{trace}(\mathbf{U}\mathbf{R}_x\mathbf{U}^T) - \ln|\det(\mathbf{U}\mathbf{R}_x\mathbf{U}^T)| - n \quad (4.3)$$

并证明  $J_1(\mathbf{U}) \geq 0$ , 当且仅当  $\mathbf{U}\mathbf{R}_x\mathbf{U}^T = \mathbf{I}$  时等号成立。利用相对梯度的概念, Cardoso等人<sup>[22]</sup>推导出下面的自适应串行白化算法:

$$\mathbf{U}_t = \mathbf{U}_{t-1} + \eta_t \cdot [\mathbf{I} - \mathbf{v}_t \mathbf{v}_t^T] \mathbf{U}_{t-1} \quad (4.4)$$

其中  $\eta_t > 0$  是学习速率参数,  $\mathbf{v}_t = \mathbf{U}_{t-1} \mathbf{x}_t$  表示白化后的输出信号向量。

算法(4.4)是一种随机梯度算法,属于LMS型学习算法的范畴,其学习速率  $\eta_t$  需要精心选择,以获得满意的收敛速度和稳态性能。若学习速率  $\eta_t$  偏小,则算法收敛慢;若  $\eta_t$  偏大,则算法收敛快,但稳态性能差,甚至可能发散。众所周知,RLS型的学习算法<sup>[62,71,133]</sup>因收敛快,跟踪能力强而得到了广泛的应用。为此,我们用指数加窗的相关函数矩阵

$$\hat{\mathbf{R}}_x = (1 - \lambda) \cdot \sum_{i=1}^t \lambda^{t-i} \mathbf{x}_i \mathbf{x}_i^T \quad (4.5)$$

逼近真实值  $\mathbf{R}_x$ , 其中  $0 < \lambda < 1$  是遗忘因子,且接近于1。不难看出,当  $t$  趋向于无穷大时,式(4.5)是  $\mathbf{R}_x$  的一个无偏估计,因为

$$\lim_{t \rightarrow \infty} E\{\hat{\mathbf{R}}_x\} = (1 - \lambda) \lim_{t \rightarrow \infty} \sum_{i=1}^t \lambda^{t-i} E\{\mathbf{x}_i \mathbf{x}_i^T\} = \mathbf{R}_x \lim_{t \rightarrow \infty} (1 - \lambda^t) = \mathbf{R}_x.$$

用  $\hat{\mathbf{R}}_x$  代替  $\mathbf{R}_x$ , 代价函数(4.3)变作

$$J(\mathbf{U}) = \text{trace}(\mathbf{U} \hat{\mathbf{R}}_x \mathbf{U}^T) - \ln |\det(\mathbf{U} \hat{\mathbf{R}}_x \mathbf{U}^T)| - n \quad (4.6)$$

计算其梯度,得

$$\nabla J(\mathbf{U}) = 2 \cdot \mathbf{U} \hat{\mathbf{R}}_x - 2 \cdot [\mathbf{U} \hat{\mathbf{R}}_x \mathbf{U}^T]^{-1} \mathbf{U} \hat{\mathbf{R}}_x \quad (4.7)$$

在上式两端同时右乘以矩阵  $\mathbf{U}^T$ , 并令  $\nabla J(\mathbf{U}) = \mathbf{O}$ , 则  $\mathbf{U} = [\hat{\mathbf{R}}_x \mathbf{U}^T]^{-1}$ 。现在,我们希望利用  $\hat{\mathbf{R}}_x$  和  $\mathbf{U}_{t-1}$  来更新  $\mathbf{U}_t$ , 于是有

$$\mathbf{U}_t = [\hat{\mathbf{R}}_x \mathbf{U}_{t-1}^T]^{-1} \quad (4.8)$$

利用式(4.5)得

$$\begin{aligned} \mathbf{U}_t &= \left[ (1 - \lambda) \sum_{i=1}^t \lambda^{t-i} \mathbf{x}_i \mathbf{x}_i^T \mathbf{U}_{t-1}^T \right]^{-1} \\ &= [\lambda \mathbf{U}_{t-1}^{-1} + (1 - \lambda) \mathbf{x}_t \mathbf{x}_t^T]^{-1} \end{aligned} \quad (4.9)$$

运用矩阵求逆引理(参阅附录B),可以推导出下面的RLS型自适应算法:

$$\mathbf{U}_t = \frac{1}{\lambda} \left[ \mathbf{U}_{t-1} - \frac{\mathbf{v}_t \mathbf{v}_t^T}{\beta + \mathbf{v}_t^T \mathbf{v}_t} \mathbf{U}_{t-1} \right] \quad (4.10)$$

式中  $\beta = \frac{\lambda}{1-\lambda}$ 。

式(4.10)是一种新的白化算法。这一算法的平衡点为  $E\{\mathbf{v}_t \mathbf{v}_t^T\} = \mathbf{I}$ ，证明如下。在(4.10)的两端同时右乘以矩阵  $\lambda \mathbf{U}_{t-1}^{-1}$ ，得

$$\lambda \mathbf{U}_t \mathbf{U}_{t-1}^{-1} = \mathbf{I} - \frac{\mathbf{v}_t \mathbf{v}_t^T}{\frac{\lambda}{1-\lambda} + \mathbf{v}_t^T \mathbf{v}_t}$$

当算法达到平衡点时，有  $\lim_{t \rightarrow \infty} \mathbf{U}_{t-1} = \lim_{t \rightarrow \infty} \mathbf{U}_t$ ，从而有  $\lim_{t \rightarrow \infty} \mathbf{U}_t \mathbf{U}_{t-1}^{-1} = \mathbf{I}$ 。将这一结果代入上式的极限中，并加以整理，可得到

$$\mathbf{v}_t \mathbf{v}_t^T = [\lambda + (1-\lambda) \mathbf{v}_t^T \mathbf{v}_t] \cdot \mathbf{I}$$

两边同取数学期望，并注意到遗忘因子  $\lambda$  小于1且接近于1，故有  $E\{\mathbf{v}_t \mathbf{v}_t^T\} = \mathbf{I}$ 。这表明，算法(4.10)的平衡点是  $E\{\mathbf{v}_t \mathbf{v}_t^T\} = \mathbf{I}$ 。这一性质保证算法收敛后的矩阵  $\mathbf{U}_t$ ，可使输出  $\mathbf{v}_t = \mathbf{U}_{t-1} \mathbf{x}_t$  为白色信号，这正是我们所希望的。

### §4.3 预白化的盲信号分离RLS算法

盲信号分离的间接求解法把总的分离矩阵  $\mathbf{B}$  分成  $\mathbf{U}$  和  $\mathbf{W}$  两个部分。上一节讨论了白化矩阵  $\mathbf{U}$  的自适应更新，本节研究  $\mathbf{W}$  的自适应学习算法。

#### 4.3.1 非线性PCA对比函数

为应付幅度不确定性，盲信号分离假定各个源信号分量具有单位功率，因此为使输出  $\mathbf{y}_t = \mathbf{B} \mathbf{x}_t = \mathbf{W} \mathbf{U} \mathbf{x}_t = \mathbf{W} \mathbf{v}_t$  是源信号向量的某个拷贝，要求  $\mathbf{y}_t$  是白色信号，即

$$E\{\mathbf{y}_t \mathbf{y}_t^T\} = \mathbf{W} E\{\mathbf{v}_t \mathbf{v}_t^T\} \mathbf{W}^T = \mathbf{W} \mathbf{W}^T = \mathbf{I}$$

从而  $\mathbf{W}$  是标准正交矩阵。

与分离矩阵  $\mathbf{B}$  相类似，正交矩阵  $\mathbf{W}$  的算法构造也是首先选择合适的对比函数，然后采用优化方法来搜索对比函数的极大值。在第二章中业已看到，确定正交矩阵  $\mathbf{W}$  也就是预白化盲信号分离的对比函数有很多种。为了推导出一种RLS型的盲信号分离算法，这里我们主要关心下面的非线性主分量分析(PCA)对比函数<sup>[89,93,96,122,124,167]</sup>：

$$J_2(\mathbf{W}) = -E \left\{ \left\| \mathbf{v}_i - \mathbf{W}^T \mathbf{g}(\mathbf{W} \mathbf{v}_i) \right\|^2 \right\} \quad (4.11)$$

式中 $\|\cdot\|$ 表示向量的2-范数,  $\mathbf{v}_i = \mathbf{U} \mathbf{x}_i$ 表示观测信号向量白化预处理后的信号向量, 而 $\mathbf{g}(\mathbf{y}) = [g_1(y_1), \dots, g_n(y_n)]^T$ 是一个非线性的列向量, 其中的非线性函数 $g_i(\cdot)$ 也就是盲信号分离的激励函数, 它需要根据算法的局部稳定性条件作适当选择。

因为 $\mathbf{W}$ 是正交矩阵, 所以利用向量的2-范数对正交变换的不变性, 有

$$\left\| \mathbf{v}_i - \mathbf{W}^T \mathbf{g}(\mathbf{W} \mathbf{v}_i) \right\| = \left\| \mathbf{W} \mathbf{v}_i - \mathbf{W} \mathbf{W}^T \mathbf{g}(\mathbf{W} \mathbf{v}_i) \right\| = \left\| \mathbf{y}_i - \mathbf{g}(\mathbf{y}_i) \right\|$$

进而式(4.11)等价于

$$J_{2a}(\mathbf{W}) = -\sum_{i=1}^n E \left\{ \left[ y_i - g_i(y_i) \right]^2 \right\} \quad (4.12)$$

由式(4.12)不难看出非线性PCA对比函数(4.11)与其它对比函数的联系。首先取激励函数为下面的非线性奇函数:

$$g_i(y_i) = \begin{cases} y_i^2 + y_i, & y_i \geq 0 \\ -y_i^2 + y_i, & y_i < 0 \end{cases} \quad (4.13)$$

并代入到(4.12)中, 得

$$J_{2b}(\mathbf{W}) = -\sum_{i=1}^n E \left\{ \left[ y_i - y_i \pm y_i^2 \right]^2 \right\} = -\sum_{i=1}^n E \{ y_i^4 \} \quad (4.14)$$

显然, 这就是分离亚高斯信号的高阶矩对比函数(2.16)。

其次, 若选择 $g_i(\cdot)$ 满足下面的关系式:

$$\ln |p_i(y_i)| = \alpha - \frac{1}{\beta} \cdot [y_i - g_i(y_i)]^2 \quad (4.15)$$

式中 $\alpha$ 是任意常数,  $\beta$ 是任意正常数, 而 $p_i(\cdot)$ 表示第 $i$ 个源信号的概率密度函数, 则对比函数(4.12)变作

$$J_{2c}(\mathbf{W}) = -\alpha\beta + \beta \sum_{i=1}^n E \left\{ \ln |p_i(y_i)| \right\} \quad (4.16)$$

因为 $\ln |\det(\mathbf{W})| = 0$ , 所以(4.16)与预白化盲信号分离的极大似然对比函数(2.10)相等价。事实上, 式(4.15)为基于非线性PCA对比函数(4.11)的盲信号分离算法选择最优激励函数提供了理论依据。

最后, 式(4.12)与盲均衡中的Bussgang算法<sup>[71,97,177]</sup>的对比函数非常相像。若 $p_i(\cdot)$



的一阶导数记为  $p'_i(\cdot)$ ，并取

$$g_i(y_i) = \frac{-E\{|y_i|^2\} p'_i(y_i)}{p_i(y_i)} \quad (4.17)$$

则 (4.12) 就是 Lambert 等人<sup>[97]</sup>提出的盲均衡对比函数。因为对于预白化的盲信号分离问题有  $E\{|y_i|^2\} = 1$ ，式 (4.17) 实际上就是盲信号分离的分值函数。

### 4.3.2 一般梯度的预白化盲分离算法

从对比函数 (4.11) 出发，Karhunen 和 Oja 等人<sup>[92,94,96,122]</sup>推导出了一大类称为非线性 PCA 子空间学习规则的盲信号分离算法：

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \eta_i \mathbf{g}(y_i) [\mathbf{v}_i^T - \mathbf{g}^T(y_i) \mathbf{W}_i] \quad (4.18)$$

式中  $\eta_i > 0$  是学习速率参数。

从对比函数 (4.12) 出发，Lambert 等人<sup>[97]</sup>推导出另一类盲信号分离算法：

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \eta_i \mathbf{v}_i [\mathbf{y}_i - \mathbf{g}(y_i)] \quad (4.19)$$

不难看出，算法 (4.18) ~ (4.19) 都可归类于 LMS 型的算法。如前所述，它们需要精心选择学习速率  $\eta_i$  以获得满意的收敛速度和稳态性能。

PCA 算法的非线性推广能够实现预白化的盲信号分离。受这个思想启发，Pajunen 等人<sup>[96,124]</sup>在 Yang<sup>[151,152]</sup>的 PAST 算法中引入非线性，从而提出了下面的 RLS 型的盲信号分离算法：

$$\begin{aligned} \mathbf{z}_i &= \mathbf{g}(y_i) = \mathbf{g}(\mathbf{W}_{i-1} \mathbf{v}_i) \\ \mathbf{h}_i &= \mathbf{P}_{i-1} \mathbf{z}_i \\ \mathbf{m}_i &= \mathbf{h}_i / (\lambda + \mathbf{z}_i^T \mathbf{h}_i) \\ \mathbf{P}_i &= \frac{1}{\lambda} \text{Tri}[\mathbf{P}_{i-1} - \mathbf{m}_i \mathbf{h}_i^T] \\ \mathbf{W}_i &= \mathbf{W}_{i-1} + \mathbf{m}_i [\mathbf{v}_i^T - \mathbf{z}_i^T \mathbf{W}_{i-1}] \end{aligned} \quad (4.20)$$

其中  $\text{Tri}[\cdot]$  表示这样的—个算子，它只计算其自变量的上三角部分的元素，然后将这些元素拷贝给下三角部分，从而使得  $\mathbf{P}_i$  是一个对称矩阵。

实现主子空间分析 (PSA: Principle Subspace Analysis) 的PAST算法具有运算量小, 收敛快等优点, 但它不能保证权矩阵的标准正交性。可以预见, 作为其非线性推广的预白化盲分离算法 (4.20) 也不可能做到这一点。另一方面, 无论是LMS型的算法 (4.18) 和 (4.19) 还是RLS型的算法 (4.20), 它们都基于一般梯度。众所周知, 自然梯度因为考虑了对比函数的参数结构, 从而代表了真正意义上的最陡上升方向, 因此比一般梯度更适合盲信号分离。那么, 对比函数 (4.11) 具有什么样的参数结构呢? 能否推导出一种利用这种参数结构的盲信号分离RLS算法? 答案是肯定的。

### 4.3.3 自然梯度的预白化盲分离新算法

非线性PCA对比函数 (4.11) 的优化实际上是一个约束问题, 因为矩阵  $\mathbf{W}$  是一个标准正交矩阵。定义在正交矩阵空间上的函数存在两种可能的流形结构。一种是对任意Hermitian矩阵  $\mathbf{M}$  都有  $J(\mathbf{W}) = J(\mathbf{M}\mathbf{W})$  的Grassmann流形, 另一种是  $J(\mathbf{W}) \neq J(\mathbf{M}\mathbf{W})$  的Stiefel流形。由于激励函数  $g_i(\cdot)$  是非线性函数, 显然对比函数 (4.11) 具有Stiefel流形结构<sup>[58,60]</sup>。下面, 我们推导利用这种Stiefel流形结构的预白化盲信号分离算法。

为推导RLS型的盲信号分离算法, 我们用指数加窗平均逼近 (4.11) 中的数学期望, 从而有下面的对比函数<sup>[93,124,167]</sup>:

$$J(\mathbf{W}_l) = -\sum_{i=1}^l \lambda^{l-i} \|\mathbf{v}_i - \mathbf{W}_l^T \mathbf{g}(\mathbf{W}_l \mathbf{v}_i)\|^2 \quad (4.21)$$

式中遗忘因子  $0 < \lambda < 1$ , 且接近于1。

用  $\mathbf{z}_i = \mathbf{g}(\mathbf{W}_{l-1} \mathbf{v}_i)$  来近似未知项  $\mathbf{g}(\mathbf{W}_l \mathbf{v}_i)$ , 则对比函数 (4.21) 对矩阵  $\mathbf{W}_l$  的一般梯度等于

$$\nabla J(\mathbf{W}_l) = -\sum_{i=1}^l \lambda^{l-i} \{-\mathbf{z}_i \mathbf{v}_i^T + \mathbf{z}_i \mathbf{z}_i^T \mathbf{W}_l\} \quad (4.22)$$

Edelman等人<sup>[60]</sup>推导了正交约束下的自然梯度, 特别地, 在Stiefel流形结构下对比函数的自然梯度和一般梯度存在以下关系<sup>[9,58,60]</sup>:

$$\tilde{\nabla} J(\mathbf{W}_l) = \mathbf{W}_l \mathbf{W}_l^T \cdot \nabla J(\mathbf{W}_l) - \mathbf{W}_l \cdot [\nabla J(\mathbf{W}_l)]^T \cdot \mathbf{W}_l \quad (4.23)$$

结合式 (4.22) 和 (4.23), 对比函数 (4.21) 对  $\mathbf{W}_l$  的自然梯度等于<sup>[167]</sup>

$$\tilde{\nabla} J(\mathbf{W}_l) = -\sum_{i=1}^l \lambda^{l-i} \{-\mathbf{W}_l \mathbf{W}_l^T \mathbf{z}_i \mathbf{v}_i^T + \mathbf{W}_l \mathbf{v}_i \mathbf{z}_i^T \mathbf{W}_l\} \quad (4.24)$$

基于 (4.24), 我们可以推导出两种盲信号分离算法, 一种是LMS型的算法, 另

一种是RLS型的算法。首先令(4.24)中的遗忘因子 $\lambda=0$ ，并重新将 $\mathbf{z}_t = \mathbf{g}(\mathbf{W}_{t-1}\mathbf{v}_t)$ 记作 $\mathbf{g}(\mathbf{y}_t) = \mathbf{g}(\mathbf{W}_t\mathbf{v}_t)$ ，则可以得到随机自然梯度

$$\tilde{\nabla}J(\mathbf{W}_t) = \mathbf{W}_t\mathbf{W}_t^T\mathbf{g}(\mathbf{y}_t)\mathbf{v}_t^T - \mathbf{y}_t\mathbf{g}^T(\mathbf{y}_t)\mathbf{W}_t \quad (4.25)$$

和对应的随机自然梯度学习算法

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \eta_t [\mathbf{W}_t\mathbf{W}_t^T\mathbf{g}(\mathbf{y}_t)\mathbf{v}_t^T - \mathbf{y}_t\mathbf{g}^T(\mathbf{y}_t)\mathbf{W}_t] \quad (4.26)$$

其中学习速率参数 $\eta_t > 0$ 。由于 $\mathbf{W}_t$ 是正交矩阵，即 $\mathbf{W}_t^T\mathbf{W}_t = \mathbf{W}_t\mathbf{W}_t^T = \mathbf{I}$ ，因此上面的算法可简化为

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \eta_t [\mathbf{g}(\mathbf{y}_t)\mathbf{v}_t^T - \mathbf{y}_t\mathbf{g}^T(\mathbf{y}_t)] \quad (4.27)$$

显然，它与Douglas<sup>[58]</sup>提出的预白化盲信号分离的自稳定算法具有相同的形式，因此算法(4.27)也是自稳定的。不过两者的推导过程完全不同。此外，Douglas<sup>[58]</sup>研究了形如(4.27)的算法的局部稳定性条件，它由下式给出：

$$\eta_t [E\{s_i g_i(s_i) - g'_i(s_i)\} + E\{s_j g_j(s_j) - g'_j(s_j)\}] > 0 \quad (4.28)$$

其中 $g'_i(\cdot)$ 表示函数 $g_i(\cdot)$ 的一阶导数， $s_i$ 是第 $i$ 个源信号分量 $s_{i,j}$ 的简写。

此外，对(4.27)中的 $\mathbf{g}(\mathbf{y}_t)\mathbf{v}_t^T$ 右乘以 $\mathbf{W}_t^T\mathbf{W}_t$ ，有

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \eta_t [\mathbf{g}(\mathbf{y}_t)\mathbf{y}_t^T - \mathbf{y}_t\mathbf{g}^T(\mathbf{y}_t)]\mathbf{W}_t \quad (4.29)$$

它是一种等变化的算法，除了学习速率参数相差一个符号外，与Cardoso等人<sup>[22]</sup>提出的正交矩阵的串行更新算法（EASI算法的一个组成部分）在形式上完全一样，当然两者的推导过程不同。

最后我们指出，算法(4.27)和(4.18)都可以用来搜索对比函数(4.11)的极大值，不同的是非线性PCA子空间算法(4.18)基于一般梯度，而(4.27)基于自然梯度。因为自然梯度比一般梯度更适合于盲信号分离，因此可以期望算法(4.27)表现要优于算法(4.18)。

尽管仿真表明<sup>[58,66,96]</sup>恰当选择学习速率 $\eta_t$ ，LMS型的算法(4.27)实现盲信号分离的结果令人满意，但本节的重点在于，从自然梯度(4.24)出发推导一种新颖的RLS型的盲信号分离算法。

在(4.24)中利用性质 $\mathbf{W}_t\mathbf{W}_t^T = \mathbf{I}$ ，并用 $\mathbf{y}_t = \mathbf{W}_{t-1}\mathbf{v}_t$ 来近似 $\mathbf{W}_t\mathbf{v}_t$ ，则根据 $\tilde{\nabla}J(\mathbf{W}_t) = \mathbf{0}$ 可以得到 $t$ 时刻的最优矩阵

$$\mathbf{W}_{\text{opt},l} = \left[ \sum_{i=1}^l \lambda^{l-i} \mathbf{y}_i \mathbf{z}_i^T \right]^{-1} \cdot \left[ \sum_{i=1}^l \lambda^{l-i} \mathbf{z}_i \mathbf{v}_i^T \right] \quad (4.30)$$

记

$$\mathbf{P}_l = \left[ \sum_{i=1}^l \lambda^{l-i} \mathbf{y}_i \mathbf{z}_i^T \right]^{-1} \quad (4.31)$$

$$\mathcal{H}_l = \sum_{i=1}^l \lambda^{l-i} \mathbf{z}_i \mathbf{v}_i^T \quad (4.32)$$

则  $\mathcal{H}_l = \lambda \mathcal{H}_{l-1} + \mathbf{z}_l \mathbf{v}_l^T$ ,  $\mathbf{P}_l^{-1} = \lambda \mathbf{P}_{l-1}^{-1} + \mathbf{y}_l \mathbf{z}_l^T$ , 因为  $\mathbf{P}_l$  可运用矩阵求逆引理<sup>[176]</sup>来更新, 从而最优矩阵(4.30)可用递归方法计算, 进而得到下面的RLS型的盲信号分离算法(RLS-I算法):

$$\begin{aligned} \mathbf{y}_l &= \mathbf{W}_{l-1} \mathbf{v}_l \\ \mathbf{z}_l &= \mathbf{g}(\mathbf{y}_l) \\ \mathbf{Q}_l &= \mathbf{P}_{l-1} / (\lambda + \mathbf{z}_l^T \mathbf{P}_{l-1} \mathbf{y}_l) \\ \mathbf{P}_l &= \frac{1}{\lambda} [\mathbf{P}_{l-1} - \mathbf{Q}_l \mathbf{y}_l \mathbf{z}_l^T \mathbf{P}_{l-1}] \\ \mathbf{W}_l &= \mathbf{W}_{l-1} + [\mathbf{P}_l \mathbf{z}_l \mathbf{v}_l^T - \mathbf{Q}_l \mathbf{y}_l \mathbf{z}_l^T \mathbf{W}_{l-1}] \end{aligned} \quad (4.33)$$

比较算法(4.27)和(4.33)不难看出, 两者在形式上有相似之处, 都基于自然梯度, 也都可以实现对比函数(4.11)和(4.21)的极大化。但区别在于前者仅使用当前时刻的观测样本参与迭代更新, 从而属于LMS型的算法, 而后者不仅使用当前时刻的样本, 还通过  $\mathbf{P}_l$  和  $\mathbf{Q}_l$  使用了过去时刻的样本, 因此属于RLS型的算法。可以预见, 算法(4.33)比(4.27)收敛要快得多。

我们指出, 算法(4.20)和(4.33)都属于RLS型的盲信号分离算法, 但不同的是, 所提出的新算法(4.33)基于自然梯度(4.24), 而Pajunen等人的算法(4.20)则基于一般梯度(4.22)。因为自然梯度利用了对比函数的Stiefel流形结构信息, 因此可以预见, RLS型的新算法(4.33)的性能要强于RLS型的算法(4.20)。

关于上述几种算法的运算量, 除了共同的  $n$  次非线性运算外, 每次迭代中, LMS型的算法(4.18)和(4.19)需要  $3n^2 + n$  次乘法运算; (4.27)需要  $5n^2 + n$  次; (4.29)需要  $4n^2 + n$  次, 而RLS型的算法(4.20)需要  $5n^2 + 3n$  次; (4.33)需要  $9n^2 + 2n$  次。不难看出, 尽管算法(4.33)的运算最大, 但相对而言, 仍然是可以接受的。

最后, 对RLS-I算法(4.33), 我们给出下面的定理。

**定理4.1** 基于自然梯度的盲信号分离RLS算法(4.33)收敛的平衡点满足  $E\{\mathbf{W}_t \mathbf{W}_t^T\} = \mathbf{I}$  和  $E\{\mathbf{z}_t \mathbf{y}_t^T\} = E\{\mathbf{y}_t \mathbf{z}_t^T\}$ , 且收敛过程与输入数据的特征值分布无关。

为方便证明, 我们记

$$\Phi_t = \mathbf{Q}_t \mathbf{y}_t \mathbf{z}_t^T \quad (4.34)$$

$$\Pi_t = \mathbf{P}_t \mathbf{z}_t \mathbf{v}_t^T \quad (4.35)$$

$$\Theta_t = \mathbf{W}_t \mathbf{W}_t^T - \mathbf{I} \quad (4.36)$$

根据文献[62, 133], 当  $t \rightarrow \infty$  时, 式(4.31)中的  $\mathbf{P}_t$  可视为准静态矩阵  $\mathbf{P}_{\text{sta}}$ 。因为  $\mathbf{P}_t^{-1} = \lambda \mathbf{P}_{t-1}^{-1} + \mathbf{y}_t \mathbf{z}_t^T$ , 故对其两边求数学期望并作极限运算, 有

$$\mathbf{P}_{\text{sta}} = (1 - \lambda) \cdot [E\{\mathbf{y}_t \mathbf{z}_t^T\}]^{-1} \quad (4.37)$$

进而根据  $\mathbf{Q}_t$  的定义, 有

$$\lim_{t \rightarrow \infty} \mathbf{Q}_t = \lim_{t \rightarrow \infty} \frac{\mathbf{P}_{\text{sta}}}{\lambda + \text{trace}(\mathbf{P}_{\text{sta}} \mathbf{y}_t \mathbf{z}_t^T)} = \frac{\mathbf{P}_{\text{sta}}}{\lambda + (1 - \lambda)n} \approx \mathbf{P}_{\text{sta}} \quad (4.38)$$

最后面的约等式成立是因为遗忘因子  $\lambda$  小于1且非常接近于1。

利用性质(4.37), 并注意到  $\Phi_t = \mathbf{I} - \lambda \mathbf{P}_t \mathbf{P}_{t-1}^{-1}$  和  $\mathbf{W}_{\text{opt}} = [E\{\mathbf{y}_t \mathbf{z}_t^T\}]^{-1} \cdot E\{\mathbf{z}_t \mathbf{v}_t^T\}$ , 不难证明

$$\lim_{t \rightarrow \infty} E\{\Pi_t\} = \mathbf{P}_{\text{sta}} E\{\mathbf{z}_t \mathbf{v}_t^T\} = (1 - \lambda) \mathbf{W}_{\text{opt}} \quad (4.39)$$

$$\lim_{t \rightarrow \infty} E\{\Phi_t\} = (1 - \lambda) \mathbf{I} \quad (4.40)$$

在算法(4.33)的两端同乘以  $\mathbf{W}_t^T$  并减去单位矩阵  $\mathbf{I}$ , 整理可得

$$\begin{aligned} \Theta_t = & \Theta_{t-1} - \Theta_{t-1} \Phi_t^T - \Phi_t \Theta_{t-1} - \Phi_t - \Phi_t^T + \mathbf{W}_{t-1} \Pi_t^T + \Pi_t \mathbf{W}_{t-1}^T \\ & - \Phi_t \mathbf{W}_{t-1} \Pi_t^T - \Pi_t \mathbf{W}_{t-1}^T \Phi_t^T + \Phi_t [\Theta_{t-1} + \mathbf{I}] \Phi_t^T + \Pi_t \Pi_t^T \end{aligned} \quad (4.41)$$

