

# 独立成分分析に基づくブラインド信号源分離

担当教員：北村大地

## 1 実験の背景と目的

信号源分離とは、複数の信号が混合された状態で観測した信号から、混合する前の信号（信号源, source）を推定する技術である。例えば、複数の人間が同時に話している音響信号から、個々の人の音声信号を分離・抽出することや、複数の方向から同時に到来する電波の内、特定の電波のみを受信すること等に応用される。これらの応用例を図1に示す。但し、ここでは観測に用いるセンサ（マイクロホンやアンテナ等）は、混合している信号源数と同じ数だけ用意していることを仮定している。

信号源分離を実現するには、各信号源の到来方向や信号の学習データ等、何らかの事前情報を学習または活用する手法が考えられる。しかしながら、そのような手法は用途が著しく制限されてしまい、一般的な応用には向かない。そこで、事前情報を全く用いない信号源分離技術が確立している。この技術は、事前情報が全く与えられない（盲目的である）という状況から、ブラインド信号源分離（blind source separation: BSS）と呼ばれている。

本実験では、BSS技術の一例として、独立成分分析（independent component analysis: ICA）<sup>\*1</sup>と呼ばれる統計数理アルゴリズムを取り上げ、なぜ事前情報を用いないまま信号源が分離できるのか、その原理を理解することを目的とする。また、ICAのプログラムを実装し、擬似的に混合した音声信号が実際に分離されることについて確かめる。さらに、現実における複数音声の混合信号を録音してICAを適用し、その結果について考察することで、より理解を深めることを目的とする。



図1 信号源分離を適用する状況: (a) 複数の人が同時に話している録音信号, (b) 複数の電波の受信信号。

<sup>\*1</sup> P. Comon, “Independent component analysis, a new concept?,” *Signal Processing*, vol. 36, no. 3, pp.287–314, 1994.

## 2 ICA の基本原理

### 2.1 信号の確率モデル

統計数理アルゴリズムである ICA では、信号の統計的性質、即ち信号の確率モデルを仮定する。これは、ICA における信号源分離の原理を理解するうえで非常に重要な概念であるため、本節ではまず信号の確率モデルについて説明する。

今、時間信号  $s(t)$  を考える。ここで、 $s(t)$  は  $t = 1, 2, \dots, T$  の離散時間のそれぞれに対して振幅値が定義されているデジタル信号である。例として、女性音声のデジタル信号を示したものが図 2 である。音声は振幅 0 を中心として正負に変動する信号であることが、図 2 から確認できる。また、図 2 の信号の  $t = 1, 2, \dots, T$  の全時刻に関して、特定の範囲の振幅値の出現回数を数

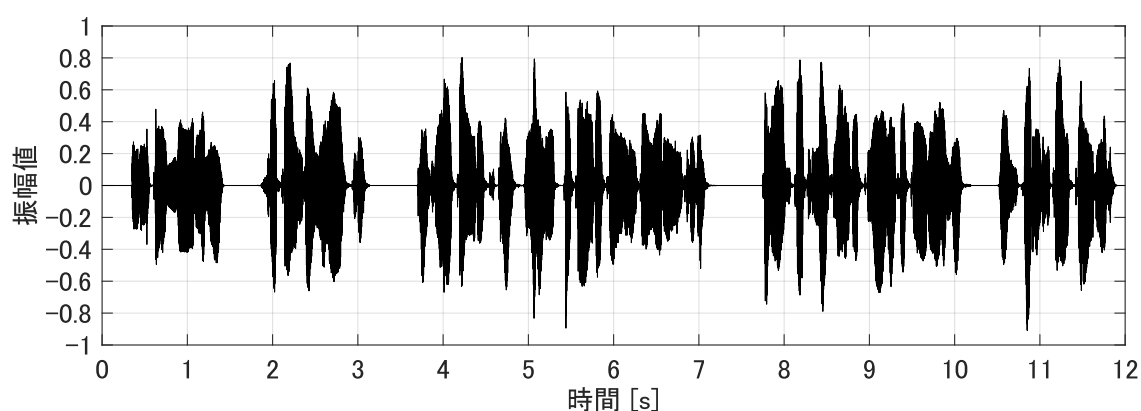


図 2 女性音声のデジタル信号。サンプリング周波数は 44.1 kHz であり全体が 12 秒の長さであることから、この波形には  $T = 44100 \times 12 = 529200$  点の振幅値が含まれている。

