

## 研究速報

適応的なロバスト Super-Exponential 法を用いた  
瞬時混合ブラインド音源分離

伊藤 雅紀<sup>†a)</sup> (正員) 河本 満<sup>††,†††</sup> (正員)  
大西 昇<sup>†</sup> (正員) 井上雄二郎<sup>†††</sup> (正員)

Adaptive Robust Super-Exponential Methods for Blind Source  
Separation of Instantaneous Mixture

Masanori ITO<sup>†a)</sup>, Mitsuru KAWAMOTO<sup>††,†††</sup>,  
Noboru OHNISHI<sup>†</sup>, and Yujiro INOUE<sup>†††</sup>, Members

<sup>†</sup> 名古屋大学大学院情報科学研究科, 名古屋市

Graduate School of Information Science, Nagoya University,  
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8603 Japan

<sup>††</sup> 産業技術総合研究所情報技術研究部門, つくば市

Information Technology Research Institute, Advanced In-  
dustrial Science and Technology, Central2, 1-1-1 Umezono,  
Tsukuba-shi, 305-8568 Japan

<sup>†††</sup> 理化学研究所バイオ・ミメティックコントロール研究センター, 名古屋市  
Bio-mimetic Control Research Center, RIKEN, 2271-130  
Anagahora, Shimoshidami, Moriyama-ku, Nagoya-shi, 463-  
0003 Japan

<sup>††††</sup> 島根大学総合理工学部, 松江市

Interdisciplinary Faculty of Sciency and Engineering, Shi-  
mane University, 1060 Nishikawatsu-cho, Matsue-shi, 690-  
8504 Japan

a) E-mail: ito@ohnishi.m.is.nagoya-u.ac.jp

あらまし ブラインド音源分離の分野で, “Super-Exponential” 法 (SEM) と呼ばれるアルゴリズムが存在する。これまでに, SEM を用いたブラインド分離アルゴリズムは提案されているが, 従来の SEM では, 二次の統計量と高次の統計量を用いているためガウス雑音に対して敏感であるという欠点がある。この欠点を補うために高次の統計量のみを用いたロバスト SEM (RSEM) が提案された。そこで本論文では, RSEM を適応的に実装する手法 (ARSEM) を提案し, それらの性能を従来の高次統計量を用いる手法とシミュレーション実験で比較することにより, ARSEM の有効性を示す。

キーワード ブラインド音源分離, Super-Exponential 法, ガウス雑音, オンライン処理

### 1. まえがき

本論文では, 統計的に独立な音源からの瞬時混合に対するブラインド音源分離問題を扱う。この問題を解くために, Super-Exponential 法 (SEM) [1]~[3] を用いる。この SEM の利点は, 指数の指数乗, つまり超指数 (Super-Exponential) のオーダーで学習が収束することである。しかしながら, これまでに提案されてきた SEM の多くは, ガウス雑音に対して敏感であるという欠点があった。その理由は, 高次の統計量と

ともに二次の統計量を用いているからである。そこで, この欠点を解決するために高次のキュムラントのみを用いたガウス雑音に対して頑健な SEM (RSEM) が提案された [4]。しかしながら, [4] では, SEM との比較で, RSEM の性能を評価しており, 他の高次統計量を用いる方法との比較は, いっさい行われていない。また, RSEM は高次統計量のみを用いる方法で, その統計量を精度良く推定するためには, 多くのデータを必要とするので, オンライン処理で果たしてうまく動くかどうか不明である。本論文では, RSEM を適応的なオンライン処理で実行する ARSEM を提案し, この ARSEM と他の高次統計量を用いた手法と比較することにより, ARSEM がブラインド音源分離に適用可能であるかどうかを検証する。

## 2. Super-Exponential アルゴリズム

### 2.1 問題の定式化

本論文では, 次の  $n$  入力  $m$  出力 ( $m \leq n$ ) の時不変な多入力多出力 (MIMO) システムによって混合された信号の分離を考える。

$$y(t) = Hs(t) + n(t) \quad (1)$$

ただし,  $y(t)$  は  $m$  次元の観測信号ベクトル,  $s(t)$  は  $n$  次元の源信号ベクトル,  $H$  は  $m$  行  $n$  列の混合行列,  $n(t)$  は  $m$  次元のノイズベクトルである。

