混合ガウスモデルを用いたオンライン・ブラインド信号分離

大 畑 正 志*・徳 成 剛*・松 岡 清 利*

Online Blind Signals Separation with Gaussian Mixture Models

Masashi OHATA*, Tsuyoshi TOKUNARI* and Kiyotoshi MATSUOKA*

This paper proposes a new online algorithm for blind separation of mixed signals. The algorithm does not any knowledge on statistical properties of the source signals. It estimates the distribution of sources based on Gaussian mixture model and separates the observed signals. Although similar algorithms can be seen in literature, they are batch type and are not applicable for nonstationary sources. Since the proposed algorithm is an online algorithm, it is applicable for nonstationary sources as well as stationary ones.

Key Words: blind separation of sources , Gaussian mixture model , online algorithm , natural learning

1. はじめに

人は、大勢の人が話している中でも特定の人の声を聞きわけることができる。このような現象をカクテル・パーティー効果という。この場合、話者から聞き手への音声の混合特性がわかっていないにもかかわらず、音声の分離が行われている。このように、複数の信号が混ざり合った観測信号のみから、原信号を推定することをブラインド信号分離という。

ブラインド信号分離の代表的なアルゴリズムとして、分離器の出力の確率密度関数が原信号の確率密度関数と一致するように分離器を求めるアルゴリズムがある^{1),5)}。このアルゴリズムは、原信号の確率分布について sub-Gaussian または super-Gaussian といった統計的性質を事前に必要とする。これに対して原信号の統計的性質を必要とせず、分離器出力の高次統計量を用いるアルゴリズムも提案されている。その一つは、分離器出力の高次統計量から原信号が sub-Gaussian 信号または super-Gaussian 信号であるかを推定し、信号分離を実現するものである⁶⁾。もう一つは、分離器出力の確率密度関数を高次統計量を用いて表現し、その高次統計量を求めることで出力の確率分布を推定し信号分離を実現している³⁾。

その他に、原信号の確率分布にパラメータを含んだモデルを採用し、このパラメータの推定と分離器の決定を同時に行うアルゴリズムが提案されている。Xiao ら⁸⁾は確率分布のモデルとして混合ガウスモデルを採用し、EM アルゴリズムによりパラメータを推定している。EM アルゴリズムはパラ

メータをバッチ的に推定するため、非定常な信号を含んでいる場合には適さない。

本論文では、オンラインで原信号の確率密度関数の推定と分離器の決定を同時に行うアルゴリズムを提案する。原信号の確率分布の推定には、Xiao らと同様に混合ガウスモデルを使用し、確率勾配法および natural learning⁷⁾の考え方に従って、そのパラメータをオンラインで推定する。そのようにして得られるアルゴリズムは、不安定現象を起こし、分離器の出力が乱れる。この現象の考察とその対策を示し、修正後のアルゴリズムが非定常な原信号を含む場合においても有効であることをコンピュータシミュレーションにて示す。

2. 問 題

N 個の信号源からそれぞれ信号 $s_1(t), \dots, s_N(t)$ ($t = \dots, -1, 0, 1, \dots$) が発生し (以後、原信号と呼ぶ)、同数のセンサにより混ざり合った信号 $x_1(t), \dots, x_N(t)$ ($t = \dots, -1, 0, 1, \dots$) を観測するものとする (以後、観測信号と呼ぶ)。このとき原信号は平均値 0 の非ガウス信号であり、統計的に互いに独立であると仮定する。原信号と観測信号の列ベクトル表現をそれぞれ $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$, $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ とする (ここで、 T は行列あるいはベクトルの転置を表す)。

$\mathbf{s}(t)$ が $\mathbf{x}(t)$ に変換される過程は、各原信号が定数倍され、加算される混合過程であるとすれば、 $N \times N$ 正則行列 $\mathbf{A} = [a_{ij}]$ を用いてつぎのように表現できる (以後、混合行列と呼ぶ)。

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

原信号 $s_i(t)$ の確率密度関数を $P_{S_i}(s_i)$ としたとき、 $\mathbf{s}(t)$ の結合確率密度関数 $P_S(\mathbf{s})$ は、信号の独立性から

* 九州工業大学工学部 北九州市戸畑区水仙町 1-1

* Kyushu Institute of Technology, Sensui-sho 1-1, Tobata-ku, Kitakyushu City

(Received October 8, 1999)

(Revised August 24, 2000)

$$P_S(s) = \prod_{i=1}^N P_{S_i}(s_i) \quad (2)$$

である。

混合過程 (1) に対して、次式により原信号を推定する。

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \quad (3)$$

ここで、 $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]^T$ は $N \times N$ 行列である (以後、分離行列と呼ぶ)。 $\mathbf{y}(t) = [y_1(t), \dots, y_N(t)]^T$ は分離器の出力を表し、第 j 番目の出力 $y_j(t)$ ($j = 1, \dots, N$) は、 $y_j(t) = \mathbf{w}_j^T \mathbf{x}(t)$ で与えられる。このとき、 $\mathbf{x}(t)$ の確率密度関数 $P_X(\mathbf{x})$ と $\mathbf{W}$ を用いれば、 $\mathbf{x}(t)$ の確率密度関数 $P_Y(\mathbf{y})$ は、

$$P_Y(\mathbf{y}) = |\det \mathbf{W}|^{-1} P_X(\mathbf{x}) \quad (4)$$

である。

分離行列 $\mathbf{W}$ が混合行列 $\mathbf{A}$ の逆行列となると、分離器の出力 $\mathbf{y}(t)$ は $\mathbf{s}(t)$ になる。しかし、一般にブラインド信号分離には、本質的に避けることのできない2つの曖昧さがある。一つは信号源と分離器の出力の対応の曖昧さで、もう一つは原信号の大きさについての曖昧さである。よって、ブラインド信号分離は DPA^{-1} を推定することである。行列 $\mathbf{D}$ および $\mathbf{P}$ は、それぞれ任意の正則対角行列および任意の順列行列を表す。