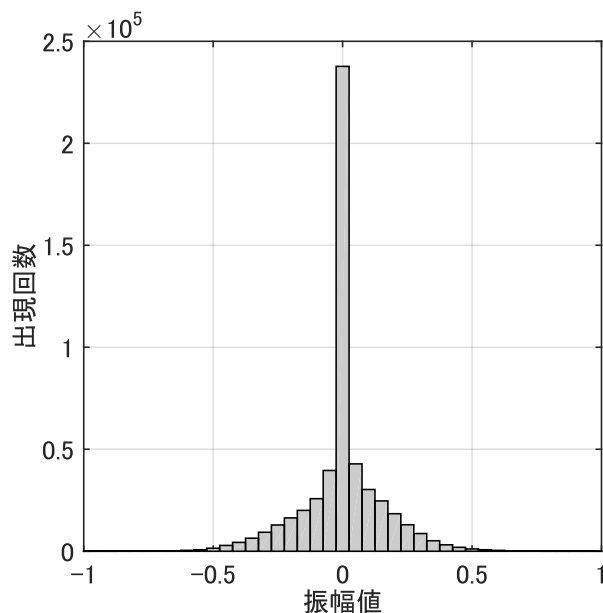


図 3 女性音声のデジタル信号の振幅値に関する出現回数の棒グラフ。1 本の棒グラフの横幅は 0.05 として描いている。0 付近の出現回数は多く、絶対値の大きい振幅値の出現回数は少ない。

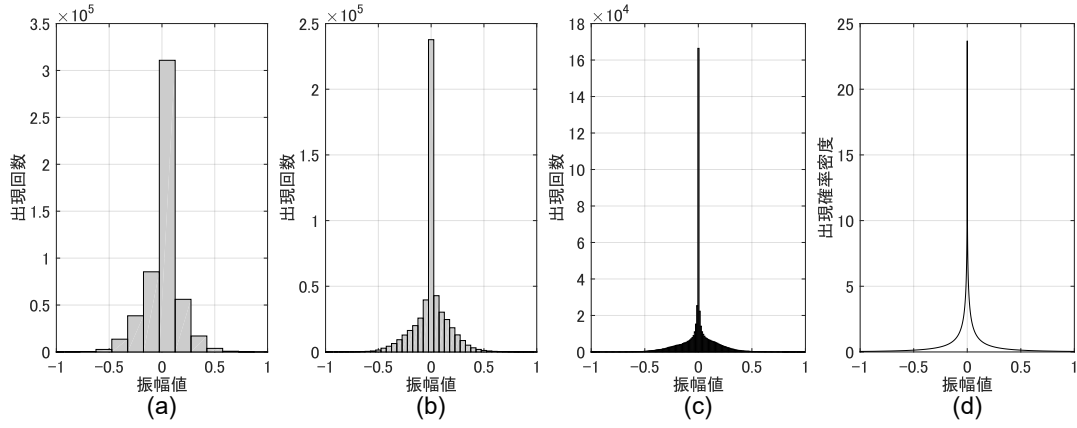


図 4 振幅値出現回数の棒グラフの横幅を変化させたときの概形の変化 (a)～(c) と、極限を考えたときの確率密度関数 (d)。

え上げて棒グラフにしたものが図 3 である。この棒グラフも、0 を中心としておよそ左右対象の形状をしており、正負の値がおおよそ等しい回数で出現していることが分かる。さらに、0 付近の振幅値の出現回数は非常に多く、逆に絶対値 ( $|s(t)|$ ) の大きな振幅値の出現回数は少なくなるという傾向が確認できる。