式 (1) を分離するために, 次のような  $m$  入力 1 出力 (MISO) の分離フィルタ  $w_l$  を考える。

$$z_l(t) = w_l^T y(t), \quad l = 1, 2, \dots, n \quad (2)$$

ここで,  $z_l(t)$  は  $l$  番目の出力信号で,  $w_l$  は  $m$  次元の列ベクトルである。式 (1) を式 (2) に代入すると,

$$\begin{aligned} z_l(t) &= w_l^T Hs(t) + w_l^T n(t) \\ &= g_l^T s(t) + w_l^T n(t), \quad l = 1, 2, \dots, n \end{aligned} \quad (3)$$

と書くことができる。ただし,  $g_l = H^T w_l$  であるような  $n$  次元列ベクトルである。ブラインド分離を実現するためには,  $w_l$  は次のような条件を満たす  $\tilde{w}_l$  にしなければならない。

$$\tilde{g}_l = H^T \tilde{w}_l = \tilde{\delta}_l, \quad l = 1, 2, \dots, n \quad (4)$$

$\tilde{\delta}_l$  は  $n$  次元列ベクトルで, 要素  $\delta_{lr}$  ( $r = 1, 2, \dots, n$ ) は  $\rho_l$  番目の要素を除いてすべて 0 である。 $\rho_l$  は 1 から  $n$  の整数のいずれかであり,  $\rho_i \neq \rho_j$  ( $i \neq j$ ) である。

## 2.2 RSEM

Super-Exponential 法では、ブライント分離を実現するフィルタ  $w_l$  を見つけるために、次のような 2 段階処理を行う。

$$w_l^{[1]} = R^\dagger d_l, \quad l = 1, 2, \dots, n \quad (5)$$

$$w_l^{[2]} = w_l^{[1]} / \sqrt{w_l^{[1]T} w_l^{[1]}}, \quad l = 1, 2, \dots, n \quad (6)$$

ここで、 $\dagger$  は擬似逆行列演算を表す。 $d_l$  はその要素  $d_{lj}$  が次の式で表されるような  $n$  次元列ベクトルである。

$$d_{lj} = \text{cum} \{z_l(t), z_l(t), z_l(t), y_j(t)\} \quad (7)$$

これは、 $z_l(t)$  と  $y_j(t)$  の四次のクロスキュムラントを表している。 $R$  として、従来の SEM では、観測信号の二次のキュムラントである共分散行列を用いていたが、提案手法では、次のような四次のキュムラントのみを用いる。

$$R = \sum_{i=1}^m C_{y,i} \quad (8)$$

$C_{y,i}$  は、

$$[C_{y,i}]_{q,r} = \text{cum} \{y_q(t), y_r(t), y_i(t), y_i(t)\} \quad (9)$$

である。ただし、 $[x]_{q,r}$  は行列  $x$  の  $(q, r)$  成分を表している。したがって、四次のキュムラントのみを用いているので、三次以上のキュムラントが 0 であるガウス雑音には左右されずに、分離信号を求めることができる。ガウス雑音に対して頑健であるので、Robust Super-Exponential Method (RSEM) [4] と呼ぶ。

## 2.3 ARSEM

ここで、式 (5)、(6) の 2 段階処理を適応的に実装してオンライン処理を行うことが可能な、RSEM の適応版 (ARSEM) を提案する。このとき、更新式は時間  $t$  の関数になるので、それを明記すると、式 (5)、(6) は次のように書き換えることができる。

$$w_l^{[1]}(t) = \bar{R}^\dagger(t) \bar{d}_l(t), \quad l = 1, 2, \dots, n \quad (10)$$

$$w_l^{[2]}(t) = w_l^{[1]}(t) / \sqrt{w_l^{[1]T}(t) w_l^{[1]}(t)}, \quad l = 1, 2, \dots, n \quad (11)$$

ただし、 $\bar{R}(t)$  は時刻  $t$  における  $R$  の推定量で、次の移動平均から求めることができる。

$$\bar{R}(t) = \alpha \bar{R}(t-1) + (1-\alpha) U(t) V(t) \quad (12)$$

ただし、 $V(t) = y(t) y^T(t)$  で定義される行列であり、

$$U(t) = V(t) - 2\bar{V}(t) - \text{tr}\{\bar{V}(t)\} I \quad (13)$$

$$\bar{V}(t) = \beta \bar{V}(t-1) + (1-\beta) V(t) \quad (14)$$

とする。この導出は付録に示す。同様に、 $\bar{d}_l(t)$  も次式の移動平均で推定する。

$$\bar{d}_l(t) = \alpha \bar{d}_l(t-1) + (1-\alpha) (z_l^2(t) - 3r_l(t)) z_l(t) y(t) \quad (15)$$

$$r_l(t) = \beta r_l(t-1) + (1-\beta) z_l^2(t) \quad (16)$$

ただし、 $\alpha, \beta$  は  $0 < \beta < \alpha < 1$  かつ、ともに 1 に近いような忘却係数である。 $r_l(t)$  は  $z_l(t)$  の分散を移動平均により推定した値である。