3. Bell の方法

本論文で提案するアルゴリズムは、Bell and Sejnowski¹⁾ (以下、Bell の方法と呼ぶ) の方法と関連深いので、その方法について簡単に説明する。

原信号の確率密度関数 $P_S(s)$ が事前にわかっていると仮定したとき、出力 $\mathbf{y}(t)$ の確率密度関数 $P_Y(\mathbf{y})$ を $P_S(s)$ にする $\mathbf{W}$ を決定すれば、分離器は原信号 $\mathbf{s}(t)$ を出力する。そこで、確率密度関数 $P_Y(\mathbf{y})$ と $P_S(s)$ の近さの尺度として、Kullback-Leibler のダイバージェンスを採用し、これを損失関数 $L_B(\mathbf{W})$ とする。

$$L_B(\mathbf{W}) = \int P_Y(\mathbf{y}) \log \frac{P_Y(\mathbf{y})}{P_S(\mathbf{y})} d\mathbf{y} \quad (5)$$

この関数は常に非負であり、 $P_Y(\mathbf{y}) = P_S(\mathbf{y})$ となると、最小値0をとる。したがって、これを $\mathbf{W}$ について最小化すれば、目的の分離行列 $\mathbf{W}^*$ を求めることができる。

式 (4) から関数 $L_B(\mathbf{W})$ は、

$$L_B(\mathbf{W}) = -\log |\det \mathbf{W}| - \sum_{i=1}^N E_X[\log P_{S_i}(\mathbf{w}_i^T \mathbf{X})] + E_X[\log P_X(\mathbf{X})]$$

と表現できる。ここで、 $E_X[\cdot]$ は確率密度関数 $P_X(\mathbf{x})$ についての期待値を表す。上式右辺第3項は、 $\mathbf{W}$ に無関係な定数であるから無視してよい。

時刻 t における分離行列 $\mathbf{W}^*$ の推定量を $\mathbf{W}(t)$ とし、その修正量を $\Delta \mathbf{W}(t) = \mathbf{W}(t+1) - \mathbf{W}(t)$ とする。損失関数

$L_B(\mathbf{W})$ の $\mathbf{W}$ についての最小化に最急降下法を用いると、 $\mathbf{W}(t)$ の学習則は次式ようになる。

$$\Delta \mathbf{W}(t) = \alpha \{ [\mathbf{W}(t)^T]^{-1} - E_X[\varphi(\mathbf{Y}(t)) \mathbf{X}(t)^T] \}$$

ただし、列ベクトル $\varphi(\mathbf{y})$ は、

$$\varphi(\mathbf{y}) \equiv [\varphi_1(y_1), \dots, \varphi_N(y_N)]^T$$

$$\varphi_i(y_i) \equiv -\frac{d}{dy_i} \log P_{S_i}(y_i)$$

であり、 α は学習係数を表し小さな正数である。これに確率勾配法を適用して期待値を瞬時値に置き換えれば、学習則は

$$\Delta \mathbf{W}(t) = \alpha \{ [\mathbf{W}(t)^T]^{-1} - \varphi(\mathbf{y}(t)) \mathbf{y}(t)^T \} \quad (6)$$

となる。さらに、natural learning を適用して⁷⁾、式 (6) の右辺に正定値行列 $\mathbf{W}(t)^T \mathbf{W}(t)$ を右から掛けたものを修正量とすれば、学習則：

$$\Delta \mathbf{W}(t) = \alpha \{ \mathbf{I} - \varphi(\mathbf{y}(t)) \mathbf{y}(t)^T \} \mathbf{W}(t) \quad (7)$$

を得る。ここで、 $\mathbf{I}$ は $N \times N$ の単位行列である。文献 1) と 5) では、関数 $\varphi_i(y_i)$ ($i = 1, \dots, N$) として、原信号 $s_i(t)$ が super-Gaussian 信号であるときは、 $\varphi_i(y_i) = \tanh(y_i/2)$ を、sub-Gaussian 信号であるときは、 $\varphi_i(y_i) = y_i^3$ を使用している。

4. 原信号の確率密度関数の推定と分離行列の決定

4.1 損失関数

一般に、原信号の確率密度関数が事前にはわからないから、パラメータを含んだモデルを導入し、パラメータの推定と分離行列の決定を行う必要がある。このとき、確率密度関数のモデルはいろいろな確率分布を表現でき、簡単であることが望ましい。そこで、本論文では混合ガウスモデルを採用する。

混合ガウスモデル $Q_{S_i}(s_i|\theta_i)$ は、次式で与えられる。

$$Q_{S_i}(s_i|\theta_i) = \sum_{k=1}^K c_{ik} \phi(s_i|\mu_{ik}, \sigma_{ik}) \quad (8)$$

$$\phi(s_i|\mu_{ik}, \sigma_{ik}) = \frac{1}{\sqrt{2\pi}\sigma_{ik}} \exp \left\{ -\frac{(s_i - \mu_{ik})^2}{2\sigma_{ik}^2} \right\} \quad (9)$$

ただし、 θ_i は、

$$\theta_i = [c_{i1}, \dots, c_{iK}, \mu_{i1}, \dots, \mu_{iK}, \sigma_{i1}, \dots, \sigma_{iK}]$$

である。 c_{ik} は $\sum_{k=1}^K c_{ik} = 1, c_{ik} \geq 0$ を、 σ_{ik} は $\sigma_{ik} > 0$ を満たすものとする。

確率密度関数 $P_{S_i}(s_i)$ をモデル $Q_{S_i}(s_i|\theta_i)$ で置き換えれば、原信号 $\mathbf{s}(t)$ の確率密度関数 $P_S(s)$ は、

$$Q_S(s|\theta) \equiv \prod_{i=1}^N Q_{S_i}(s_i|\theta_i) \quad (10)$$

に置き換わることになる。ただし、 $\theta = \{\theta_1, \dots, \theta_N\}$ とする。損失関数 $L_B(\mathbf{W})$ において、この置き換えを行い、それを $L_1(\mathbf{W}, \theta)$ とすれば、

$$L_1(\mathbf{W}, \theta) \equiv \int P_Y(\mathbf{y}|\mathbf{W}) \log \frac{P_Y(\mathbf{y}|\mathbf{W})}{Q_S(\mathbf{y}|\mathbf{W}, \theta)} d\mathbf{y} \quad (11)$$