このような「信号の振幅値に対する出現回数」は、「各振幅値がどの程度出現しやすいかという確率（頻度）」として解釈することもできる。例えば、図 2 に示す女性の音声信号は、図 3 から分かる通り「0 付近の振幅値を高い確率で発生する確率モデルから生成された信号」と捉えることができ、これは信号  $s(t)$  の統計的な性質を表現していることになる。

出現回数の棒グラフの横幅を変化させた場合、図 4(a)–(c) に示すように全体的な形状が変化する。棒グラフの横幅を極限まで小さくした場合は、図 4(d) のような振幅値の確率密度関数が現れる。これは、（離散的な横幅ではなく）連続的な振幅値に対する出現確率密度を表現している。改めて、信号  $s(t)$  は「図 4(d) の確率密度関数からランダムに生成された信号」と考えることができ、この確率密度関数を信号  $s(t)$  の生成モデルと呼び、以後  $p(s)$  と表記する。 $p(s)$  は確率密度関数であるので、次の性質を満たす。

$$p(s) \geq 0 \quad (1)$$

$$\int_{-\infty}^{\infty} p(s) ds = 1 \quad (2)$$

即ち、確率密度値は必ず 0 以上であり、全範囲で積分した確率値（全事象が起こる確率）は 1 である。これは確率の公理として与えられている。

良く知られた確率密度関数としてガウス分布（Gaussian distribution, 正規分布とも呼ばれる）がある。これは次式で表される。

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(z - \mu)^2}{2\sigma^2}\right) \quad (3)$$

ここで、 $z$  は確率変数である。ガウス分布は平均  $\mu$  と分散  $\sigma^2$  をパラメータとして持ち、図 5 に示す概形をしている。一般に、測定誤差やノイズ信号等の生成モデルはガウス分布になることが知られている（ガウス分布からランダムに生成された値をとる）。逆に、音声や電波等、情報として価値のある信号の生成モデルは通常ガウス分布にはならない。確かに、図 2 の女性の音声信号の生成モデルは、(0～12 秒という短い範囲では) 図 4(d) に示す分布となっており、最頻値がガ

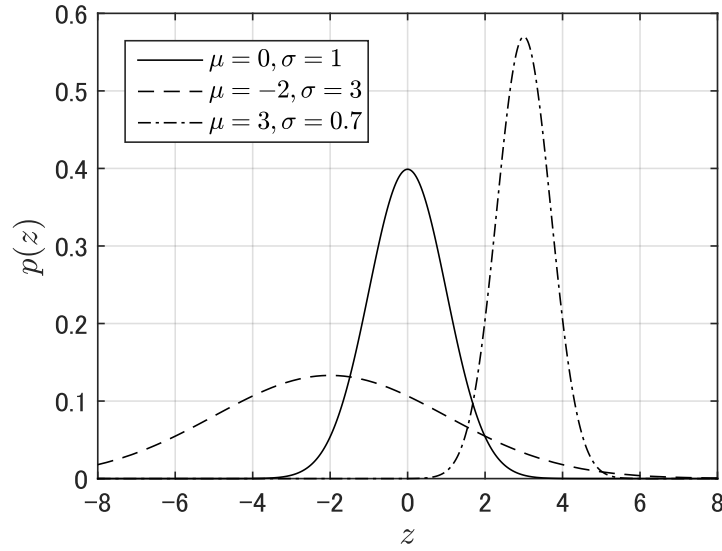


図5 ガウス分布の概形．ガウス分布は平均  $\mu$  を中心として線対称な形状をしており，分散  $\sigma^2$  が大きい程広がった分布となる．

ウス分布よりも随分尖った形状をしていることが確認できる．このように，最頻値がガウス分布よりも急尖的で（spiky）両脇の裾野が長く尾を引く（heavy-tailed）ような分布は**優ガウス分布**（super-Gaussian distribution, leptokurtic distribution）と呼ばれ，逆に最頻値がフラットで両脇の裾野がすぐに0に落ちるような分布は**劣ガウス分布**（sub-Gaussian distribution, platykurtic distribution）と呼ばれる．一般に音声信号は優ガウスな統計的性質を持つ．

