

## 加权 Möller 算法的广义加权规则

杜柏阳<sup>1</sup> 孔祥玉<sup>1</sup> 冯晓伟<sup>1</sup> 高迎彬<sup>2</sup> 曹泽豪<sup>1</sup>

**摘 要** 提出一套针对加权 Möller 算法的广义加权规则用于探讨改进并行提取次成分分析的理论和方法问题. 该规则仅在加权 Möller 算法上引入一个加权规则参数, 通过调节参数后算法性能上的变化, 实现加权 Möller 算法稳定性在动力学层面上的分析, 探讨加权参数变化对算法稳定性的影响. 基于常微分方程方法对所提出规则下的加权 Möller 算法进行稳定性证明, 并分析其中关键函数的性质. 最后, MATLAB 仿真验证了所提出规则的性能和算法性质.

**关键词** 多个次成分, 广义加权规则, 并行提取, Möller 算法

**引用格式** 杜柏阳, 孔祥玉, 冯晓伟, 高迎彬, 曹泽豪. 加权 Möller 算法的广义加权规则. 自动化学报, 2020, 46(1): 193–199

**DOI** 10.16383/j.aas.c180012

## Generalized Weighted Rules on Modified Möller Algorithm

DU Bo-Yang<sup>1</sup> KONG Xiang-Yu<sup>1</sup> FENG Xiao-Wei<sup>1</sup>  
GAO Ying-Bin<sup>2</sup> CAO Ze-Hao<sup>1</sup>

**Abstract** A generalized weighted rule for weighted Möller algorithm is introduced to explore the modified minor component analysis by parallel extraction in theory and application. The proposed rule helps to analyze the stability in the dynamic of the algorithm in aspects of different properties, through altering the generalized weighted parameter, which is the only alterable parameter adopted in the weighted Möller algorithm. The stability analysis of the weighted Möller algorithm modified by the proposed generalized weighted rule is evaluated based on ordinary differential equation. And the analysis of the key functions is given. Finally, the properties of the modified algorithms and the ability of the proposed rule are verified by simulations in MATLAB.

**Key words** Multiple minor components, generalized weighted rule, parallel extraction, Möller algorithm

**Citation** Du Bo-Yang, Kong Xiang-Yu, Feng Xiao-Wei, Gao Ying-Bin, Cao Ze-Hao. Generalized weighted rules on modified Möller algorithm. *Acta Automatica Sinica*, 2020, 46(1): 193–199

在信息处理领域, 次成分分析 (Minor component analysis, MCA) 广泛应用于总体最小二乘<sup>[1]</sup>, 波达方向估计<sup>[2]</sup>, 图像处理<sup>[3]</sup>, 频率估计<sup>[4]</sup>, 语音增强<sup>[5]</sup> 和目标跟踪<sup>[6]</sup> 等问题. 通常, 次成分 (Minor component, MC) 是指在输入信号的自相关矩阵中前  $r$  个最小特征值对应的特征向量. 由这些特征向量张成的特征子空间被称为次子空间 (Minor subspace, MS).

对 MCA 问题研究的诸多方法中, 基于 Hebbian 规则的方法是近些年对该问题研究的热点. 许多学者从

不同的角度对这个问题展开了研究, 提出许多非常有趣的算法. 当前 MCA 主要有三种类型, 次成分序贯提取算法, 次子空间提取算法和次成分并行提取算法. 早期研究集中在次成分序贯提取, 比如改进 Hebbian 算法的 Xu 算法<sup>[7]</sup>, 针对高维数据的 MCA EXIN 算法<sup>[8]</sup>, 从自稳定性角度提出的 Möller 算法<sup>[9]</sup> 等. 在提取次子空间的研究中, Kong 算法<sup>[10–11]</sup> 能够在保证算法的效率的同时, 对算法的自稳定性展开深入的探讨. 虽然上述算法均具有良好的性能, 但要实现提取多个次成分时, 算法的计算成本问题就显现出来<sup>[12]</sup>.

次成分并行提取算法则很好地解决了这个问题, 比如利用 PCA 网络改进的 AMMD (Adaptive multiple minor directions) 算法<sup>[13]</sup>, 但是该算法必须预先估计信号自相关矩阵的最小特征值. 为避免预先估计, 实现自适应地提取, 需要采用直接设计 MCA 网络的方法. 主要方式有两种, 一种是通过设计转换机制将次子空间提取算法改造为次成分并行提取算法<sup>[14]</sup>; 另外一种就是通过设计加权信息准则, 利用其梯度流, 实现次成分的并行提取<sup>[15]</sup>, 或者直接通过加权对算法改进, 得到次成分并行提取算法<sup>[16]</sup>. 其中第二种方法相比第一种方法, 可减少一些计算过程, 因此对其研究很有意义. 作者团队前期已经对并行提取算法开展了研究工作<sup>[17]</sup>.

当前, 采用不同加权方式对应的并行算法效果不尽相同<sup>[18]</sup>, 因而对一定加权规则下不同算法的性能研究就十分必要. 当前对这个问题的研究还比较少. 因此, 在本文中, 作者主要针对改进型 Möller 算法为基础的并行提取算法, 设计了一种广义加权规则. 进而分析了在这种加权规则下, 加权规则参数变化时的算法稳定性问题. 广义加权规则在理论上更进一步探讨算法的“速度–稳定性”问题, 加深对基于 Hebbian 规则 MCA 算法稳定性的理解; 在方法上也可应用于对于其他的并行次成分提取算法的改进.