を得る。これを $\mathbf{W}$ と θ について最小化し、信号分離を実現する。文献8)では、EM アルゴリズムを含んだバッチ処理によりこの損失関数を最小化しているが、本論文では確率勾配法を適用しオンラインで最小化する。

$\mathbf{w}_i^T$, μ_{ik} および σ_{ik} ($k = 1, \dots, K$) が、それぞれ $b_i \mathbf{w}_i^T$, $b_i \mu_{ik}$ および $b_i \sigma_{ik}$ に変化したとき、損失関数の値は変化しない。だから、 $\mathbf{w}_i$ のスケールすなわち $y_i(t)$ の大きさは決まらない。 $Q_{S_i}(s_i|\theta_i)$ の分散が一定のもとで、 $\mathbf{w}_i$ と θ_i を推定すれば、このスケールの不確実性は解決するであろう。しかし、この分散は次式に示すように複雑であるから、オンラインアルゴリズムに適さない。

$$E_{S_i}[S_i^2] = \sum_{k=1}^K c_{ik} (\mu_{ik}^2 + \sigma_{ik}^2) - \left(\sum_{k=1}^K c_{ik} \mu_{ik} \right)^2$$

そこで、 θ_i に対してではなく出力 $y_i(t)$ に対して拘束条件を設ける。本論文では、拘束条件として $E_{Y_i}[Y_i^2] = 1$ を採用する。 $y_i(t)$ ($i = 1, \dots, N$) が統計的に互いに独立であるとき、この拘束条件は $\mathbf{y}(t)$ の共分散行列 $E_{\mathbf{Y}}[\mathbf{Y}\mathbf{Y}^T]$ が単位行列 $\mathbf{I}$ になることに等しい。これは次に示す関数 $L_2(\mathbf{W})$ が最小になるとき実現される。

$$L_2(\mathbf{W}) \equiv \int P_{\mathbf{Y}}(\mathbf{y}|\mathbf{W}) \log \frac{P_{\mathbf{Y}}(\mathbf{y}|\mathbf{W})}{\Phi(\mathbf{y})} d\mathbf{y} \quad (12)$$

ここで、 $\Phi(\mathbf{y})$ は平均ベクトル $\mathbf{0}$ で共分散行列 $\mathbf{I}$ の N 次元のガウス分布を表す。

$$\Phi(\mathbf{y}) \equiv \frac{1}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}\mathbf{y}^T \mathbf{y}\right) \quad (13)$$

関数 $L_2(\mathbf{W})$ は $P_{\mathbf{Y}}(\mathbf{y}|\mathbf{W})$ と $\Phi(\mathbf{y})$ の Kullback-Leibler ダイバージェンスで、常に非負である。これが最小となるのは、出力 $\mathbf{y}(t)$ の共分散行列が $\mathbf{I}$ になるときである。また、 $L_2(\mathbf{W})$ は 2 次統計量のみを含むから、この関数の最小化だけで $\mathbf{W}$ を決定することはできない。そこで、 $L_1(\mathbf{W}, \theta)$ と $L_2(\mathbf{W})$ を同時に最小にする $\mathbf{W}$ と θ を求める。これは、 $L_G(\mathbf{W}, \theta) \equiv L_1(\mathbf{W}, \theta) + \eta L_2(\mathbf{W})$ を $\mathbf{W}$ と θ について最小化することで実現できる。ただし、 η は正数とする。

$$L_G(\mathbf{W}, \theta) = \int P_{\mathbf{Y}}(\mathbf{y}|\mathbf{W}) \log \frac{P_{\mathbf{Y}}(\mathbf{y}|\mathbf{W})}{Q_{\mathbf{S}}(\mathbf{y}|\mathbf{W}, \theta)} d\mathbf{y} + \eta \int P_{\mathbf{Y}}(\mathbf{y}|\mathbf{W}) \log \frac{P_{\mathbf{Y}}(\mathbf{y}|\mathbf{W})}{\Phi(\mathbf{y})} d\mathbf{y} \quad (14)$$

これを評価することにより、 $L_1(\mathbf{W}, \theta)$ と $L_2(\mathbf{W})$ をそれぞれ最小にする $\mathbf{W}$ が定まり、それにもなって θ が決まる。

4.2 パラメータに対する natural learning

Amari⁷⁾ が提案した natural learning を関数 $L_G(\mathbf{W}, \theta)$ の最小化に適用する。この方法は、正則行列の空間および確率密度関数の全体に対して適当な計量を導入して、効率的な学習を行うアルゴリズムである。

本論文では、 $\mathbf{W}$ と θ についての natural learning を別々に求める。 $\mathbf{W}$ に対しては、3 章と同様に求める。 θ に対しては、以下のように求める。

損失関数 $L_G(\mathbf{W}, \theta)$ を θ について整理すると次式のようにになる。

$$L_G(\mathbf{W}, \theta) = - \sum_{i=1}^N E_{Y_i} [\log Q_{S_i}(Y_i|\theta_i)] + (1 + \eta) E_{\mathbf{Y}} [\log P_{\mathbf{Y}}(\mathbf{Y})] - \eta E_{\mathbf{Y}} [\log \Phi(\mathbf{Y})]$$

$\mathbf{W}$ は一定であると仮定すれば、第 1 項以外は θ に無関係であるから、第 1 項をパラメータ θ について最小化すればよい。また、パラメータ θ_i と θ_m ($m \neq i$) は、独立なパラメータであるから、 θ_i についての natural learning を導けばよい。そこで、各 θ_i についての損失関数 $l_i(\theta_i)$ を次式のように定める。

$$l_i(\theta_i) \equiv -E_{Y_i} [\log Q_{S_i}(Y_i|\theta_i)] \quad (15)$$

この関数が最小になる θ_i^* を求めるとき、パラメータ $\{c_{ik}|k=1, \dots, K\}$ は、拘束条件 $\sum_{k=1}^K c_{ik} = 1$ を満たさなければならない。したがって、このパラメータの自由度は、 $(K-1)$ である。そこで、 c_{ik} ($k=1, \dots, K-1$) を自由パラメータとして、 c_{iK} を $c_{iK} = 1 - \sum_{k=1}^{K-1} c_{ik}$ で置き換えて、 c_{ik} ($k=1, \dots, K-1$)、 μ_{ik} および σ_{ik} ($k=1, \dots, K$) について学習則を求める。その後、拘束条件 $\sum_{k=1}^K c_{ik} = 1$ から、 c_{iK} についての学習則を求める。