对上式两端求数学期望, 运用性质(4.39)~(4.40), 并忽略  $1 - \lambda$  的二次项(上式后四项), 有

$$E\{\Theta_t\} = (2\lambda - 1) E\{\Theta_{t-1}\} - 2(1 - \lambda) \mathbf{I} + (1 - \lambda) E\{\mathbf{W}_{t-1} \mathbf{W}_{\text{opt}}^T + \mathbf{W}_{\text{opt}} \mathbf{W}_{t-1}^T\} \quad (4.42)$$

因为考虑的是局部收敛性, 所以上式中的未知项  $E\{\mathbf{W}_{t-1} \mathbf{W}_{\text{opt}}^T\}$  和  $E\{\mathbf{W}_{\text{opt}} \mathbf{W}_{t-1}^T\}$  可用



$E\{\mathbf{W}_{\text{opt}}\mathbf{W}_{\text{opt}}^T\}=\mathbf{I}$  来近似, 从而式 (4.42) 可写作

$$E\{\Theta_t\} \approx (2\lambda - 1)E\{\Theta_{t-1}\} \quad (4.43)$$

由于系数  $2\lambda - 1 < 1$ , 只取决于遗忘因子  $\lambda$ , 因此  $\Theta_t$  将收敛为零矩阵。进而由 (4.36) 知  $E\{\mathbf{W}_t\mathbf{W}_t^T\}=\mathbf{I}$ , 即算法 (4.33) 中的权矩阵  $\mathbf{W}_t$  将收敛至正交矩阵, 且收敛过程与输入数据的特征值分布无关。

此外, 算法 (4.33) 的平衡点对应于

$$E\{\mathbf{P}_t\mathbf{z}_t\mathbf{v}_t^T - \mathbf{Q}_t\mathbf{y}_t\mathbf{z}_t^T\mathbf{W}_{t-1}\} = \mathbf{O} \quad (4.44)$$

在上式两端同时右乘以矩阵  $\mathbf{W}_{t-1}^T$  并作极限运算, 得

$$\mathbf{P}_{\text{sta}}E\{\mathbf{z}_t\mathbf{y}_t^T\} = E\{\mathbf{Q}_t\mathbf{y}_t\mathbf{z}_t^T\mathbf{W}_{t-1}\mathbf{W}_{t-1}^T\} \approx \mathbf{P}_{\text{sta}}E\{\mathbf{y}_t\mathbf{z}_t^T\} \quad (4.45)$$

显然, 除  $E\{\mathbf{W}_t\mathbf{W}_t^T\}=\mathbf{I}$  外, 算法的平衡点还满足  $E\{\mathbf{z}_t\mathbf{y}_t^T\} = E\{\mathbf{y}_t\mathbf{z}_t^T\}$ 。证明完毕。 ■

我们指出, 若非线性函数  $\mathbf{z}_t = \mathbf{g}(\mathbf{y}_t)$  选择恰当, 则  $E\{\mathbf{z}_t\mathbf{y}_t^T\} = E\{\mathbf{y}_t\mathbf{z}_t^T\}$  意味着输出信号  $\mathbf{y}_t$  在二阶和某个高阶之前是彼此统计独立的, 从而达到信号分离的目的。

#### 4.3.4 性能仿真

为了验证本节提出的预白化的盲信号分离RLS算法 (4.33) 的有效性, 我们对以下信号进行了仿真实验<sup>[3,153,156]</sup>:

- S1: 符号信号  $\text{sign}(\cos(2\pi 155t))$ ;
- S2: 高频正弦信号  $\sin(2\pi 800t)$ ;
- S3: 低频正弦信号  $\sin(2\pi 90t)$ ;
- S4: 相位调制信号  $\sin(2\pi 300t + 6\cos(2\pi 60t))$ ;
- S5: 在  $[-1,1]$  均匀分布的随机噪声信号。

以速率10kHz对传感器接收信号采样以产生离散观测信号向量  $\mathbf{x}_t$ 。因为算法要求观测数据必须经过白化预处理, 为简单起见, 假定混合矩阵是  $\mathbf{A}$  正交矩阵, 其元素在每次运行中都是随机生成的 (先随机产生一个非奇异矩阵, 再作正交分解即可获得需要的正交矩阵  $\mathbf{A}$ )。

作为比较, 我们同时运行自稳定算法 (4.27)<sup>[58]</sup>、串行更新算法 (4.29)<sup>[22]</sup>、Pajunen 等人的基于一般梯度的RLS型算法 (4.20)<sup>[124]</sup> 和本节提出的基于自然梯度的RLS型算法 (4.33)。算法的性能用串音误差<sup>[3,78,153,156]</sup>

$$E_{\text{cl}} = \sum_{k=1}^n \left( \sum_{l=1}^n \frac{|c_{kl}|}{\max_p |c_{kp}|} - 1 \right) + \sum_{l=1}^n \left( \sum_{k=1}^n \frac{|c_{kl}|}{\max_p |c_{pl}|} - 1 \right) \quad (4.46)$$

来衡量，式中  $\mathbf{C}_l = [c_{kl}] = \mathbf{W}_l \mathbf{A}$ 。此外，我们定义性能指数

$$E_{\text{dev}} = \|\mathbf{W}_l \mathbf{W}_l^T - \mathbf{\Lambda}\|_F \quad (4.47)$$

来衡量  $\mathbf{W}_l$  偏离正交矩阵的程度，其中  $\|\cdot\|_F$  表示Frobenius矩阵范数，而  $\mathbf{\Lambda} = \text{diag}(\mathbf{W}_l \mathbf{W}_l^T)$  表示由  $\mathbf{W}_l \mathbf{W}_l^T$  的对角元素组成的对角矩阵。之所以取对角矩阵  $\mathbf{\Lambda}$  而不是单位矩阵  $\mathbf{I}$  是因为盲信号分离问题存在固有的幅度不确定性（仿真表明，上述四种算法中，除Pajunen等人<sup>[124]</sup>的RLS型算法（4.20）中的  $\mathbf{W}_l \mathbf{W}_l^T$  收敛至对角矩阵外，其它三种算法都收敛为单位矩阵）。

四种算法使用相同的初始矩阵  $\mathbf{W}_0 = \mathbf{I}$ 。为兼顾收敛速度和稳态性能，LMS型的算法（4.27）和（4.29）的学习速率分别取  $200 \times T$  和  $-60 \times T$ ，其中  $T = 10^{-4}$  为采样周期。RLS型的算法（4.20）和（4.33）使用相同的遗忘因子  $\lambda = 0.983$ 。此外，根据局部稳定性条件（4.28），除串行更新算法（4.29）取激励函数  $g_l(y_l) = y_l^3$  外，其它三种算法都取  $g_l(y_l) = \tanh(y_l)$ 。

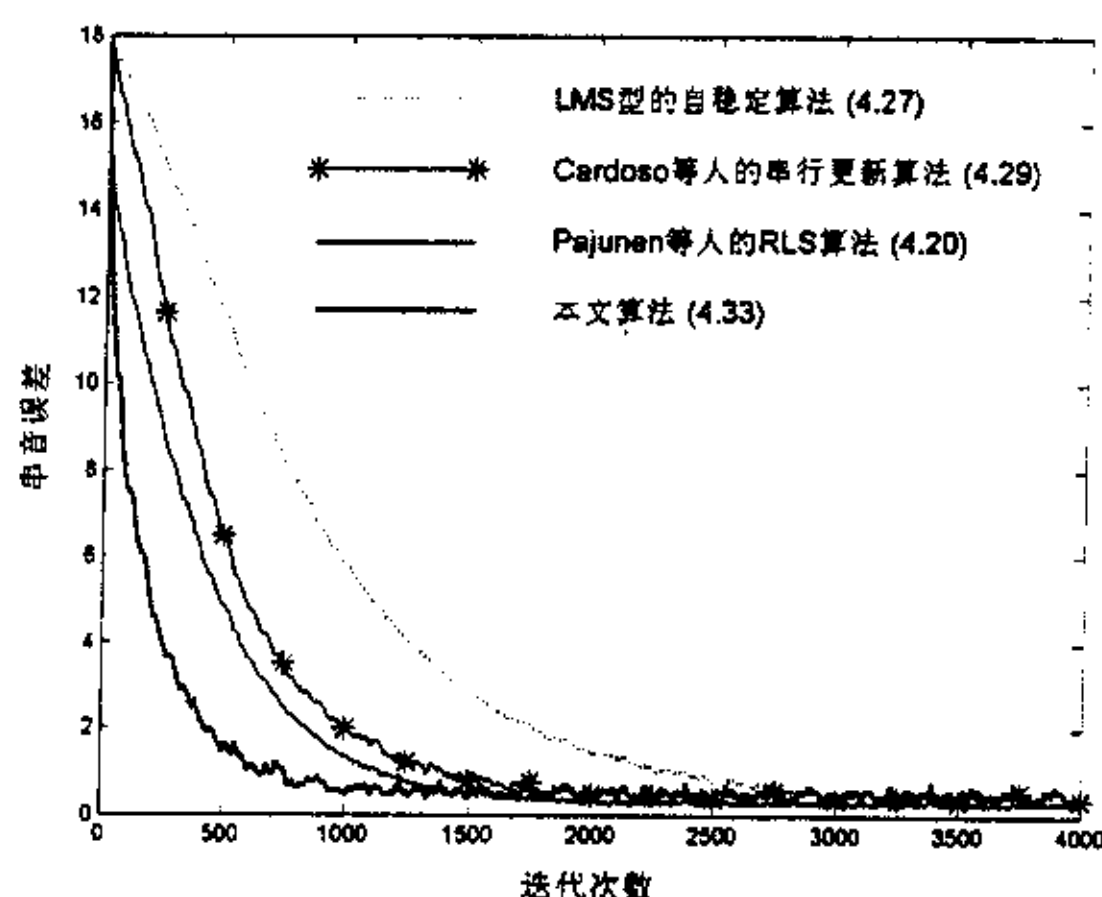


图4.1 串音误差  $E_{\text{cl}}$  的均值

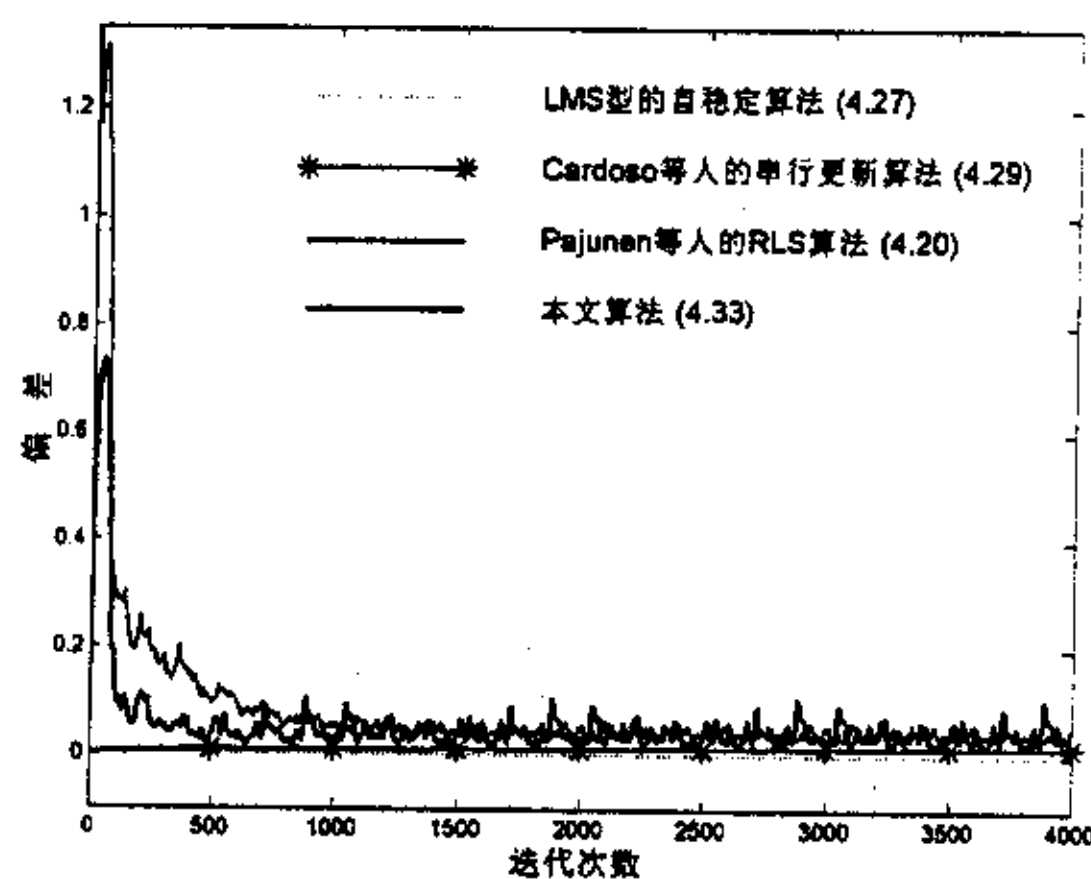


图4.2 偏差  $E_{\text{dev}}$  的均值

独立运行2000次试验（每次运行中的混合矩阵  $\mathbf{A}$  和噪声源信号  $\mathbf{S5}$  都是随机生成的），平均的串音误差  $E_{\text{cl}}$  和分离矩阵偏离正交矩阵的性能指数  $E_{\text{dev}}$  分别如图4.1和图4.2所示。不难看出：

1. RLS型的盲信号分离算法（4.20）和（4.33）收敛比LMS型的盲信号分离算法

(4.27) 和 (4.29) 快;

2. 与基于一般梯度的盲信号分离RLS算法 (4.20) 相比, 本节提出的基于自然梯度的RLS型算法 (4.33) 能很快实现分离矩阵的标准正交化, 且收敛速度也明显快, 但稳态性能稍差。

除了上述两点外, 需要补充说明的是, 四种算法中的RLS型算法 (4.20) 和 (4.33) 以及LMS型的自稳定算法 (4.27) 能保证分离矩阵稳定在正交矩阵附近, 但串行更新算法 (4.29) 的分离矩阵则逐步偏离正交矩阵, 参见图4.2, 因此最终不可避免地 (大概20000次迭代后) 发散。这一现象最早由EASI算法的缔造者Cardoso和Laheld<sup>[22]</sup>在计算机仿真时发现, 并提出了一种自稳定的EASI算法来解决这一问题。后来, Douglas<sup>[58]</sup>从理论上加以分析, 提出了另一种解决方案, 也就是自稳定的盲分离算法 (4.27)。

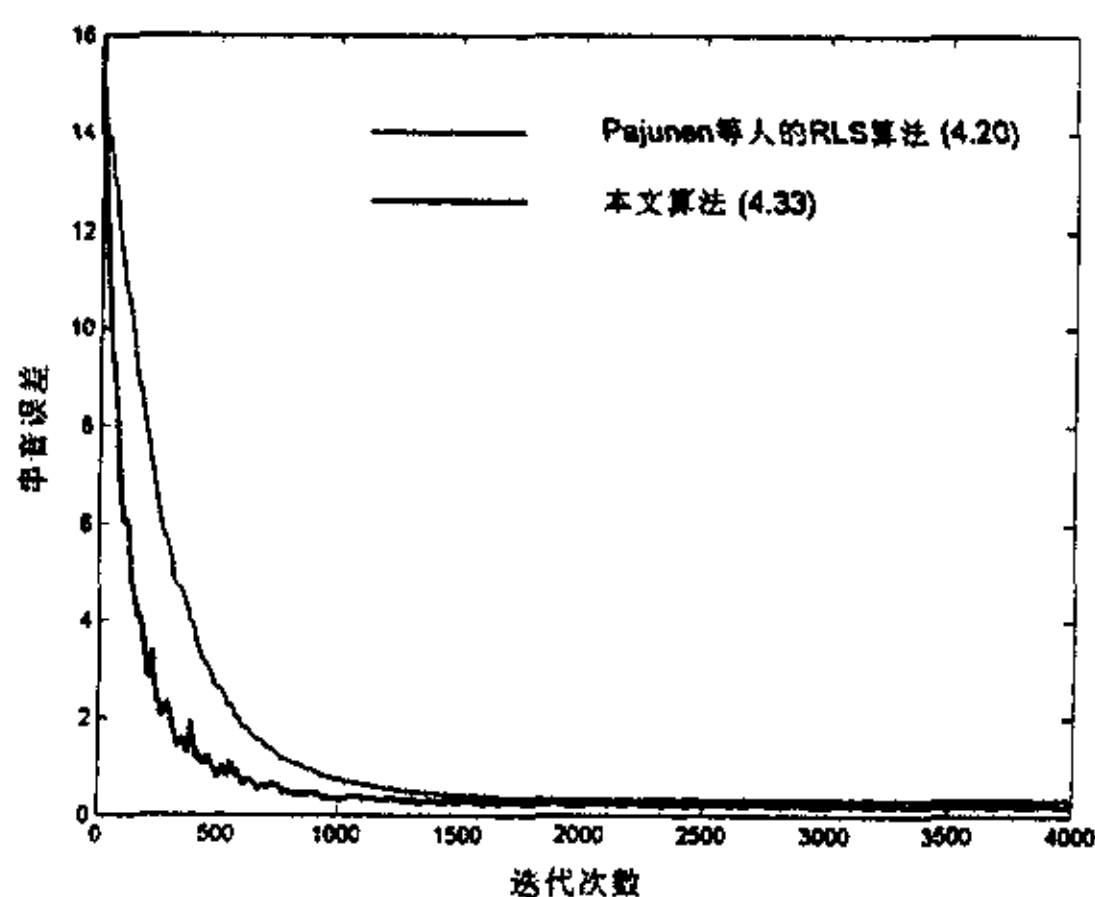


图4.3 串音误差  $E_d$  的均值, 时变遗忘因子

与LMS型算法使用时变学习速率参数一样, RLS型的算法也可以使用时变遗忘因子来加快算法的收敛, 改进算法的稳态性能。仿真中取  $\lambda$  从初始值0.96分别经800次迭代 (本文算法 (4.33)) 和1300次迭代 (Pajunen等人的RLS型算法 (4.20)) 线性增加到0.996并稳定在0.996, 次数不同是因为两种算法的收敛速度不一样所致。同样, 作2000次独立运行, 平均串音误差如图4.3所示。可以看到两者的收敛速度和稳态性能都得到了明显改善, 而基于自然梯度的本文算法 (4.33) 表现仍然要强于基于Pajunen等人的基于一般梯度的RLS型算法 (4.20)。这与理论分析的结果是一致的。

## §4.4 一种新的盲信号分离RLS算法

### 4.4.1 算法推导

在本章的第2节我们提出了一种RLS型的预白化算法(4.10)，在第3节推导了一种基于自然梯度的预白化的盲信号分离RLS算法(4.33)，即RLS-I算法。如果观测信号未作白化预处理，则盲信号分离需要级联运算(4.10)和(4.33)。事实上，两种RLS型的算法可以合而为一，从而推导出一种新的盲信号分离RLS算法(RLS-II算法)。

将算法(4.10)和(4.33)的两端相乘，经简单整理有

$$\mathbf{B}_t = \frac{1}{\lambda} [\mathbf{I} - \mathbf{Q}_t \mathbf{y}_t \mathbf{z}_t^T] \{ \mathbf{B}_{t-1} + \tilde{\alpha}_t \cdot [\mathbf{P}_{t-1} \mathbf{z}_t - (1-\lambda) \mathbf{y}_t] \mathbf{v}_t^T \mathbf{U}_{t-1} \} \quad (4.48)$$

其中系数  $\tilde{\alpha}_t = [\lambda + (1-\lambda) \mathbf{v}_t^T \mathbf{v}_t]^{-1}$  是一个标量。

根据定理4.1，当迭代次数  $t$  足够多时，可以实现  $\mathbf{W}_{t-1}^T \mathbf{W}_{t-1} \approx \mathbf{I}$ 。因此，对于足够大的  $t$ ，有

$$\mathbf{v}_t^T \mathbf{U}_{t-1} \approx \mathbf{v}_t^T \cdot \mathbf{W}_{t-1}^T \mathbf{W}_{t-1} \cdot \mathbf{U}_{t-1} = \mathbf{y}_t^T \mathbf{B}_{t-1} \quad (4.49)$$

$$\mathbf{v}_t^T \mathbf{v}_t \approx \mathbf{v}_t^T \cdot \mathbf{W}_{t-1}^T \mathbf{W}_{t-1} \cdot \mathbf{v}_t = \mathbf{y}_t^T \mathbf{y}_t \quad (4.50)$$

将(4.49)~(4.50)代入到式(4.48)中，我们得到下面的算法(RLS-II算法)：

$$\mathbf{y}_t = \mathbf{B}_{t-1} \mathbf{x}_t$$

$$\mathbf{z}_t = \mathbf{g}(\mathbf{y}_t)$$

$$\alpha_t = [\lambda + (1-\lambda) \mathbf{y}_t^T \mathbf{y}_t]^{-1}$$

$$\mathbf{Q}_t = \mathbf{P}_{t-1} / (\lambda + \mathbf{z}_t^T \mathbf{P}_{t-1} \mathbf{y}_t)$$

$$\mathbf{P}_t = \frac{1}{\lambda} [\mathbf{P}_{t-1} - \mathbf{Q}_t \mathbf{y}_t \mathbf{z}_t^T \mathbf{P}_{t-1}]$$

$$\mathbf{B}_t = \frac{1}{\lambda} [\mathbf{I} - \mathbf{Q}_t \mathbf{y}_t \mathbf{z}_t^T] \{ \mathbf{B}_{t-1} + \alpha_t \cdot [\mathbf{P}_{t-1} \mathbf{z}_t - (1-\lambda) \mathbf{y}_t] \mathbf{y}_t^T \mathbf{B}_{t-1} \} \quad (4.51)$$

算法(4.51)是一种新的RLS型的盲信号分离算法，它与Pajunen等人<sup>[96,124]</sup>的RLS型算法(4.20)以及本章第3节提出的RLS-I算法(4.33)的区别在于：

- Pajunen 等人的 RLS 型算法(4.20)基于一般梯度，而新算法(4.51)基于自

然梯度, 由于自然梯度比一般梯度更适合盲信号分离<sup>[3,7,153]</sup>, 所以可以期望新算法 (4.51) 工作更优越;

- RLS 型的盲信号分离算法 (4.20) 和 (4.33) 必须先对观测信号进行白化预处理才能工作, 而新算法 (4.51) 不需要, 因为它隐含执行了 RLS 型的白化算法 (4.10)。

对于RLS-II算法 (4.51), 我们有下面的定理。

**定理 4.2** RLS型的盲信号分离算法 (4.51) 是等变化的, 且满足分离矩阵的非奇异性条件。

证明: 在算法 (4.51) 的两端同时右乘以混合矩阵  $\mathbf{A}$ , 并记矩阵函数

$$\mathbf{F}(\mathbf{y}_t) = \frac{1}{1-\lambda} \left\{ [\mathbf{I} - \mathbf{Q}_t \mathbf{y}_t \mathbf{z}_t^T] [\mathbf{I} + \alpha_t (\mathbf{P}_{t-1} \mathbf{z}_t - (1-\lambda) \mathbf{y}_t) \mathbf{y}_t^T] - \lambda \mathbf{I} \right\} \quad (4.52)$$

则经过整理可得

$$\mathbf{C}_t = \mathbf{C}_{t-1} + \mu \mathbf{F}(\mathbf{y}_t) \mathbf{C}_{t-1} \quad (4.53)$$

其中参数  $\mu = \lambda^{-1} - 1$ , 而  $\mathbf{C}_t = \mathbf{B}_t \mathbf{A}$  代表混合—分离合成系统。由构造 (4.52) 知,  $\mathbf{F}(\mathbf{y}_t) = \mathbf{F}(\mathbf{C}_{t-1} \mathbf{s}_t)$  取决于遗忘因子  $\lambda$ , 合成系统  $\mathbf{C}_{t-1}$  和源信号  $\mathbf{s}_t$ , 而与单独的混合矩阵  $\mathbf{A}$  以及分离矩阵  $\mathbf{B}_{t-1}$  无关。因此,  $\mathbf{C}_t$  的更新不依赖于混合矩阵  $\mathbf{A}$ , 由文献[22]或者式 (3.33) 知, 算法 (4.51) 是等变化的。

另一方面, 算法 (4.51) 的连续形式的更新方程是  $d\mathbf{B}/dt = \mu \mathbf{F}(\mathbf{y}) \mathbf{B}$ , 类似于文献[153,156]的证明, 有

$$\frac{d \det(\mathbf{B})}{dt} = \text{trace} \left[ \frac{\partial \det(\mathbf{B})}{\partial \mathbf{B}^T} \cdot \frac{d\mathbf{B}}{dt} \right] = \det(\mathbf{B}) \cdot \text{trace} [\mu \mathbf{F}(\mathbf{y})]$$

从而

$$\det(\mathbf{B}_t) = \det(\mathbf{B}_0) \cdot \exp \left\{ \mu \int_0^t \text{trace} [\mathbf{F}(\mathbf{y}(\tau))] d\tau \right\} \quad (4.54)$$

因此, 若初始值  $\mathbf{B}_0$  非奇异, 则  $\mathbf{B}_t$  必定非奇异, 故算法 (4.51) 满足分离矩阵的非奇异性条件。证明完毕。 ■

我们指出, 除了RLS型的算法 (4.10) 和 (4.33) 的组合外, LMS型的算法 (4.4) 和 (4.27) 也可以结合在一起。若忽略学习速率  $\eta_t$  的高阶无穷小项, 继续用  $\mathbf{y}_t^T \mathbf{B}_{t-1}$  来近似  $\mathbf{v}_t^T \mathbf{U}_{t-1}$ , 即可得到



$$\mathbf{B}_t = \mathbf{B}_{t-1} + \eta_t [\mathbf{I} - \mathbf{y}_t \mathbf{y}_t^T + \mathbf{g}(\mathbf{y}_t) \mathbf{y}_t^T - \mathbf{y}_t \mathbf{g}^T(\mathbf{y}_t)] \mathbf{B}_{t-1} \quad (4.55)$$

显然, 除了非线性的激励函数相差一个符号外, 算法(4.55)与著名的EASI算法<sup>[22]</sup>具有相同的形式。关于非线性PCA子空间学习算法与EASI算法的联系, Karhunen等人文献[96]中作了比较详细的讨论, 这里我们借助于自然梯度更加明确了这种联系。

#### 4.4.2 性能仿真

为了验证本节提出的盲信号分离RLS算法(4.51)的有效性, 我们仍然对以下信号进行仿真实验<sup>[3,153,156]</sup>:

- S1: 符号信号  $\text{sign}(\cos(2\pi 155t))$ ;
- S2: 高频正弦信号  $\sin(2\pi 800t)$ ;
- S3: 低频正弦信号  $\sin(2\pi 90t)$ ;
- S4: 相位调制信号  $\sin(2\pi 300t + 6\cos(2\pi 60t))$ ;
- S5: 在 $[-1,1]$ 均匀分布的随机噪声信号。

仿真中, 混合矩阵 $\mathbf{A}$ 的各元素为在 $[-1,1]$ 区间均匀分布的随机数。以速率10kHz对传感器接收信号进行采样以产生离散观测信号 $\mathbf{x}_t$ 。算法的性能用串音误差<sup>[3,78,153,156]</sup>

$$E_{ci} = \sum_{p=1}^N \left( \sum_{q=1}^N \frac{|c_{pq}|}{\max_l |c_{pl}|} - 1 \right) + \sum_{q=1}^N \left( \sum_{p=1}^N \frac{|c_{pq}|}{\max_l |c_{lq}|} - 1 \right) \quad (4.56)$$

来衡量, 与(4.46)不同, 这里 $\mathbf{C}_t = [\mathbf{c}_{pq}] = \mathbf{B}_t \mathbf{A}$ 。

在算法(4.51)中, 遗忘因子 $\lambda = 0.993$ , 初始矩阵 $\mathbf{P}_0 = \mathbf{I}$ 。作为比较, 我们同时运行Pajunen等人<sup>[124]</sup>的RLS型算法(4.20)和Cardoso等人<sup>[22]</sup>的LMS型算法(4.55)。因为算法(4.20)只能实现预白化的盲信号分离, 所以用本章第2节提出的白化算法(4.10)对观测数据进行白化预处理, 遗忘因子 $\lambda = 0.993$ 。为获得满意的收敛速度和稳态性能, 算法(4.55)取学习速率 $\eta = 60 \times T$ , 其中 $T = 10^{-4}$ 为采样周期。三种算法取相同的初始矩阵 $\mathbf{B}_0 = \mathbf{I}$ , 并根据稳定性条件(4.28)取相同的激励函数 $g_t(\mathbf{y}_t) = \tanh(\mathbf{y}_t)$ 。