この手法では、信号を一つずつ取り出すため、取り出す信号が異なる音源になるようにする必要がある。そこで、出力信号が四次の意味で無相関になるように、次式で示す Gram-Schmidt のような直交化を行う。

$$w_i(t) = w_i(t) - \sum_{j=1}^{i-1} \frac{w_i^T(t) \bar{R}(t) w_j(t)}{w_j^T(t) \bar{R}(t) w_j(t)} w_j(t) \quad (17)$$

これを 2 段階処理の真ん中に挟むことで、取り出される信号が互いに独立であるようにする。

## 3. シミュレーション実験

ここでは  $m = n = 2$ 、すなわち、2 入力 2 出力の混合システムの分離を考える。

### 3.1 RSEM のガウス雑音への頑健性

まずは、バッチ処理で他の手法と比較して RSEM がガウス雑音に対して頑健に動作することを確認するために分離実験を行った。混合行列は、

$$H = \begin{bmatrix} 1.0 & 0.5 \\ 0.4 & 1.0 \end{bmatrix} \quad (18)$$

という値を用いて、源信号には 10 秒の長さの音声信号を用い、混合信号には雑音を付加した。その信号対雑音比 (SNR) は 5 dB であった。ただし、 $\sigma_{y_i}$  を混合信号の標準偏差、 $\sigma_n$  雑音信号の標準偏差として、 $\text{SNR} = 10 \log_{10}(\sigma_{y_i}^2 / \sigma_n^2)$  と定義する。分離フィルタの初期値は、 $w_1 = [1 \ 0]^T, w_2 = [0 \ 1]^T$  である。分離性能は multichannel intersymbol-interference (MISI) を用いて評価した。

$$\text{MISI} = \sum_{i=1}^2 \frac{\sum_{j=1}^2 g_{ij}^2 - \max_j g_{ij}^2}{\max_j g_{ij}^2}$$

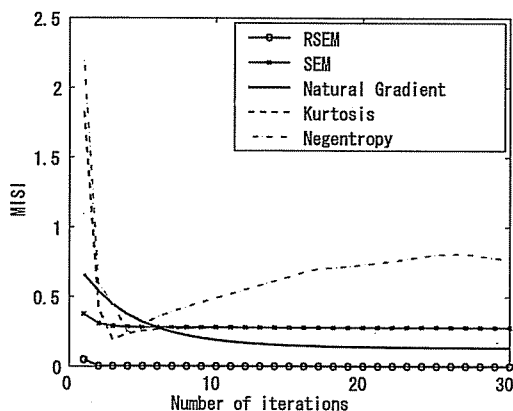


図1 バッチ処理における各分離アルゴリズムの性能比較  
Fig. 1 The performance comparison of batch processing algorithms.

$$+ \sum_{j=1}^2 \frac{\sum_{i=1}^2 g_{ij}^2 - \max_i g_{ij}^2}{\max_i g_{ij}^2} \quad (19)$$

このMISIが0に近づくほど、所望の解に近い分離フィルタが得られていることを表している。RSEMと比較するものとして、従来のSEM [3]、自然こう配アルゴリズム [5]、尖度の絶対値を最大化するFastICA [7]、ネグントロピーを最大化するFastICA [6]を選んだ。

図1に結果を示す。これは、4種類の音源を用いて、その組合せの6通りの平均の結果を示している。図中、横軸が繰返し回数、縦軸がMISIである。黒の線と○で描かれているRSEMは5dBと大きなガウス雑音が付加された場合でも、それに左右されずに、望ましい解に収束することが分かる。またそれはわずか2回のループで収束しており高速に収束していることが分かる。しかし、他のアルゴリズムでは、分離処理が収束した場合でもMISIの値は0より大きく、望ましい解には収束していない。これは、ガウス雑音が付加されたために、正しい解が得られていないからである。