パラメータベクトル θ_i' を

$$\theta_i' \equiv [c_{i1}, \dots, c_{i(K-1)}, \mu_{i1}, \dots, \mu_{iK}, \sigma_{i1}, \dots, \sigma_{iK}]$$

とし、これに対するフィッシャーの情報行列を $I_{S_i}(\theta_i')$ とする。この行列を確率密度関数

$$Q_{S_i}(s_i|\theta_i') = \sum_{k=1}^{K-1} c_{ik} \{ \phi(s_i|\mu_{ik}, \sigma_{ik}) - \phi(s_i|\mu_{iK}, \sigma_{iK}) \} + \phi(s_i|\mu_{iK}, \sigma_{iK})$$

からなる多様体における計量行列として定義し、そのもとでの最急降下法を考える。 θ_i' および c_{iK} の修正量をそれぞれ $\Delta\theta_i'$ および Δc_{iK} とすれば、パラメータ θ_i に対する natural learning は、次式のようになる。

$$\Delta\theta_i' \propto I_{S_i}(\theta_i')^{-1} \frac{\partial}{\partial\theta_i'} \log Q_{S_i}(y_i|\theta_i')$$

$$\Delta c_{iK} \propto - \sum_{k=1}^{K-1} \Delta c_{ik}$$

しかし、混合ガウスモデルのフィッシャーの情報行列 $I_{S_i}(\theta_i')$ は非常に複雑であるから、オンラインアルゴリズムに適さない。そこで $I_{S_i}(\theta_i')$ の代わりに、他の計量行列を用いることにする。

混合ガウスモデルのパラメータは EM アルゴリズムにより推定できることが、よく知られている。このアルゴリズムでは、確率変数 S_i は確率 c_{ik} で選択されたガウス分布 $\phi(s_i|\mu_{ik}, \sigma_{ik})$ に従うと考えている。この考え方に従い、確率変数 $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iK})$ を導入する。

$$\mathbf{Z}_i \in \{e_1, \dots, e_K\}$$

$$e_k = (0, \dots, 0, 1, 0, \dots, 0) \quad (k = 1, \dots, K)$$

この確率変数がガウス分布の選択を担い、 $\phi(s_i|\mu_{ik}, \sigma_{ik})$ を選択する確率は $\Pr(\mathbf{Z}_i = e_k) = c_{ik} \quad (k = 1, \dots, N)$ である。確率変数 $S_i, \mathbf{Z}_i$ の結合確率密度関数 $Q_{S_i, \mathbf{Z}_i}(s_i, \mathbf{z}_i|\theta'_i)$ は、

$$Q_{S_i, \mathbf{Z}_i}(s_i, \mathbf{z}_i|\theta'_i) = \sum_{k=1}^{K-1} c_{ik} z_{ik} \phi(s_i|\mu_{ik}, \sigma_{ik}) + z_{iK} \left(1 - \sum_{k=1}^{K-1} c_{ik}\right) \phi(s_i|\mu_{iK}, \sigma_{iK})$$

である。そのフィッシャーの情報行列 $I_{S_i, \mathbf{Z}_i}(\theta'_i)$ は、 $I_{S_i}(\theta'_i)$ に比べて容易に計算できるから、これを natural learning の計量行列に採用する。確率密度関数 $Q_{S_i, \mathbf{Z}_i}(s_i, \mathbf{z}_i|\theta'_i)$ についてのフィッシャーの情報行列は、次式のようになる。

$$I_{S_i, \mathbf{Z}_i}(\theta'_i) = \begin{bmatrix} \mathbf{C}_i & 0 & 0 \\ 0 & \mathbf{M}_i & 0 \\ 0 & 0 & \mathbf{\Sigma}_i \end{bmatrix} \in \mathbf{R}^{D \times D}$$

($D = 3K - 1$)

ただし、行列 $\mathbf{C}_i$ 、 $\mathbf{M}_i$ および $\mathbf{\Sigma}_i$ は、

$$\mathbf{C}_i \equiv \begin{bmatrix} \frac{1}{c_{i1}} + \frac{1}{c_{iK}} & \frac{1}{c_{iK}} & \dots & \frac{1}{c_{iK}} \\ \frac{1}{c_{iK}} & \ddots & & \vdots \\ \vdots & & \ddots & \frac{1}{c_{iK}} \\ \frac{1}{c_{iK}} & \dots & \frac{1}{c_{iK}} & \frac{1}{c_{i(K-1)}} + \frac{1}{c_{iK}} \end{bmatrix}$$

$\in \mathbf{R}^{(K-1) \times (K-1)}$

$$\mathbf{M}_i \equiv \text{diag} \left\{ \frac{c_{i1}}{\sigma_{i1}^2}, \dots, \frac{c_{iK}}{\sigma_{iK}^2} \right\} \in \mathbf{R}^{K \times K}$$

$$\mathbf{\Sigma}_i \equiv \text{diag} \left\{ \frac{2c_{i1}}{\sigma_{i1}^2}, \dots, \frac{2c_{iK}}{\sigma_{iK}^2} \right\} \in \mathbf{R}^{K \times K}$$