## 2.2 信号源の混合モデルと分離方法

本節では、「複数の信号源が混合する」という現象を，数学的にモデル化する．今，2つの信号源  $s_1(t)$  及び  $s_2(t)$  があり，その混合信号を2つのセンサで観測するという状況を考える．信号源が音声，センサがマイクロホンであった場合の例を図6に示している．センサで観測された信号を  $x_1(t)$  及び  $x_2(t)$  とすると，2つの信号源の混合現象は次の連立方程式でモデル化できる．

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \quad (4)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \quad (5)$$

ここで，信号が伝搬する際に乗じられる係数  $a_{mn}$ （ $m$  はセンサの番号， $n$  は信号源の番号）は，時刻  $t$  には依存しないことを仮定する．即ち，信号源の位置，センサの位置，信号の伝搬速度等が時間に依らず一定であることを仮定している．

簡略化して表記するために，伝搬係数  $a_{mn}$  をまとめた行列を次のように定義する．

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad (6)$$

行列  $\mathbf{A}$  は混合行列と呼ばれる．行列  $\mathbf{A}$  を用いると，式 (4) 及び (5) の連立方程式は，次式のように書き直すことができる．

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} \quad (7)$$

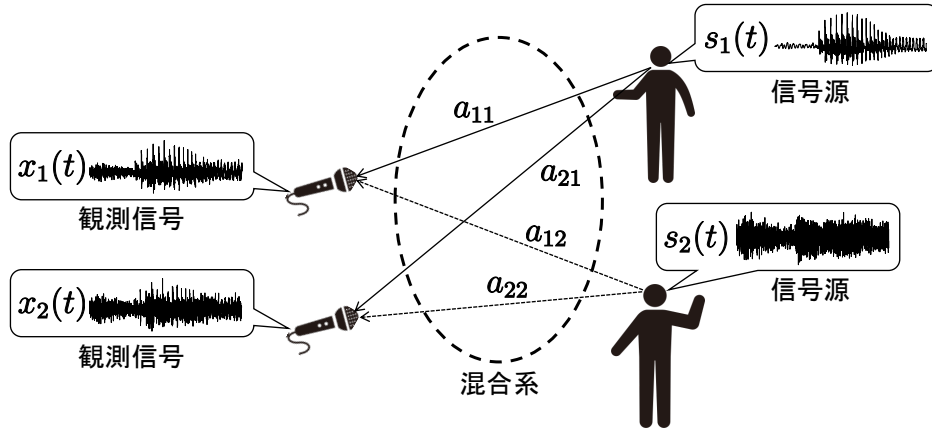


図6 2つの音声信号の混合と観測.

また、観測信号ベクトル  $\mathbf{x}(t) = (x_1(t), x_2(t))^T$  及び信号源ベクトル  $\mathbf{s}(t) = (s_1(t), s_2(t))^T$  を用いると、次式となる.

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (8)$$

式 (8) において、もし混合行列  $\mathbf{A}$  の行列式が非零 ( $a_{11}a_{22} - a_{12}a_{21} \neq 0$ ) であれば、 $\mathbf{A}$  の逆行列が存在することになる. このとき、観測信号  $\mathbf{x}(t)$  と混合行列の逆行列  $\mathbf{A}^{-1}$  を用いることで、分離信号  $y_1(t)$  及び  $y_2(t)$  を次のように得ることができる.

$$\begin{aligned} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} \\ &= \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} \end{aligned}$$

上式は、 $\mathbf{W} = \mathbf{A}^{-1}$  なる分離行列と分離信号ベクトル  $\mathbf{y}(t) = (y_1(t), y_2(t))^T$  を用いると

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \quad (9)$$

と書ける. さらに、分離行列  $\mathbf{W}$  の各要素を

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \quad (10)$$

と表し、 $\mathbf{W}$  の各行を縦ベクトルとして  $\mathbf{w}_1 = (w_{11}, w_{12})^T$  及び  $\mathbf{w}_2 = (w_{21}, w_{22})^T$  と定義する (即ち  $\mathbf{W} = (\mathbf{w}_1 \mathbf{w}_2)^T$ ) と、式 (9) は各分離信号の式として書き直すこともできる.