### 3.2 ARSEMの性能評価

次に、提案手法であるARSEMの有効性を確認するために、オンライン処理の実験を行った。この場合も、源信号には音声信号を用いた。混合信号にはガウス雑音を付加しており、そのSNRは5, 10, ∞(dB)と変化した(SNR=∞(dB)の場合は、ノイズなし)。混合行列 $H$ は式(18)と同じである。忘却係数は $\alpha = 0.99995$ ,  $\beta = 0.9999$ とした。

ARSEMと比較するものとして、バッチ処理のときと同様に、従来のSEM [3]と、自然こう配アルゴリズ

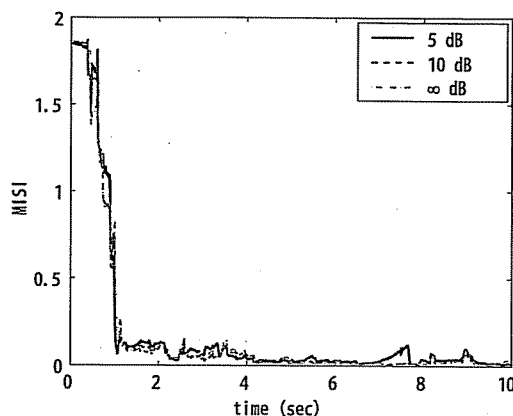


図2 ARSEMの分離性能  
Fig. 2 The separation performance of ARSEM.

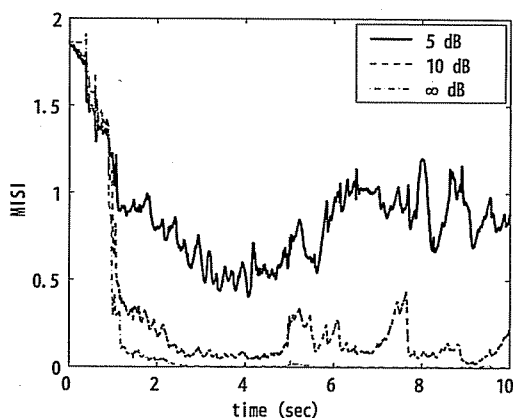


図3 SEMの分離性能  
Fig. 3 The separation performance of SEM.

ム [5]、尖度の絶対値を最大化するこう配法アルゴリズム [7]、ネグントロピーの最大化こう配法アルゴリズム [7]の分離性能を調べた。

図2から図6にARSEM、従来のSEM、自然こう配アルゴリズム、尖度最大化、ネグントロピー最大化アルゴリズムの6通りの試行の平均の結果をそれぞれ示す。図中では、横軸が時間、縦軸がMISIを表し、5dBが黒い実線、10dBが濃い灰色の破線、∞dBが薄い灰色の鎖線で描かれている。これらの結果からARSEMは、ガウス雑音に影響されることなく、分離開始から、大体2秒後には分離処理が収束し、望ましい分離フィルタを得ることが可能であることが分かる。

他の手法では、雑音のレベル (dB) が下がるに従い、正しい収束解が得られにくいことが見て取れる。従来のSEMでは、雑音を付加しないときには、ARSEM

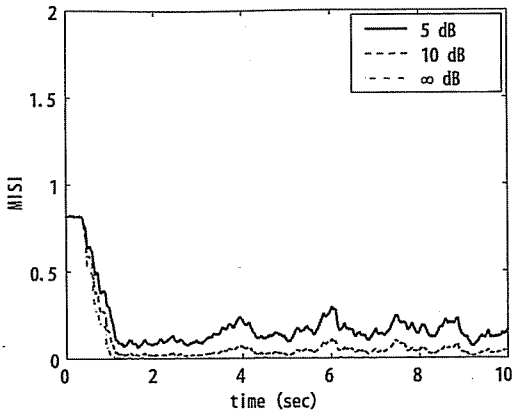


図4 自然こう配法の分離性能  
Fig.4 The separation performance of natural gradient algorithm.

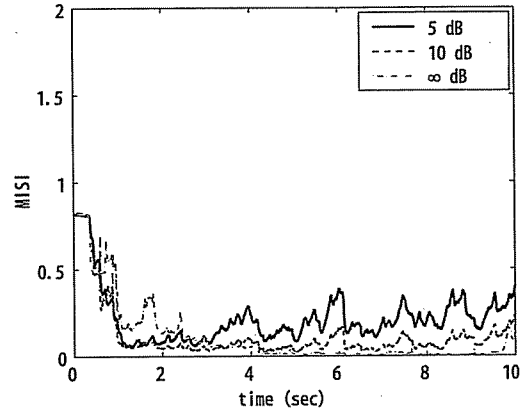


図6 ネグエントロピー最大化アルゴリズムの分離性能  
Fig.6 The separation performance of maximization of negentropy.