である。diag $\{\dots\}$ は、括弧内の要素を対角要素に持ち、そのほかの要素はすべて0である対角行列を表す。この情報行列を計量行列にしたとき、 θ'_i と c_{iK} についての学習則は、それぞれ

$$\Delta \theta'_i \propto I_{S_i, \mathbf{Z}_i}(\theta'_i)^{-1} \frac{\partial}{\partial \theta'_i} \log Q_{S_i}(y_i|\theta'_i)$$

$$\Delta c_{iK} \propto - \sum_{k=1}^{K-1} \Delta c_{ik}$$

である。

5. 確率密度関数の推定を含むアルゴリズム

これまでのことから、損失関数 $L_G(\mathbf{W}, \theta)$ をパラメータ θ と分離行列 $\mathbf{W}$ について最小化するオンラインアルゴリズムを具体的な数式で示す。 σ_{ik} は常に正でなければならないから、 $\sigma_{ik} = \exp(\xi_{ik})$ と変数変換し、 ξ_{ik} に対して学習則を求める。いま、時刻 t における分離行列とパラメータの推定量をそれぞれ $\mathbf{W}(t)$ 、 $c_{ik}(t)$ 、 $\mu_{ik}(t)$ および $\xi_{ik}(t)$ とする。 $d\sigma_{ik}/d\xi_{ik} = \sigma_{ik}$ を考慮すると、各パラメータの学習則はそ

れぞれ

$$\Delta \mathbf{W}(t) = \beta_W \mathbf{f}(t) \mathbf{W}(t) \quad (16)$$

$$\mathbf{f}(t) = (\eta + 1) \mathbf{I} - \{\varphi(t) + \eta \mathbf{y}(t)\} \mathbf{y}(t)^T$$

$$\Delta c_{ik}(t) = \beta_\theta \{h_{ik}(t) - c_{ik}(t)\} \quad (17)$$

$$\Delta \mu_{ik}(t) = \beta_\theta \frac{h_{ik}(t)}{c_{ik}(t)} \{y_i(t) - \mu_{ik}(t)\} \quad (18)$$

$$\Delta \xi_{ik}(t) = \beta_\theta \frac{h_{ik}(t)}{2c_{ik}(t)} \left\{ \frac{[y_i(t) - \mu_{ik}(t)]^2}{e^{2\xi_{ik}(t)}} - 1 \right\} \quad (19)$$

($i = 1, \dots, N; k = 1, \dots, K$)

となる。ただし、 $h_{ik}(t)$ 、および $\varphi(t)$ は次式を表す。

$$h_{ik}(t) \equiv \frac{c_{ik}(t) \phi(y_i(t) | \mu_{ik}(t), e^{\xi_{ik}(t)})}{\sum_{l=1}^K c_{il}(t) \phi(y_i(t) | \mu_{il}(t), e^{\xi_{il}(t)})} \quad (20)$$

$$\mathbf{h}_i(t) \equiv [h_{i1}(t), \dots, h_{iK}(t)]^T \quad (21)$$

$$\mathbf{r}_i(t) \equiv [r_{i1}(t), \dots, r_{iK}(t)]^T \quad (22)$$

$$r_{ik}(t) \equiv e^{-2\xi_{ik}(t)} (y_i(t) - \mu_{ik}(t)) \quad (23)$$

$$\varphi(t) \equiv [\mathbf{h}_1(t)^T \mathbf{r}_1(t), \dots, \mathbf{h}_N(t)^T \mathbf{r}_N(t)]^T \quad (24)$$

($i = 1, \dots, N; k = 1, \dots, K$)

そして、 β_W は $\mathbf{W}(t)$ に対する学習係数、 β_θ は c_{ik} 、 μ_{ik} および ξ_{ik} ($i = 1, \dots, N; k = 1, \dots, K$) の学習係数を表す。

初期値 $c_{ik}(0)$ ($i = 1, \dots, N; k = 1, \dots, K$) が、

$$\sum_{k=1}^K c_{ik}(t) = 1, 0 \leq c_{ik}(t) \quad (25)$$

($i = 1, \dots, N; k = 1, \dots, K$)

を満たせば、 $h_{ik}(0)$ は $0 \leq h_{ik}(0) \leq 1$ を満たす。 $0 < \beta_\theta < 1$ に対して、任意の時刻 t において学習則 (17) は拘束条件 (25) を常に満たす。

6. シミュレーション

前章で求めたアルゴリズムを用いて、信号源が2つの場合についてシミュレーションを行った。混合行列 $\mathbf{A}$ を

$$\mathbf{A} = \begin{bmatrix} 2.0 & 1.0 \\ -1.0 & 0.6 \end{bmatrix}$$

に設定した。原信号 $s_1(t)$ は確率密度関数

$$P_{S_1}(s_1) = \frac{1}{8} \phi(s_1| -2, 0.2) + \frac{3}{4} \phi(s_1|0, 0.2) + \frac{1}{8} \phi(s_1|2, 0.2)$$

に従う super-Gaussian 信号を、 $s_2(t)$ は 確率密度関数

$$P_{S_2}(s_2) = \frac{1}{4} \phi(s_2| -3, 0.2) + \frac{1}{4} \phi(s_2| -1, 0.2) + \frac{1}{4} \phi(s_2|1, 0.2) + \frac{1}{4} \phi(s_2|3, 0.2)$$

に従う sub-Gaussian 信号を用いた。 $Q_{S_i}(s_i|\theta_i)$ は $K = 5$ の混合ガウスモデルを用いた。さらに、区間 $(-1, 1)$ の一様分布から取り出した乱数を分離行列の各要素の初期値に設定した。 θ に対する初期値は、確率密度関数 $Q_{S_i}(s_i|\theta_i)$ ($i = 1, 2$)

が区間 $(-2.5, 2.5)$ で一様分布に近くなるように設定した。

$$\begin{aligned} c_{ik}(0) &= 0.2 \\ \xi_{ik}(0) &= -\log(2.0\sqrt{2.0\log(2.0)}) \\ \mu_{i1}(0) &= -2.0, \mu_{i2}(0) = -1.0, \mu_{i3}(0) = 0.0 \\ \mu_{i4}(0) &= 1.0, \mu_{i5}(0) = 2.0 \\ (i &= 1, 2; k = 1, \dots, 5) \end{aligned}$$