$$y_1(t) = \mathbf{w}_1^T \mathbf{x}(t) \quad (11)$$

$$y_2(t) = \mathbf{w}_2^T \mathbf{x}(t) \quad (12)$$

以上の議論より、観測信号  $\mathbf{x}(t)$  に対して信号源分離を行うには、混合行列  $\mathbf{A}$  の逆行列である分離行列  $\mathbf{W}$  を推定すればよいことが分かる. 音源数が3個以上の場合でも、観測するためのセンサが同じ数だけあれば混合行列  $\mathbf{A}$  が正方行列になるため、やはり分離行列  $\mathbf{W} = \mathbf{A}^{-1}$  が存在すれば信号源分離が達成される. しかしながら、BSS では混合行列  $\mathbf{A}$  が未知であるため、 $\mathbf{x}(t)$  のみから

$\mathbf{W}$  を求めることは不可能である。ICA では、信号源の混合モデル式 (8) の仮定の他に、信号そのものの統計的なモデル ( $p(s_1)$  及び  $p(s_2)$  に対する仮定) を導入することで、分離行列  $\mathbf{W}$  が推定できるようになる。

## 2.3 統計的独立性

ICA による信号源分離を理解する上での重要な概念として、統計的独立性がある。これは、複数の確率変数がお互いの情報を全く持っていないという状態である。例えば、大小 2 つのサイコロを振ったときの出目を考えたとき、大きなサイコロの出目が 6 であったからといって小さなサイコロの出目の傾向や確率が変わることはない。これは、「2 つのサイコロの出目という確率変数が統計的に独立」という事実を示している。本節では、この統計的独立性について数学的に定義し、説明する。

2 つの確率変数  $z_1$  と  $z_2$  及びそれらの生成モデル  $p(z_1)$  と  $p(z_2)$  を考える。確率変数  $z_1$  と  $z_2$  が同時に起こる確率を同時分布と呼び、 $p(z_1, z_2)$  と表記する。同時分布は複数の確率変数の直積から成る空間上で定義される確率密度関数である。また、「 $z_2$  の値が  $z_2 = z_2$  に確定した下での  $z_1$  の起こる確率」及び「 $z_1$  の値が  $z_1 = z_1$  に確定した下での  $z_2$  の起こる確率」を条件付き分布と呼び、それぞれ  $p(z_1|z_2 = z_1)$  及び  $p(z_1|z_2 = z_2)$  と表記する。ここで、 $z_1$  及び  $z_2$  は確率変数がとったある値（定数）を表す。条件付き分布は省略して  $p(z_1|z_2)$  と表記すること多いが、これは単に「 $z_2$  が何らかの値に確定した下での  $z_1$  の確率分布」を表している。

同時分布  $p(z_1, z_2)$  は  $z_1$  と  $z_2$  が同時にある値をとる確率であるので、2 つの確率変数  $z_1$  と  $z_2$  の関係性に依らず、条件付き分布を用いて次式のように分解することができる。

$$p(z_1, z_2) = p(z_1|z_2)p(z_2) \quad (13)$$

$$= p(z_2|z_1)p(z_1) \quad (14)$$

一方、確率変数  $z_1$  と  $z_2$  がもし統計的に独立ならば場合、同時分布に対して次式の分解が成立する。

$$p(z_1, z_2) = p(z_1)p(z_2) \quad (15)$$

式 (13) 及び (14) と式 (15) を比較すると、結局「2 つの確率変数  $z_1$  と  $z_2$  が統計的に独立」とは、次式が成り立つこととも解釈できる。