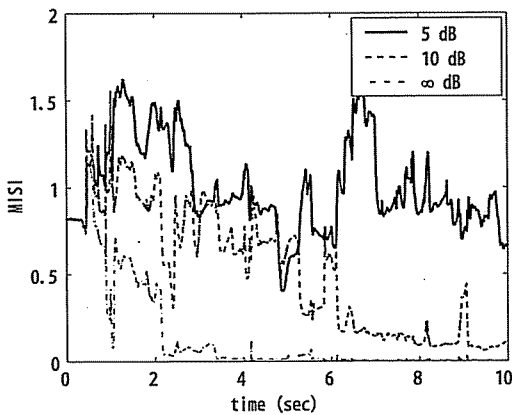


図5 尖度最大化アルゴリズムの分離性能  
Fig.5 The separation performance of maximization of kurtosis.

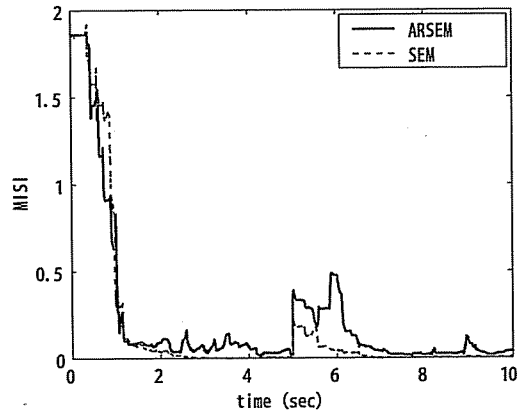


図7 混合行列が変化するときの分離性能 (ARSEM, SEM)  
Fig.7 The separation performance of time-variant system. (ARSEM, SEM)

と同程度の性能が得られていたが、雑音が付加されると正しい解が得られていない。尖度やネグエントロピーを最大化するアルゴリズムでは、白色化をしないで分離を行わないと分離性能が悪いため、オンライン処理で分離を行いながら同時に白色化処理を行った。そのときに二次統計量をもとに処理することになるが、ガウス雑音がこの処理に影響を及ぼすため、正しく分離ができないと考えられる。ただし、SNRが $\infty$  dBであるような雑音がない場合と比較すると、ARSEMで分離を行うときは、高次の統計量を求めるために多量のサンプル点を必要とするため、他のアルゴリズムに比べ、若干収束に時間がかかることも分かる。

### 3.3 混合系の変化

最後に、オンライン処理での混合系の変化への追従

性を調べるために、混合行列がある時点で変化した場合を考える。最初の5秒間は3.1や3.2と同じ式(18)の混合行列であり、5秒の時点で、

$$H = \begin{bmatrix} 1.0 & 0.3 \\ 0.6 & 1.0 \end{bmatrix} \quad (20)$$

に変化させた。このときにガウス雑音を加えないでどのように変化するか調べたものが図7、図8である。この場合も6通りの平均である。図7では、ARSEMとSEMの分離性能をそれぞれ実線と破線で表している。図8は、実線、破線、鎖線で自然こう配、尖度、ネグエントロピーを用いた手法の結果となっている。

これらの結果から、ARSEMの追従性能は、一番早

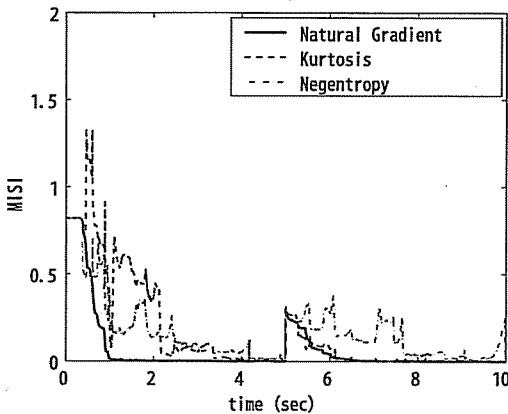


図 8 混合行列が変化するときの分離性能 (自然こう配, 尖度, ネグントロピー)

Fig. 8 The separation performance of time-variant system. (Natural Gradient, Kurtosis, Negentropy)

表 1 提案手法と他手法の比較

Table 1 Comparison of the proposed method and other methods.