$\mathbf{W}$ と θ の学習係数をそれぞれ $\beta_W = 2.5 \times 10^{-3}$, $\beta_\theta = 5.0 \times 10^{-3}$ に設定した。さらに, $y_i(t) (i = 1, \dots, N)$ に対する拘束条件への重みを $\eta = 1.0$ とした。

以上に示した設定のもとで学習則 (16)~(19) を 20,000 回計算した。このときの計算回数に対する分離器の出力 $y_1(t)$ と $y_2(t)$ の時間変化を Fig. 2(a) に, 計算終了直前 500 回に対する出力の散布図を Fig. 2(b) に, そして最終的に推定された各原信号の確率密度関数を Fig. 2(c) に示す。さらに, $c_{ik}(t)$, $\mu_{ik}(t)$, および $\sigma_{ik}(t) (= \exp(\xi_{ik}(t))) (i = 1, 2; k = 1, \dots, 5)$ の時間変化を Fig. 4(a)~(f) に示す。Fig. 2(a) から計算開始して約 500 回で分離していることがわかり, さらに Fig. 2(c) から原信号の確率密度関数が正しく推定されていることがわかる。しかし, $y_1(t)$ が約 14000 回で乱れている。これは分離行列 $\mathbf{W}(t)$ の乱れによるものと考えられ, この原因を次節で考察する。

7. 考 察

7.1 アルゴリズムの特異現象

前章で示したシミュレーションのような特異現象の原因は, 分離行列 $\mathbf{W}(t)$ の突発的な変動であると考えられる。Fig. 2(a) と Fig. 4 から, この現象と対応してパラメータ $\sigma_{13}(t) = \exp(\xi_{13}(t))$ が急激に変化していることがわかる。 $\sigma_{13}(t)$ の変化が, この特異現象を引き起こす根本的な原因であると推測できる。このパラメータの急激な変動が起こるメカニズムを考察する。

$\sigma_{13}(t)$ は, 計算回数 14300 回で急激に変化している。このとき ($t_0 = 14300$), $c_{13}(t_0) \simeq 0$ で, $c_{1k}(t_0) (k \neq 3)$ は 0 より十分に大きく, つぎの関係を満たす $y_1(t_0)$ を観測した。

$$\begin{aligned} \phi(y_1(t_0) | \mu_{13}(t_0), \sigma_{13}(t_0)) \\ \gg \phi(y_1(t_0) | \mu_{1k}(t_0), \sigma_{1k}(t_0)) (k \neq 3) \\ \{y_1(t_0) - \mu_{13}(t_0)\}^2 < \sigma_{13}^2(t_0) \end{aligned}$$

このとき, $h_{13}(t)/c_{13}(t_0) \simeq 1/c_{13}(t_0)$ となり, 式 (19) から, $\sigma_{13}(t_0 + 1) = \exp(\xi_{13}(t_0 + 1))$ は 0 へ近づく。したがって, $y_1(t_0 + 1)$ と学習則 (17) により, $\mathbf{W}(t_0)$ が大きく変化する。

確率勾配法により, $\mathbf{W}$ と $\theta_i (i = 1, \dots, N)$ の推定を行うため, その時々観測信号の影響を受け, 上記のような特異な現象を引き起こす。この特異な現象を防ぐ方法の一つとしては, EM アルゴリズムのようにパラメータ $\theta_i (i = 1, \dots, N)$ の推定をバッチ的に行うことである。本論文の場合であれば, 分離行列 $\mathbf{W}^*$ とパラメータ θ^* を同時に求めず, ある一定時間ごとに, 交互に $\mathbf{W}$ と θ をバッチ的に最適な値を求め

ることで, ここで示した学習時の特異な現象を避けることができるであろう。しかし, バッチ的な処理を持ち込むと, 原信号が非定常であるとき, パラメータ θ の推定が信号の変化に追従できなくなる。

オンラインで上記の現象に対処する方法としては, つぎのような対策が考えられる。 $c_{ik}(t)$ が非常に小さくなりゼロに近づくことで, この特異現象が起こるのであるから, $c_{ik}(t)$ をゼロに近づかせない拘束条件を設けることで, この特異な現象は回避できよう。すなわち, $c_{ik}(t) (i = 1, \dots, N; k = 1, \dots, K)$ が $L_G(\mathbf{W}, \theta)$ を最小にし, それぞれが $1/K$ にもっとも近くなるようなペナルティー項を付け加える。 γ を小さな正数としたとき, この拘束条件を組み込んだ学習則は, 次式のようにになる。

$$\Delta c_{ik}(t) = \beta_\theta \{h_{ik}(t) - c_{ik}(t) + \gamma g(c_{ik}(t))\} \quad (26)$$

$$g(c_{ik}(t)) \equiv \frac{1}{K} - c_{ik}(t) \quad (27)$$

$$(i = 1, \dots, N; k = 1, \dots, K)$$

$0 < \gamma < 1$ に対して, $c_{ik}(t)$ の初期値がすべて $\gamma/K(\gamma+1) < c_{ik}(0)$ を満たせば, この学習則に従う $c_{ik}(t)$ は常に条件 (25) と次の条件を満たす。

$$\frac{\gamma}{K(\gamma+1)} \leq c_{ik}(t)$$

よって, 学習則 (17) を式 (26) と式 (27) に変更することで, 特異な現象を防ぐことができるであろう。このことを確かめるために 6 章で行ったシミュレーションと同じ条件で, 上式の学習則を用いてシミュレーションを行った。ただし, $\gamma = 0.1$ に設定した。原信号が Fig. 3(a) に示すように推移するとき, 出力 $y_1(t)$ と $y_2(t)$ は Fig. 3(b) となり, 原信号と出力の推定誤差は Fig. 3(c) となる。原信号と出力の間にはスケールの違いがあるため, 次式に示す量を誤差信号 $e(t)$ とし, これを表示した。

$$\begin{aligned} e(t) &= \text{diag}\{v_{11}(t), v_{22}(t)\} \mathbf{y}(t) \\ &\quad - \text{diag}\{2.0, 0.6\} \mathbf{s}(t) \end{aligned} \quad (28)$$