## 1 问题描述

次成分分析并行提取算法的基本学习规则一般表达为

$$W(k+1) = W(k) + \eta \Delta W(k) \quad (1)$$

其中,  $\eta$  为学习因子, 且满足  $\eta \in (0, 1)$ ;  $W(k)$  为权值矩阵, 且满足  $W(k) \in \mathbf{R}^{n \times r}$ ,  $n$  为输入信号的维数,  $r$  为需要提取的次成分的个数,  $k$  为当前的迭代次数;  $\Delta W(k)$  为算法的梯度方向, 满足  $\Delta W(k) = \partial J / \partial W|_{W=W(k)}$ ,  $J$  是一个信息准则. 对于输入信号  $\mathbf{x}$  而言, 其自相关矩阵  $R = E[\mathbf{x} \cdot \mathbf{x}^T]$  是对称矩阵, 可通过矩阵特征值分解可得

$$R = U \Lambda U^T \quad (2)$$

其中,  $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$ , 由矩阵  $R$  的特征向量组成,  $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  由矩阵  $R$  的特征值组成. 为方便描述, 令  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ , 那么前  $r$  个特征值  $\lambda_1, \lambda_2, \dots, \lambda_r$  对应的特征向量  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r$  称为信号  $\mathbf{x}$  的前  $r$  个次成分. 在现实过程中, 输入向量的自相关矩阵  $R$  不可直接获得, 只得通过可行的输入向量  $\mathbf{x}$  实时估计  $\hat{R}$  获得,

$$\hat{R}(k+1) = \frac{k}{(k+1)} \hat{R}(k) + \frac{\mathbf{x}(k+1)\mathbf{x}^T(k+1)}{(k+1)} \quad (3)$$

易知  $\lim_{k \rightarrow \infty} \hat{R}(k) = R$ , 因此采用式 (4) 估计是可行的. 次成分并行提取, 就是在迭代算法中, 通过构造合适的  $\Delta W$ , 使  $W$  收敛到稳定的状态. 这个过程中, 状态矩阵  $W$  就是对多个次成分的估计. 本文主要讨论在加权矩阵在广义加权规则里变化的情况下, 算法的收敛稳定性问题.

收稿日期 2018-01-09 录用日期 2018-05-29  
Manuscript received January 9, 2018; accepted May 29, 2018  
国家自然科学基金 (61374120, 61673387) 资助  
Supported by National Natural Science Foundation of China (61374120, 61673387)  
本文责任编辑 谭莹  
Recommended by Associate Editor TAN Ying  
1. 西安高科技研究所 西安 710025 2. 中国石油大学 (北京) 北京 100100  
1. The Xi'an Researching Institute of High Technology, Xi'an 710025 2. China University of Petroleum (Beijing), Beijing 100100

为后续的讨论方便, 将需要用到的引理列写如下,

**引理 1**<sup>[18]</sup>. 有  $\Lambda, \Gamma \in \mathbf{R}^{n \times n}$ ,  $\Lambda$  为对角阵,  $\pi$  为一个置换,  $\pi^2 = \mathbf{1}_n$ . 若  $\Gamma$  的  $(i, \pi(i))$  元素不为零, 其他元素均为零, 那么  $\Gamma^T \Lambda \Gamma$  为对角阵. 此时,  $\Gamma$  为非酉置换矩阵.

**引理 2**<sup>[18]</sup>. 若  $\Gamma \in \mathbf{R}^{n \times n}$  为非酉置换矩阵, 其对应置换为  $\pi$ . 则特征值

$$\xi = \begin{cases} \pm[\Gamma]_{i, \pi(i)}, & i \neq \pi(i) \\ [\Gamma]_{i, \pi(i)}, & i = \pi(i) \end{cases}, \quad i = 1, 2, \dots, n$$

## 2 广义加权规则

在加权 Möller 自适应算法的基础上, 给出如下广义加权规则

$$\Delta W(k) = W(k) W^T(k) R W(k) D^{-p} - R W(k) \left( 2 W^T(k) W(k) D^{-p} + D^{1-p} \right) \quad (4)$$

其中, 加权矩阵  $D = \text{diag}\{d_1, d_2, \dots, d_r\}$ , 且满足  $d_1 < d_2 < \dots < d_r$ . 该算法能够自适应地实现次成分的并行提取.  $p$  是一个标量, 不同的  $p$  对应不同的算法学习行为. 对于每一个具体的  $p$ , 相应的算法步骤可参考如下:

**步骤 1.** 已知确定的  $p$  并且初始化学习因子  $\eta$  和加权矩阵  $D$ .

**步骤 2.** 计算自相关矩阵  $\hat{R}$ . 将  $\mathbf{x}(k)$  代入式 (4).

**步骤 3.** 次成分提取, 将  $\hat{R}$  代入式 (3), 回到步骤 2, 直到  $W$  稳定到特定矩阵.

该广义加权规则下对应不同  $p$  的学习行为均包含共同的驻点. 这个特点可以通过定理 1 证实.

**定理 1.** 对于加权规则下的式 (4), 其极值点为  $W = \bar{P} D^{\frac{1}{2}}$ , 其中  $\bar{P} \in \mathbf{R}^{n \times n}$  为包含  $R$  中前  $r$  个对应次成分的矩阵.

**证明.** 不妨对  $W$  进行 SVD 分解, 表达为  $W = P \Sigma Q^T$ , 其中  $\Sigma \in \mathbf{R}^{r \times r}$  为各元素均为正数的对角阵,  $Q \in \mathbf{R}^{r \times r}$  为正交阵,  $P \in \mathbf{R}^{n \times r}$  包含  $r$  个正交向量. 因为稳定点处  $dW/dt = 0_{n \times r}$ , 式 (4) 微分形式为

$$\frac{dW}{dt} = W W^T R W D^{-p} + R W D^{1-p} - 2 R W W^T W D^{-p} \quad (5)$$

将  $W$  的 SVD 分解代入式 (5) 中, 得

$$\frac{d(P \Sigma Q^T)}{dt} = P \Sigma^2 P^T R P \Sigma Q^T D^{-p} - 2 R P \Sigma^3 Q^T D^{-p} + R P \Sigma Q^T D^{1-p} \quad (6)$$

已知  $Q^T Q = I_r$ ,  $P^T P = I_r$ ,  $I_r \in \mathbf{R}^{r \times r}$  为单位阵, 那么分别对式 (6) 左乘  $P^T$ , 右乘  $D^{p-1} Q$  可得

$$0_{n \times r} = \Sigma^2 P^T R P \Sigma Q^T D^{-1} Q - 2 P^T R P \Sigma^3 Q^T D^{-1} Q + R P \Sigma \quad (7)$$

为方便表达, 令  $A = P^T R P$ . 因为  $P$  为列满秩,  $R$  为满秩方阵, 可知  $A$  为满秩矩阵. 因此式 (7) 可以改写为

$$2 \Sigma^2 - \Sigma^{-1} A^{-1} \Sigma^2 A \Sigma = Q^T D Q \quad (8)$$

由于  $Q^T D Q$  和  $\Sigma^2$  均为对称矩阵, 可得

$$\Sigma^{-1} A^{-1} \Sigma^2 A \Sigma = (\Sigma^{-1} A^{-1} \Sigma^2 A \Sigma)^T \quad (9)$$