独立运行2000次, 图4.4给出了三种算法的平均串音误差曲线图(如果同时运行白化算法(4.10)和预白化的盲分离算法(4.33), 获得的收敛曲线与本节提出的RLS型算法(4.51)相重合, 这说明(4.51)隐含执行了白化算法(4.10), 且算法推导过程所作的近似(4.49)~(4.50)是合理的)。不难看出: RLS-II算法(4.51)不需要对观测数据进行白化预处理就能够实现信号的盲分离, 而且收敛速度明显比LMS型的算法(4.55)快。由于自然梯度比一般梯度更适合盲信号分离<sup>[3,7,153]</sup>, 基于自然梯度的

新算法(4.51)比基于一般梯度的RLS算法(4.20)收敛快。三者的稳态性能相当。

为提高算法的收敛速度,改进算法的稳态性能,LMS型盲信号分离算法可以使用时变学习速率参数,典型的是时间递减学习速率<sup>[156]</sup>:

$$\eta_t = \begin{cases} \eta_0, & t \leq T_0 \\ \eta_0 e^{-T_d(t-T_0)}, & t > T_0 \end{cases} \quad (4.57)$$

式中 $\eta_0$ ,  $T_0$ 及 $T_d$ 是适当选择的常数。仿真试验取 $\eta_0 = 150T$ ,  $T_0 = 600$ 和 $T_d = 16T$ 。

与此相类似,RLS型算法中的遗忘因子 $\lambda$ 也可以是时变的。仿真中取 $\lambda$ 从初始值0.985经950次(本节提出的算法(4.51))和1650次(Pajunen等人的RLS算法<sup>[124]</sup>)迭代线性递增到0.9995并稳定在0.9995,次数不同是因为两种RLS算法收敛速度不一样所致。同样,作2000次独立运行,平均性能如图4.5所示。可以看到,三者的收敛速度和稳态性能都得到了明显改善,而算法(4.51)表现仍然是最好的。

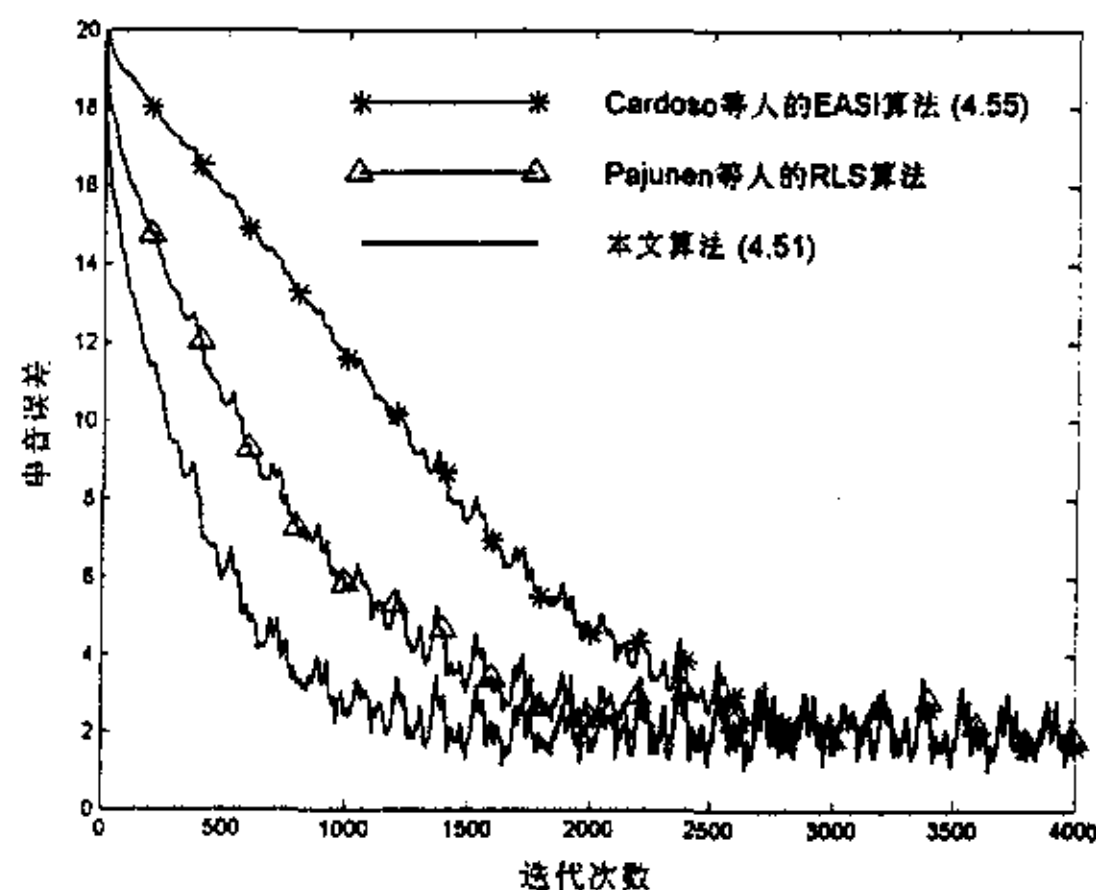


图4.4 2000次独立运行串音误差均值图,  
固定遗忘因子和固定学习速率

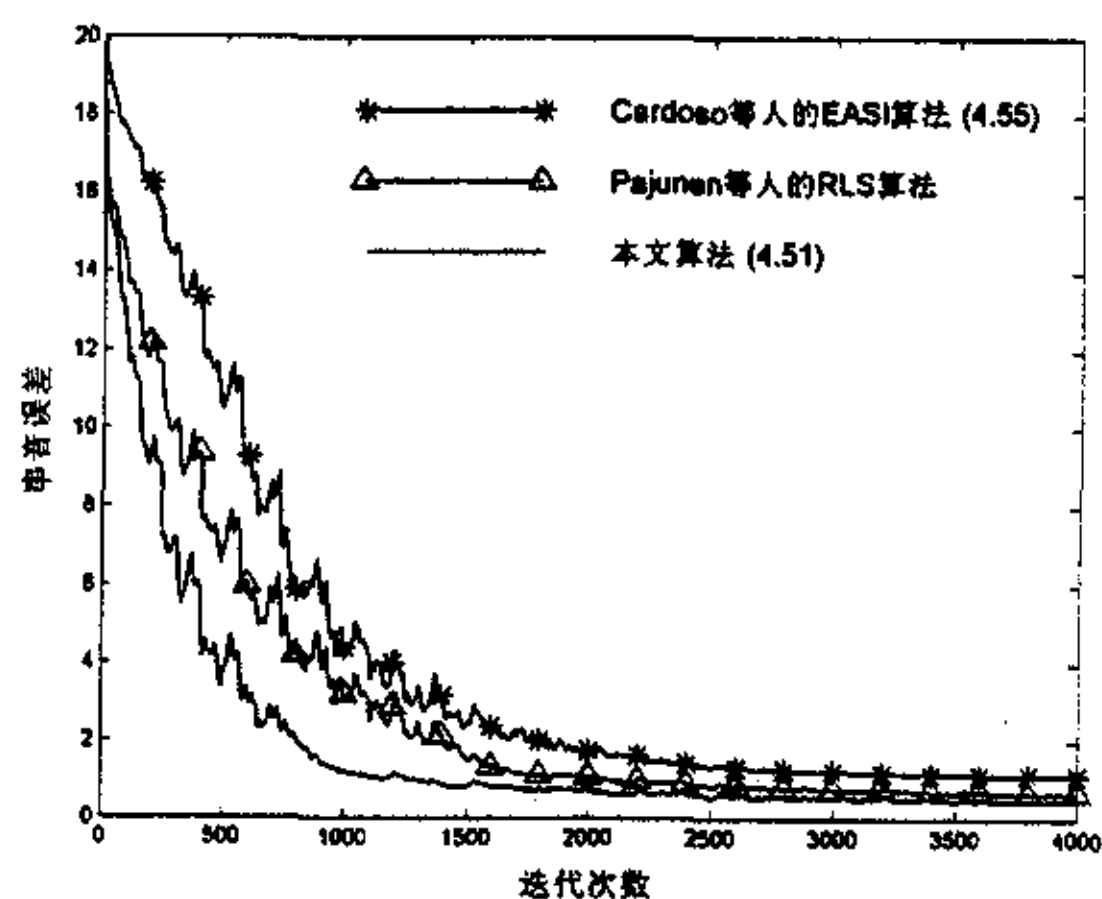


图4.5 2000次独立运行串音误差均值图,  
时间递增遗忘因子和时间递减学习速率

使用时变学习速率的LMS型算法除了参数选择的难度大外,另一个显著缺点是:若 $\eta_t$ 已降至很小,仍有信号未被分离,则这些信号的提取将非常困难。换句话说,它不能保证信号的完全分离,这也正是图4.5中的LMS型算法(4.55)稳态值稍高于其它两种算法的原因之所在。相比之下,RLS型的算法则不存在这些不足。

## § 4.5 小 结

信号处理的滤波器理论表明,RLS型的算法因使用更多的观测样本学习,从而比LMS型的算法具有更快的收敛速度和更好的跟踪能力。另一方面,自然梯度因为考虑

了对比函数的参数结构，它代表了真正意义上的最陡上升方向，因此比一般梯度更适合于盲信号分离。基于这两方面的考虑，本章推导了两种基于自然梯度的盲信号分离RLS算法，具体内容包括：

(1) 推导了一种RLS型的预白化算法，其平衡点的分析表明它的输出是我们期望的白色信号向量。

(2) 从非线性PCA对比函数出发，结合分离矩阵的标准正交约束，推导出了一种新颖的自然梯度，并由此得到了两类算法。一类是LMS型的算法，它与已有的非线性PCA子空间算法、Douglas的自稳定算法、以及Cardoso等人的串行更新算法相联系。另一类是RLS型的算法(RLS-I算法)，它与Pajunen等人提出的基于一般梯度的盲信号分离RLS算法相联系。理论证明RLS-I算法能实现预白化的盲信号分离，而大量的计算机仿真也表明RLS-I算法较Pajunen等人的RLS算法工作更优越。

(3) 将提出的RLS型的预白化算法和RLS-I算法相结合，经过合理近似，我们推导出一种新的盲信号分离RLS算法(RLS-II算法)，因为它隐含执行了白化运算，因此不需要对观测数据进行白化预处理就能实现信号的盲分离。理论证明RLS-II算法是等变化的，并满足分离矩阵的非奇异性条件，大量的计算机仿真试验也验证了RLS-II算法的有效性。

## 第五章 混合亚高斯和超高斯信号的盲分离

### §5.1 引言

理论研究表明,盲信号分离的最优激励函数是取决于源信号概率密度函数的分值函数,但在盲信号分离问题中,源信号的概率分布总是未知的。盲信号分离的局部稳定性分析表明,使用的激励函数即使不是分值函数,但若保证由激励函数刻画的算法搜索方向与由分值函数刻画的算法搜索方向成锐角,换句话说,始终保证算法沿对比函数的上升方向(尽管不是最陡上升方向)进行搜索,则算法收敛同样导致盲信号分离。那么,在什么样的条件下,两种搜索方向成锐角呢?盲信号分离的局部稳定性条件回答了这个问题。这一部分内容已经在第二章作了介绍。

当前,激励函数的选择主要有两种方法。一种是所谓的“自顶而底”的方法,它首先为源信号的概率分布建立模型,比如广义高斯分布模型和混合高斯分布模型等,然后根据观测信号样本估计模型参数,并由此确定激励函数。另一种是所谓的“自底而顶”的方法,它直接根据算法的局部稳定性条件选择激励函数。如果模型的参数估计准确,第一种方法能够获得渐进最优甚至最优的激励函数即分值函数,但缺点在于参数估计需要大量的观测样本,实时性差,算法相对复杂,运算量大。第二种方法的激励函数形式简单,算法也容易实现。以算法(2.26)为例,对亚高斯信号三次函数满足算法的局部稳定性条件;对超高斯信号,符号函数和双曲正切函数满足条件。最近,Mathis和Douglas证明<sup>[114]</sup>,不存在一个固定(不含自适应参数)的非线性函数(确切地说,是不存在一个固定的估计函数),能够实现任意分布源信号的盲分离。因此第二种方法的缺点是它只能分离服从特定分布的源信号。

为了在保留算法简单结构和小运算量的同时,实现任意分布源信号的盲分离,研究人员进行了大量积极的探索,提出了许多专门算法,包括峰度最大化算法<sup>[54]</sup>、推广的ICA算法<sup>[68,101]</sup>、灵活的ICA算法<sup>[37]</sup>、源自适应的ICA算法<sup>[148]</sup>、阈值ICA算法<sup>[113]</sup>和级联神经网络盲提取算法<sup>[141]</sup>等。本章选用归一化峰度的绝对值之和作为预白化盲信号分离的对比函数,结合正交约束下的自然梯度和现有的串行白化算法,推导一种具有信号自适应激励函数的盲信号分离算法。

本章的内容组织如下。首先在第2节介绍信号的亚高斯和超高斯分类方法;第3节介绍激励函数选择的理论依据:局部稳定性条件、渐进方差和稳健性,并简单证明分值函数是最优的激励函数;第4节推导一种稳健的混合亚高斯和超高斯信号的盲分离

算法，并将新算法与相关算法作比较；第5节给出计算机仿真，最后在第6节对本章内容作小结。

## §5.2 信号分类

信号处理中的信号有多种分类方法<sup>[175]</sup>，比如平稳信号与非平稳信号；周期信号与非周期信号；离散时间信号与连续时间信号；随机信号与确定性信号，等等。在盲信号分离研究中，经常将信号分成三类：高斯信号，亚高斯信号和超高斯信号。

亚高斯信号和超高斯信号是根据概率密度函数来定义的。设信号  $s$  经过归一化（零均值和单位方差）处理得到的信号为  $\bar{s}$ ，并记  $\bar{s}$  的概率密度函数为  $p(x)$ ，标准高斯分布的概率密度函数为  $g(x)$ ，则有下列的定义。

**定义 5.1**<sup>[110]</sup> 信号  $s$  称为超高斯信号，若  $\exists x_0 \in \mathbb{R}^+$ ，使得  $\forall x \geq x_0, p(x) > g(x)$ ；信号  $s$  称为亚高斯信号，若  $\exists x_0 \in \mathbb{R}^+$ ，使得  $\forall x \geq x_0, p(x) < g(x)$ 。

记  $\text{cum}_p[s]$  表示信号  $s$  的  $p$  阶累积量，参阅附录A。定义归一化的峰度<sup>[54,141]</sup>

$$\kappa_4(s) = \frac{\text{cum}_4[s]}{\{\text{cum}_2[s]\}^2} \quad (5.1)$$

并由此将信号分成以下三类：

- 1) 正峰度信号， $\kappa_4(s) > 0$ ；
- 2) 负峰度信号， $\kappa_4(s) < 0$ ；
- 3) 零峰度信号， $\kappa_4(s) = 0$ 。

常见的正峰度信号如部分语音信号，负峰度信号如数字通信中的部分调制信号，而最常见的零峰度信号就是我们熟悉的高斯信号。需要指出的是，高斯信号一定是零峰度信号，但零峰度信号却未必是高斯信号。

Mansour等人<sup>[110]</sup>证明，若概率密度函数  $p(x)$  和  $g(x)$  只有两个交点，也就是说方程  $f(x) = g(x)$  只有两个解，则正峰度信号  $\kappa_4(s) > 0$  与超高斯信号的定义等价，负峰度信号  $\kappa_4(s) < 0$  与亚高斯信号的定义等价。从这个意义上讲，正峰度信号又称超高斯信号，负峰度信号又称亚高斯信号。但当方程  $f(x) = g(x)$  的解多于两个时，这种直观的理解不再成立。换句话说，超高斯信号的归一化峰度值可能小于零，而亚高斯信号的归一化峰度值也可能大于零。

应当指出的是，在盲信号分离研究中习惯认为  $f(x) = g(x)$  只有两个解，从而对超



(亚)高斯信号与正(负)峰度信号不加区分。此外,在 $f(x)=g(x)$ 只有两个解的条件下,如果用广义高斯分布模型<sup>[177]</sup>

$$p(s, \alpha) = \frac{\alpha}{2\lambda\Gamma(\alpha^{-1})} \exp\left[-|\lambda^{-1}s|^\alpha\right] \quad (5.2)$$

来逼近信号 $s$ 的概率分布,其中 $\alpha$ 和 $\lambda$ 为模型参数, $\Gamma(\cdot)$ 表示Gamma函数,则也可以根据高斯指数 $\alpha$ ,将信号分成三类: $\alpha=2$ 为高斯信号; $\alpha<2$ 为超高斯信号; $\alpha>2$ 为亚高斯信号。比较典型的超高斯信号是 $\alpha=1$ 的拉普拉斯(Laplacian)分布信号,典型的亚高斯信号是 $\alpha=+\infty$ 的均匀分布信号。

### §5.3 激励函数的选择依据

第二章介绍的局部稳定性条件和稳态均方误差分析为盲信号分离激励函数的选择提供了理论依据。前者决定盲信号分离算法能否工作,后者决定算法的工作效果,两者的最佳结合点就是分值函数。

#### 5.3.1 渐近方差

设第 $i$ 个源信号的概率密度函数为 $p_i(s_i)$ ,与之相对应的激励函数为 $g_i(s_i)$ ,并记 $g_i(s_i)$ 的一阶导函数为 $g'_i(s_i)$ ,原函数族为 $\mathcal{G}_i(s_i)$ ,即 $\mathcal{G}'_i(s_i)=g_i(s_i)$ 。考虑分离矩阵 $\mathbf{B}$ 中与 $s_i$ 相对应的列 $\mathbf{b}_i$ ,Hyvarinen等人<sup>[80,89]</sup>证明, $\mathbf{b}_i$ 的估计值 $\hat{\mathbf{b}}_i$ 的渐进协方差矩阵的迹(trace)等于

$$\mathcal{V}_g = C(\mathbf{A}) \frac{E\{g_i^2(s_i)\}E\{s_i^2\} - [E\{s_i g_i(s_i)\}]^2}{[E\{s_i g_i(s_i) - g'_i(s_i)\}]^2} \quad (5.3)$$

其中 $C(\mathbf{A})$ 是一个仅依赖于混合矩阵 $\mathbf{A}$ 的常数。

把迹(5.3)看成是激励函数 $g_i(s_i)$ 的函数,并在任意非线性函数构成的函数空间进行搜索,Hyvarinen等人<sup>[80,89]</sup>证明当且仅当 $\mathcal{G}_i(s_i)$ 取下面的形式:

$$\mathcal{G}_i(s_i) = c_1 \log p_i(s_i) + c_2 s_i^2 + c_3 \quad (5.4)$$

其中 $c_1$ 、 $c_2$ 和 $c_3$ 为任意常数,迹(5.3)取得最小值。

显然,分值函数满足条件(5.4),因此它刻画的盲信号分离算法的分离矩阵具有最小的渐进估计方差。

### 5.3.2 稳健性

盲信号分离激励函数的选择需要考虑的另一个因素是对异常值的稳健性。稳健性又称鲁棒性 (robustness), 它意味着观测数据的单个较大的异常值对估计子的性能不产生实质性的影响。盲信号分离问题的估计子可以是分离矩阵  $\mathbf{B}$  也可以是分离矩阵的某一系列  $\mathbf{b}_i$ 。

**定义5.2**<sup>[89]</sup> 估计子  $\hat{\theta}$  称为  $\theta$  的M-估计子, 若  $\hat{\theta}$  是方程

$$E\{\zeta(\mathbf{r}, \theta)\} = 0 \quad (5.5)$$

的解, 其中  $\mathbf{r}$  是一个随机向量,  $\zeta(\cdot)$  是估计子的定义函数。

若某估计子是M-估计子, 则其稳健性可通过影响函数 (IF: Influence Function) 来分析。直观地理解, 影响函数就是用来描述单个观测样本对估计子性能的影响情况的函数。为使估计子  $\hat{\theta}$  具有良好的稳健性, 换句话说, 即使是比较大的异常观测样本, 它对估计子的影响也是有限的, 从而对估计子的性能不产生实质性的影响, 我们希望影响函数是随机向量  $\mathbf{r}$  的有界函数。

**定义5.3**<sup>[89]</sup>  $\theta$  的M-估计子  $\hat{\theta}$  称为B-鲁棒的估计子, 若影响函数  $\text{IF}(\mathbf{r}, \hat{\theta})$  是随机向量  $\mathbf{r}$  的有界函数, 即  $\exists M \in \mathbb{R}^+$ , 使得

$$\sup_{\mathbf{r}} \|\text{IF}(\mathbf{r}, \hat{\theta})\| < M \quad (5.5)$$

成立, 其中  $\sup(\cdot)$  表示上确界算子, 而  $\|\cdot\|$  表示向量的2-范数。

可以证明<sup>[89]</sup>定义5.2中的M-估计子  $\hat{\theta}$  的影响函数为

$$\text{IF}(\mathbf{r}, \hat{\theta}) = \Upsilon \zeta(\mathbf{r}, \hat{\theta}) \quad (5.6)$$

其中  $\Upsilon$  是与随机向量  $\mathbf{r}$  不相关的任意可逆矩阵。

以提取单个源信号的串行算法为例, 预白化盲信号提取的估计子是  $\theta = (\mathbf{w}_i, \beta)$ <sup>[89]</sup>, 其中  $\mathbf{w}_i$  表示正交分离矩阵  $\mathbf{W}$  的某一系列,  $\beta$  是拉格朗日 (lagrangian) 乘子, 而随机向量  $\mathbf{r} = \mathbf{U}\mathbf{x}$ , 表示观测数据经白化预处理获得的向量。令  $h_i(y_i) = y_i g_i(y_i)$ ,  $g_i(\cdot)$  是算法的激励函数, Hyvarinen等人<sup>[89]</sup>推导了估计子的影响函数, 并证明其2-范数的平方等于

$$\|\text{IF}(\mathbf{r}, \hat{\theta})\|^2 = \|\Upsilon \zeta(\mathbf{r}, (\mathbf{w}_i, \beta))\|^2 = C_1 \rho^{-2} h_i^2(\mathbf{w}_i^T \mathbf{r}) + C_2 h_i(\mathbf{w}_i^T \mathbf{r}) + C_3 \quad (5.7)$$

式中  $C_1$ 、 $C_2$  和  $C_3$  是与随机向量  $\mathbf{r}$  无关的任意常数, 而  $\rho = \mathbf{w}_i^T \mathbf{r} / \|\mathbf{r}\|$  表示向量  $\mathbf{w}_i$  与  $\mathbf{r}$  夹角的余弦。

由式 (5.7) 不难看出, 估计子  $\hat{\mathbf{w}}_i$  的稳健性取决于激励函数  $g_i(\cdot)$ 。  $h_i(y_i) = y_i g_i(y_i)$  随自变量  $y_i$  增加越缓慢,  $\hat{\mathbf{w}}_i$  的稳健性也就越高。因为通常情况下, 异常值与信号向量  $\mathbf{r}$  几乎是正交的<sup>[89]</sup>, 因此对于较大的异常值, 参数  $\rho$  一般很小, 从而影响函数 (5.7) 是一个无界函数, 估计子  $\hat{\mathbf{w}}_i$  不是 B-鲁棒的估计子。

关于稳健性, 我们有下面的定理。

**定理 5.1<sup>[86,89]</sup>** 设  $\mathbf{r} = \mathbf{U}\mathbf{x}_i$  表示观测数据经稳健的白化预处理获得的白色信号向量。估计子  $\hat{\mathbf{w}}_i$  的影响函数 (5.7) 是  $\mathbf{r}$  的无界函数。若  $h_i(y_i) = y_i g_i(y_i)$  是有界函数, 则  $\forall \varepsilon > 0$ , 在集合  $\{\mathbf{r} | \hat{\mathbf{w}}_i^T \mathbf{r} > \varepsilon\}$  中, 影响函数 (5.7) 是有界的。

定理 5.1 说明, 为使观测数据的异常值对分离矩阵的估计性能影响尽可能地小, 激励函数  $g_i(\cdot)$  的选择应该满足  $h_i(y_i) = y_i g_i(y_i)$  是有界函数, 或至少保证  $h_i(y_i)$  随着  $|y_i|$  的增加而不会很快增加。

显然, 如果激励函数是盲信号分离的分值函数

$$\varphi_i(y_i) = -\frac{p'_i(y_i)}{p_i(y_i)} \quad (5.8)$$

则  $h_i(y_i) = y_i \varphi_i(y_i)$  是有界的, 或至少不会随  $|y_i|$  的增加而很快增加, 因为在统计意义下  $E\{h_i(y_i)\} = 0$ 。相比之下, 三次激励函数  $\varphi_i(y_i) = y_i^3$  不满足上述条件, 因此对应算法的稳健性要差许多。

作为总结, 盲信号分离激励函数的选择依据是: 满足局部稳定性条件、提供尽可能小的渐进方差和良好的稳健性。其中渐进方差与第二章介绍的算法的稳态性能即源信号的恢复质量密切相关, 而这对于梯度自适应 (批处理) 算法, 可以通过降低学习速率参数 (增加批处理的块长) 来改善, 因此局部稳定性条件和稳健性条件就成为激励函数选择的决定性因素。

## §5.4 一种稳健的盲信号分离算法

本节, 我们利用归一化峰度的绝对值之和作为预白化盲信号分离的对比函数, 推导一种具有信号自适应激励函数的盲信号分离算法, 它能实现混合峰度信号的盲分离, 并具有令人满意的收敛性和稳健性。

### 5.4.1 算法推导

在第四章中，我们指出分离矩阵  $\mathbf{B}$  的确定有两种方法：直接法和间接法。下面采用间接法推导新算法，它先求矩阵  $\mathbf{U}$  对观测数据进行预白化，再求正交矩阵  $\mathbf{W}$ ，并作用于白化后的数据向量以提取源信号，从而总的分离矩阵  $\mathbf{B} = \mathbf{W}\mathbf{U}$ 。

有多种算法求解白化矩阵  $\mathbf{U}$ ，如基于特征值分解的批处理算法<sup>[89]</sup>，基于解相关的自适应算法<sup>[55]</sup>，基于随机相对梯度的LMS型算法<sup>[22]</sup>和第四章提出的RLS型的预白化算法等。这里假定观测信号与源信号的数目相等，并使用Cardoso和Laheld提出的LMS型的串行白化算法更新白化矩阵，即<sup>[22]</sup>

$$\mathbf{U}_{i+1} = \mathbf{U}_i + \eta_i [\mathbf{I} - \mathbf{v}_i \mathbf{v}_i^T] \mathbf{U}_i \quad (5.9)$$

其中  $\eta_i > 0$  是学习步长参数， $\mathbf{v}_i = \mathbf{U}_i \mathbf{x}_i$  表示白化后的输出信号向量。

同样的，求解正交矩阵  $\mathbf{W}$  的自适应算法也是多种多样，如非线性PCA算法<sup>[96,122]</sup>，EASI算法中的串行更新算法<sup>[22]</sup>、自稳定的LMS型算法<sup>[58]</sup>和RLS型的算法<sup>[124,167]</sup>等。为实现混合峰度信号的盲分离，我们选用归一化峰度的绝对值之和<sup>[116]</sup>

$$J(\mathbf{y}_i; \mathbf{W}) = \frac{1}{4} \sum_{i=1}^n |\kappa_4(y_{i,j})| \quad (5.10)$$

作为对比函数，推导正交矩阵  $\mathbf{W}$ 。

记  $m_{pi} = E\{y_{i,j}^p\}$  表示  $y_{i,j}$  的  $p$  阶矩。因为  $\mathbf{W}$  是正交矩阵， $\mathbf{v}_i$  是白色信号向量，所以  $\mathbf{y}_i = \mathbf{W}\mathbf{v}_i$  也是白色信号向量，从而有

$$\kappa_4(y_{i,j}) = \frac{\text{cum}_4[y_{i,j}]}{\{\text{cum}_2[y_{i,j}]\}^2} = \frac{m_{4i} - 3m_{2i}^2}{m_{2i}^2} \quad (5.11)$$

将式 (5.11) 代入到 (5.10)，不难证明对比函数  $J(\mathbf{y}_i; \mathbf{W})$  对  $\mathbf{W}$  的第  $i$  行第  $j$  列元素  $\mathbf{W}_{ij}$  的偏导数等于

$$\nabla_{\mathbf{W}_{ij}} J(\mathbf{y}_i; \mathbf{W}) = \frac{\partial}{\partial \mathbf{W}_{ij}} J(\mathbf{y}_i; \mathbf{W}) = E\{\varphi_i(y_{i,j}) \mathbf{v}_{j,i}\} \quad (5.12)$$

其中非线性函数

$$\varphi_i(y_{i,j}) = \text{sgn}[\kappa_4(y_{i,j})] \cdot \frac{m_{2i} y_{i,j}^3 - m_{4i} y_{i,j}}{m_{2i}^3} \quad (5.13)$$