ここで, $v_{kk}(t) (k = 1, 2)$ は行列 $\mathbf{W}(t)$ の逆行列 $\mathbf{W}(t)^{-1}$ の第 k 行第 k 列の要素である。 c_{ik} の学習則を式 (17) から式 (26), (27) に変更しても, 原信号の推定に対しては何ら問題はないことは Fig. 3 から明らかである。さらに $c_{ik}(t)$, $\mu_{ik}(t)$, $\sigma_{ik}(t) = \exp(\xi_{ik}(t)) (i = 1, 2; k = 1, \dots, 5)$ を Fig. 5(a)~(f) に示す。これらの図から, c_{ik} は, 0 に近づかず, $\mu_{ik}(t)$ と $\sigma_{ik}(t)$ は緩やかに変化しており, 目的を達成している。2 つの原信号がともに super-Gaussian 信号や sub-Gaussian 信号である場合でも, 同様に上記のような現象は起こらなかった。

7.2 学習係数

本論文で提案するアルゴリズムでは, 二つの学習係数を用いる。一つは分離行列 $\mathbf{W}$ を決定するための β_W であり, もう一つは原信号の確率分布を推定するための β_θ である。これらの学習係数の設定のしかたにより, ここで提案したアル

ゴリズムの収束性が左右される．そこで，この説では学習係数 β_W と β_θ の設定方法に対する指針を与える．

出力 $y_i(t) (i = 1, \dots, N)$ の確率密度関数を $P_{Y_i}(y_i)$ としたとき，損失関数 $L_G(\mathbf{W}, \theta)$ は，次式のように表現できる．

$$\begin{aligned} L_G(\mathbf{W}, \theta) = & \int P_Y(\mathbf{y}|\mathbf{W}) \log \frac{P_Y(\mathbf{y}|\mathbf{W})}{\prod_{i=1}^N P_{Y_i}(y_i|\mathbf{W})} d\mathbf{y} \\ & + \eta \int P_Y(\mathbf{y}|\mathbf{W}) \log \frac{P_Y(\mathbf{y}|\mathbf{W})}{\Phi(\mathbf{y}|\mathbf{W})} d\mathbf{y} \\ & + \sum_{i=1}^N \int P_{Y_i}(y_i|\mathbf{W}) \log \frac{P_{Y_i}(y_i|\mathbf{W})}{Q_{S_i}(y_i|\mathbf{W}, \theta_i)} dy_i \end{aligned}$$

もし $L_G(\mathbf{W}, \theta)$ を θ について最小化し， $P_{Y_i}(y_i) \approx Q_{S_i}(y_i|\mathbf{W}, \theta_i^*)$ となるならば，損失関数 $L_G(\mathbf{W}, \theta^*)$ は，

$$\begin{aligned} L_G(\mathbf{W}, \theta^*) \approx & \int P_Y(\mathbf{y}|\mathbf{W}) \log \frac{P_Y(\mathbf{y}|\mathbf{W})}{\prod_{i=1}^N P_{Y_i}(y_i|\mathbf{W})} d\mathbf{y} \\ & + \eta \int P_Y(\mathbf{y}|\mathbf{W}) \log \frac{P_Y(\mathbf{y}|\mathbf{W})}{\Phi(\mathbf{y}|\mathbf{W})} d\mathbf{y} \quad (29) \end{aligned}$$

と近似できる．この式の右辺第1項は，出力 $y_i(t) (i = 1, \dots, N)$ の相互情報量であるから，上式は拘束条件付きの出力 $\mathbf{y}(t)$ の相互情報量を表す．このことから，損失関数 $L_G(\mathbf{W}, \theta)$ を $\mathbf{W}, \theta$ について最小化するとき， θ についての最小化が $\mathbf{W}$ についての最小化とくらべて速ければ，この最小化は近似的に相互情報量の最小化に等しくなる．学習係数の観点から見れば，このことは β_θ を β_W より大きくすることに相当する．このとき， $Q_{S_i}(y_i|\theta_i)$ は，出力 $y_i(t)$ の確率分布 $P_{Y_i}(y_i)$ を推定するために導入したモデルであると考えられることもできる．

7.3 非定常な原信号の信号分離

この節では，本論文で提案するオンラインアルゴリズムが非定常な原信号の場合にも有効であることを示す．原信号 $s_1(t)$ は，super-Gaussianからsub-Gaussianへ，さらにsuper-Gaussianへと変化する非定常な非 Gauss 信号でつぎの確率密度関数に従う信号とした．

$$\begin{aligned} P_{S_1}(s_1(t)) = & p_0(t)\phi(s_1(t)|-2, 0.2) \\ & + (1 - 2p_0(t))\phi(s_1(t)|0, 0.2) \\ & + p_0(t)\phi(s_1(t)|2, 0.2) \end{aligned}$$

ただし， $p_0(t)$ はFig. 1に示したタイムスケジュールに従うものとし， $s_2(t)$ は次式に示す正弦波とした．

$$s_2(t) = 2\sin(0.002\pi t)$$

このときの原信号の時間変化をFig. 6(a)に示す．

6章と同じ混合行列 $\mathbf{A}$ ，パラメータ c_{ik} ， μ_{ik} ， $\xi_{ik} (i = 1, 2; k = 1, \dots, 5)$ と分離行列 $\mathbf{W}$ の初期値を使用し，学習則(16)，(18)，(19)，(26)および(27)に基づいて学習を行った．ただし，学習係数は $\beta_W = 1.0 \times 10^{-4}$ と $\beta_\theta = 1.0 \times 10^{-3}$ とした．このときの分離器の出力 $y_i^*(t) (i = 1, 2)$ の時間経過をFig. 6(b)に示す． $y_i^*(t) (i = 1, 2)$ は

$$\mathbf{y}^*(t) = \text{diag}\{v_{11}(t), v_{22}(t)\} \mathbf{y}(t)$$

Fig. 1 The time schedule of the nonstationary source $s_1(t)$

の各要素を表す．ただし， $v_{kk}(t) (k = 1, 2)$ は $\mathbf{W}(t)^{-1}$ の第 k 行第 k 列の要素とする．

本論文で提案するアルゴリズムは，分離器の各出力の分散を1にするから，Fig. 6(a)に示した原信号に対して，出力はその分散が1となるような変化をする．分離器の出力が原信号となっていることを明確にするため， $y_i^*(t) (i = 1, 2)$ を表示した．また，原信号と出力との間の誤差 $e(t)$ を式(28)で定義し，この変化をFig. 6(c)に表示する．これらの図から出力 $y_1^*(t)$ が，super-Gaussianからsub-Gaussianさらにsuper-Gaussianへと変化する， $y_2^*(t)$ は原信号と同一の周波数を持つ正弦信号となっていることがわかる．このことから本論文で提案するアルゴリズムは，原信号が非定常な場合でも有効であることがわかる．