由此可以导出

$$\Sigma^2 A \Sigma^2 A = A \Sigma^2 A \Sigma^2 \quad (10)$$

不妨设  $[A \Sigma^2 A]_{ij} = b_{ij}$ ,  $[\Sigma^2]_{ij} = \sigma_{ij}^2$ , 展开式 (10) 有  $\sigma_i^2 b_{ij} = \sigma_j^2 b_{ji}$ . 若  $i \neq j$ ,  $b_{ij} = 0$ , 即  $A \Sigma^2 A$  为对角矩阵. 应用引理 1, 可知  $A$  为非酉置换矩阵. 又已知  $A$  为对称矩阵, 应用引理 2,  $A$  为对角矩阵, 且对角元素均为正. 则

$$\Sigma^{-1} A^{-1} \Sigma^2 A \Sigma = \Sigma^2 \quad (11)$$

并且  $P$  为矩阵  $R$  的特征向量. 将式 (11) 代入式 (8), 可知  $Q^T D Q = \Sigma^2$ , 再次应用引理 1, 则可知  $Q$  为非酉置换矩阵. 另外,  $Q^T D^{\frac{1}{2}} = \Sigma Q^T$ . 那么,

$$W = P Q^T D^{\frac{1}{2}} = \bar{P} D^{\frac{1}{2}} \quad (12)$$

其中,  $\bar{P} = P Q^T$ .  $\square$

**注 1.** 该稳定点对于任意  $p$  取值对应的算法中,  $W = \bar{P} D^{\frac{1}{2}}$  均为算法的极值点.

## 3 稳定性分析

在已知算法存在极值点的情况下, 要确定算法在极值点处的稳定性, 还得结合  $W^T(k) W(k)$  在  $F$  范数上的稳定性, 因此有如下定理.

**定理 2.** 对于加权 Möller 算法而言, 若与自相关矩阵的熵相关的范数有界  $\|D^{1-p} W^T R W\| / \|W^T W\| \leq M$ ,  $M \in \mathbf{R}$ , 且学习因子  $\eta$  足够小的情况下, 那么  $W^T(k) W(k)$  在  $F$  范数意义上是稳定性的.

**证明.** 在学习因子足够小的情况下, 小到可以忽略二阶以上项的影响,

$$\begin{aligned} W^T(k+1) W(k+1) &= \\ &= (W(k) + \eta \Delta W(k))^T (W(k) + \eta \Delta W(k)) = \\ &= W^T(k) W(k) + \eta \left\{ W^T(k) R W(k) D^{1-p} + \right. \\ &= W^T(k) W(k) W^T(k) R W(k) D^{-p} + \\ &= D^{-p} W^T(k) R W(k) W^T(k) W(k) - \\ &= 2 W^T(k) R W(k) W^T(k) W(k) D^{-p} - \\ &= 2 D^{-p} W^T(k) W(k) W^T(k) R W(k) + \\ &= \left. D^{1-p} W^T(k) R W(k) \right\} \quad (13) \end{aligned}$$

邻近时刻的  $F$  范数变化情况,

$$\begin{aligned} \frac{\|W(k+1)\|_F^2}{\|W(k)\|_F^2} &= \frac{\text{tr}(W^T(k+1) W(k+1))}{\text{tr}(W^T(k) W(k))} \approx \\ &= 1 + 2\eta \text{tr} \left\{ D^{1-p} W^T(k) R W(k) \right\} \times \\ &= \left( \text{tr} \left( W^T(k) W(k) \right) \right)^{-1} - \\ &= 2\eta \text{tr} \left\{ D^{-p} W^T(k) W(k) W^T(k) R W(k) \right\} \times \\ &= \left( \text{tr} \left( W^T(k) W(k) \right) \right)^{-1} \quad (14) \end{aligned}$$

因为  $\|D^{1-p} W^T R W\| / \|W^T W\|$  有界, 可得

$$\frac{\|W(k+1)\|_F^2}{\|W(k)\|_F^2} \begin{cases} > 1, & \text{tr}(W^T(k)W(k)) < \text{tr}(D) \\ = 1, & \text{tr}(W^T(k)W(k)) = \text{tr}(D) \\ < 1, & \text{tr}(W^T(k)W(k)) > \text{tr}(D) \end{cases} \quad (15)$$

由式 (15) 可知, 状态矩阵的  $F$  范数是收敛且有界的.  $\square$

**注 2.** 利用定理 1 可知,  $W = \bar{P}D^{\frac{1}{2}}$  为算法的稳定点, 再结合定理 2 的结论, 容易推理得,  $W = \bar{P}D^{\frac{1}{2}}$  点不仅是算法的驻点, 还是一个极小值点. 但是, 采用一阶近似方法, 如果学习因子  $\eta$  较大, 收敛速度较快, 但是容易导致数值发散. 因而, 需要讨论加权规则参数  $p$  和学习因子  $\eta$  满足速度-稳定性条件的可行域. 具体就是针对  $\eta \|D^{1-p} W^T R W\| / \|W^T W\|$  有上界, 推导出  $p$  和  $\eta$  的范围. 选取邻近可行域边界的参数时, 算法收敛速度较快, 在蒙特卡洛数值仿真中以大概率出现数值发散. 在可行域中, 选取远离边界的参数时, 算法计算时间会加长, 但是在蒙特卡洛仿真中以小概率出现数值发散.

**推论 1.** 若加权规则参数  $p$  和学习因子  $\eta$  满足有界条件  $p \in (p_{\min}, p_{\max})$ ,  $\eta \in (0, \eta_{\max})$ , 那么  $\eta \|D^{1-p} W^T R W\| / \|W^T W\|$  有界.