就是盲信号分离中所谓的激励函数。

若记列向量  $\Phi(y_i) = [\varphi_1(y_{1,i}), \dots, \varphi_n(y_{n,i})]^T$ , 则可将 (5.12) 写成矩阵形式, 有

$$\nabla J(y_i; \mathbf{W}) = E\{\Phi(y_i) \mathbf{v}_i^T\} \quad (5.14)$$

因为  $\mathbf{W}$  是正交矩阵, 所以可以利用正交约束下的自然梯度<sup>[9,58,60,167]</sup>

$$\tilde{\nabla} J = \nabla J \cdot \mathbf{W}^T \mathbf{W} - \mathbf{W} \cdot [\nabla J]^T \cdot \mathbf{W} \quad (5.15)$$

来搜索对比函数 (5.10) 的极大值, 得

$$\tilde{\nabla} J(y_i; \mathbf{W}) = E\{\Phi(y_i) \mathbf{y}_i^T - \mathbf{y}_i \Phi^T(y_i)\} \cdot \mathbf{W}$$

从而有下面的随机自然梯度学习算法:

$$\mathbf{W}_{i+1} = \mathbf{W}_i + \eta_i [\Phi(y_i) \mathbf{y}_i^T - \mathbf{y}_i \Phi^T(y_i)] \mathbf{W}_i \quad (5.16)$$

其中  $\eta_i > 0$  是学习速率参数。

Cardoso 等人<sup>[22]</sup>以及 Douglas<sup>[58]</sup>研究了形如 (5.16) 的盲信号分离算法的局部稳定性条件, 它由下式给出:

$$\eta_i [E\{y_i \varphi_i(y_i)\} - E\{\varphi'_i(y_i)\} E\{y_i^2\} + E\{y_j \varphi_j(y_j)\} - E\{\varphi'_j(y_j)\} E\{y_j^2\}] > 0 \quad (5.17)$$

其中  $\varphi'_i$  表示激励函数  $\varphi_i$  的一阶导数,  $y_i$  是  $y_{i,i}$  的简写。

根据激励函数 (5.13), 不难证明

$$E\{y_i \varphi_i(y_i)\} = \text{sgn}[\kappa_4(y_i)] \cdot m_{2i}^{-3} \cdot E\{m_{2i} y_i^4 - m_{4i} y_i^2\} = 0 \quad (5.18)$$

$$E\{\varphi'_i(y_i)\} = \text{sgn}[\kappa_4(y_i)] \cdot m_{2i}^{-3} \cdot E\{3m_{2i} y_i^2 - m_{4i}\} = -m_{2i}^{-1} |\kappa_4(y_i)| \quad (5.19)$$

因此, 如果最多只有一个源信号是零峰度值信号, 则算法 (5.16) 的局部稳定性条件必定成立。

此外, 根据式 (5.18), 激励函数满足估计子的有界性条件 (随着  $|y_i|$  的增加, 函数  $h_i(y_i) = y_i \varphi_i(y_i)$  不会很快增加), 因此算法 (5.16) 工作是稳健的。

我们指出, 归一化峰度的绝对值之和 (5.10) 是盲信号分离的对比函数的前提是观测信号经过白化预处理, 或者混合矩阵约定为正交矩阵。如果观测信号未作白化预处理, 则需要级联运行白化算法和上面的信号分离算法 (5.16)。事实上, LMS 型的预白化算法 (5.9) 和正交矩阵  $\mathbf{W}$  的更新算法 (5.16) 可以合而为一, 两端相乘, 并



忽略学习步长 $\eta_i$ 的二阶无穷小项, 有

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i [\mathbf{I} - \mathbf{y}_i \mathbf{y}_i^T + \Phi(\mathbf{y}_i) \mathbf{y}_i^T - \mathbf{y}_i \Phi^T(\mathbf{y}_i)] \mathbf{B}_i \quad (5.20)$$

算法(5.20)是一种新的盲信号分离算法, 尽管它与Cardoso等人提出的EASI算法<sup>[22]</sup>具有相同的形式, 但两者的核心部分即激励函数完全不同。我们提出的激励函数形如(5.13), 它是信号自适应的, 稳健性好, 并能实现混合峰度信号的盲分离; 而EASI算法的激励函数取 $\varphi(y_i) = y_i^3$ 或者 $\varphi(y_i) = -y_i^3$ , 前者只能实现正峰度信号的盲分离, 后者则只能实现负峰度信号的盲分离。

#### 5.4.2 与相关算法的比较

为实现混合峰度信号的盲分离, 现有文献提出了许多专门算法, 下面我们将新算法(5.20)与其中的典型算法作比较。

首先是推广的ICA算法, 它是Lee等人<sup>[101,68]</sup>结合Bell等人<sup>[14]</sup>的信息最大化算法和Amari等人<sup>[7]</sup>的自然梯度算法提出的, 算法形式如下:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i [\mathbf{I} - \Phi(\mathbf{y}_i) \mathbf{y}_i^T] \mathbf{B}_i \quad (5.21)$$

激励函数有两种选择, 它们是

$$\varphi_i(y_{i,j}) = \text{sgn}[\kappa_4(y_{i,j})] \cdot \tanh(y_{i,j}) + y_{i,j} \quad (5.22a)$$

$$\varphi_i(y_{i,j}) = 2 \tanh(y_{i,j} + \rho_i) + 2 \tanh(y_{i,j} - \rho_i) - 2 \tanh(y_{i,j}) \quad (5.22b)$$

式中, 参数 $\rho_i = 1 - \text{sgn}[\kappa_4(y_{i,j})]$ 。

为源信号建立广义高斯分布模型(5.2), 并根据系统输出信号的归一化峰度的估计值选择高斯指数, Choi等人<sup>[37]</sup>提出了灵活的ICA算法。算法的形式如(5.21), 激励函数

$$\varphi_i(y_{i,j}) = |y_{i,j}|^{\alpha-1} \cdot \text{sgn}[y_{i,j}] \quad (5.23)$$

其中高斯指数 $\alpha$ 的取值为

$$\alpha = \begin{cases} 4 & , \quad \kappa_4(y_{i,j}) < 0 \\ 1 & , \quad 0 \leq \kappa_4(y_{i,j}) < 20 \\ 0.8 & , \quad \kappa_4(y_{i,j}) \geq 20 \end{cases}$$

与所提出的新算法(5.20)一样,上述两种算法都是等变化的,也都满足分离矩阵的非奇异性条件,但由于使用的激励函数的形式有限(2到3种选择),换句话说,激励函数不是信号自适应的,因此相应算法的性能受限制。相比之下,算法(5.20)使用了信号自适应的激励函数(5.13),对异常值的稳健性高,其性能也优于上述两种算法。计算机仿真验证了这一点。

用混合高斯分布模型

$$f(y) = \sum_{k=1}^M \pi_k \cdot \frac{1}{\sigma_k \sqrt{2\pi}} \exp \left[ -\frac{(y - \mu_k)^2}{2\sigma_k^2} \right] \quad (5.24)$$

为源信号建模, Vlassis等人<sup>[148]</sup>提出了源自自适应的ICA算法:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i \left[ \mathbf{I} - E \{ \Phi(\mathbf{y}_i) \mathbf{y}_i^T \} \right] \mathbf{B}_i \quad (5.25)$$

这是一种批处理算法。如果高斯核函数的均值等间隔排列,并具有相同的方差 $\sigma$ ,则算法的激励函数

$$\varphi_i(y_{i,l}) = \frac{1}{\sigma^2} \left\{ y_{i,l} - \frac{\sum_{k=1}^M \pi_k \mu_k \exp \left[ -0.5\sigma^{-2} (y_{i,l} - \mu_k)^2 \right]}{\sum_{k=1}^M \pi_k \exp \left[ -0.5\sigma^{-2} (y_{i,l} - \mu_k)^2 \right]} \right\} \quad (5.26)$$

源自自适应的ICA算法的优点在于对异常值的稳健性高,并有望获得渐进最优的稳态性能。但作为一种批处理算法,尽管激励函数(5.26)可通过FFT快速算法来估计,其运算量仍然相当可观(高斯核函数的个数 $M$ 取值一般超过100)。最关键的是,这种方法容易陷于局部极值点<sup>[148]</sup>,因此需要用其它混合峰度信号的盲分离算法启动才能正常工作。

从广义高斯分布模型(5.2)出发, Mathis等人<sup>[113]</sup>推导出拉普拉斯分布源信号(一种超高斯信号)的分值函数是符号函数,并通过极限运算证明均匀分布源信号(一种亚高斯信号)的分值函数是一种阶梯函数。由于符号函数是一种特殊的阶梯函数,因此可以将两者统一,从而得到阈值ICA算法,形如(5.21),激励函数

$$\varphi_i(y_{i,l}) = \begin{cases} 0, & |y_{i,l}| < c_i \\ \zeta_i \operatorname{sgn}[y_{i,l}], & |y_{i,l}| \geq c_i \end{cases} \quad (5.27)$$

其中参数 $\zeta_i$ 和 $c_i$ 用梯度自适应算法获得。这种方法的优点是激励函数形式简单,容易在DSP上实现,并且对异常值具有较高的稳健性。但缺点在于需要用梯度算法更新参

数 $\zeta_i$ 和 $c_i$ ，若这两个梯度算法的步长参数选择不当，则难以获得满意的收敛性能。

由归一化峰度的绝对值对比函数

$$J_{KMA}(\mathbf{w}_i) = |\kappa_4(y_{i,i})| \quad (5.28)$$

出发，Ding等人<sup>[54]</sup>提出了峰度最大化算法。这种算法能够实现混合峰度信号的盲提取，但作为一种串行算法，它不能同时分离所有的源信号，并且使用Gram-Schmidt类正交化技术实现压缩映射(deflation)会导致误差积累，因此随着信号个数的增加，信号的恢复质量将显著恶化<sup>[66,141]</sup>。

从最大熵逼近的负熵对比函数出发，即

$$J(\mathbf{y}_i; \mathbf{B}) = \sum_{i=1}^n \left[ E\{L(y_{i,i})\} - E\{L(y_{gau})\} \right]^2 \quad (5.29)$$

式中 $y_{gau}$ 表示标准的高斯随机变量，而

$$L(u) = \frac{1}{a} \log \cosh(au), \quad 1 \leq a \leq 2 \quad (5.30)$$

Hyvarinen等人<sup>[79,86]</sup>提出了快速定点ICA盲信号提取算法：

$$\begin{aligned} \mathbf{w}_{i,j+1} &= E\left\{\mathbf{v}_i L'(\mathbf{w}_{i,j}^T \mathbf{v}_i)\right\} - E\left\{L''(\mathbf{w}_{i,j}^T \mathbf{v}_i)\right\} \cdot \mathbf{w}_{i,j} \\ \mathbf{w}_{i,j+1} &= \mathbf{w}_{i,j+1} / \|\mathbf{w}_{i,j+1}\| \end{aligned} \quad (5.31)$$

其中 $L'$ 和 $L''$ 分别表示函数 $L$ 的一阶导数和二阶导数， $\mathbf{v}_i = \mathbf{U}\mathbf{x}_i$ 表示观测数据白化预处理后的输出信号向量。虽然收敛快，稳健性好，但作为一种批处理串行算法，快速定点ICA算法也使用Gram-Schmidt类正交化技术实现压缩映射过程，因此具有与峰度最大化算法类似的缺点。

利用稳定性条件(2.39)和(2.40)的互补性，Hoff等人<sup>[78]</sup>提出了切换ICA算法。对于一种确定性的激励函数，算法(2.26)和(2.27)中必定有一个是局部稳定的，因此，通过在两者之间的切换，可实现任意分布的非高斯信号的盲分离。对现有的单一峰度信号的盲分离算法作适当改造，可获得一大类切换ICA算法。需要说明的是，上述思想是由Cichocki等人<sup>[43]</sup>最早提出的，他们的算法采用形式

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i [\mathbf{I} - \mathbf{f}(\mathbf{y}_i) \mathbf{g}^T(\mathbf{y}_i)] \mathbf{B}_i \quad (5.32)$$

而非线性函数取作

$$f_i(y_{i,t}) = \begin{cases} \tanh(10y_{i,t}), & \kappa_4(y_{i,t}) \geq 0.1 \\ y_{i,t}, & \kappa_4(y_{i,t}) < 0.1 \end{cases}$$

$$g_i(y_{i,t}) = \begin{cases} \tanh(10y_{i,t}), & \kappa_4(y_{i,t}) < -0.1 \\ y_{i,t}, & \kappa_4(y_{i,t}) \geq -0.1 \end{cases} \quad (5.33)$$

与前面提到的推广的ICA算法和灵活的ICA算法一样, 切换ICA算法因为使用的激励函数不是信号自适应的, 从而算法的收敛性能有限。

最后, 需要提到的是级联神经网络盲提取算法, 它是Thawonmas等人<sup>[141]</sup>根据归一化峰度的绝对值对比函数(5.28)提出的, 形式如下:

$$\mathbf{b}_{i,t+1} = \mathbf{b}_{i,t} + \eta_{i,t} \varphi_i(\mathbf{b}_{i,t}^T \mathbf{x}_t^{(i)}) \mathbf{x}_t^{(i)} \quad (5.34)$$

式中 $\mathbf{b}_{i,t}$ 表示分离矩阵 $\mathbf{B}_t$ 的第 $i$ 列, 即提取第 $i$ 个源信号的权向量,  $\eta_{i,t} > 0$ 是与之对应的学习速率参数。算法(5.34)中的激励函数 $\varphi_i(\cdot)$ 与(5.13)相同, 数据向量 $\mathbf{x}_t^{(i)} = \mathbf{x}_t$ 就是观测信号向量, 而

$$\mathbf{x}_t^{(i+1)} = \mathbf{x}_t^{(i)} - \mathbf{w}_t^{(i)} \mathbf{b}_{i,t}^T \mathbf{x}_t^{(i)}, \quad i = 2, \dots, n \quad (5.35)$$

是观测信号向量经压缩映射处理获得的信号向量。算法(5.35)中的权向量 $\mathbf{w}_t^{(i)}$ 通过梯度算法

$$\mathbf{w}_t^{(i)} = \mathbf{w}_t^{(i)} + \tilde{\eta}_{i,t} \mathbf{x}_t^{(i+1)} \mathbf{b}_{i,t}^T \mathbf{x}_t^{(i)} \quad (5.36)$$

自适应更新。

显然, 级联神经网络盲提取算法(5.34)是一种串行算法, 与同为串行算法的快速定点ICA算法(5.31)相比, 其突出优点是不需要对观测数据进行白化预处理, 而且信号的恢复质量不会随着信号个数的增加而显著恶化。级联盲提取算法的不足之处在于作为一种串行算法, 它不能同步提取所有的源信号, 也不具有等变化性。因为压缩映射过程用梯度算法(5.35)~(5.36)来实现, 因此必须确保该过程收敛即一个信号被完全分离后才能进行另外一个后续信号的提取, 从而导致算法收敛较慢。此外, 若压缩映射过程中的步长参数 $\tilde{\eta}_{i,t}$ 选择不当, 则同样会导致误差积累, 难以获得满意的分离结果。

有必要指出, 具有激励函数(5.13)的本文算法(5.20)从形式上看, 似乎是Cardoso等人<sup>[22]</sup>的EASI算法与Thawonmas等人<sup>[141]</sup>的激励函数的一种简单结合, 但这种结合是经过严格推导得到的, 并且推导过程与文献[22]和[141]不同。

## §5.5 性能仿真

为了验证本章提出的混合亚高斯和超高斯信号的盲分离算法 (5.20) 的有效性, 我们对以下信号进行了仿真试验:

- S1: 符号信号  $\text{sgn}(\cos(2\pi 155t))$ ;
- S2: 低频正弦信号  $\sin(2\pi 90t)$ ;
- S3: 在  $[-1,1]$  均匀分布的随机噪声信号;
- S4: 参数为1的指数分布信号;
- S5: 参数为1的Rayleigh分布信号。

上述5个信号中, S1~S3为亚高斯信号, 归一化的峰度值分别等于-2, -1.5和-1.2; S4~S5为超高斯信号, 归一化的峰度值分别等于6和0.25 (注: 实际值等于  $(-6\pi^2 + 24\pi - 16)/(\pi^2 - 8\pi + 16) \approx 0.2451$ )。

仿真中, 混合矩阵  $\mathbf{A}$  的元素为在  $[-1,1]$  区间均匀分布的随机数。以速率10kHz对传感器接收信号进行采样以产生离散观测信号  $\mathbf{x}_t$ 。盲信号分离算法的性能用“串音”误差<sup>[3,78,153,156]</sup>

$$E_{ci} = \sum_{p=1}^n \left( \sum_{q=1}^n \frac{|c_{pq}|}{\max_l |c_{pl}|} - 1 \right) + \sum_{q=1}^n \left( \sum_{p=1}^n \frac{|c_{pq}|}{\max_l |c_{lq}|} - 1 \right) \quad (5.37)$$

来衡量, 其中  $\mathbf{C}_i = [c_{pq}] = \mathbf{B}_i \mathbf{A}$  表示混合-分离合成系统的传递矩阵。

作为比较, 我们同时运行本文算法 (5.20), 推广的ICA算法<sup>[101]</sup> (5.22a) 和 (5.22b) 以及灵活的ICA算法<sup>[37]</sup> (5.23)。几种算法中用到的二阶和四阶矩用递推公式<sup>[37]</sup>

$$m_{pi,t+1} = (1-\lambda)m_{pi,t} + \lambda y_{i,t}^p, \quad p=2,4 \quad (5.38)$$

更新, 式中  $\lambda$  是泄漏因子 (leakage factor), 在仿真中取  $\lambda = 0.005$ 。归一化的峰度值  $\kappa_4(y_{i,t})$  根据式 (5.11) 求出。由于迭代次数较少时, 高阶矩的估计误差大, 为减少这种影响, 我们给激励函数 (5.13) 加约束条件  $-1 \leq \varphi_i(y_{i,t}) \leq 1$  (这种约定是可行的, 因为  $E\{\varphi_i(y_{i,t})y_{i,t}\} = 0$ )。所有算法使用相同的步长参数  $\eta = 0.004$ 。采用5000个样本, 图 5.1 给出了500次独立运行 (每次运行中的混合矩阵  $\mathbf{A}$  和随机信号 S3~S5 都是随机产生的) 平均“串音”误差曲线图。图中的四条曲线从上而下, L1 表示推广的ICA算法 (5.22a), L2 表示推广的ICA算法 (5.22b), L3 表示灵活的ICA算法 (5.23), L4 表示本文算法。不难看出: 本文提出的新算法 (5.20) 能够实现混合亚高斯和超高斯信号的盲分离, 其收敛速度明显比推广的ICA算法和灵活的ICA算法快, 稳态性能 (表征



了信号的恢复质量) 也好。

为考虑异常值对算法性能的影响, 我们在第1501个样本中, 给确定性信号S1和S2分别加异常值10和20, 其它条件不变。同样作500次独立运行, 平均性能如图5.2所示 (由于灵活的ICA算法不够稳健, 500次中大概3~4次发散, 作统计时这些数据丢弃不用)。可以看到, 从第1501次迭代开始, 四种算法的性能都发生了变化, 而本文算法 (5.20) 表现仍然是最好的。

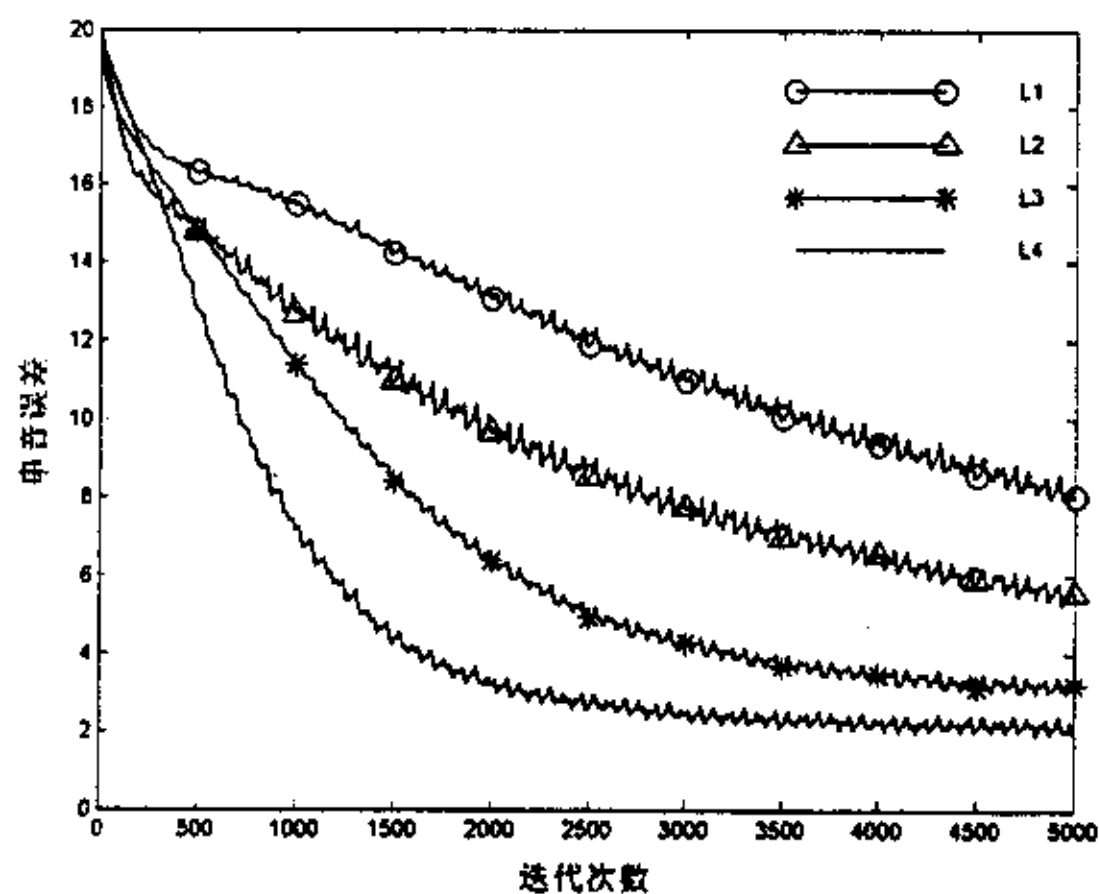


图5.1 串音误差均值图, 无异常值

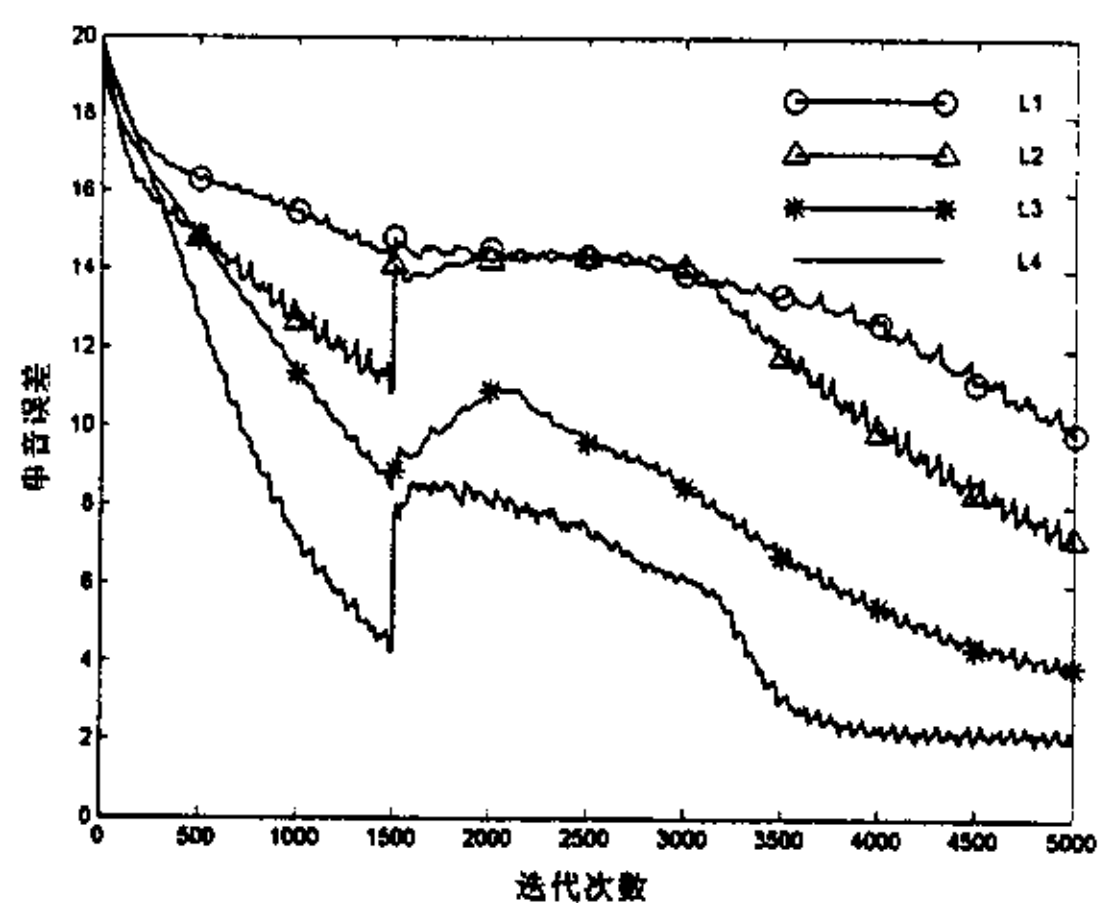


图5.2 串音误差均值图, 有异常值

## §5.6 小 结

盲信号分离的最优激励函数——分值函数依赖于未知的源信号的概率分布。盲信号分离的局部稳定性研究表明, 当源信号全部由超高斯信号或者全部由亚高斯信号组成时, 用简单的非线性函数替代原有的分值函数, 则算法也能够实现盲信号分离。但当源信号含混合峰度信号时, 用简单非线性函数替代分值函数的做法不能奏效。从这个意义上讲, 混合亚高斯和超高斯信号的盲分离是盲信号分离研究的难点之一。为解决这个问题, 本章提出了一种稳健的新算法, 具体内容如下:

(1) 给出了亚高斯信号和超高斯信号的定义, 明确了亚(超)高斯信号与负(正)峰度信号之间的联系和区别。

(2) 介绍了激励函数选择的理论依据: 局部稳定性条件, 渐进方差和稳健性。并简单证明分值函数是最优的激励函数。

(3) 用归一化峰度的绝对值之和出作为预白化盲信号分离的对比函数, 推导了一种新的等变化的盲信号分离算法。新算法的激励函数是信号自适应的, 并自动满足算



法的局部稳定性条件和稳健性条件。

(4) 分析比较了新算法与现有典型的混合亚高斯和超高斯信号的盲分离或盲提取算法的不同，并用大量的计算机仿真验证了新算法的有效性：能够实现混合峰度信号的并行盲分离，并对观测信号的异常值具有较高的稳健性。

## 第六章 超定盲信号分离

### §6.1 引言

根据观测信号和源信号个数的多少, 盲信号分离问题可以分成以下三种情况:

- 正定 (determined) 盲信号分离: 观测信号与源信号的数目相等;
- 超定 (over-determined) 盲信号分离: 观测信号比源信号的个数多;
- 欠定 (under-determined) 盲信号分离: 观测信号比源信号的个数少。

正定盲信号分离又称完备 (complete) 的盲信号分离, 而超定盲信号分离又称欠完备 (under-complete) 的盲信号分离, 欠定盲信号分离又称超完备 (over-complete) 的盲信号分离。

Cao等人<sup>[18]</sup>以及Li等人<sup>[106]</sup>的研究表明, 为确保所有的源信号都是可以提取的, 盲信号分离问题要求混合矩阵是列满秩的, 换句话说, 观测信号的个数应该不少于源信号的个数。对于观测信号比源信号个数少的欠定情况<sup>[9,18,89,99,100,104~106]</sup>, 因为只有部分信号是可以提取的, 而当前尚不存在一种并行算法, 能够实现这些可提取的信号同步分离 (注: 但确实存在串行算法, 可以将这些信号依次提取出来, 感兴趣的读者可参阅文献[18]和[106]等), 因此, 盲信号分离一般只讨论前两种情况。

当观测信号不少于源信号的个数时, 一个很现实的问题是源信号个数的有效确定。为简单起见, 多数盲信号分离文献假定源信号的个数是已知的, 并且与接收天线的个数相等, 即通常只讨论完备的盲信号分离问题。然而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 (比如在无线通信系统中, 各小区随时会有手机用户的加入, 也随时会有手机用户的撤出), 从而混合矩阵是非方阵和不可逆的。因此, 超定盲信号分离问题的研究具有重大的现实意义。

超定盲信号分离问题的研究最早是由Zhang、Amari和Choi等人<sup>[9,28,36,164]</sup>发起的。特别地, Zhang等人<sup>[164]</sup>从李群 (Lie group) 和Stiefel流形出发, 提出了一种超定盲信号分离的对比函数, 并推导了相应的自然梯度算法。