$$p(z_1|z_2) = p(z_1) \quad (16)$$

$$p(z_2|z_1) = p(z_2) \quad (17)$$

これらの式は、 $z_1$  の値が先に何かに定まろうとも  $z_2$  の確率分布は変わらず、逆に  $z_2$  の値が先に何かに定まろうとも  $z_1$  の確率分布は変わらないということを示している。即ち、本節冒頭のサイコロの出目の例と同様に、確率変数  $z_1$  と  $z_2$  が互いに何も影響しないということである。逆に、2 つの確率変数を「 $z_1$  が大きなサイコロの出目、 $z_2$  が大きなサイコロと小さなサイコロの出目の合計値」と定義した場合は、これらの確率変数は明らかに独立にはならず、互いに相関関係が生じる。なぜなら、「 $z_1$  の値（大きなサイコロの出目）が大きくなったとき  $z_2$  の値（大小サイコロの出目の和）も大きくなる」という傾向が生まれるからである。このように独立でない確率変数に対しては、式 (15) が成立しないため、式 (15) は統計的独立性の定義式と考えることができる。

## 2.4 混合信号の統計的性質

前節の統計的独立性が、時間信号の生成モデルに対して成り立つか否かについて考えていこう。信号源  $s_1(t)$  及び  $s_2(t)$  の生成モデル  $p(s_1)$  及び  $p(s_2)$  を考えたとき、通常信号源は互いに無関係である。例えば、一人目の話している内容から二人目の話している内容を信号として予測するようなことは通常できないはずである。そのため、信号源  $s_1(t)$  と  $s_2(t)$  は統計的に独立とみなすことができ、次式が成立する。

$$p(s_1, s_2) = p(s_1)p(s_2) \quad (18)$$

しかし、混合信号  $x_1(t)$  及び  $x_2(t)$  は、どちらも信号源  $s_1(t)$  及び  $s_2(t)$  を両方含む信号である。 $x_1(t)$  と  $x_2(t)$  における  $s_1(t)$  と  $s_2(t)$  の混合割合 ( $a_{m1}$  と  $a_{m2}$  の比) は異なるが、それでも図 6 に示すように  $x_1(t)$  と  $x_2(t)$  の 2 つの波形はよく似ており、これらは強い相関関係を持つ非独立な信号のはずである。

以上より、「信号源  $s_1(t)$  及び  $s_2(t)$  は互いに独立で式 (18) が成立するが、混合信号  $x_1(t)$  及び  $x_2(t)$  は非独立である」という統計的仮定を置くことができる。ICA では、信号源間の統計的独立性として式 (18) が成り立つことを仮定し、混合行列  $\mathbf{A}$  を事前に知ることなく分離行列  $\mathbf{W}$  を推定できる。

## 2.5 独立性最大化による分離行列推定問題の定式化

今、仮に混合信号を完全に分離する分離行列  $\mathbf{W}$  が推定できた場合、式 (9) で得られる分離信号  $y_1(t)$  及び  $y_2(t)$  は混合前の信号源  $s_1(t)$  及び  $s_2(t)$  と同じく統計的に独立な信号である。逆に言えば、 $y_1(t)$  及び  $y_2(t)$  が統計的に独立となるように、変数である分離行列  $\mathbf{W}$  を決めることができれば、信号源分離が達成される。従って ICA による BSS は、統計的独立性の定義式 (18) より、次式が成立するような  $\mathbf{W}$  を決める問題と解釈できる。

$$p(y_1, y_2) = p(y_1)p(y_2) \quad (19)$$

ここで、 $p(y_1)$  及び  $p(y_2)$  はそれぞれ分離信号  $y_1(t)$  及び  $y_2(t)$  の生成モデルであり、 $p(y_1, y_2)$  は同時分布である。

上記の問題を定式化する。分離信号  $y_1(t)$  及び  $y_2(t)$  ができるだけ独立となるような分離行列  $\mathbf{W}$  を求める問題は、次式で与えられる。

$$\min_{\mathbf{W}} \mathcal{J}(\mathbf{W}) \quad (20)$$

$$\mathcal{J}(\mathbf{W}) = \mathcal{D}_{\text{KL}} [p(y_1, y_2) \| p(y_1)p(y_2)] \quad (21)$$