**证明.** 应用瑞利熵性质可知,

$$\begin{aligned} \eta \max \{ \lambda_1 d_1^{1-p}, \lambda_2 d_2^{1-p}, \dots, \lambda_r d_r^{1-p} \} &\geq \\ \eta \frac{\|D^{1-p} W^T R W\|}{\|W^T W\|} &\geq \\ \eta \min \{ \lambda_1 d_1^{1-p}, \lambda_2 d_2^{1-p}, \dots, \lambda_r d_r^{1-p} \} & \end{aligned}$$

已知, 对于确定信号而言, 自相关矩阵的特征值也是确定的, 即  $\lambda_i$  确定的, 并且假设算法中  $d_i$  确定的. 不妨设  $\delta w = \eta \lambda d^{1-p}$ , 若存在  $\max(\delta w)$  和  $\min(\delta w)$  分别在  $\eta \lambda_i d_i^{1-p}$  最大值和最小值区间以外, 则  $\eta \|D^{1-p} W^T R W\| / \|W^T W\|$  有界. 为此, 先分别对两个变量求导,

$$\frac{\partial(\delta w)}{\partial(\eta)} = \lambda d^{1-p} > 0 \quad (16)$$

$$\frac{\partial(\delta w)}{\partial(p)} = \eta(1-p)\lambda d^{-p} \begin{cases} \geq 0, & p \geq 1 \\ < 0, & p < 1 \end{cases} \quad (17)$$

默认学习因子  $\eta \in \mathbf{R}^+$ , 由上述公式易知  $\delta w$  在  $\eta \in (0, \eta_{\max})$  范围内是凸函数, 且  $p = 1$  时函数轨迹处于脊线. 那么  $\delta w_{\max} = \eta_{\max} \lambda$ ,  $\delta w_{\min} \in \{\eta_{\max} \lambda d^{1-p_{\min}}, \eta_{\max} \lambda d^{1-p_{\max}}, 0\}$ .  $\square$

**注 3.** 在实际应用中, 若  $\|W(k)\|_F^2$  改变的比率过大容易出现数值发散, 因此需要考虑对  $\delta w$  的值有一定的约束, 进而逆向推导出加权规则参数  $p$  和学习因子  $\eta$  的取值范围. 不妨令  $\delta w \in [0, 0.5]$ , 假设中  $\Sigma = \text{diag}\{[1 \ 2 \ 3]\}$  和  $D = \text{diag}\{[1 \ 2 \ 4]\}$ , 那么  $p$  和  $\eta$  的取值范围如图 1 所示. 其中点分布越密集表示距离稳定边界越远, 算法越容易稳定, 相反分布越稀疏表示距离稳定边界越近, 算法越容易发散. 此时, 对应的  $\delta w$  的变化如图 2 所示.

**注 4.** 易知改进 Möller 加权算法是所提出加权规则下的算法当  $p = 1$  时的特例. 如图 2 所示深色线条, 并且随着  $\eta$  逐渐从大到小变化,  $\delta w$  也是从大到小变化显著, 这也表明算法迭代的总体增量在不断减小, 算法的稳定性能在不断增强.

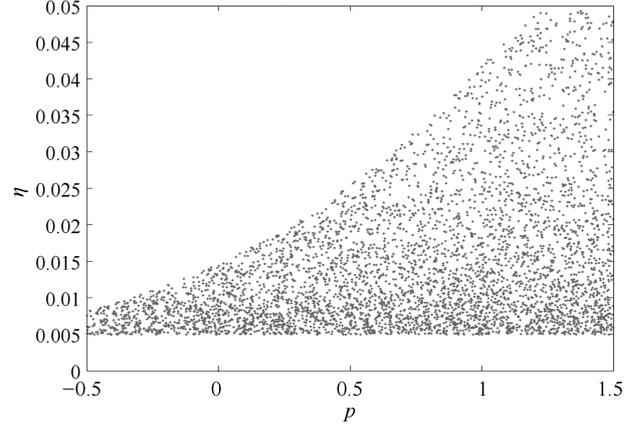


图 1  $p$  和  $\eta$  的取值范围  
Fig. 1 Ranges of  $p$  and  $\eta$

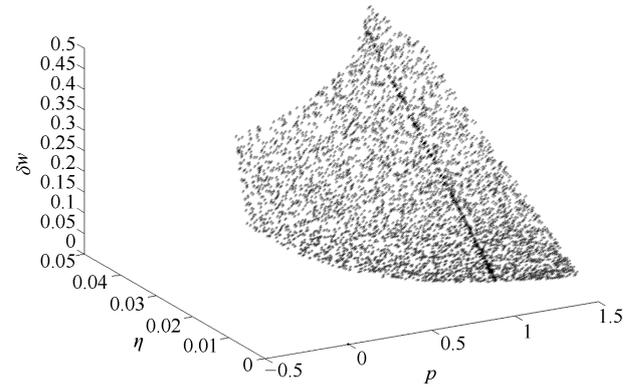


图 2  $\delta w$  的变化,  $p = 1$  处用深色标示  
Fig. 2 Change of  $\delta w$  where  $p = 1$  is marked in deep color

**注 5.** 考虑到公式  $\delta w = \eta \lambda d^{1-p}$  中包含了信号特征值  $\lambda$ , 那么对不同类型的信号, 加权规则需要有相应的改变. 这是因为不同类型的信号代表了不同的特征值, 加权规则的参数需要相应地调整. 在并行计算过程中, 多个特征值差异过大 (比如几个数量级的差别) 容易导致算法发散, 因此需要协调几个  $\delta w$  的范围. 分析图 2 可知, 为了使算法适应当前的信号, 可以通过降低  $p$  值的同时, 尽可能提高  $\eta$ , 然后对比多次测试的实验效果, 从而获得当前意义下最优的算法参数. 另外, 如果考虑采用优化过程选取合适的取值, 那么需要针对特定的优化目标和问题要求, 在本规则的可收敛稳定的范围内, 优化得出合适的取值.

#### 4 仿真实验

根据注 4 可知,  $p = 1$  时算法的稳定性已经在文献 [16] 验证. 这里主要考虑对于不同的  $p$  和  $\eta$  时算法的稳定性. 为保证算法具有可比性, 在生成数据的时候对信号特征统一规定, 认为各个仿真算例中信号的自相关矩阵特征值为为同一组特定值, 但是这组特定值为任意选取的, 在这里考虑  $\Sigma = \text{diag}\{[1 \ 2 \ 3]\}$  和  $D = \text{diag}\{[1 \ 2 \ 4]\}$ . 另外, 表征算法性能的测量指标主要是考虑方向余弦、估计模值, 特征值, 以及各个次成分的正交性. 方向余弦主要表现次成分方向的收敛速度, 表达为

$$\text{DireCou}(i, k) = \frac{\|\mathbf{w}_i^T(k)\mathbf{u}_i\|}{\|\mathbf{w}_i(k)\| \cdot \|\mathbf{u}_i\|} \quad (18)$$

其中,  $w_i$  和  $u_i$  分别是信号特征中相互对应的估计和真实的次成分, 易知在两个方向相同时该函数值为 1.