与他们的不同, 本章第2节从分离矩阵的奇异值分解出发, 提出一种新的基于独立分量分析的超定盲信号分离对比函数, 并推导一般梯度学习算法。然后, 借助于相对梯度<sup>[22]</sup>证明, 完备盲信号分离的自然梯度算法同样适合于超定盲信号分离问题, 并用大量的计算机仿真试验证实了这一点。第3节介绍一种Ye等人<sup>[157]</sup>提出的适用于无噪声情况的源信号个数的有效确定方法, 并结合第2节的研究成果, 推导源信号

个数未知甚至动态变化情况下的盲信号分离算法。第4节给出混合矩阵列满秩这一盲信号分离基本假设的理论依据，并简单叙述了病态盲信号分离问题。最后，第5节是本章的小结部分。

## §6.2 超定盲信号分离

考虑不含噪声的线性盲信号分离模型：

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \quad t = 1, 2, \dots \quad (6.1)$$

其中  $\mathbf{x}_t$  和  $\mathbf{s}_t$  分别表示  $m$  维的观测信号向量和  $n$  维的源信号向量。与前面几章假定混合矩阵  $\mathbf{A}$  是非奇异方阵不同，这里只约定  $\mathbf{A}$  是  $m \times n$  的列满秩矩阵， $m \geq n$ 。

为简单起见，暂时假设源信号的数目  $n$  是已知的。这样，盲信号分离的目的是确定一个  $n \times m$  的行满秩矩阵  $\mathbf{B}$ ，使得系统输出

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t \quad (6.2)$$

是源信号向量  $\mathbf{s}_t$  的一个拷贝。

### 6.2.1 对比函数

因为输出信号向量  $\mathbf{y}_t$  是源信号  $\mathbf{s}_t$  各分量的一个线性组合，而  $\mathbf{s}_t$  的分量相互独立，因此若  $\mathbf{y}_t$  的分量也彼此独立，则  $\mathbf{y}_t$  必定是  $\mathbf{s}_t$  的一个拷贝。这正是独立分量分析的核心思想。显然，衡量输出  $\mathbf{y}_t$  各分量是否相互独立的最直接的目标函数是  $\mathbf{y}_t$  的互信息。利用互信息与微分熵<sup>[49]</sup>的关系，有下面的对比函数<sup>[48]</sup>：

$$J(\mathbf{y}_t; \mathbf{B}) = -I(y_{1,t}, \dots, y_{n,t}; \mathbf{B}) = -\sum_{i=1}^n H(y_{i,t}; \mathbf{B}) + H(\mathbf{y}_t; \mathbf{B}) \quad (6.3)$$

其中

$$H(\mathbf{y}_t; \mathbf{B}) = -\int p(\mathbf{y}_t; \mathbf{B}) \ln p(\mathbf{y}_t; \mathbf{B}) d\mathbf{y}_t \quad (6.4)$$

$$H(y_{i,t}; \mathbf{B}) = -\int p_i(y_{i,t}; \mathbf{B}) \ln p_i(y_{i,t}; \mathbf{B}) dy_{i,t} \quad (6.5)$$

分别表示  $\mathbf{y}_t$  的联合微分熵和边缘微分熵。

由矩阵的奇异值分解<sup>[176]</sup>知，对于  $n \times m$  的行满秩矩阵  $\mathbf{B}$ ，一定存在  $n \times n$  的正交矩阵  $\mathbf{U}$  和  $m \times m$  的正交矩阵  $\mathbf{V}$ ，使得

$$\mathbf{B} = \mathbf{U}[\Lambda : \mathbf{O}] \mathbf{V}^T \quad (6.6)$$

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

将  $\mathbf{V}$  分块, 并记前  $n$  列构成的子矩阵为  $\mathbf{V}_1$ , 后  $m-n$  列构成的子矩阵为  $\mathbf{V}_2$ , 则

$$\mathbf{y}_i = \mathbf{B} \mathbf{x}_i = \mathbf{U}[\Lambda : \mathbf{O}] \mathbf{V}^T \mathbf{x}_i = \mathbf{U} \Lambda \mathbf{V}_1^T \mathbf{x}_i \quad (6.7)$$

由上式以及微分熵的定义 (6.4), 很容易证明

$$H(\mathbf{y}_i; \mathbf{B}) = \ln |\det(\mathbf{U} \Lambda)| + H(\mathbf{V}_1^T \mathbf{x}_i) \quad (6.8)$$

对于行满秩矩阵  $\mathbf{B}$ , 由矩阵奇异值分解<sup>[176]</sup>的性质知,

$$\det(\Lambda) = \sqrt{\det(\mathbf{B} \mathbf{B}^T)} \quad (6.9)$$

将 (6.8) ~ (6.9) 代入到式 (6.3), 得对比函数

$$J(\mathbf{y}_i; \mathbf{B}) = -\sum_{i=1}^n H(\mathbf{y}_{i,i}; \mathbf{B}) + \frac{1}{2} \ln |\det(\mathbf{B} \mathbf{B}^T)| + H(\mathbf{V}_1^T \mathbf{x}_i) \quad (6.10)$$

根据熵与互信息的关系以及微分熵对正交变换的不变性<sup>[49]</sup>, 我们有

$$H(\mathbf{V}_1^T \mathbf{x}_i) = H(\mathbf{x}_i) + I(\mathbf{V}_1^T \mathbf{x}_i; \mathbf{V}_2^T \mathbf{x}_i) - H(\mathbf{V}_2^T \mathbf{x}_i) \quad (6.11)$$

因为分离矩阵  $\mathbf{B} = \mathbf{U}[\Lambda : \mathbf{O}] \mathbf{V}^T = \mathbf{U} \Lambda \mathbf{V}_1^T + \mathbf{U} \mathbf{O} \mathbf{V}_2^T = \mathbf{U} \Lambda \mathbf{V}_1^T$ , 所以  $\mathbf{B}$  不依赖于子矩阵  $\mathbf{V}_2$ , 从而可以任意选择  $\mathbf{V}_2$ , 使得  $H(\mathbf{V}_1^T \mathbf{x}_i)$  不依赖于分离矩阵  $\mathbf{B}$ 。具体地, 对于无噪声的线性混合模型 (6.1), 我们选择  $\mathbf{V}_2$  为  $\mathbf{R}_x = \mathbf{A} E\{\mathbf{s}_i \mathbf{s}_i^T\} \mathbf{A}^T = \mathbf{A} \mathbf{A}^T$  的零空间的标准正交基构成的矩阵, 则  $\mathbf{A} \mathbf{A}^T \mathbf{V}_2 = \mathbf{0}$ , 两端同时左乘以列满秩矩阵  $\mathbf{A}$  的广义逆矩阵  $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  有  $\mathbf{A}^T \mathbf{V}_2 = \mathbf{0}$ , 进而  $\mathbf{V}_2^T \mathbf{x}_i = \mathbf{V}_2^T \mathbf{A} \mathbf{s}_i = \mathbf{0}$ , 显然  $H(\mathbf{V}_1^T \mathbf{x}_i) = H(\mathbf{x}_i)$  不依赖于分离矩阵  $\mathbf{B}$ 。对于观测数据含噪声的情况, 如果噪声功率足够低, 能确保  $\mathbf{V}_1$  张成观测数据的信号子空间,  $\mathbf{V}_2$  张成噪声子空间, 从而  $I(\mathbf{V}_1^T \mathbf{x}_i; \mathbf{V}_2^T \mathbf{x}_i) = I(\mathbf{V}_1^T \mathbf{s}_i; \mathbf{V}_2^T \mathbf{n}_i) = 0$ , 则  $H(\mathbf{V}_1^T \mathbf{x}_i) = H(\mathbf{x}_i) - H(\mathbf{V}_2^T \mathbf{x}_i)$  也不依赖于  $\mathbf{B}$ 。

因为  $H(\mathbf{V}_1^T \mathbf{x}_i)$  不是分离矩阵  $\mathbf{B}$  的函数, 因此可以从 (6.10) 中省去, 从而有下面的超定盲信号分离的对比函数:

$$J(\mathbf{y}_i; \mathbf{B}) = -\sum_{i=1}^n H(\mathbf{y}_{i,i}; \mathbf{B}) + \frac{1}{2} \ln |\det(\mathbf{B} \mathbf{B}^T)| \quad (6.12)$$

显然, 通常的完备盲信号分离的对比函数

$$J(\mathbf{y}_i; \mathbf{B}) = -\sum_{i=1}^n H(y_{i,j}; \mathbf{B}) + \ln |\det(\mathbf{B})| \quad (6.13)$$

是 (6.12) 在分离矩阵  $\mathbf{B}$  取可逆方阵情况下的一个特例。

此外, Zhang 等人<sup>[164]</sup>提出了对比函数

$$J(\mathbf{y}_i; \mathbf{B}) = -\sum_{i=1}^n H(y_{i,j}; \mathbf{B}) + \ln |\det(\mathbf{B}\mathbf{E}^T)| \quad (6.14)$$

其中  $\mathbf{E}$  表示李群  $Gl(n, m)$  下的一个恒等算子。因为  $\ln |\det(\mathbf{B}\mathbf{E}^T)| = \frac{1}{2} \ln |\det(\mathbf{B}\mathbf{B}^T)|$ , 所以两种对比函数 (6.12) 和 (6.14) 实际是等价的。

### 6.2.2 算法推导

以对比函数 (6.12) 为出发点, 下面我们推导分离矩阵  $\mathbf{B}$  的自适应更新, 即梯度学习算法。对比函数  $J(\mathbf{y}_i; \mathbf{B})$  的全微分

$$dJ(\mathbf{y}_i; \mathbf{B}) = J(\mathbf{y}_i; \mathbf{B} + \boldsymbol{\varepsilon}) - J(\mathbf{y}_i; \mathbf{B}) = \sum_{i=1}^n \sum_{j=1}^m \frac{\partial}{\partial \mathbf{B}_{ij}} J(\mathbf{y}_i; \mathbf{B}) d\mathbf{B}_{ij} \quad (6.15)$$

其中  $\boldsymbol{\varepsilon}$  表示分离矩阵  $\mathbf{B}$  的微小扰动矩阵, 而  $\mathbf{B}_{ij}$  表示  $\mathbf{B}$  的第  $i$  行第  $j$  列元素。

利用性质 (3.16), 有

$$\frac{\partial}{\partial \mathbf{W}_{ij}} \ln |\det(\mathbf{W})| = [\mathbf{W}^{-1}]_{ji}$$

因此

$$\begin{aligned} \frac{\partial}{\partial \mathbf{B}_{ij}} \ln |\det(\mathbf{B}\mathbf{B}^T)| &= \sum_{p,q=1}^n \frac{\partial}{\partial [\mathbf{B}\mathbf{B}^T]_{pq}} \ln |\det(\mathbf{B}\mathbf{B}^T)| \cdot \frac{\partial}{\partial \mathbf{B}_{ij}} [\mathbf{B}\mathbf{B}^T]_{pq} \\ &= \sum_{p,q=1}^n \left[ (\mathbf{B}\mathbf{B}^T)^{-1} \right]_{pq} \cdot \frac{\partial}{\partial \mathbf{B}_{ij}} \sum_{l=1}^m \mathbf{B}_{pl} \mathbf{B}_{ql} \\ &= \sum_{p,q=1}^n \left[ (\mathbf{B}\mathbf{B}^T)^{-1} \right]_{pq} \cdot (\mathbf{B}_{qj} \delta_{pi} + \mathbf{B}_{pj} \delta_{qi}) \\ &= 2 \left[ (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} \right]_{ij} \end{aligned} \quad (6.16)$$

此外,

$$\begin{aligned}\frac{\partial}{\partial \mathbf{B}_{ij}} \sum_{l=1}^n H(y_{l,i}; \mathbf{B}) &= -\frac{\partial}{\partial \mathbf{B}_{ij}} \sum_{l=1}^n E \left\{ \ln |p_l(y_{l,i}; \mathbf{B})| \right\} \\ &= -E \left\{ \frac{p'_l(y_{l,i}; \mathbf{B})}{p_l(y_{l,i}; \mathbf{B})} x_{j,l} \right\}\end{aligned}\quad (6.17)$$

其中  $p'_l$  表示  $p_l$  的一阶导数。记  $\varphi_i = -p'_i/p_i$  ( $i=1, \dots, n$ ) 表示分值函数, 则有

$$dJ(\mathbf{y}_i; \mathbf{B}) = \sum_{i=1}^n \sum_{j=1}^m \left( -E \{ \varphi_i x_{j,i} \} + \left[ (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} \right]_{ij} \right) d\mathbf{B}_{ij} \quad (6.18)$$

从而对比函数 (6.12) 的一般梯度等于

$$\nabla J(\mathbf{y}_i; \mathbf{B}) = -E \{ \Phi(\mathbf{y}_i) \mathbf{x}_i^T \} + (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} \quad (6.19)$$

式中列矢量  $\Phi(\mathbf{y}_i) = [\varphi_1, \dots, \varphi_n]^T$ 。分离矩阵  $\mathbf{B}$  的随机一般梯度更新算法为

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i \cdot \left[ (\mathbf{B}_i \mathbf{B}_i^T)^{-1} \mathbf{B}_i - \Phi(\mathbf{y}_i) \mathbf{x}_i^T \right] \quad (6.20)$$

其中  $\eta_i > 0$  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<sup>[14,89,153]</sup>是式 (6.20) 在行满秩矩阵  $\mathbf{B}$ , 取非奇异方阵情况下的特例。

如果在微分公式 (6.15) 中, 分离矩阵取相对扰动值  $\varepsilon \mathbf{B}$  而非绝对扰动  $\varepsilon$ , 则可以获得对比函数的相对梯度  $\hat{\nabla} J(\mathbf{y}_i; \mathbf{B})$ 。Cardoso 和 Laheld<sup>[22]</sup>给出了相对梯度和一般梯度的关系式:

$$\hat{\nabla} J(\mathbf{y}_i; \mathbf{B}) = \nabla J(\mathbf{y}_i; \mathbf{B}) \cdot \mathbf{B}^T \quad (6.21)$$

证明  $\hat{\nabla} J(\mathbf{y}_i; \mathbf{B}_i) \mathbf{B}_i$  对应于对比函数的上升方向, 并提出下面的相对梯度学习算法,

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i \hat{\nabla} J(\mathbf{y}_i; \mathbf{B}_i) \mathbf{B}_i \quad (6.22)$$

将式 (6.19) 和 (6.21) 代入到上式, 不难得到超定盲信号分离的相对梯度算法:

$$\mathbf{B}_{i+1} = \mathbf{B}_i + \eta_i \left[ \mathbf{I} - \Phi(\mathbf{y}_i) \mathbf{y}_i^T \right] \mathbf{B}_i \quad (6.23)$$

显然, 它与通常的完备盲信号分离的自然梯度算法<sup>[7,89,153]</sup>具有相同的形式。但不同的是, 式 (6.23) 中的分离矩阵  $\mathbf{B}_i$  是行满秩的, 而不再要求是非奇异方阵。

我们注意到, 算法 (6.23) 是 Zhang 等人<sup>[164]</sup>借助于李群的概念首先得到的, 但区别在于:

- 1) 对比函数 (6.12) 是从分离矩阵的奇异值分解出发, 并经过严格推导得出的,



而Zhang等人以直观的方式提出对比函数 (6.14)，没有严格证明；

2) Zhang等人推导算法 (6.23) 要借助于李群和Stiefel流形的概念，而我们的方法不需要。相比之下，我们的方法也更容易理解。

与一般梯度算法 (6.20) 相比，自然梯度算法 (6.23) 具有不需要矩阵求逆，满足串行更新 (serial updating) 规则，提供盲信号分离的均匀收敛性等显著特点外，另一个突出优点在于它满足分离矩阵的行满秩要求，换句话说，如果初始分离矩阵行满秩，则当前时刻的分离矩阵也必定是行满秩的。这一特点是算法稳定工作的前提条件，证明如下。

设  $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^T \mathbf{Y}) = \sum_{i,j} \mathbf{X}_{ij} \mathbf{Y}_{ij}$ ， $|\mathbf{B}\mathbf{B}^T| = \det(\mathbf{B}\mathbf{B}^T)$ ，类似于文献[153, 156, 178]，

不难证明

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

其中  $\mathbf{F}(\mathbf{y}_t) = \mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T$ 。因为

$$|\mathbf{B}_t \mathbf{B}_t^T| = |\mathbf{B}_0 \mathbf{B}_0^T| \exp \left\{ 2 \int_0^t \eta(\xi) \text{trace}[\mathbf{F}(\mathbf{y}(\xi))] d\xi \right\}$$

所以若初始分离矩阵  $\mathbf{B}_0$  行满秩，则分离矩阵  $\mathbf{B}_t$  一定是行满秩的。证明完毕。 ■

### 6.2.3 性能仿真

为了验证超定盲信号分离学习算法 (6.23) 的有效性，我们对以下信号<sup>[3, 153, 156]</sup>进行了仿真试验：

- S1: 符号信号  $\text{sgn}(\cos(2\pi 155t))$ ;
- S2: 高频正弦信号  $\sin(2\pi 800t)$ ;
- S3: 低频正弦信号  $\sin(2\pi 90t)$ ;
- S4: 相位调制信号  $\sin(2\pi 300t - 6 \cos(2\pi 60t))$ ;
- S5: 幅度调制信号  $\sin(2\pi 9t) \sin(2\pi 300t)$ ;
- S6: 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 $s_i$ 。图6.1是某一次运行中的六个源信号的波形图（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<sup>[3,78,153,156]</sup>

$$E_{cl} = \sum_{p=1}^n \left( \sum_{q=1}^n \frac{|c_{pq}|}{\max_l |c_{pl}|} - 1 \right) + \sum_{q=1}^n \left( \sum_{p=1}^n \frac{|c_{pq}|}{\max_l |c_{lq}|} - 1 \right) \quad (6.24)$$

来衡量，其中 $\mathbf{C}_l = [c_{pq}] = \mathbf{B}_l \mathbf{A}$ 表示混合—分离合成系统的传递矩阵。

在算法(6.23)中取激励函数 $\varphi_i = y_{i,j}^3$ ，固定步长 $\eta_i = 40 \times T$ （ $T = 10^{-4}$ 为采样周期），初始分离矩阵 $\mathbf{B}_0$ 为 $6 \times 9$ 的随机矩阵，其元素在 $[-1,1]$ 区间均匀分布。图6.2给出了200次独立运行（每次运行中的随机信号S6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(6.23)性能的影响，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6.2中给出。图6.3至图6.6给出了其中某一次运行的“串音”误差及各自的分离结果。

经过比较不难看出：在无噪声或者噪声足够低的环境下，算法(6.23)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，从而对比函数的最大化会导致提取噪声和部分信号被抑制。