ここで、式 (20) は「関数  $\mathcal{J}(\mathbf{W})$  の値を最小化する変数  $\mathbf{W}$  を探す」という問題を表している\*2。また  $\mathcal{D}_{\text{KL}} [p(z) \| q(z)]$  は、引数の 2 つの分布  $p(z)$  及び  $q(z)$  間の距離を測る関数であり、カルバック-ライブラ・ダイバージェンス (Kullback-Leibler divergence: KL divergence) と呼ばれ、次式

---

\*2 一見すると関数  $\mathcal{J}(\mathbf{W})$  は変数  $\mathbf{W}$  を含んでいないように見えるが、分離信号  $y_1(t)$  及び  $y_2(t)$  は式 (9) で定義されるため、 $\mathbf{W}$  を含んでいる点に注意する。

で定義されている.

$$\mathcal{D}_{\text{KL}} [p(z)||q(z)] = \int_{-\infty}^{\infty} p(z) \log \frac{p(z)}{q(z)} dz \quad (22)$$

従って, 式 (21) は

$$\mathcal{J}(\mathbf{W}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log \frac{p(y_1, y_2)}{p(y_1)p(y_2)} dy_1 dy_2 \quad (23)$$

と書きなおすことができ, 式 (20) は「同時分布  $p(y_1, y_2)$  と分布積  $p(y_1)p(y_2)$  の分布間距離を最も小さくする分離行列  $\mathbf{W}$  を推定する」という問題であることが分かる. 分離の結果  $y_1(t)$  と  $y_2(t)$  が完全に独立となった場合, 式 (19) が成立し, 分布間距離は最小値  $\mathcal{D}_{\text{KL}} [p(y_1, y_2)||p(y_1)p(y_2)] = 0$  となる.

式 (23) は次のように変形できる.

$$\begin{aligned} \mathcal{J}(\mathbf{W}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log p(y_1, y_2) dy_1 dy_2 - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log p(y_1)p(y_2) dy_1 dy_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log p(y_1, y_2) dy_1 dy_2 \\ &\quad - \left[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log p(y_1) dy_1 dy_2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log p(y_2) dy_1 dy_2 \right] \end{aligned} \quad (24)$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log p(y_1, y_2) dy_1 dy_2 \\ &\quad - \left[ \int_{-\infty}^{\infty} p(y_1) \log p(y_1) dy_1 + \int_{-\infty}^{\infty} p(y_2) \log p(y_2) dy_2 \right] \end{aligned} \quad (25)$$

ここで, 式 (24) から式 (25) の変形には, 次式に示す同時分布の周辺化 (積分消去) を用いた.

$$\int_{-\infty}^{\infty} p(z_1, z_2) dz_2 = p(z_1) \quad (26)$$

さらに, 式 (9) の関係が確率変数ベクトル  $\mathbf{y} = (y_1, y_2)^T$  及び  $\mathbf{x} = (x_1, x_2)^T$  にも成立することを考えると,

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (27)$$

が成立する. このように  $\mathbf{W}$  で線形変換する前 ( $\mathbf{x}$ ) と後 ( $\mathbf{y}$ ) の確率変数を考えたとき, それぞれの同時分布  $p(\mathbf{y}) = p(y_1, y_2)$  と  $p(\mathbf{x}) = p(x_1, x_2)$  の間には, 次式が成立する<sup>\*3</sup>.

$$p(\mathbf{y}) = \frac{1}{|\det \mathbf{W}|} p(\mathbf{x}) \quad (28)$$

$$dy_1 dy_2 = |\det \mathbf{W}| dx_1 dx_2 \quad (29)$$

---

<sup>\*3</sup> 式 (28) は  $p(\mathbf{x})$  と  $p(\mathbf{y})$  がいずれも確率密度関数であり  $\int_{-\infty}^{\infty} p(\mathbf{y}) d\mathbf{y} = \int_{-\infty}^{\infty} p(\mathbf{x}) |\partial \mathbf{y} / \partial \mathbf{x}| d\mathbf{x} = 1$  を満たすことから導かれる. ここで,  $|\partial \mathbf{y} / \partial \mathbf{x}|$  は線形変換のヤコビアンである.  $\mathbf{y} = \mathbf{W}\mathbf{x}$  ならば  $|\partial \mathbf{y} / \partial \mathbf{x}| = |\det \mathbf{W}|$  であり, これは式 (29) そのものである.