估计模值主要表现估计次成分模值变化情况, 表达为

$$Norm(i, k) = \|w_i(k)\| \quad (19)$$

其中,  $w_i$  含义同方向余弦.

特征值综合表现估计次成分表达信号特征的精度, 表达为

$$EstiMinorVal(i, k) = w_i^T(k) \hat{R}(k) w_i(k) \quad (20)$$

其中,  $\hat{R}(k)$  是通过式 (4) 估计的信号自相关矩阵.

正交性主要表现并行提取过程中各个估计次成分之间的独立性, 表达为

$$Orth(i, j, k) = w_j^T(k) w_i(k) \quad (21)$$

其中,  $i \neq j$ .

范数变化率主要表达算法在迭代过程中的波动情况, 表达为

$$\delta w(i, k) = \frac{\|w_i(k)\|_F}{\|w_i(k-1)\|_F} \quad (22)$$

仿真算法均在 MATLAB 下采用蒙特卡洛实验方法进行, 算例 1 和算例 2 中, 实验数据的生成方式与文献 [16] 相同, 实验结果是通过 20 次独立重复实验平均获得.

#### 4.1 算例 1

为表达广义加权规则的影响, 主要考虑  $p$  对算法性能的影响, 不妨固定学习因子, 令  $\eta = 0.015$ . 为说明广义加权规则的对比性, 在图 1 中  $\eta = 0.015$  对应  $p$  可取值范围内, 分别选取  $p = 0.6$  和  $p = 1.5$ , 对比这两种情况下的算法各项性能.

综合方向余弦和估计模值 (图 3 和图 4), 算法中对次成分的估计一般都在  $10^3$  附近到达稳定状态, 并且综合特征值和正交性 (图 5 和图 6) 可知, 每一个次成分都能够对应到不同的特征值, 并且最终不同的次成分互不相关. 这验证了加权广义规则下的改进 Möller 自适应算法对应不同的  $p$  值均能够完成提取次成分的任务, 不仅如此, 还能够完成同时并行提取的任务.

另外, 广义加权规则下不同的  $p$  值, 对应的  $F$  范数变化率也不相同, 具体地, 图 7 中  $F$  范数变化率对应于  $p = 0.6$  时在  $(0 \sim 8) \times 10^{-3} + 1$  范围内波动, 对应于  $p = 1.5$  时则在  $(-25 \sim 10) \times 10^{-3} + 1$  范围内波动. 两个范围相差 4 倍左右, 这体现了算法稳定程度的差异.

这个差异体现在图 3 中就是方向余弦的斜率在  $p = 0.6$  相较  $p = 1.5$  更加平滑; 在图 4 中就是估计模值的波动范围在  $p = 0.6$  相较  $p = 1.5$  更加有限; 同样, 在图 5 中就是特征值估计的波动范围在  $p = 0.6$  相较  $p = 1.5$  有限的更加明显.

这个就验证了对于不同的  $p$  值, 算法的性能不相同, 具体而言, 对于使算法在迭代中  $F$  范数的变化率的波动范围较大的  $p$  值, 对应着算法在方向余弦、模值估计和特征值估计都表现出较弱的稳定性.

#### 4.2 算例 2

为了直观体现出本文规则的效果, 在图 1 中  $\eta = 0.015$  对应  $p$  可取值范围内, 分别选取  $p = 0.6, p = 0.8, p = 1.0, p = 1.2$  和  $p = 1.5$ . 对比各个算法中平均方向余弦、平均模值变化和平均特征值.

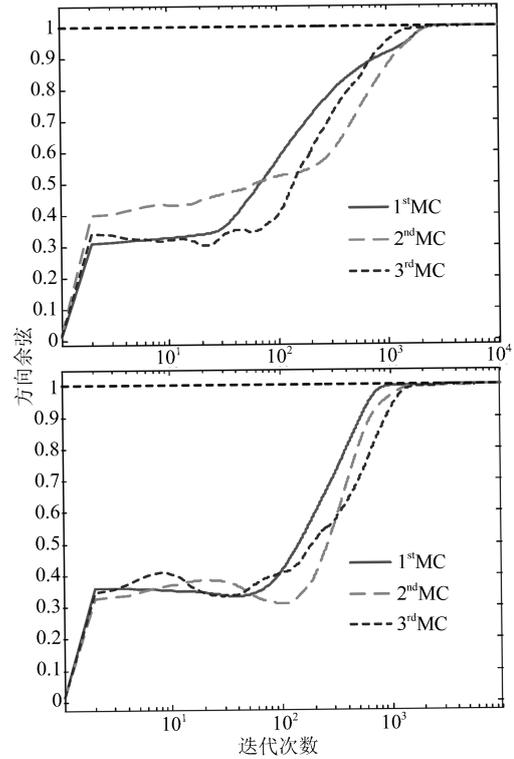


图 3 方向余弦曲线 (上为  $p = 0.6$ , 下为  $p = 1.5$ )  
Fig. 3 Curves of the direction cosine ( $p = 0.6$  is on the top and  $p = 1.5$  is at the bottom)

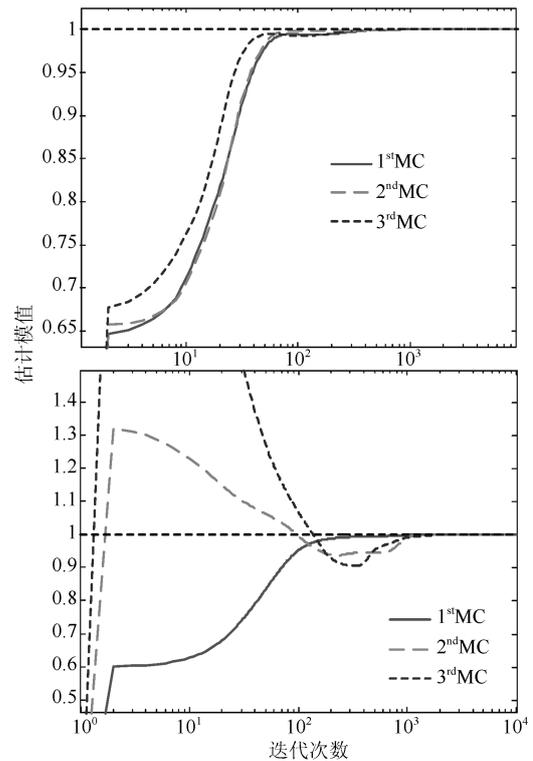


图 4 估计模值曲线 (上为  $p = 0.6$ , 下为  $p = 1.5$ )  
Fig. 4 Curves of the estimated norm ( $p = 0.6$  is on the top and  $p = 1.5$  is at the bottom)