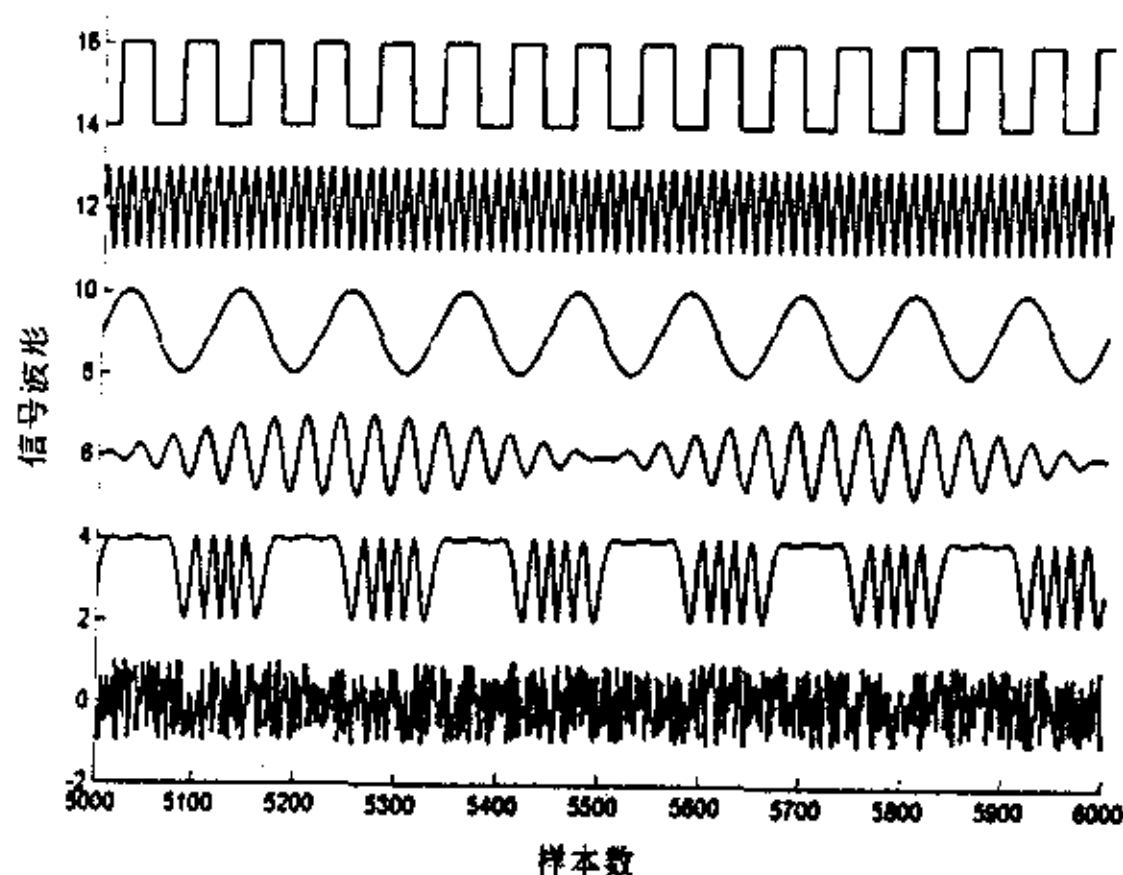


图 6.1 源信号波形图

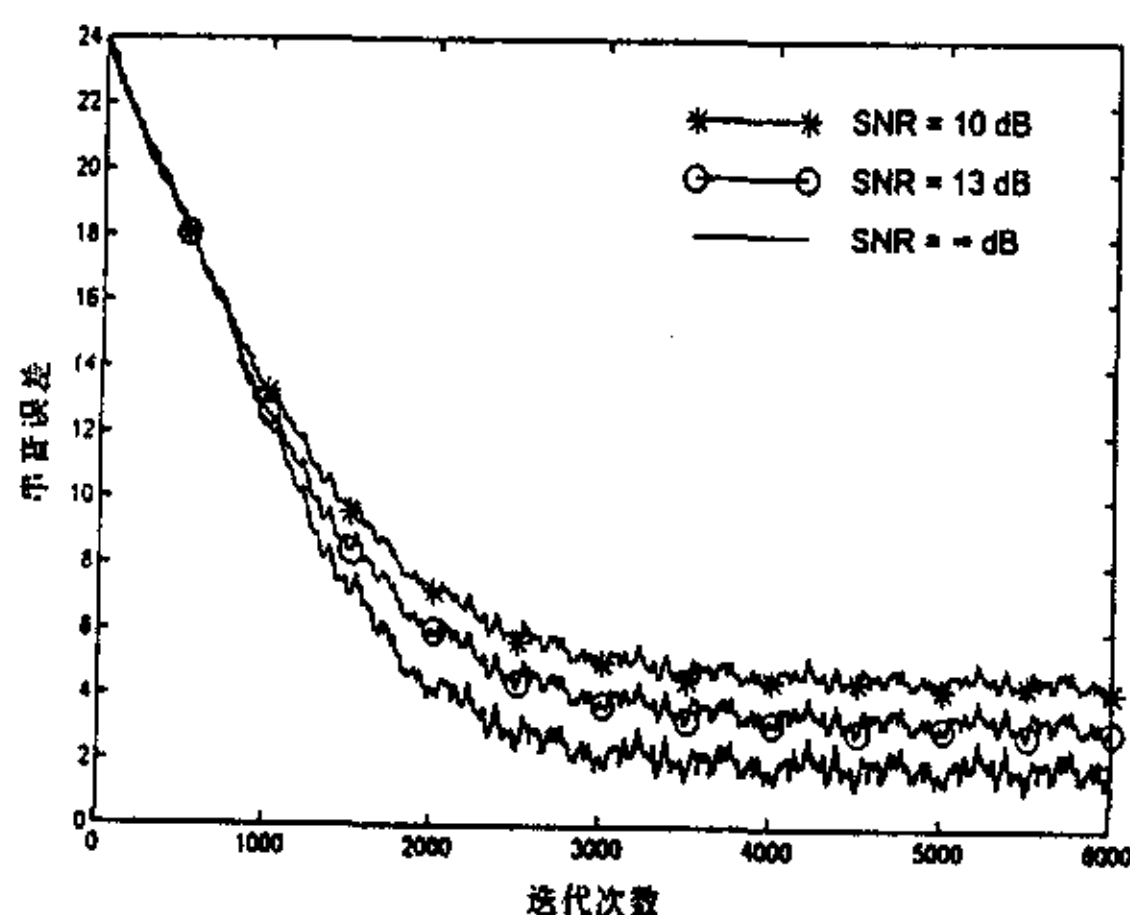


图 6.2 200次独立运行串音误差均值图

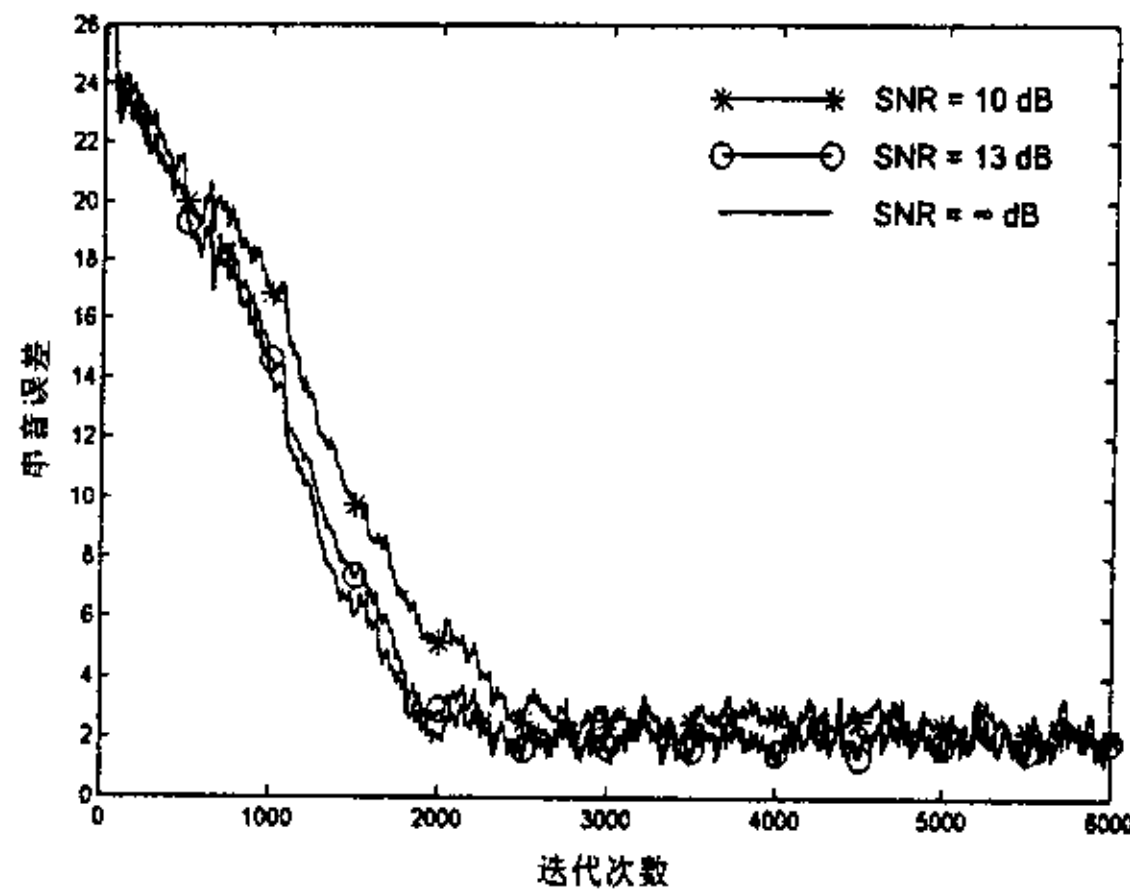
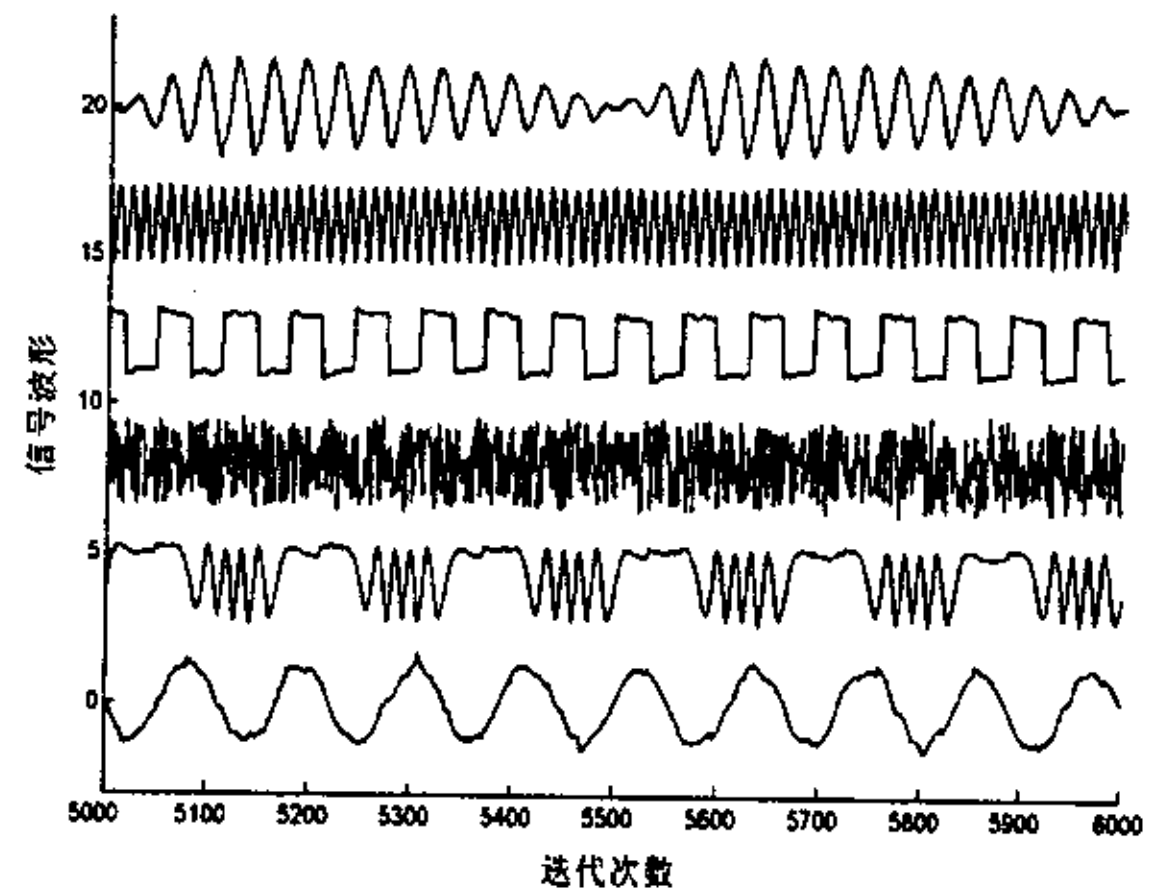
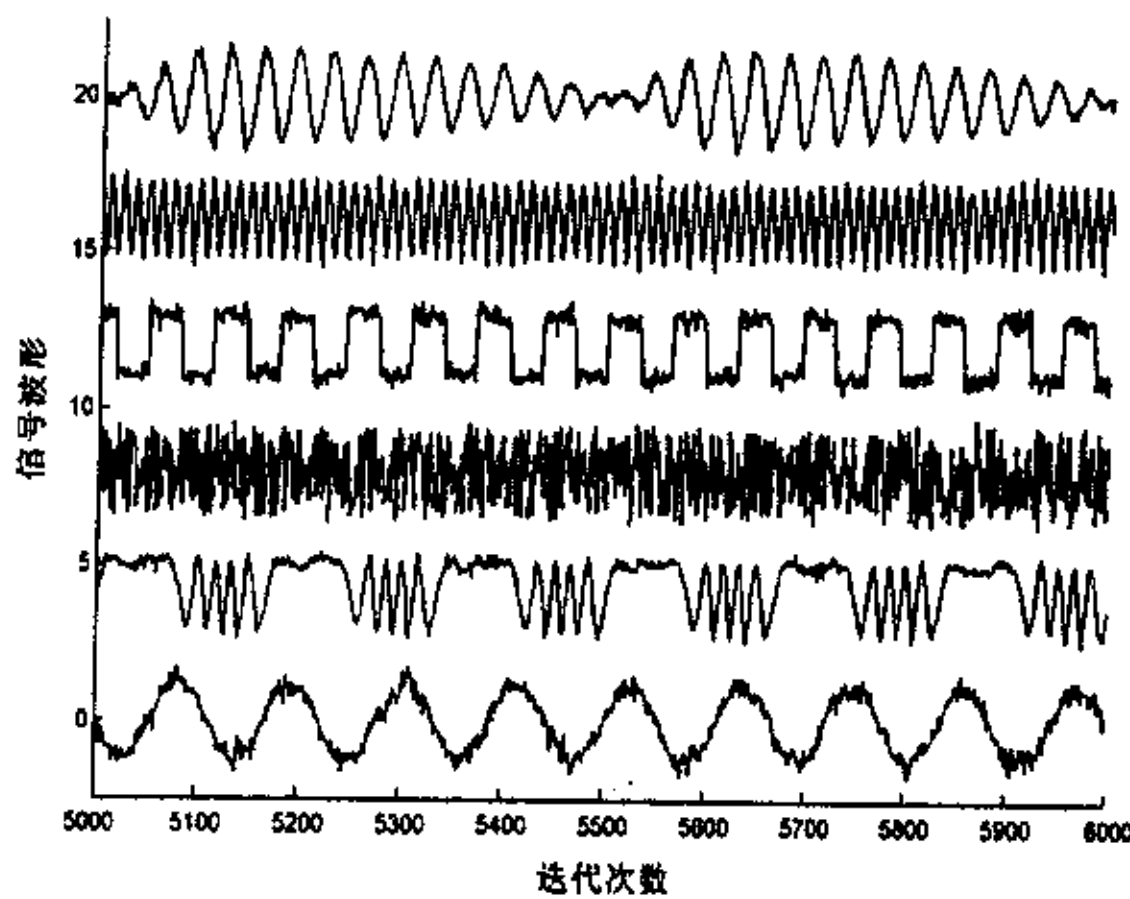
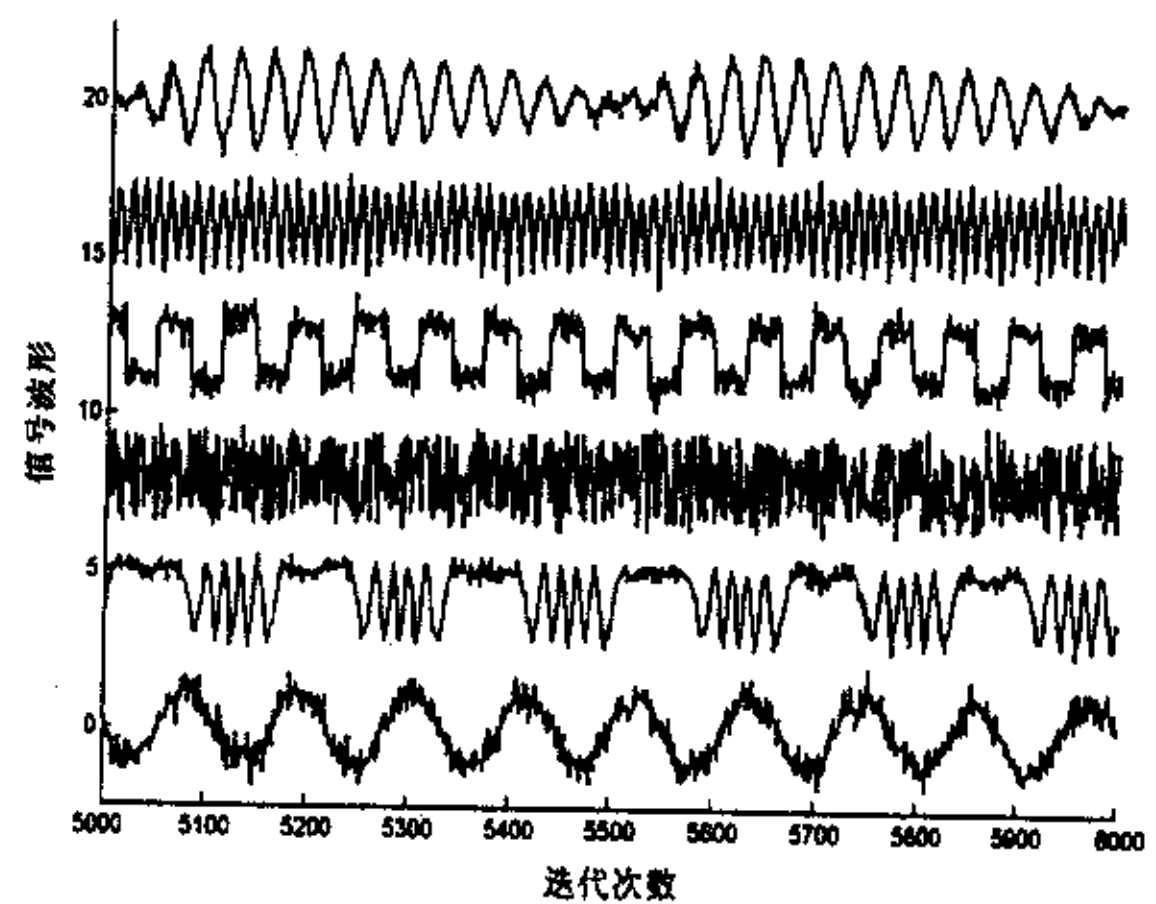


图 6.3 某一次运行串音误差曲线图

图 6.4 无噪声分离结果,  $\text{SNR}=\infty$ 图 6.5 分离结果,  $\text{SNR}=13 \text{ dB}$ 图 6.6 分离结果,  $\text{SNR}=10 \text{ dB}$ 

### §6.3 未知源个数的盲信号分离

本章研究观测信号个数比源信号个数多的超定盲信号分离问题。上一节具体研究了源信号个数已知情况下的超定盲信号分离对比函数,并推导了相应的自然梯度学习算法。下面,我们讨论源信号个数未知甚至动态变化(仍然满足观测信号的个数不少于源信号的个数这一前提条件)的一般情况。

### 6.3.1 算法实现

因为源信号的个数  $n$  即混合矩阵  $\mathbf{A}$  的列数未知, 我们不可能像上一节那样, 设计一个  $n \times m$  的行满秩矩阵, 使得网络输出是源信号向量的某个拷贝。那么, 这种情况下的盲信号分离问题如何求解呢? 通常的做法是先对观测数据进行白化, 并根据观测样本的自相关函数矩阵估计源信号的个数, 然后采用某种算法, 如 Douglas<sup>[59]</sup> 的自稳定算法、Karhunen 等人<sup>[92,96,122]</sup> 的非线性 PCA 算法、Pajunen 等人<sup>[124]</sup> 的基于一般梯度的 RLS 算法以及作者新近提出的基于自然梯度的 RLS 算法<sup>[167]</sup> 等, 来确定一个正交矩阵从而实现盲信号分离。因为自相关函数的估计需要一定数量的观测样本, 否则不能保证估计的精度, 所以这种方法实时性差, 难以应付源信号个数动态变化的复杂情况。

为克服上述方法实时性差的缺点, 我们用一个  $m \times m$  的 ( $m$  代表观测信号的个数, 它是已知的) 非奇异矩阵  $\mathbf{B}$  (仍称为分离矩阵) 作用于观测数据向量, 使系统输出

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t \quad (6.25)$$

含希望恢复的源信号向量。因为在观测信号无噪声情况下,

$$\mathbf{y}_t = \mathbf{B}\mathbf{A} \cdot \mathbf{s}_t \quad (6.26)$$

的各分量是源信号  $\mathbf{s}_t$  各分量的线性组合, 因此当  $m > n$  时, 可以通过对  $\mathbf{B}$  的初等行变换将  $\mathbf{y}_t$  中除源信号外的其它  $m - n$  个分量置零, 从而将分离矩阵  $\mathbf{B}$  分成两个部分。一部分是混合矩阵  $\mathbf{A}$  的广义逆矩阵与广义交换矩阵的乘积, 它对应的输出向量是源信号向量的某个拷贝, 另一部分是  $\mathbf{A}$  的转置的零空间与广义交换矩阵的乘积, 它对应于  $m - n$  个零信号输出分量。

记  $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  表示列满秩矩阵  $\mathbf{A}$  的广义逆矩阵,  $\mathcal{N}(\mathbf{A}^T)$  表示  $\mathbf{A}$  的转置的零空间, 根据以上分析, 有<sup>[157]</sup>

$$\mathbf{B} = \mathbf{D}\mathbf{P} \cdot \left[ (\mathbf{A}^+)^T, \mathcal{N}(\mathbf{A}^T) \right]^T \quad (6.27)$$

其中  $\mathbf{D}$  表示非奇异对角矩阵,  $\mathbf{P}$  表示交换矩阵, 两者的乘积构成广义交换矩阵以描述盲信号分离问题的信号幅度不确定性和信号排序不确定性。

因为第一部分即混合矩阵  $\mathbf{A}$  的广义逆可采用上一节介绍的自然梯度算法 (6.23) 自适应获得, 因此源信号个数未知情况下的盲信号分离问题就转化为  $\mathcal{N}(\mathbf{A}^T)$  的确定问题 (它包含了源信号个数的有效确定问题)。

**定理 6.1<sup>[157]</sup>** 若  $m \times n$  的混合矩阵  $\mathbf{A}$  列满秩, 源信号向量  $\mathbf{s}_t$  是各态历经的平稳随机过程, 则  $\mathbf{A}^T$  与  $m$  个观测信号向量  $\{\mathbf{x}_t = \mathbf{A}\mathbf{s}_t | t = 1, \dots, m\}$  构成的矩阵  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_m]^T$  具

有相同的零空间。

证明 盲信号分离中的各个源信号分量相互独立,从而彼此之间不存在任何函数关系,又因为  $\mathbf{s}_l$  是各态历经的平稳随机过程,所以  $m \times n$  的矩阵  $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_m]^T$  是列满秩的。

对任意  $m$  维的列向量  $\mathbf{b} = [b_1, \dots, b_m]^T$ , 若  $\mathbf{b} \in \mathcal{N}(\mathbf{A}^T)$ , 则  $\mathbf{A}^T \mathbf{b} = \mathbf{0}$ , 进而有  $\mathbf{Xb} = [\mathbf{x}_1, \dots, \mathbf{x}_m]^T \mathbf{b} = [\mathbf{As}_1, \dots, \mathbf{As}_m]^T \mathbf{b} = \mathbf{SA}^T \mathbf{b} = \mathbf{0}$ , 因此  $\mathbf{b}$  属于矩阵  $\mathbf{X}$  的零空间, 即  $\mathbf{b} \in \mathcal{N}(\mathbf{X})$ 。

另一方面,  $\forall \mathbf{b} \in \mathcal{N}(\mathbf{X})$ , 有  $\mathbf{Xb} = [\mathbf{x}_1, \dots, \mathbf{x}_m]^T \mathbf{b} = [\mathbf{As}_1, \dots, \mathbf{As}_m]^T \cdot \mathbf{b} = \mathbf{SA}^T \mathbf{b} = \mathbf{0}$ 。因为  $\mathbf{S}$  是列满秩矩阵, 所以在  $\mathbf{SA}^T \mathbf{b} = \mathbf{0}$  的两端同时左乘以  $\mathbf{S}$  的广义逆矩阵  $\mathbf{S} = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T$  有  $\mathbf{A}^T \mathbf{b} = \mathbf{0}$ , 故  $\mathbf{b} \in \mathcal{N}(\mathbf{A}^T)$ 。 ■

根据上面的定理, 如果各个源信号分量不仅相互独立, 而且是各态历经的平稳随机过程, 则对于无噪声的线性盲信号分离模型, 我们可以从  $m$  个观测信号样本得到  $\mathbf{A}^T$  的零空间, 并由此估计出源信号的个数  $n$ 。由于用这种方法估计源信号的个数需要的样本很少 (因为  $m$  即是观测信号的个数, 通常很小), 从而这种方法对源信号个数动态变化时的盲信号分离问题也同样适用。

### 6.3.2 性能仿真

为验证定理6.1, 也为了进一步说明超定盲信号分离算法 (6.23) 的有效性, 我们对以下五个源信号进行计算机仿真 (不考虑高频正弦信号是因为它在波形上接近于均匀分布的噪声信号):

- S1: 符号信号  $\text{sgn}(\cos(2\pi 155t))$ ;
- S2: 低频正弦信号  $\sin(2\pi 90t)$ ;
- S3: 相位调制信号  $\sin(2\pi 300t - 6 \cos(2\pi 60t))$ ;
- S4: 幅度调制信号  $\sin(2\pi 9t) \sin(2\pi 300t)$ ;
- S5: 在  $[-1, 1]$  均匀分布的随机噪声信号。

为作图时清晰起见, 只采用5个接收天线, 混合矩阵  $\mathbf{A}$  的元素为在  $[-1, 1]$  区间均匀分布的随机数。以10kHz的速率对接收天线输出采样以产生离散观测信号  $\mathbf{x}_l$ 。算法性能用归一化的“串音”误差<sup>[3,78,153,156]</sup>

$$E_{ci} = \frac{1}{m} \sum_{p=1}^m \left( \sum_{q=1}^n \frac{|c_{pq}|}{\max_l |c_{pl}|} - 1 \right) + \frac{1}{n} \sum_{q=1}^n \left( \sum_{p=1}^m \frac{|c_{pq}|}{\max_l |c_{lq}|} - 1 \right) \quad (6.28)$$



来衡量, 其中  $C_i = [c_{pq}] = B_i A$  表示整个混合-分离系统的传递矩阵。

在Matlab语言环境下进行计算机仿真, 并用其提供的库函数null.m求解混合矩阵  $A$  的转置的零空间  $\mathcal{N}(A^T)$ , 根据定理6.1, 它实际就等于  $m$  个观测信号向量构成的矩阵  $X = [x_1, \dots, x_m]^T$  的零空间, 并由此估计进而跟踪源信号的个数  $n$ 。采用算法(6.23)学习混合矩阵  $A$  的广义逆矩阵  $A^+$ , 初始分离矩阵  $B_0 = 0.5 \times I$ , 激励函数采用Yang和Amari提出的自适应非线性函数<sup>[153]</sup>, 即

$$\varphi_i(y_{i,l}) = f_i(\kappa_{3i}, \kappa_{4i}) y_{i,l}^2 + g_i(\kappa_{3i}, \kappa_{4i}) y_{i,l}^3 \quad (6.29)$$

式中  $\kappa_{3i} = E\{y_{i,l}^3\}$  表示信号  $y_{i,l}$  的偏度(skewness),  $\kappa_{4i} = E\{y_{i,l}^4\} - 3$  表示信号的(未归一化的)峰度(kurtosis), 而

$$f_i(\kappa_{3i}, \kappa_{4i}) = -\frac{1}{2}\kappa_{3i} + \frac{9}{4}\kappa_{3i}\kappa_{4i} \quad (6.30)$$

$$g_i(\kappa_{3i}, \kappa_{4i}) = -\frac{1}{6}\kappa_{4i} + \frac{3}{2}\kappa_{3i}^2 + \frac{3}{4}\kappa_{4i}^2 \quad (6.31)$$

偏度和峰度用下面的公式更新:

$$\kappa_{3i,l+1} = \kappa_{3i,l} - \mu_l \cdot (\kappa_{3i,l} - y_{i,l}^3) \quad (6.32)$$

$$\kappa_{4i,l+1} = \kappa_{4i,l} - \mu_l \cdot (\kappa_{4i,l} - y_{i,l}^4 + 3) \quad (6.33)$$

式中  $\mu_l$  为偏度和峰度更新的步长参数。仿真中取  $\mu_l = 0.004$ 。

算法(6.23)中的学习速率取时变函数<sup>[156]</sup>:

$$\eta_l = \begin{cases} \eta_0 & , \quad t \leq T_0 \\ \eta_0 \exp[-T_d(t - T_0)] & , \quad t > T_0 \end{cases} \quad (6.34)$$

式中  $\eta_0$ ,  $T_0$  及  $T_d$  是适当选择的常数。仿真中分别取  $\eta_0 = 0.01$ ,  $T_0 = 400$  和  $T_d = 0.0015$ 。

考虑源信号个数未知且动态变化情况下的盲信号分离(源信号个数未知却固定不变是它的一种特殊情况)。仿真使用了6000个样本。其中, 在前2000个样本源信号由符号信号S1、相位调制信号S3和随机噪声信号S5组成; 从第2001个样本开始增加两个源信号: 正弦信号S2和幅度调制信号S4; 而从第4001个样本开始撤出相位调制信号S3。图6.7给出了某一次运行中各时段的源信号波形图。独立运行200次(每次运行中的噪声信号S5和混合矩阵  $A$  都是随机产生的), 得到的平均串音误差曲线见图6.8。图6.9和图6.10分别给出其中某一次运行的性能曲线与对应的分离结果。

我们强调, 每增加一个源信号, 就相当于混合矩阵增加了一列, 该新增列的元素

也随机产生,而减少一个源信号,则混合矩阵中与之对应的列被删除。所有这些情况,包括混合矩阵  $A$ 、源信号、源信号的个数  $n$  及其动态变化情况对算法都是未知的(算法仅利用观测信号样本及其维数  $m$  进行信号分离)。

由图6.7至图6.10,不难看出:1) 对于混合矩阵列满秩,源信号是各态历经的平稳随机过程条件下的盲信号分离,根据定理6.1只需要少量样本就能估计出源信号的个数进而检测出其动态变化情况;2) 算法(6.23)能有效地实现超定盲信号分离,它对应的输出分量是我们期望提取的源信号,除此之外的输出分量是零信号。