Method	Convergence Speed	Robustness with Gaussian noise
ARSEM	△	○
SEM	○	×
Natural Gradient	○	△
Kurtosis	△	×
Negentropy	×	△

○ .. Good, △ .. Not so bad, × .. Bad

く学習が収束している自然こう配法より若干劣る追従性能を示しているが, ARSEM は, たとえ混合行列が変化したとしても, 十分オンラインで処理できるアルゴリズムであることが分かる. ARSEM で若干性能が劣るのは, 式 (10) で分離フィルタを求める際に, 四次のキウムラント行列と, 四次のキウムラントベクトルというように, 高次の統計量の積から計算するために, 外れ値が存在するとその影響を大きく受けるためだと思われる.

#### 4. む す び

ガウス雑音に対して頑健な Super-Exponential アルゴリズムを適応的に実装する手法を提案し, その有効性を検討した. 瞬時混合における音声信号の分離のシミュレーション実験の結果から, 提案アルゴリズムが雑音に対して頑健に, 望ましい収束解を得ることができると示した. ただし, 高次の統計量を求めるために多くのデータを必要とするため, 収束に時間がかかるという欠点も存在する. 表 1 に今回の実験結果に

ついてまとめる. 今後の課題としては, 畳込み混合への拡張などが挙げられる.

#### 文 献

- [1] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. Inf. Theory*, vol.39, no.2, pp.504-519, March 1993.
- [2] Y. Inouye and K. Tanebe, "Super-exponential algorithms for multichannel blind deconvolution," *IEEE Trans. Signal Process.*, vol.48, no.3, pp.881-888, March 2000.
- [3] M. Martone, "Fast adaptive super-exponential multistage beamforming for cellular base-station transceivers with antenna arrays," *IEEE Trans. Veh. Technol.*, vol.48, no.4, pp.1017-1028, July 1999.
- [4] M. Kawamoto, K. Kohno, and Y. Inouye, "Robust super-exponential methods for deflationary blind source separation of instantaneous mixtures," *IEEE Trans. Singal Process.*, vol.53, no.5, pp.1933-1937, May 2005.
- [5] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, ed. D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, vol.8, pp.757-763, The MIT Press, 1996.
- [6] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol.10, no.3, pp.626-634, 1999.
- [7] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

#### 付 録

式 (12) の導出を以下に示す. 簡単のため,  $m = 2$  の場合について考えると, 式 (8) の行列  $R$  は次のように表現できる.

$$R = \sum_{i=1}^2 \begin{bmatrix} \text{cum}\{y_1, y_1, y_i, y_i\} & \text{cum}\{y_1, y_2, y_i, y_i\} \\ \text{cum}\{y_2, y_1, y_i, y_i\} & \text{cum}\{y_2, y_2, y_i, y_i\} \end{bmatrix}$$

したがって,  $R$  は

$$R = \sum_{i=1}^2 \left\{ \begin{bmatrix} E[y_1^2 y_i^2] & E[y_1 y_2 y_i^2] \\ E[y_2 y_1 y_i^2] & E[y_2^2 y_i^2] \end{bmatrix} - 2 \begin{bmatrix} E[y_1 y_i] E[y_1 y_i] & E[y_1 y_i] E[y_2 y_i] \\ E[y_2 y_i] E[y_1 y_i] & E[y_2 y_i] E[y_2 y_i] \end{bmatrix} - \begin{bmatrix} E[y_1^2] E[y_i^2] & E[y_1 y_2] E[y_i^2] \\ E[y_2 y_1] E[y_i^2] & E[y_2^2] E[y_i^2] \end{bmatrix} \right\} \\ = E[yy^T yy^T] - 2E[yy^T] E[yy^T] - \text{tr}\{E[yy^T]\} E[yy^T] \quad (A.1)$$

と書き直すことができる。式 (A.1) は,  $m \geq 3$  の場合にも成り立つ。

$$R_1 = yy^T yy^T - 2E[yy^T]yy^T - \text{tr}\{E[yy^T]\}yy^T \quad (\text{A.2})$$

とすれば,  $R = E[R_1]$  が成立する。ここで,

$$R_1 = (yy^T - 2E[yy^T] - \text{tr}\{E[yy^T]\})yy^T \quad (\text{A.3})$$

となるので,  $V = yy^T$  とすれば, 式 (12) の移動平均で  $R$  を推定することができる。

(平成 17 年 8 月 10 日受付, 18 年 1 月 10 日最終原稿受付)