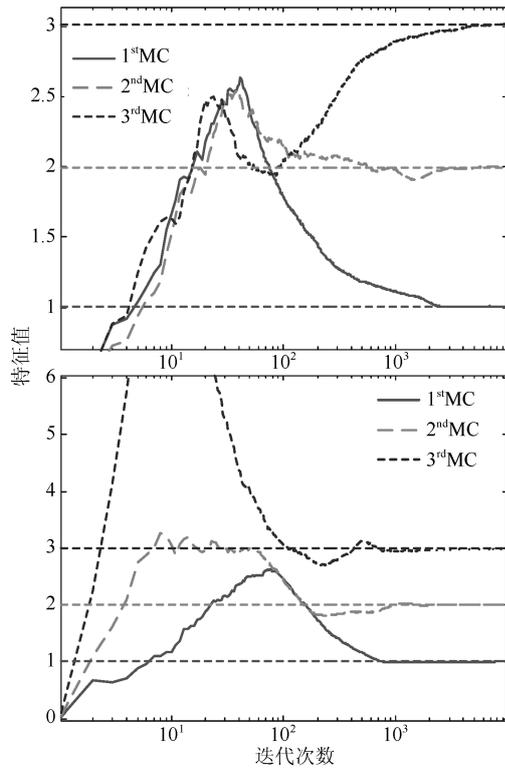


图 5 特征值变化 (上为  $p = 0.6$ , 下为  $p = 1.5$ )

Fig. 5 Curves of the eigenvalues ( $p = 0.6$  is on the top and  $p = 1.5$  is at the bottom)

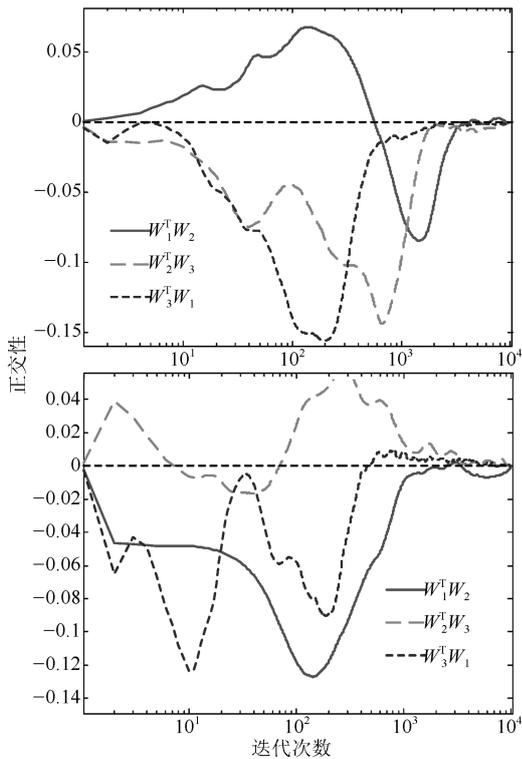


图 6 正交性变化 (上为  $p = 0.6$ , 下为  $p = 1.5$ )

Fig. 6 Curves of the orthogonality ( $p = 0.6$  is on the top and  $p = 1.5$  is at the bottom)

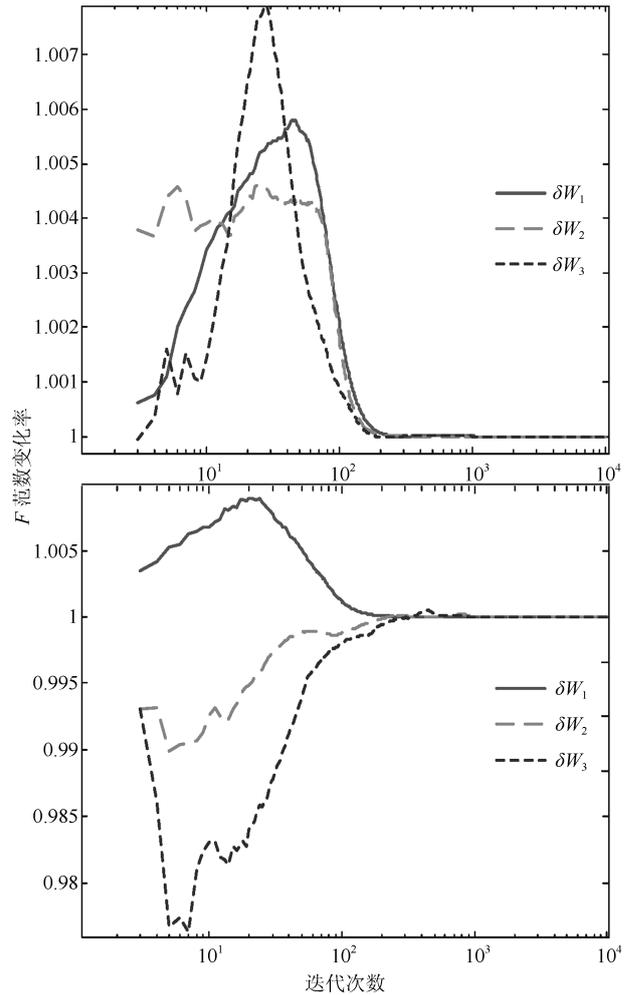


图 7  $\delta w$  值变化 (上为  $p = 0.6$ , 下为  $p = 1.5$ )

Fig. 7 Curves of  $\delta w$  ( $p = 0.6$  is on the top and  $p = 1.5$  is at the bottom)

平均方向余弦为  $r$  个次成分的方向余弦的平均值.

$$DM_p(k) = \frac{1}{r} \sum_{i=1}^r DireCou_p(i, k) \quad (23)$$

其中,  $DireCou$  与式 (18) 中含义相同.