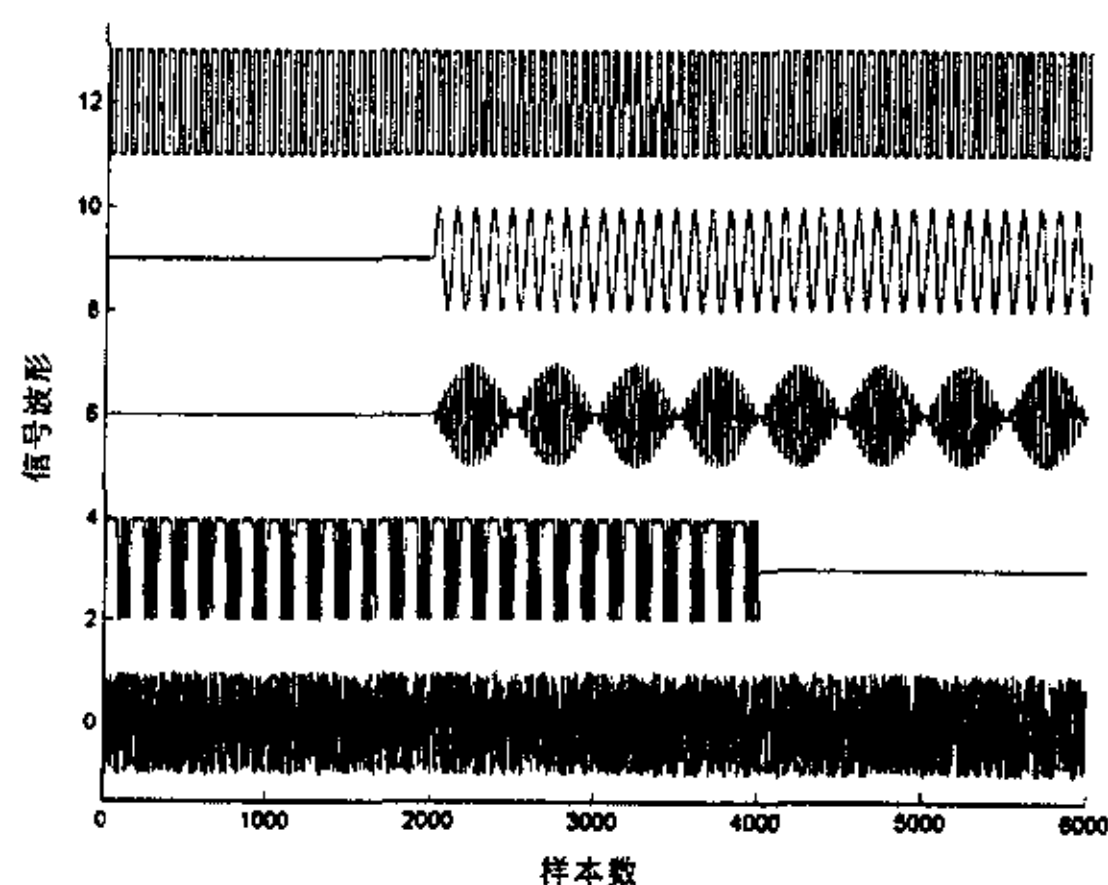


图 6.7 源信号波形图

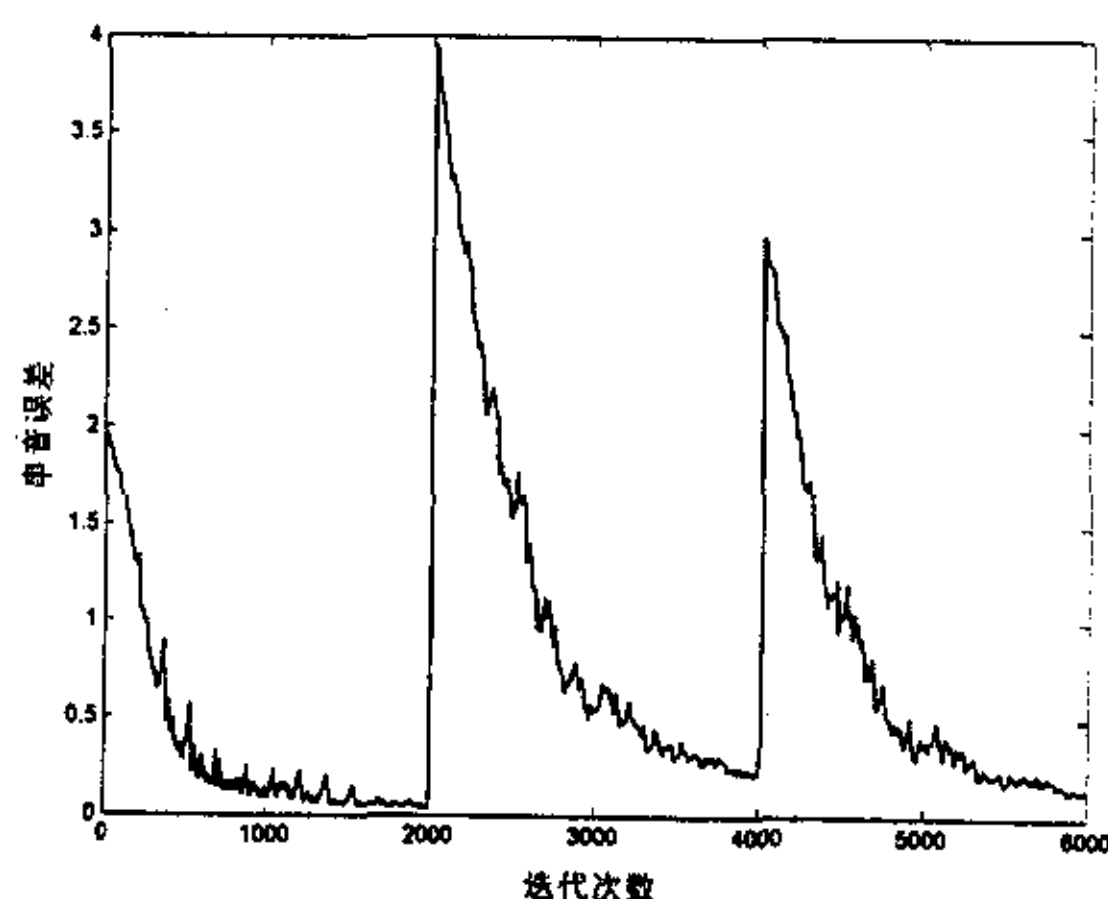


图6.8 200次独立运行串音误差均值图

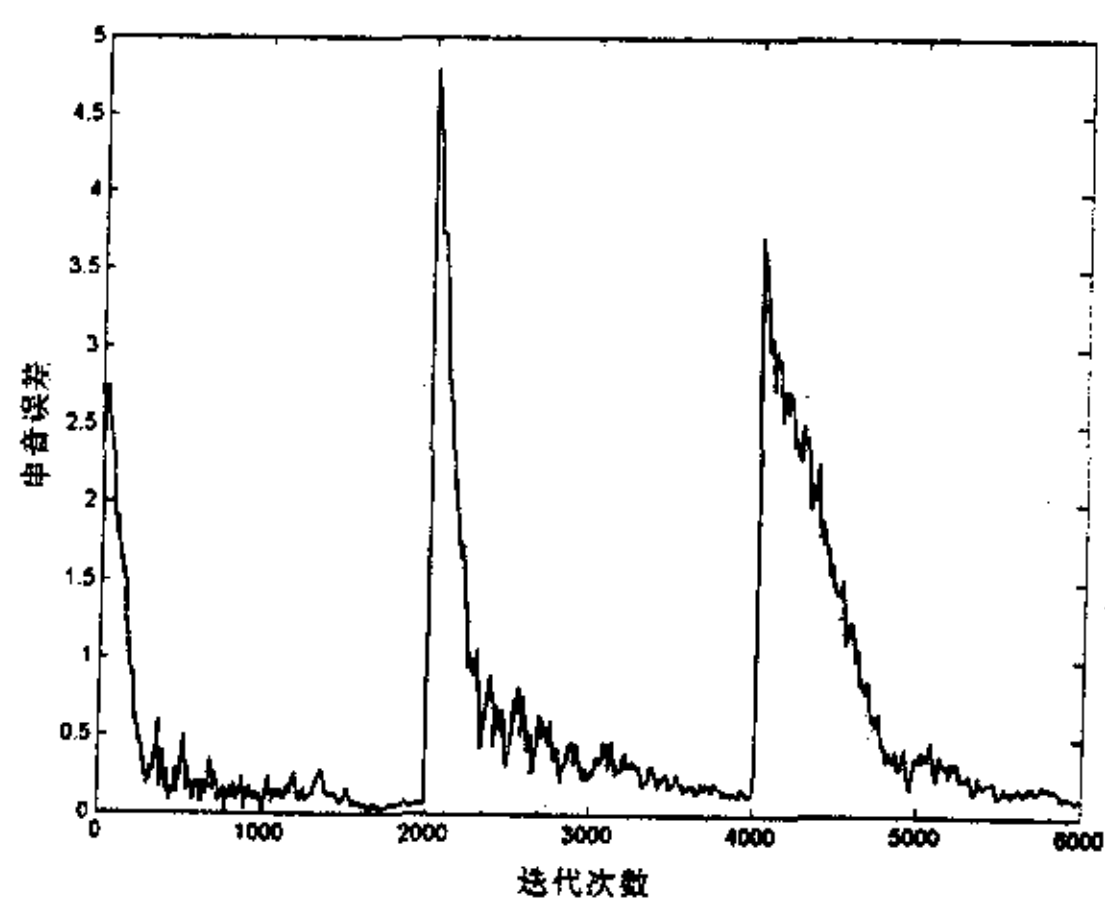


图6.9 某一次运行串音误差曲线图

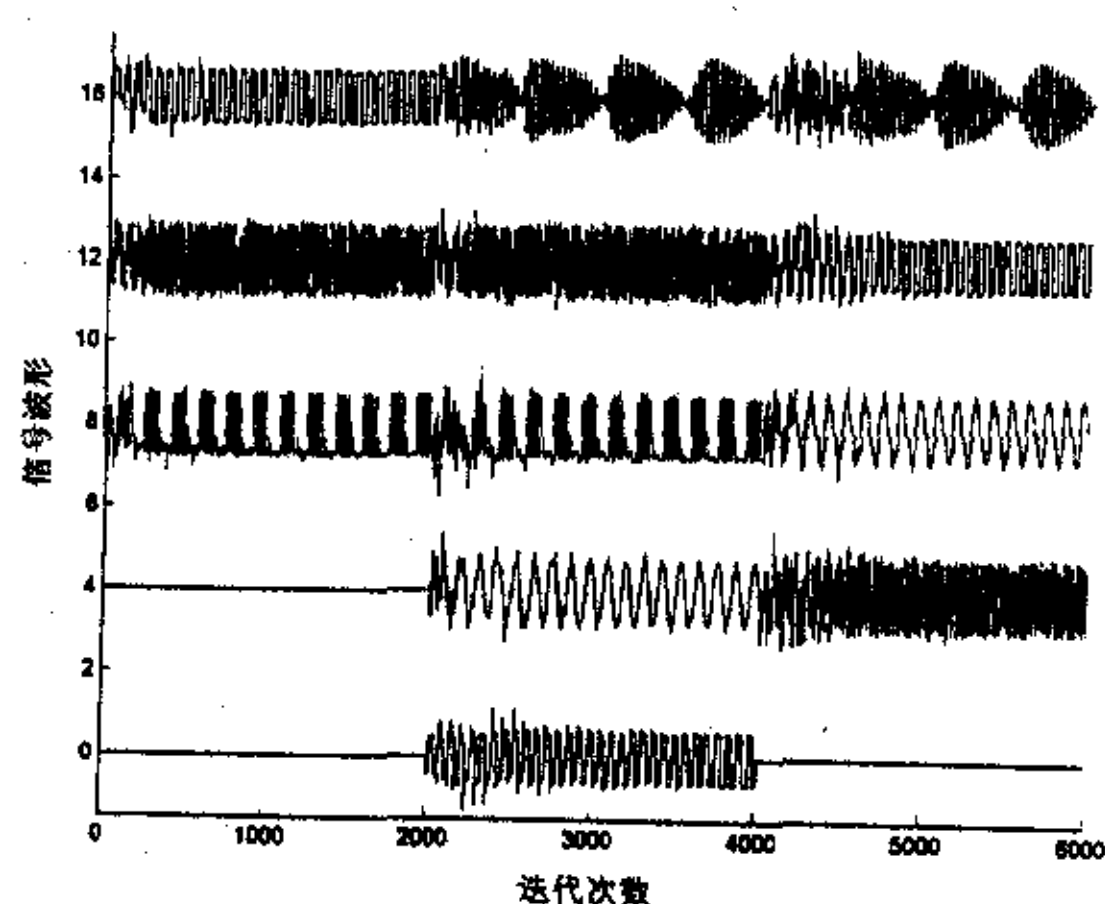


图6.10 系统输出信号波形图

### §6.4 病态盲信号分离

在本章前两节的讨论中, 我们假定混合矩阵  $\mathbf{A}$  是列满秩矩阵, 并强调这是盲信号分离问题的一个基本假设条件, 它保证所有的源信号分量都是可分离或可提取的。为什么? 当混合矩阵  $\mathbf{A}$  不是列满秩而是列亏损矩阵时, 盲信号分离问题情况又如何呢? 下面的定理回答了这些问题。

**定理 6.2<sup>[106]</sup>** 当且仅当在  $m \times n$  的混合矩阵  $\mathbf{A}$  中, 存在一个  $n-p$  列元素构成的  $m \times (n-p)$  的子矩阵  $\mathbf{A}_1$  和由其它  $p$  列元素构成的  $m \times p$  的子矩阵  $\mathbf{A}_2$ , 且满足条件  $\text{rank}(\mathbf{A}_1) < \text{rank}(\mathbf{A}_2)$  时, 存在一个  $m \times m$  的非奇异分离矩阵  $\mathbf{B}$ , 使得  $\mathbf{y}_i = \mathbf{B}\mathbf{x}_i = \mathbf{B}\mathbf{A}\mathbf{s}_i$  中的一个分量是某  $p$  个源信号的线性混合, 而其它  $m-1$  个分量中不含这  $p$  个源信号。

**定理 6.3<sup>[106]</sup>** 若在  $m \times n$  的混合矩阵  $\mathbf{A}$  中, 最多存在  $L$  个  $m \times (n-1)$  的子矩阵  $\mathbf{A}_i$ , 使得  $\text{rank}(\mathbf{A}_i) < \text{rank}(\mathbf{A})$ , 则理论上最多有  $L$  个源信号是可提取的。

这两个定理的证明参阅文献[106]。根据定理6.3不难看出, 若混合矩阵  $\mathbf{A}$  是列满秩矩阵, 则所有的源信号分量都是可以分离或者可以提取的。反之, 若混合矩阵  $\mathbf{A}$  是列亏损的, 则只有部分信号能够被提取。

混合矩阵  $\mathbf{A}$  是列亏损矩阵时的盲信号分离问题习惯称为病态盲信号分离, 它具体包括以下三种情况:

- C1) 观测信号的个数少于源信号的个数;
- C2) 观测信号与源信号的个数相等, 但混合矩阵  $\mathbf{A}$  是奇异矩阵;
- C3) 观测信号的个数多于源信号的个数, 但混合矩阵  $\mathbf{A}$  是列亏损的。

此外, 在第二章中我们看到, 观测信号含加性噪声时的盲信号分离问题(假设加性噪声各分量彼此独立, 并与所有的源信号分量也相互独立)是观测信号比源信号个数少这种情况的一个特例, 因此也属于病态盲信号分离问题的范畴。

最后我们强调, 尽管病态盲信号分离问题中并不是所有的源信号分量能够被分离或被提取的, 但如果混合矩阵的秩不等于1且不存在由全零元素构成的列(如果存在, 它相当于该列对应的源信号分量不起作用, 从而可将该列从混合矩阵中删除), 则根据定理6.3, 理论上至少有一个源信号分量是可以提取的。应该用什么样的方法提取可提取的源信号分量, 以及其它的病态盲信号分离问题该如何解决, 这是当前盲信号分离研究有待攻克的难点之一, 感兴趣的读者可参阅文献[18]和[106]。

## §6.5 小 节

根据观测信号与源信号个数的多少,可以将盲信号分离问题分成欠定,正定和超定三种情况。由于观测信号比源信号个数少的欠定盲信号分离问题是病态的,换句话说,只有部分源信号分量是可以分离或提取的,因此本章主要围绕正定和超定两种情况展开,内容主要涉及以下几个方面:

(1) 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适用于正定和超定两种情况的对比函数,并严格推导了对应的随机梯度学习算法;

(2) 借助于相对梯度的概念,理论证明完备盲信号分离的自然梯度算法同样适用于超定盲信号分离问题,并通过大量的计算机仿真验证了这一点;

(3) 在混合矩阵列满秩,源信号是各态历经的平稳随机过程条件下,介绍了一种Ye等人<sup>[157]</sup>提出的源信号个数的有效确定方法,结合我们提出的超定盲信号分离的自然梯度算法,成功解决了源信号个数未知甚至动态变化情况下超定盲信号分离的自适应并行实现问题;

(4) 简单介绍了几种病态盲信号分离问题,解释了混合矩阵列满秩是并行盲信号分离的一个基本假设条件的理论原因。

## 第七章 总结和展望

### §7.1 总 结

盲信号分离问题因为在无线数据通信、雷达和声纳、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,从而引起了信号处理学界和神经网络学界的共同兴趣。本论文围绕这一热点课题展开,并把研究重点放在自适应并行盲信号分离问题上,主要内容可概括如下:

1. 从信号模型出发,分析了盲信号分离问题所固有的两种不确定性以及实现盲信号分离的四个基本假设条件,介绍了盲信号分离算法的构造框架,研究了盲信号分离的对比函数理论和局部稳定性理论,并列举了若干常用的对比函数,总结了不同形式盲信号分离算法的局部稳定性条件,最后给出了算法的性能评价测度。这部分内容是盲信号分离研究的基础。

2. 自适应并行盲信号分离主要有三种技术:熵最大化分析法,非线性主分量分析法和独立分量分析法。无论采用那种技术,最终的学习算法都可以归纳为一种随机梯度即LMS型的算法,因此需要选择最优的学习速率参数,以解决算法收敛速度和稳态性能之间的矛盾。为此,本文定义了两个信号之间、一个信号与其它多个信号以及所有信号之间的相依性测度,并根据相依性测度提出一种分阶段学习的新思想。分阶段学习将学习速率参数由一维标量推广到二维学习速率矩阵,并把整个信号分离过程分成三个阶段进行,而学习速率矩阵在不同阶段有不同的形式:在初始阶段取常数矩阵,在捕捉阶段取对角矩阵,在跟踪阶段则为一般矩阵。由于学习速率矩阵的元素由信号的分离状态决定,从而既加快了信号分离前半阶段的收敛,又提高了后半阶段的跟踪和恢复的精度。以盲信号分离的主流技术,独立分量分析法中的自然梯度算法为例,结合分阶段学习思想得到的新算法,理论上具有等变化性和分离矩阵的非奇异性这两个关键性能。此外,理论分析比较和大量的计算机仿真都表明,与现有的固定学习速率以及不同形式的时变学习速率相比,分阶段学习的学习速率是最优的,表现为它具有最快的收敛速度,最好的稳态性能,信道与/或信道变化时,工作最稳定。

3. 分阶段学习的思想很好地解决了算法收敛速度和稳态性能之间的矛盾,但它需要自适应更新信号的相依性测度,导致算法的运算量增加。信号处理的滤波器理论



表明, RLS型的算法因使用更多的观测样本学习, 从而比LMS型的算法具有更快的收敛速度和更好的跟踪能力。为此, 本文从非线性主分量分析的对比函数出发, 结合分离矩阵的正交约束条件, 推导了一种新的基于自然梯度的RLS型算法(RLS-I算法), 理论证明它能实现预白化的盲信号分离, 而大量的计算机仿真也表明它比Pajunen等人提出的基于一般梯度的RLS型的盲信号分离算法优越。为解决白化预处理中学习速率参数的选择这一问题, 本文还推导了一种RLS型的预白化算法, 将之与RLS-I算法相结合, 经过合理近似又得到了一种新的盲信号分离RLS算法(RLS-II算法)。因为该算法隐含执行了预白化运算, 从而不需要对观测数据进行白化预处理就能实现信号的盲分离。理论证明RLS-II算法是等变化的, 并满足分离矩阵的非奇异性条件。

4. 盲信号分离的半参数统计理论表明, 实现盲信号分离的最优激励函数是分值函数, 它依赖于源信号的概率分布。尽管局部稳定性分析指出, 若源信号只含超高斯信号或者只含亚高斯信号, 则使用一个简单的非线性函数代替分值函数, 也能保证分离矩阵始终沿对比函数的上升方向搜索, 从而算法收敛同样导致盲信号分离。但问题是, 若源信号由超高斯信号和亚高斯信号混合而成, 则上述方法无法工作。为解决这个问题, 本文从归一化峰度的绝对值之和对比函数出发, 推导出一种新颖的盲信号分离算法, 它能实现混合亚高斯和超高斯信号的盲分离。理论证明新算法所提供的信号自适应的激励函数自动满足算法的局部稳定性条件和稳健性条件, 而理论比较和大量的计算机仿真也都表明, 新算法较现有的同类算法优越。

5. 根据观测信号与源信号个数的多少, 盲信号分离问题可以分成欠定, 正定和超定三种情况。由于观测信号比源信号个数少的欠定盲信号分离是病态的, 换句话说, 只有部分源信号分量是可以分离或提取的, 因此本文重点研究正定和超定两种情况, 并从分离矩阵的奇异值分解出发, 提出了一种新的基于独立分量分析的盲信号分离对比函数, 经过严格推导, 得出了对应的随机梯度学习算法。然后, 借助于相对梯度的概念, 理论证明完备盲信号分离的自然梯度算法同样适用于超定盲信号分离问题。在混合矩阵列满秩, 源信号是各态历经的平稳随机过程条件下, 本文还介绍了一种由Ye等人提出的源信号个数的有效确定方法, 结合前面的超定盲信号分离的自然梯度算法, 成功解决了源信号个数未知甚至动态变化情况下的超定盲信号分离问题。

## §7.2 展 望

近十多年来, 盲信号分离无论在理论和算法还是在实际应用方面都获得了长足的发展。本文就其中的自适应并行处理问题作了一些工作, 取得了一些进展, 但工作是

初步的,还有很多问题有待于进一步研究和解决。

1. 首先是自适应并行算法的全局收敛性分析问题。盲信号分离问题的并行在线实现算法是多种多样的,但在理论上,这些算法最多是局部收敛的。目前尚不存在一种自适应并行算法可以证明具有全局收敛性。因为盲信号分离问题利用的信息非常有限,而且算法中的激励函数又是一种高度非线性的函数,因此客观地讲,全局收敛性分析的难度相当大,但只有真正解决了这个问题,自适应并行盲信号分离算法在理论上才是完善的。

2. 当前的自适应并行盲信号分离算法通常不考虑噪声,或者假定噪声非常小,足可以忽略不计。但在实际应用中,观测信号总要受噪声的污染,某些情况下噪声还比较大,甚至远远强于信号。因此一定强度的噪声,尤其是色噪声对算法性能的影响是必须考虑的又一个难题,这个问题不解决,自适应并行盲信号分离的应用范围将受很大的限制。

3. 为保证所有的源信号分量都是可以分离的,盲信号分离问题约定混合矩阵列满秩。当混合矩阵是列亏损矩阵时,理论研究表明只有部分信号是可以提取的。这些可以提取的信号如何同步分离,更进一步地,混合矩阵动态变化,包括各种病态情况下的盲信号分离如何实现,它们构成了一个新的研究方向。

4. 盲信号分离问题固有的信号排序不确定性导致输出信号的排列顺序是不可预测的,但在很多场合,我们希望恢复出的源信号按照一定的顺序排列,或者希望一个甚至多个感兴趣的信号在特定的位置输出,解决这两个问题具有重大的现实意义,也必将推动盲信号分离的广泛应用。

5. 现有的盲信号分离集中讨论线性瞬时混合模型,而对存在延时与/或卷积的非瞬时混合模型研究较少。由于这涉及到多通道盲反卷积系统的稳定性和相位不确定性问题,尤其是源信号个数未知以及观测信号带噪声的情况,问题将更加复杂。此外,尽管已有少量文献开始了对非线性混合模型的研究,但考虑的情况都极特殊,因此更一般情况下的非线性混合模型是盲信号分离研究面临的又一大挑战。

6. 盲信号分离问题只利用观测数据以及独立性假设恢复源信号,而在很多实际应用中,源信号除了相互独立外,还有其它先验知识可以利用,如非平稳性或循环平稳特性、周期/非周期性、恒模特性以及有限字符特性等。考虑这些先验信息的半盲信号分离问题研究是实用而有趣的。

7. 除了在线盲信号分离外，批处理即离线盲信号分离的研究也非常活跃，但无论在线还是离线盲信号分离问题，它们都假定观测信号的样本足够多，通常情况也如此，但在个别应用场合，样本数可能非常有限。因此，有限样本的盲信号分离问题也是一个潜在的研究课题。

8. 最后，技术的发展离不开应用的土壤，将现有的盲信号分离理论和算法应用于解决实际工程问题也是十分有意义的。

## 附录A 高阶累积量

本文的许多章节都用到了高阶累积量, 为便于阅读, 下面给出其定义。  
若随机过程  $x(t)$  的  $k$  阶矩和第一特征函数分别定义为

$$\text{mom}_k[x] = E\{x^k(t)\} \quad (\text{A.1})$$

$$\Phi(\varpi) = E\{\exp[j\varpi x(t)]\} \quad (\text{A.2})$$

其中  $j = \sqrt{-1}$  表示虚数单位, 则随机过程  $x(t)$  的  $k$  阶累积量

$$\text{com}_k[x] = j^{-k} \left[ \frac{d^k}{d\varpi^k} \ln(\Phi(\varpi)) \right]_{\varpi=0} \quad (\text{A.3})$$

并习惯称  $\Psi(\varpi) = \ln[\Phi(\varpi)]$  为  $x(t)$  的第二特征函数。

根据上面的定义, 不难推导由矩计算累积量的一般公式。特别地, 对于零均值的平稳随机过程  $x(t)$ , 其一阶累积量也等于零, 而

$$\text{com}_2[x(t), x(t+\tau)] = E\{x(t)x(t+\tau)\} = R_x(\tau) \quad (\text{A.4})$$

$$\text{com}_3[x(t), x(t+\tau_1), x(t+\tau_2)] = E\{x(t)x(t+\tau_1)x(t+\tau_2)\} \quad (\text{A.5})$$

$$\begin{aligned} \text{com}_4[x(t), x(t+\tau_1), x(t+\tau_2), x(t+\tau_3)] &= E\{x(t)x(t+\tau_1)x(t+\tau_2)x(t+\tau_3)\} \\ &\quad - R_x(\tau_1)R_x(\tau_2-\tau_3) - R_x(\tau_2)R_x(\tau_3-\tau_1) - R_x(\tau_3)R_x(\tau_1-\tau_2) \end{aligned} \quad (\text{A.6})$$

关于累积量, 还有下面几个重要的性质。

**性质1** 若  $\alpha_i, \lambda_i (i=1, \dots, k)$  是常数,  $x_i (i=1, \dots, k)$  是随机变量, 则

$$\text{cum}[\lambda_1 x_1 + \alpha_1, \dots, \lambda_k x_k + \alpha_k] = \left( \prod_{i=1}^k \lambda_i \right) \text{cum}[x_1, \dots, x_k] \quad (\text{A.7})$$

**性质2** 累积量对于其变量具有可加性, 即

$$\text{cum}[x_1 + y_1, x_2, \dots, x_k] = \text{cum}[x_1, x_2, \dots, x_k] + \text{cum}[y_1, x_2, \dots, x_k] \quad (\text{A.8})$$

**性质3** 若随机变量  $\{x_i\}$  与随机变量  $\{y_i\}$  独立, 则

$$\text{cum}[x_1 + y_1, \dots, y_k + x_k] = \text{cum}[x_1, \dots, x_k] + \text{cum}[y_1, \dots, y_k] \quad (\text{A.9})$$

**性质4** 如果  $k$  个随机变量  $\{x_i\}$  的一个子集同其它部分独立, 则

$$\text{cum}[x_1, \dots, x_k] = 0 \quad (\text{A.10})$$

**性质5** 高斯随机变量的  $k (k > 2)$  阶累积量恒等于零。

## 附录B 矩阵求逆引理

在推导RLS型的预白化算法以及RLS型的盲信号分离算法时，经常要用到下面的矩阵求逆引理。

设  $\mathbf{A}$  是  $n \times n$  的非奇异矩阵， $\mathbf{C}$  是  $m \times m$  的非奇异矩阵， $\mathbf{B}$  和  $\mathbf{D}$  是两个任意的  $n \times m$  的矩阵，若  $\mathbf{D}^T \mathbf{A}^{-1} \mathbf{B} + \mathbf{C}^{-1}$  的逆矩阵存在，则增广矩阵  $\mathbf{A} + \mathbf{B} \mathbf{C} \mathbf{D}^T$  具有逆矩阵，且

$$[\mathbf{A} + \mathbf{B} \mathbf{C} \mathbf{D}^T]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} [\mathbf{D}^T \mathbf{A}^{-1} \mathbf{B} + \mathbf{C}^{-1}]^{-1} \mathbf{D}^T \mathbf{A}^{-1} \quad (\text{B.1})$$

特别地，若  $\mathbf{C}$  是一个常数  $\lambda$ ， $\mathbf{B}$  和  $\mathbf{D}$  退化为两个  $n \times 1$  的列向量  $\mathbf{b}$  和  $\mathbf{d}$ ，并假设  $\lambda \mathbf{d}^T \mathbf{A}^{-1} \mathbf{b} + 1 \neq 0$ ，则  $\mathbf{A} + \lambda \mathbf{b} \mathbf{d}^T$  具有可逆矩阵，且

$$[\mathbf{A} + \lambda \mathbf{b} \mathbf{d}^T]^{-1} = \mathbf{A}^{-1} - \lambda \cdot \frac{\mathbf{A}^{-1} \mathbf{b} \mathbf{d}^T \mathbf{A}^{-1}}{1 + \lambda \mathbf{d}^T \mathbf{A}^{-1} \mathbf{b}} \quad (\text{B.2})$$



## 致 谢

当我的博士论文搁笔之际，我轻轻地松了一口气。而恰恰在这个时候，我的心情难以平静，因为在近五年的硕博连读学习生涯中，我所感受到的要远比这篇博士论文厚重得多，难以尽言。

深深地感谢我的导师——张贤达特聘教授对我的教诲和指导！正是由于他这几年来学术上的悉心指导和生活上的亲切关怀，才使我得以顺利地完成博士期间的学习和研究任务，并使我最终能够完成本文的写作。张老师以渊博的知识、敏捷的思维和对前沿方向的洞察力，带领我逐步走向科学研究的道路，使我的研究范围逐步扩大、内容逐步深入。张老师不仅仅教给我科学知识和科研方法，更重要的是他正直的人品、严谨的治学态度、实事求是的工作作风、忘我的工作精神和学术上孜孜以求的进取精神，都潜移默化地影响着我，这一切都使我获益匪浅。

由衷地感谢中科院保铮院士。从他的谆谆教诲中，我感受到老一辈科学家崇高的思想境界和优秀品质，以及严谨的，一丝不苟的治学作风和对事业的不懈追求。

感谢实验室的吴顺君教授、张守宏教授、焦李成教授、廖桂生教授和许进教授等给予的指导和帮助；感谢实验室的陈瑞林、张淑红、李兰、詹志伟、刘华锐、白洁老师等给予的支持和帮助；感谢雷达信号处理重点实验室提供的良好工作环境和一流的科研条件。

感谢冯大政教授、水鹏朗教授和欧阳缙教授，与他们的多次畅谈和学术讨论给我很多启发，使我获益良多。

感谢我的师兄楼顺天副教授、李小军、陈建峰、冶继民和师弟韩芳明博士生给予我各方面的支持、鼓励和帮助。

感谢裴炳南、王彤、懂瑞军、刑孟道、罗斌凤、吴云涛、侯彪、黄晖、田晓华、吴莉莉、王万林、宁蔚、袁莉、周伟达、钟桦、张大朴等博士生给予我的关心和帮助。

感谢周延、丁志刚、胡涛、计世浩、薛金银、宫剑、陈建红、付庆霞以及研究生九八二班所有同学给予我的关心和帮助。

感谢吕园平、刘军红、段喜东、曹俊友、朱鹏、马强、邓旭亮、刘军、韩普等同学以及郭宝龙、陈逢时和杨剑老师的关心和帮助。

感谢栾金龙、张景元、苏宝银、姚金源、俞开江、朱飞云、杜存君、胡明等同学给我的关心、支持和帮助。

感谢岳博、石军、李景峰、张颖光、郑建忠、王峰、郭昕、黄建华同学对我的帮助，难忘与他们兄弟般的共同生活。

感谢父母多年的辛劳，感谢他们给我生活上的关怀和爱护，以及对我学业的耐心支持。感谢哥哥、妹妹以及黄薇全家对我的支持、鼓励和帮助。

衷心地感谢所有关心、爱护和帮助过我的人们！

作者：朱孝龙

二零零三年三月 于西安

## 参考文献

- [1] Amari S., "A theory of adaptive pattern classifiers," *IEEE Trans. Electronic Computers*, vol. 16, pp. 299-307, 1967.
- [2] Amari S., "Differential geometrical methods of statistics," Springer Lecture Notes in Statistics, vol.23, Bering: Springer, 1985.
- [3] Amari S., Cichocki A., and Yang H. H., "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, Touretzky D. S., Mozer M. C., and Hasselmo M. E. (eds.), Cambridge, MA: MIT press, vol. 8, pp. 757-763, 1996.
- [4] Amari S., and Cardoso J. F., "Blind source separation: semiparametric statistical approach," *IEEE Trans. Signal Processing*, vol. 45, no. 11, pp. 2692-2697, 1997.
- [5] Amari S., Chen T. P., and Cichocki A., "Stability analysis of learning algorithms for blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345-1351, 1997.
- [6] Amari S., "Neural learning in structured parameter space: Neural Riemannian gradient," in *Advances in Neural Information Processing Systems*, Mozer M. C., Jordan M. I., and Petsche T. (eds.), Cambridge, MA: MIT Press, vol.9, pp. 127-133, 1997.
- [7] Amari S., "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251-276, 1998.
- [8] Amari S., and Cichocki A., "Adaptive blind signal processing: Neural network approaches," *Proc. IEEE*, vol. 86, no. 10, pp. 2026-2048, 1998.
- [9] Amari S., "Natural gradient for over- and under-complete bases in ICA," *Neural Computation*, vol. 11, no. 8, pp. 1875-1883, 1999.
- [10] Amari S., Cichocki A., and Yang H. H., "Blind signal separation and extraction: Neural and information-theoretic approaches," in Haykin S., (ed.), *Unsupervised Adaptive Filtering*, vol. I, pp. 63-138, 2000.
- [11] Attias H., "Independent factor analysis," *Neural Computation*, vol. 11, no. 4, pp. 803-851, 1999.
- [12] Baker S. N., Kilner J. M., and Ahmed A., "Use of a blind separation algorithm to remove electrical cross-talk from multiple EMG recordings," in *Proc. J. of Physiology*, vol. 509-P, pp. 168-172, 1998.