8. 結 論

本論文では，分離器の各出力の確率密度関数に混合ガウスモデルを用いて，分離行列とその出力の確率密度関数の推定を同時に行うオンラインアルゴリズムを提案した．混合ガウスモデルは，表現が簡単で多くの確率密度関数を表現できるという特徴を持つため，確率密度関数の推定問題では有用なモデルである．natural learningに従えば，パラメータの推定は，混合ガウスモデルのフィッシャーの情報行列を必要とするが，これは複雑でオンラインアルゴリズムには適さない．そこで，我々は他の情報行列を用いた．

確率勾配法に従って学習則に期待値の代わりに瞬時値を用いたから，パラメータ $c_{ik}(t) (i = 1, \dots, N; k = 1, \dots, K)$ が0に近づいたとき，パラメータ θ の推定量が急激に変化して分離器の出力を乱した．本論文では，このような特異な現象を防ぐ対策として，パラメータ $c_{ik}(t)$ が，それぞれ零にならないような拘束条件を設け， $\mu_{ik}(t)$ と $\sigma_{ik}(t) (= \exp(\xi_{ik}(t)))$ が，大きな変動を起こさないようにした．その結果得られたアルゴリズムは，7.3章で示したように原信号が非定常信号であるときにも有効である．

(a) The outputs

(b) The distributions of the outputs

(c) The probability density function which was estimated

Fig. 2 The separator's outputs and their distributions

(a) The source signals

(b) The separators' outputs

(c) The errors between the sources and the outputs

Fig. 3 The sources, the outputs and the errors between them when the algorithm is modified

(a) The evolution of $c_{1k} (k = 1, \dots, 5)$ (b) The evolution of $c_{2k} (k = 1, \dots, 5)$ (c) The evolution of $\mu_{1k} (k = 1, \dots, 5)$ (d) The evolution of $\mu_{2k} (k = 1, \dots, 5)$ (e) The evolution of $\sigma_{1k} (k = 1, \dots, 5)$ (f) The evolution of $\sigma_{2k} (k = 1, \dots, 5)$

Fig. 4 The evolution of the parameters before the algorithm is modified

(a) The evolution of $c_{1k}(k = 1, \dots, 5)$ (b) The evolution of $c_{2k}(k = 1, \dots, 5)$ (c) The evolution of $\mu_{1k}(k = 1, \dots, 5)$ (d) The evolution of $\mu_{2k}(k = 1, \dots, 5)$ (e) The evolution of $\sigma_{1k}(k = 1, \dots, 5)$ (f) The evolution of $\sigma_{2k}(k = 1, \dots, 5)$

Fig. 5 The evolution of the parameters when the algorithm is modified

Fig. 6 The separators' outputs for the nonstationary sources

参考文献

- 1) A.J.Bell and T.J.Sejnowski: An information-maximisation approach to blind separation and blind deconvolution, *Neural Computation*, **7**-6, 1129/1159(1995)
- 2) L.Xu and M.I.Jordan: On Convergence Properties of the EM Algorithm for Gaussian Mixtures, *Neural Computation*, **8**-1, 129/151(1996)
- 3) S.Amari, A.Cichocki and H.H.Yang: A New Learning Algorithm for Blind Signal Separation, *Advances in Neural Information Processing Systems*, **8**, 757/763(1996)
- 4) S. Amari and J. F. Cardoso: Blind Source Separation-Semiparametric Statistical Approach, *IEEE Transaction on Signal Processing*, **45**-11, 2692/2700(1997)
- 5) S.Amari, T.-P.Chen, and A.Cichocki: Stability Analysis of Adaptive Blind Source Separation, *Neural Networks*, **10**-8, 1345/1351(1997)
- 6) S.C.Douglas, A.Cichocki, H.H.Yang and S.Amari: Multichannel Blind Deconvolution and Equalization Using the Natural Gradient, *Proc. of IEEE International Workshop on Wireless Communication*, 101/104(1997)
- 7) S.Amari: Natural Gradient works Efficiently in Learning, *Neural Computation*, **10**-2, 251/276(1998)
- 8) Y.Xiao and H.Guangrui: Speech Separation based on the GMM pdf Estimation and the Feedback Architecture, *Proc. of ICA International Workshop on Theory II*, 353/357(1999)

[著者紹介]

大畑正志 (学生会員)

1994年九州工業大学工学部設計生産工学科卒業。96年九州工業大学大学院工学研究科博士前期課程修了。同年三菱電機(株)入社。98年九州工業大学大学院工学研究科博士後期課程入学。現在に至る。信号処理、ニューラルネットワークの研究に従事。

徳成剛 (正会員)

1985年九州工業大学工学部制御工学科卒業。87年同大学大学院工学研究科修士課程修了。同年東陶機器(株)商品開発本部入社。91年九州工業大学工学部, 97年同大学地域共同研究センター, 99年同大学工学部助手, 現在に至る。主にニューラルネット, 画像計測, ブラインド信号分離の研究に従事。

松岡清利 (正会員)

1971年九州工業大学工学部機械工学科卒業。76年東京大学大学院博士課程修了(機械工学専攻)。同年九州工業大学講師。78年助教授91年教授。現在, ニューラルネット, 信号処理の研究に従事。工学博士。