平均估计模值为  $r$  个次成分的估计模值的平均值.

$$NM_p(k) = \frac{1}{r} \sum_{i=1}^r Norm_p(i, k) \quad (24)$$

其中,  $Norm$  与式 (19) 中含义相同.

平均特征值为  $r$  个次成分的特征值的平均值.

$$EM_p(k) = \frac{1}{r} \sum_{i=1}^r EstiMinorVal_p(i, k) \quad (25)$$

其中,  $EstiMinorVal$  与式 (20) 中含义相同.

如图 8~10 所示, 上述各个变量在收敛过程呈现出梯度变化的特点, 随着  $p$  的值从 0.6 到 1.5 变化, 方向余弦和估计模值的震荡范围逐渐变大. 可以想见, 如果输入信号逐渐变大, 在这 5 种信号中, 首先不能收敛的算法就是  $p = 1.5$  的算法, 最后不能收敛的算法是  $p = 0.6$  的算法. 这样的差异说明

在了所提出的加权规则下,  $p$  值逐渐变化, 其对应的算法性能也是逐渐变化的.

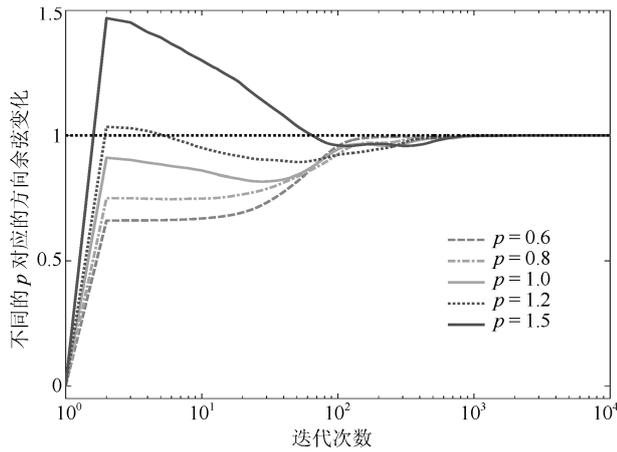


图 8 不同的  $p$  值对应的方向余弦变化

Fig. 8 Curves of the direction cosine for different  $p$

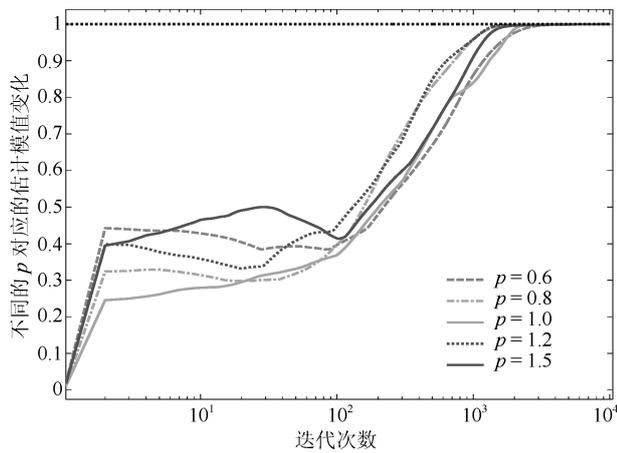


图 9 不同的  $p$  值对应的估计模值变化

Fig. 9 Curves of the estimated norm for different  $p$

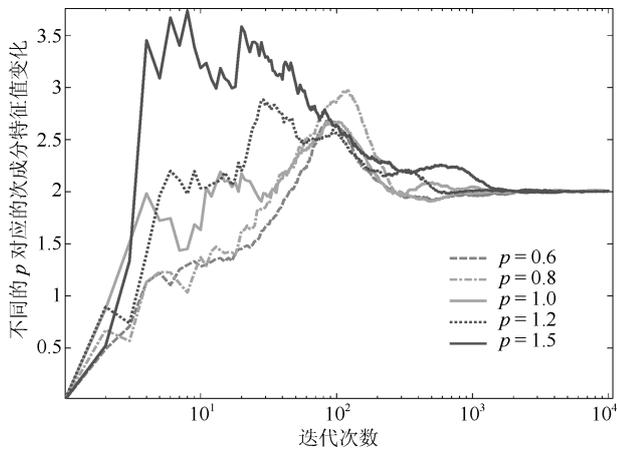


图 10 不同的  $p$  值对应的特征值变化

Fig. 10 Curves of the eigenvalues for different  $p$

4.3 算例 3

针对图片次成分信息, 检验不同的  $p$  值对应提取的次成分能力的差别. 原始的 Lenna 图像被重新分解为  $8 \times 8$  的

不重合的图像块, 然后每个图像块重新排列成 64 维的列向量. 由此可得一个 64 维的列向量集合. 对此集合去均值并归一化, 得到算法的输入数据. 采用本文提出的各个算法提出次成分, 但是只取迭代步长为 200 的估计值. 为直观体现提取次成分的差别, 取原始数据与其在次成分空间投影后数据的差, 将得到的数据恢复成图片数据. 其中  $\eta = 0.015$ . 图 11 (a) 为原始图像, 图 11 (b) ~ 图 11 (f) 分别为  $p = 0.6$  到  $p = 1.5$  的算法提取后的效果图.



图 11 去次成分的效果变化图

Fig. 11 Minor-component-reducing graphs

不妨定义重构误差为 1 减去重构图片与原始图片的 2 范数比例. 具体数据见表 1. 通过图 11 各个子图和表 1 中数据对比可知, 迭代步长未达到收敛程度时, 较大的  $p$  值对应的算法所估计的次成分偏差明显比较大. 较小的  $p$  值对应的算法收敛性能更加稳定.

表 1 不同的  $p$  对应的重构误差值

Table 1 Reconstruction error rates for different  $p$

$p$	0.6	0.8	1.0	1.2	1.5
重构误差值 (%)	16.383	20.126	22.878	29.369	31.047

5 结论

本文提出一套针对加权 Möller 算法的广义加权规则, 主要得出广义加权规则参数变化的情况下, 算法的稳定性能呈现出变化的性质. 同时, 通过常微分方程解算的方法, 结合其中关键函数性质的分析, 找出参数的变化同稳定性程度变化

的联系, 可为选择最优的加权规则参数和学习因子取值范围提供参考. 通过三个对比仿真实验验证了所提出的结论的有效性. 加权规则表明, 当前的并行 MCA 算法的适用条件还存在一个比较大的下降空间, 对于不同特点的信号类型可以使用不同的加权规则, 从而提高算法的综合性能, 改善提出次成分的效果. 并且加权规则的研究还可以推广到其他的同类别的算法中 (例如并行 Chen 算法、并行 Douglas MCA 算法以及并行 Oja-Xu 算法等), 下一步还可以推广到更加具有一般意义的加权规则的形式, 从而适用于更加广泛的并行算法改进方案中.