- 
- [13] Basak J., and Amari S., "Blind separation of uniformly distributed signals: A general approach," *IEEE Trans. Neural Networks*, vol.10, no.5, pp. 1173-1185, 1999.
- [14] Bell A. J., and Sejnowski T. J., "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129-1159, 1995.
- [15] Belouchrani A., Meraim K. A., Cardoso J. F., and Moulines E., "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434-444, 1997.
- [16] Belouchrani A., and Cichocki A., "A robust whitening procedure in blind source separation context," *Electronics Letters*, vol. 36, pp. 2050-2051, 2000.
- [17] Burel G., "Blind separation of sources: a nonlinear neural algorithm," *Neural Networks*, vol. 5, no. 6, pp. 937-947, 1992.
- [18] Cao X. R., and Liu R. W., "A general approach to blind source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 3, pp. 562-571, 1996.
- [19] Cardoso J. F., "Source separation using higher order moments," *ICASSP '89*, pp. 2109-2112, 1989.
- [20] Cardoso J. F., "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," *ICASSP '90*, pp. 2655-2658, 1990.
- [21] Cardoso J. F., and Souloumiac A., "Blind beamforming for non-Gaussian signals," *IEE Proc.-F*, vol. 140, no. 6, pp. 362-370, 1993.
- [22] Cardoso J. F., and Laheld B., "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017-3030, 1996.
- [23] Cardoso J. F., Bose S., and Friedlander B., "On optimal source separation based on second and fourth order cumulants," in *Proc. IEEE Signal Processing Workshop on SSAP*, Corfu, Greece, pp. 198-201, June 1996.
- [24] Cardoso J. F., and Souloumiac A., "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 17, pp. 161-164, 1996.
- [25] Cardoso J. F., "Infomax and maximum likelihood for source separation," *IEEE Signal Processing Letters*, vol. 4, no. 4, pp.112-114, 1997.
- [26] Cardoso J. F., "Estimating function for source separation," *ICASSP '97*, vol. 5, pp. 3449-3452, 1997.
- [27] Cardoso J. F., "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009-2025, 1998.
- [28] Cardoso J. F., "Multidimensional independent component analysis," *ICASSP '98*, pp.
-

- 1941-1944, 1998.
- [29] Cardoso J. F., "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157-192, 1999.
- [30] Cardoso J. F., "On the stability of source separation algorithms," *J. of VLSI Signal Processing*, vol. 26, no. 1, pp. 7-14, 2000.
- [31] Castedo L., Escudero C. J., and Dapena A., "A blind signal separation method for multiuser communications," *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1343-1348, 1997.
- [32] Chang C., Ding Z., Yau S. F., and Chan F. H., "A matrix-pencil approach to blind separation of colored nonstationary signals," *IEEE Trans. Signal Processing*, vol. 48, no. 3, pp. 900-907, 2000.
- [33] Charkani N., and Deville Y., "Self-adaptive separation of convolutively mixed signals with a recursive structure, Part I: Stability analysis and optimization of the asymptotic behavior," *Signal Processing*, vol. 73, pp. 225-254, 1999.
- [34] Chaumette E., Comon P., and Muller K. R., "ICA-based technique for radiating sources estimation: application to airport surveillance," *IEE Proc. -F*, vol. 140, no. 6, pp. 395-401, 1994.
- [35] Chen T. P., Hua Y., and Yan W., "Global convergence of Oja's subspace algorithm for principle component extraction," *IEEE Trans. Neural Networks*, vol. 9, no. 1, pp. 58-67, 1998.
- [36] Choi S., Cichocki A., Zhang L. Q., and Amari S., "Approximate maximum likelihood source separation using natural gradient," in *Proc. 3<sup>rd</sup> IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Taoyuan, Taiwan, pp. 20-23, 2001.
- [37] Choi S., Cichocki A., and Amari S., "Flexible independent component analysis," *J. of VLSI Signal Processing*, vol. 26, no. 1, pp. 25-38, 2000.
- [38] Cichocki A., Unbehauen R., Moszczynski L., and Rummert E., "A new on-line adaptive learning algorithm for blind separation of source signals," in *Proc. Int. Symp. on Artificial Neural Networks*, Tainan, Taiwan, pp. 406-411, 1994.
- [39] Cichocki A., Unbehauen R., and Rummert E., "Robust learning algorithm for blind separation of signals," *Electronics Letters*, vol. 30, no. 17, pp. 1386-1387, 1994.
- [40] Cichocki A., and Unbehauen R., "Robust neural networks with on-line learning for blind identification and blind separation of sources," *IEEE Trans. Circuits and Systems-I: Fundamental Theory and Applications*, vol. 43, no. 21, pp. 894-906, 1996.



- 
- [41] Cichocki A., Amari S., Adachi M., and Kasprzak W., "Self-adaptive neural networks for blind separation of sources," in *Proc. Int. Symp. on Circuits and Systems*, vol. 2, pp. 157-160, May 1996.
- [42] Cichocki A., Amari S., and Cao J., "Blind separation of delayed and convolved sources with self-adaptive learning rate," in *Proc. Int. Symp. on Nonlinear Theory and Applications (NOLTA'96)*, pp. 229-232, 1996.
- [43] Cichocki A., Sabala I., Choi S., Orsier B., and Szupiluk R., "Self-adaptive independent component analysis for sub-Gaussian and super-Gaussian mixtures with unknown number of source signals and additive noise," in *Int. Symp. on Nonlinear Theory and applications (NOLTOA'97)*, Honolulu, HI, pp. 731-734, 1997.
- [44] Cichocki A., Rhawonmas R., and Amari S., "Sequential blind signal extraction in order specified by stochastic properties," *Electronic Letters*, vol. 33, pp. 64-65, 1997.
- [45] Cichocki A., Sabala I., and Amari S., "Intelligent neural networks for blind signal processing with unknown number of sources," in *Int. Symp. on Engineering of Intelligent Systems*, Tenerife, Spain, pp. 148-154, 1998.
- [46] Cichocki A., Zhang L. Q., Choi S., and Amari S., "Nonlinear dynamic independent component analysis using state-space and neural network models," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pp. 99-104, Aussois, France, 1999.
- [47] Cohen M., and Andreou A., "Current-mode subthreshold mos implementation of the Herault-Jutten autoadaptive network," *IEEE J. Solid-State Circuits*, vol. 27, no. 5, pp. 714-727, 1992.
- [48] Comon P., "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287-314, 1994.
- [49] Cover T. M., and Thomas J. A., *Elements of information theory*. New York: Wiley, chapter 9, pp. 224-238, 1991.
- [50] Cruces S., and Castedo L., "Stability analysis of adaptive algorithm for blind source separation of convolutive mixtures," *Signal Processing*, vol. 78, no. 3, pp. 265-276, 1999.
- [51] Cruces S., Cichocki A., and Castedo-Ribas L., "An iterative inversion approach to blind source separation," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1423-1437, 2000.
- [52] Delfosse N., and Loubaton P., "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, vol. 45, pp. 59-83, 1995.
-

- 
- [53] Deville Y., "Analysis of the convergence properties of self-normalized source separation neural networks," *IEEE Trans. Signal Processing*, vol. 47, no. 5, pp.1272-1287, 1999.
- [54] Ding Z., and Nguyen T., "Stationary points of a kurtosis maximization algorithm for blind source separation and antenna beamforming," *IEEE Trans. Signal Processing*, vol. 48, no. 6, pp. 1587-1596, 2000.
- [55] Douglas S. C., and Cichocki A., "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing*, vol. 45, no. 11, pp. 2829-2840, 1997.
- [56] Douglas S. C., and Cichocki A., "Adaptive step size techniques for decorrelation and blind source separation," in *Proc. 32<sup>nd</sup> Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, New York: IEEE press, vol. 2, pp. 1191-1195, 1998.
- [57] Douglas S. C., Cichocki A., and Amari S., "A bias removal technique for blind source separation with noisy measurements," *Electronics Letters*, vol. 34, pp. 1379-1380, 1998.
- [58] Douglas S. C., "Self-stabilized gradient algorithms for blind source separation with orthogonal constraints," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1490-1497, 2000.
- [59] Douglas S. C., and Kung S. Y., "Gradient adaptive algorithm for contrast-based blind deconvolution," *J. of VLSI Signal Processing*, vol. 26, no. 1, pp. 47-60, 2000.
- [60] Edelman A., Arias T., and Smith S. T., "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, pp. 303-353, 1998.
- [61] Ehlers F., and Schuster H., "Blind separation of convolutive mixtures and an application in automatic speech recognition in a noisy environment," *IEEE Trans. Signal Processing*, vol. 45, no. 10, pp. 2608-2612, 1997.
- [62] Eleftheriou E., and Falconer D., "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoustics Speech Signal Processing*, vol. 34, no. 5, pp. 1097-1110, 1986.
- [63] Everson R. M., and Roberts S. J., "Blind source separation for nonstationary mixing," *J. of VLSI Signal Processing*, vol. 26, no. 1, pp.15-23, 2000.
- [64] Fiori S., "Blind signal processing by the adaptive activation function neurons," *Neural Networks*, vol. 13, no. 6, pp. 597-611, 2000.
- [65] Fiori S., "On blind separation of complex-valued sources by extended Hebbian learning," *IEEE Signal Processing Letters*, vol. 8, no. 8, pp. 217-220, 2001.
- [66] Giannakopoulos X., Karhunen J., and Oja E., "Experimental comparison of neural
-

- algorithms for independent component analysis and blind separation," *Int. J. of Neural Systems*, vol. 9, no. 4, pp. 99-114, 1999.
- [67] Girolami M., and Fyfe C., "An extended exploratory projection pursuit network with linear and nonlinear anti-hebbian connections applied to the cocktail party problem," *Neural Networks*, vol. 10, pp. 1607-1618, 1997.
- [68] Girolami M., "An alternative perspective on adaptive independent component analysis algorithms," *Neural Computation*, vol. 10, no. 8, pp. 2103-2114, 1998.
- [69] Girolami M., *Self-organising neural networks: independent component analysis and blind source separation*, London: Springer-Verlag, 1999.
- [70] Haykin S. (ed.), *Blind deconvolution*, Prentice Hall, 1994.
- [71] Haykin S., *Adaptive Filter Theory*, 3rd edition, Prentice Hall, 1996.
- [72] Haykin S. (ed.), *Unsupervised adaptive filtering, Vol I: Blind source separation*. New York: Wiley, 2000.
- [73] Haykin S. (ed.), *Unsupervised adaptive filtering, Vol II: Blind deconvolution*. New York: Wiley, 2000.
- [74] He Z. Y., Liu J., and Yang L. X., "Blind separation of images using edgeworth expansion based on ICA algorithm," *Chinese Journal of Electronics*, vol. 8, no. 3, pp. 278-282, 1999.
- [75] He Z. Y., Yang L., Liu J., Lu Z., He C., and Shi Y., "Blind source separation using clustering-based multivariate density estimation algorithm," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 575-579, 2000.
- [76] Herault J., and Jutten C., "Space or time adaptive signal processing by neural network model," in Denker J. S. (ed.), *Neural networks for computing: AIP conf. proceedings 151*, New York: American Institute for physics, 13-16 April, 1986.
- [77] Hild K. E., Erdogmus D., and Principe J., "Blind source separation using Renyi's mutual information," *IEEE Signal Processing Letters*, vol. 8, no.6, pp.174-176, 2001.
- [78] Hoff T. P., Lindgren A. G., and Kaelin A. N., "Transpose properties in the stability and performance of the classic adaptive algorithms for blind source separation and deconvolution," *Signal Processing*, vol. 80, no. 9, pp.1807-1822, 2000.
- [79] Hyvarinen A., and Oja E., "A Fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483-1492, 1997.
- [80] Hyvarinen A., "One-unit contrast function for independent component analysis: A statistical analysis," in *Proc. IEEE Workshop on Neural Networks for Signal Processing*, Amelia Island, Florida, Sweden, pp. 388-397, 1997.

- 
- [81] Hyvarinen A., and Oja E., "Independent component analysis by general nonlinear Hebbian-like learning rules," *Signal Processing*, vol. 64, no. 3, pp. 301-313, 1998.
- [82] Hyvarinen A., "New approximation of differential entropy for independent component analysis and project pursuit," in *Advanced in Neural Information Processing Systems*, Cambridge, MA: MIT Press, vol. 10, pp. 273-279, 1998.
- [83] Hyvarinen A., "Noisy independent component analysis, maximum likelihood estimation and competitive learning," in *Proc. Int. Joint Conf. on Neural Networks*, Alaska, pp. 2282-2286, 1998.
- [84] Hyvarinen A., "Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood," *Neurocomputing*, vol. 22, pp. 49-67, 1998.
- [85] Hyvarinen A., "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.
- [86] Hyvarinen A., "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.
- [87] Hyvarinen A., "Gaussian moments for noisy independent component analysis," *IEEE Signal Processing Letters*, vol. 6, no. 6, pp. 145-147, 1999.
- [88] Hyvarinen A., and Pajunen P., "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Networks*, vol. 12, pp. 429-439, 1999.
- [89] Hyvarinen A., Karhunen J., and Oja E., *Independent component analysis*, New York: Wiley, 2001.
- [90] Jacobs R. A., "Increased rates of convergence through learning rate adaptation," *Neural Networks*, vol. 1, pp. 295-307, 1988.
- [91] Jutten C., and Herault J., "Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1-20, 1991.
- [92] Karhunen J., and Joutsensalo J., "Representation and separation of signals using nonlinear PCA type learning," *Neural Networks*, vol. 7, pp. 113-127, 1994.
- [93] Karhunen J., and Pajunen P., "Blind source separation using least-squares type adaptive algorithms," *ICASSP'97*, pp. 3361-3364, April 1997.
- [94] Karhunen J., Oja E., Wang L., Eigario R., and Joutsensalo J., "A class of neural networks for independent component analysis," *IEEE Trans. Neural Networks*, vol. 8, no. 3, pp. 486-504, 1997.
- [95] Karhunen J., Cichocki A., Kasprzak W., and Pajunen P., "On neural blind separation with noise suppression and redundancy reduction," *Int. J. of Neural Systems*, vol. 8, no. 2, pp. 219-232, 1997.
-



- 
- [96] Karhunen J., Pajunen P., and Oja E., "The nonlinear PCA criterion in blind source separation: Relations with other approaches," *Neurocomputing*, vol. 22, no. 1, pp. 5-20, 1998.
- [97] Lambert R., *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*. Ph.D dissertation, Univ. of Southern California, Dept. of Electrical Engineering, May 1996.
- [98] Lee T. W., Bell A. J., and Lambert R., "Blind separation of delayed and convolved sources," in *Advance in Neural Information Processing Systems*, Cambridge, MA: MIT press, vol. 9, pp. 754-763, 1997.
- [99] Lee T. W., *Independent component analysis: Theory and applications*. Kluwer, 1998.
- [100] Lee T. W., Lewicki M. S., Girolami M., and Sejnowski T. J., "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Processing Letters*, vol. 6, no. 4, pp. 87-90, 1999.
- [101] Lee T. W., Girolami M., and Sejnowski T. J., "Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources," *Neural Computation*, vol. 11, no. 2, pp. 409-433, 1999.
- [102] Lee T. W., Girolami M., Bell A., and Sejnowski T. J., "A unifying information theoretic framework for independent component analysis," *Computers and Mathematics with Applications*, vol. 31, no. 11, pp. 1-11, 2000.
- [103] Lehmann E. L., *Testing Statistical Hypothesis*. New York: Wiley, 1959.
- [104] Lewick M., and Sejnowski T. J., "Learning overcomplete representations," *Neural Computation*, vol. 12, pp. 337-365, 2000.
- [105] Li Y. Q., Wang J., and Zurada J. M., "Blind extraction of singularly mixed source signals," *IEEE Trans. Neural Networks*, vol. 11, pp. 1413-1422, 2000.
- [106] Li Y. Q., and Wang J., "Sequential blind extraction of instantaneously mixed sources," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 997-1006, 2002.
- [107] Lindgren U. A., and Broman H., "Source separation using a criterion based on second-order statistics," *IEEE Trans. Signal Processing*, vol. 46, no. 7, pp. 1837-1850, 1998.
- [108] Luo J., Ling X. T., and Liu R. W., "Principal independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 4, pp. 912-917, 1999.
- [109] Macchi O., and Moreau E., "Adaptive unsupervised separation of discrete sources," *Signal Processing*, vol. 73, pp. 49-66, 1999.
- [110] Mansour A., and Jutten C., "What should we say about the kurtosis?" *IEEE Signal*
-



- Processing Letters*, vol. 6, no. 12, pp. 321-322, 1999.
- [111] Makeig S., Jung T. P., and Bell T. J., "Blind separation of auditory event-related brain responses into independent components," in *Proc. Natl. Acad. Sci.*, vol. 94, pp. 10979-10984, 1997.
- [112] Mathews V. J., and Xie Z., "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Processing*, vol. 41, pp. 2075-2087, 1993.
- [113] Mathis H., Hoff T. P., and Joho M., "Blind separation of signals with mixed kurtosis signs using threshold activation functions," *IEEE Trans. Neural Networks*, vol. 12, no. 3, pp. 618-624, 2001.
- [114] Mathis H., and Douglas S. C., "On the existence of universal nonlinearities for blind source separation," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1007-1016, 2002.
- [115] Matsuoka K., Ohya M., and Kawamoto M., "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411-419, 1995.
- [116] Moreau E., and Macchi O., "High-order contrasts for self-adaptive source separation," *Int. J. of Adaptive Control and Signal Processing*, vol. 10, pp. 19-46, 1996.
- [117] Moreau E., and Moreau N. T., "Nonsymmetrical contrasts for source separation," *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2241-2252, 1999.
- [118] Murata N., Muller K. R., Ziehe A., and Amari S., "Adaptive on-line learning in changing environments," in *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT press, pp. 599-605, 1997.
- [119] Obradovic D., and Deco G., "Infomax and ICA: is there a difference?" *Neural Computation*, vol. 10, no. 8, pp. 2085-2101, 1998.
- [120] Ohata M., and Matsuoka K., "Stability analysis of information-theoretic blind separation algorithms in the case where the sources are nonlinear processes," *IEEE Trans. Signal Processing*, vol. 50, no. 1, pp. 69-77, 2002.
- [121] Oja E., "Principal components, minor components, and linear neural networks," *Neural Networks*, vol. 5, pp. 927-935, 1992.
- [122] Oja E., "The nonlinear PCA learning rule and signal separation: Mathematical analysis," *Neurocomputing*, vol. 17, no. 1, pp. 25-46, 1997.
- [123] Pajunen P., Hyvarinen A., and Karhunen J. K., "Nonlinear blind source separation by self-organizing maps," in *Proc. Int. Conf. on Neural Information Processing*, Hong Kong, pp. 1207-1210, 1996.

- 
- [124] Pajunen P., and Karhunen J., "Least-squares methods for blind source separation based on nonlinear PCA," *Int. J. of Neural Systems*, vol. 8, pp. 601-612, 1998.
- [125] Papoulis A., *Probability, Random Variables, and Stochastic Process*. 3<sup>rd</sup> edition, New York: McGraw-Hill, pp. 190-191, 1991.
- [126] Papadias C. B., and Paulraj A., "A constant modulus algorithm for multi-user signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Letters*, vol. 4, pp. 178-181, 1997.
- [127] Papadias C. B., "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3508-3519, 2000.
- [128] Parra L., and Spence C., "On-line convolutive blind source separation of nonstationary signals," *J. of VLSI Signal Processing*, vol. 26, no. 1, pp 39-46, 2000.
- [129] Pham D. T., "Blind separation of instantaneous mixtures of sources via an independent component analysis," *IEEE Trans. Signal Processing*, vol. 44, no. 11, pp. 2768-2779, 1996.
- [130] Pham D. T., and Garrat P., "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1712-1725, 1997.
- [131] Pham D. T., and Cardoso J. F., "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 49, no. 9, pp. 1837-1848, 2001.
- [132] Pearlmutter B., and Parra L., "A context-sensitive generalization of ICA," in *Int. Conf. on Neural Information Processing*, pp. 151-157, 1996.
- [133] Poor H. V., and Wang X. D., "Code-aided interference suppression for DS/CDMA communications—Part II: Parallel blind adaptive implementations," *IEEE Trans. Comm.*, vol. 45, no. 9, pp. 1112-1122, 1997.
- [134] Ristaniemi T., *Synchronization and Blind Signal Processing in CDMA Systems*. PhD thesis, Univ. of Jyväskylä, Jyväskylä, Finland, 2000.
- [135] Smaragdis P., "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21-34, 1998.
- [136] Souloumiac A., "Blind source detection and separation using second order nonstationary signals," *ICASSP'95*, pp.1912-1915, 1995.
- [137] Taleb A., and Jutten C., "Source separation in post nonlinear mixtures: An entropy-based algorithm," *ICASSP'98*, pp. 2089-2092, 1998.
-

- 
- [138] Taleb A., and Jutten C., "Source separation in post-nonlinear mixtures," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2807-2820, 1999.
- [139] Tan Y., Wang J., and Zurada J. M., "Nonlinear blind source separation using a radial basis function network," *IEEE Trans. Neural Networks*, vol. 12, no. 1, pp. 124-134, 2001.
- [140] Thawonmas R., and Cichocki A., "Blind extraction of source signals with specified stochastic features," *ICASSP '97*, vol. 4, pp. 3353-3356, 1997.
- [141] Thawonmas R., Cichocki A., and Amari S., "A cascade neural network for blind extraction without spurious equilibria," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E81-A, no. 9, pp. 1833-1846, 1998.
- [142] Tong L., Liu R. W., Soon V. C., and Huang Y. F., "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits and Systems*, vol. 38, no. 5, pp. 499-509, 1991.
- [143] Torkkola K., "Blind separation of delayed and convolved sources" in Haykin S., (ed.), *Unsupervised Adaptive Filtering*, vol. I, pp.321-375, 2000.
- [144] Tugnait J. K., "Identification and deconvolution of multichannel nongaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. 45, pp. 658-672, 1997.
- [145] Tugnait J. K., "Adaptive blind separation of convolutive mixtures of independent linear signals," *Signal Processing*, vol. 73, no. 1, pp. 139-152, 1999.
- [146] Tugnait J. K., "On blind separation of convolutive mixtures of independent linear signals in unknown additive noise," *IEEE Trans. Signal Processing*, vol. 46, no. 11, pp. 3117-3123, 1998.
- [147] Vee van der A. J., "Blind separation of BPSK sources with residual carriers," *Signal Processing*, vol. 73, pp. 67-79, 1999.
- [148] Vlassis N., and Motomura Y., "Efficient source adaptivity in independent component analysis," *IEEE Trans. Neural Networks*, vol. 12, no. 3, pp. 559-565, 2001.
- [149] Xu L., "Temporal BYY learning for state space approach, hidden markov model, and blind source separation," *IEEE Trans. Signal Processing*, vol. 48, no. 7, pp. 2132-2144, 2000.
- [150] Xun X. A., and Douglas S. C., "Mean square error analyses of adaptive blind source separation algorithms," in *Proc. IEEE Signal Processing Workshop on Neural Networks for Signal Processing*, pp. 333-342, 2001.
- [151] Yang B., "Projection approximation subspace tracking," *IEEE Trans. Signal*
-

- Processing*, vol. 43, no. 1, pp. 95-107, 1995.
- [152] Yang B., "Asymptotic convergence analysis of the project approximation subspace tracking algorithm," *Signal Processing*, vol. 50, pp. 123-136, 1996.
- [153] Yang H. H., and Amari S., "Adaptive on-line learning algorithms for blind separation: Maximum entropy and minimum mutual information," *Neural Computation*, vol. 9, no. 5, pp. 1457-1482, 1997.
- [154] Yang H. H., Amari S., and Cichocki A., "Information back-propagation for blind separation of sources from nonlinear mixture," in *Proc. Int. Conf. on Neural Networks*, pp. 2141-2146, 1997.
- [155] Yang H. H., Amari S., and Cichocki A., "Information-theoretic approach to blind separation of sources in nonlinear mixture," *Signal Processing*, vol. 64, pp. 291-300, 1998.
- [156] Yang H. H., "Serial updating rule for blind separation derived from the method of scoring," *IEEE Trans. Signal Processing*, vol. 47, no. 8, pp. 2279-2285, 1999.
- [157] Ye J. M., and Zhang X. D., "Blind source separation with unknown and dynamically changing number of source signals," submitted to *IEEE Trans. Signal Processing*.
- [158] Yellin D., and Weinstein E., "Criteria for multichannel signal separation," *IEEE Trans. Signal Processing*, vol. 42, pp. 2158-2167, 1994.
- [159] Yellin D., and Weinstein E., "Multichannel signal separation: Methods and analysis," *IEEE Trans. Signal Processing*, vol. 44, pp. 106-118, 1996.
- [160] Yeredor A., "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Processing Letters*, vol. 7, no. 7, pp. 197-200, 2000.
- [161] Yeredor A., "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1545-1553, 2002.
- [162] Yu X., and Hu G. R., "Speech separation based on the GMM pdf estimation and the feedback architecture," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, pp. 353-357, 1999.
- [163] Zarzoso V., and Nandi A. K., "Blind separation of independent sources for virtually any source probability density function," *IEEE Trans. Signal Processing*, vol. 47, no. 9, pp. 2419-2423, 1999.
- [164] Zhang L. Q., Cichocki A., and Amari S., "Natural gradient algorithm for blind separation of overdetermined mixtures with additive noise," *IEEE Signal Processing*

- Letters*, vol.6, no.11, pp.293-295, 1999.
- [165] Zhang L. Q., and Cichocki A., "Blind separation of filtered sources using state-space approach," in *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT press, vol. 11, pp. 648-654, 1999.
- [166] Zhang L. Q., and Cichocki A., "Blind deconvolution of dynamical systems: a state-space approach," *J. of Signal Processing*, vol. 4, no. 2, pp. 111-130, 2000.
- [167] Zhu X. L., and Zhang X. D., "Adaptive RLS algorithm for blind source separation using a natural gradient," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 432-435, 2002.
- [168] 丁志刚, 朱孝龙, 焦李成, "基于独立分量分析的 DS-CDMA 系统接收机," 电子学报, vol. 11A, pp. 97-100, 2000.
- [169] 冯大政, 保铮, 张贤达, "信号盲分离问题的多阶段分解算法," 自然科学进展, vol. 12, no. 3, pp. 324-328, 2002.
- [170] 何振亚, 刘琚, 杨绿溪, "盲均衡和信道参数估计的一种 ICA 和进化计算方法," 中国科学 E 辑, vol. 30, no. 1, pp. 1-7, 2000.
- [171] 焦李成, 马海波, 刘芳, "多用户检测和独立分量分析: 进展与展望," 自然科学进展, vol. 12, no. 4, pp. 365-371, 2002.
- [172] 凌燮亭, "近场宽带信号源的盲分离," 电子学报, vol. 24, no. 7, pp. 87-92, 1996.
- [173] 刘琚, 何振亚, "盲源分离和盲反卷积," 电子学报, vol. 20, no. 4, pp. 570-576, 2002.
- [174] 张贤达, 保铮, "盲信号分离," 电子学报, vol. 12A, pp. 1766-1771, 2001.
- [175] 张贤达. 现代信号处理. 北京: 清华大学出版社, 1994.
- [176] 张贤达. 信号处理中的线性代数. 北京: 科学出版社, 1997.
- [177] 张贤达, 保铮. 通信信号处理. 北京: 国防工业出版社, 2000.
- [178] 张贤达, 朱孝龙, 保铮, "基于分阶段学习的盲信号分离," 中国科学 (E 辑) vol. 32, no. 5, pp. 693-703, 2002.



## 博士学习阶段 (合作) 发表与撰写的学术论文

1. Zhu X. L., and Zhang X. D., "Adaptive RLS algorithm for blind source separation using a natural gradient," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 432-435, 2002
2. Zhang X. D., Zhu X. L., and Bao Z., "Grading learning for blind source separation," *Science in China*, Ser. F, vol. 46, no. 1, pp. 31-44, 2003.
3. Zhu X. L., Zhang X. D., and Ye J. M., "Natural gradient-based recursive least-squares algorithm for adaptive blind source separation," *Science in China*, Ser. F, accepted for publication, 2003.
4. Zhu X. L., Chen J. F., and Zhang X. D., "Adaptive blind source separation in order specified by stochastic properties," in *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'2003)*, to appear, 2003.
5. Chen J. F., Zhu X. L., and Zhang X. D., "A multiuser detector based on blind source separation," in *Proc. Int. Conf. on Comm. Tech. (ICCT'2003)*, to appear, 2003.
6. Zhu X. L., and Zhang X. D., "A robust algorithm for blind separation of sources with mixed kurtosis signs," submitted to *Chinese J. of Electronics*.
7. Chen J. F., Zhu X. L., and Zhang X. D., "A new algorithm for joint range-DOA-frequency estimation of near-field sources," submitted to *Applied Signal Processing*.
8. 丁志刚, 朱孝龙, 焦李成, "基于独立分量分析的 DS-CDMA 系统接收机," *电子学报*, vol. 11A, pp. 97-100, 2000.
9. 张贤达, 朱孝龙, 保铮, "基于分阶段学习的盲信号分离," *中国科学 (E 辑)*, vol. 32, no. 5, pp. 693-703, 2002.
10. 朱孝龙, 张贤达, 冶继民, "基于自然梯度的递归最小二乘盲信号分离," *中国科学 (E 辑)*, 录用.
11. 朱孝龙, 张贤达, "基于奇异值分解的超定盲信号分离," *电子与信息学报*, 录用.
12. 朱孝龙, 张贤达, "基于选优估计函数的盲信号分离," *西安电子科技大学学报*, vol. 30, no. 3, 2003.