### References

- Gao K, Ahmad M O, Swamy M. Learning algorithm for total least-squares adaptive signal processing. *Electronics Letters*, 1992, **28**(4): 430–432.
- Feng D Z, Zheng W X, Jia Y. Neural network learning algorithms for tracking minor subspace in high-dimensional data stream. *IEEE Transaction on Neural Networks*, 2005, **16**(3): 513–521
- Pan Zong-Xu, Yu Jing, Xiao Chuang-Bai, Sun Wei-Dong. Spectral similarity-based super resolution for hyperspectral images. *Acta Automatica Sinica*, 2014, **40**(12): 2797–2807 (潘宗序, 禹晶, 肖创柏, 孙卫东. 基于光谱相似性的高光谱图像超分辨率算法. *自动化学报*, 2014, **40**(12): 2797–2807)
- Mathew G, Reddy VU, Development and analysis of a neural network approach to pisarenko's harmonic retrieval method. *IEEE Transaction on Signal Processing*, 1994, **42**(3): 663–667
- Cheng Ning, Lie Wen-Ju. Microphone array post-filter based on multi-statistical models and perceptual properties of human ears. *Acta Automatica Sinica*, 2010, **36**(1): 74–86 (程宁, 刘文举. 基于多统计模型和人耳听觉特性麦克风阵列后滤波语音增强算法. *自动化学报*, 2010, **36**(1): 74–86)
- Qi Mei-Bin, Yue Zhou-Long, Shu Kun, Jiang Jian-Guo. Multi-object tracking using hierarchical data association based on generalized correlation clustering graphs. *Acta Automatica Sinica*, 2017, **43**(1): 152–160 (齐美彬, 岳周龙, 疏坤, 蒋建国. 基于广义关联聚类的分层关联多目标跟踪. *自动化学报*, 2017, **43**(1): 152–160)
- Lei Xu, Erkki Oja, Ching Y Suen. Modified Hebbian learning for curve and surface fitting. *Neural Networks*, 1992, **5**(3): 441–457
- Cirincione Giansalvo, Cirincione Maurizio, Héroult Jeanny, Van Huffel Sabine, The MCA EXIN neuron for the minor component analysis. *IEEE Transaction on Neural Networks*, 2002, **13**(1): 160–187
- Ralf Möller. A self-stabilizing learning rule for minor component analysis. *International Journal of Neural Systems*, 2004, **14**(1): 1–8
- Kong Xiang-Yu, Hu Chang-Hua, Han Chong-Zhao. A self-stabilizing MSA algorithm in high-dimension data stream. *Neural Networks*, 2010, **23**(7): 865–871
- Kong X Y, Hu C H, Ma H G, Han C Z, A unified self-stabilizing neural network algorithm for principal and minor components extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(2): 185–198
- Chen T P, Shun I A, Murata N, Sequential extraction of minor components. *Neural Processing Letters*, 2001, **13**: 195–201
- Tan K K, Lv J H, Zhang Y, Huang S. Adaptive multiple minor directions extraction in parallel using a PCA neural network. *Theoretical Computer Science*, 2010, **411**: 4200–4215
- Thameri M, Abed-Meraim, K, Belouchrani A. Low complexity adaptive algorithms for principal and minor component analysis. *Digital Signal Processing*, 2013, **23**: 19–29
- Gao Y B, Kong X Y, Zhang H H, Hou L A, A weighted information criterion for multiple minor components and its adaptive extraction algorithms. *Neural Networks*, 2017, **89**: 1–10
- Gao Ying-Bin, Kong Xiang-Yu, Hu Chang-Hua, Hou Li-An. Modified Möller algorithm for multiple minor components extraction. *Control & Decision*, 2017, **32**(3): 493–497 (高迎彬, 孔祥玉, 胡昌华, 侯立安, 并行提取多个次成分的改进型 Möller 算法. *控制与决策*, 2017, **32**(3): 493–497)
- Kong Ling-Zhi, Gao Ying-Bin, Li Hong-Zeng, Zhang Hua-Peng. A fast algorithm that extracts multiple principal components in parallel. *Acta Automatica Sinica*, 2017, **43**(5): 835–842 (孔令智, 高迎彬, 李红增, 张华鹏. 一种快速的多个主成分并行提取算法. *自动化学报*, 2017, **43**(5): 835–842)
- Toshihisa T. Generalized weighted rules for principal components tracking. *IEEE Transactions on Signal Processing*, 2005, **53**(4): 1243–1253

杜柏阳 西安高科技研究所博士研究生. 主要研究方向为信号特征提取. E-mail: duboyangepgc@163.com (DU Bo-Yang Ph. D. candidate at the Xi'an Researching Institute of High Technology. His main research interest is signal feature extraction.)

孔祥玉 西安高科技研究所教授. 主要研究方向为多元信号分析, 信号处理. 本文通信作者. E-mail: xiangyukong01@163.com (KONG Xiang-Yu Professor at the Xi'an Researching Institute of High Technology. His research interest covers multivariate signal analysis, and signal processing. Corresponding author of this paper.)

冯晓伟 西安高科技研究所讲师. 主要研究方向为特征提取, 信号处理, 大型工业过程监控. E-mail: xiaowei121416@163.com (FENG Xiao-Wei Lecturer at the Xi'an Researching Institute of High Technology. His research interest covers feature extraction, signal processing and large-scale industrial process monitoring.)

高迎彬 中国石油大学 (北京) 信息科学与工程学院讲师. 主要研究方向为特征提取, 信号处理. E-mail: welcome8793@sina.com (GAO Ying-Bin Lecturer at School of Information Science and Engineering, China University of Petroleum (Beijing). His research interest covers feature extraction and signal processing.)

曹泽豪 西安高科技研究所硕士研究生. 主要研究方向为多元信号分析. E-mail: czh17782759319@163.com (CAO Ze-Hao Master student candidate at the Xi'an Researching Institute of High Technology. His research interest covers multivariate signal analysis.)