ここで、 $\det \mathbf{W}$  は行列  $\mathbf{W}$  の行列式を表す。式 (28) 及び (29) を式 (25) の第一項に代入すると、

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(y_1, y_2) \log p(y_1, y_2) dy_1 dy_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{|\det \mathbf{W}|} p(x_1, x_2) \left[ \log \frac{1}{|\det \mathbf{W}|} p(x_1, x_2) \right] |\det \mathbf{W}| dx_1 dx_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \log \frac{1}{|\det \mathbf{W}|} p(x_1, x_2) dx_1 dx_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \log p(x_1, x_2) dx_1 dx_2 - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \log |\det \mathbf{W}| dx_1 dx_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \log p(x_1, x_2) dx_1 dx_2 - \log |\det \mathbf{W}|
\end{aligned} \tag{30}$$

となり、結局  $\mathbf{W}$  を推定するために最小化する関数  $\mathcal{J}(\mathbf{W})$  は次式になる。

$$\begin{aligned}
\mathcal{J}(\mathbf{W}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) \log p(x_1, x_2) dx_1 dx_2 - \log |\det \mathbf{W}| \\
&\quad - \left[ \int_{-\infty}^{\infty} p(y_1) \log p(y_1) dy_1 + \int_{-\infty}^{\infty} p(y_2) \log p(y_2) dy_2 \right]
\end{aligned} \tag{31}$$

## 2.6 最急降下法

式 (31) の最小化問題を考える前に、一般的な関数の最小化アルゴリズムについて説明する。いま、何らかの関数  $\mathcal{C}(\boldsymbol{\theta})$  の値ができるだけ小さくなるような変数  $\boldsymbol{\theta}$  を推定する問題を考える。

$$\min_{\boldsymbol{\theta}} \mathcal{C}(\boldsymbol{\theta}) \tag{32}$$

ここで、最小化する対象となる関数  $\mathcal{C}(\boldsymbol{\theta})$  は目的関数 (objective function) やコスト関数 (cost function) と呼ばれる。また、変数  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N)^T$  は  $N$  個のスカラー変数から成るベクトルである点に注意する。上記のような最適化問題を解くためには、基本的に停留点  $d\mathcal{C}/d\boldsymbol{\theta} = \mathbf{0}$  (変数  $\boldsymbol{\theta}$  で  $\mathcal{C}(\boldsymbol{\theta})$  を微分した結果が零ベクトルとなる点) を求める必要がある。2 次方程式における解の公式のように代数解<sup>\*4</sup>として求められる場合は単純であるが、 $d\mathcal{C}/d\boldsymbol{\theta} = \mathbf{0}$  を満たす  $\boldsymbol{\theta}$  が代数的に解けない等の場合には、変数  $\boldsymbol{\theta}$  を何らかの値で初期化したうえで停留点に徐々に近づいていく最適化アルゴリズムが用いられる。

代表的な最適化アルゴリズムに最急降下法 (steepest descent) がある。これは、一階微分ベクトル<sup>\*5</sup>

$$\nabla \mathcal{C}(\boldsymbol{\theta}) = \frac{d\mathcal{C}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \tag{33}$$

$$= \begin{pmatrix} \frac{\partial \mathcal{C}(\boldsymbol{\theta}^{(l)})}{\partial \theta_1} \\ \frac{\partial \mathcal{C}(\boldsymbol{\theta}^{(l)})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \mathcal{C}(\boldsymbol{\theta}^{(l)})}{\partial \theta_N} \end{pmatrix} \tag{34}$$

<sup>\*4</sup> このような解は閉形式解 (closed-form solution) と呼ばれる。

<sup>\*5</sup> スカラー場に対する勾配 (ベクトル場) と同じものである。

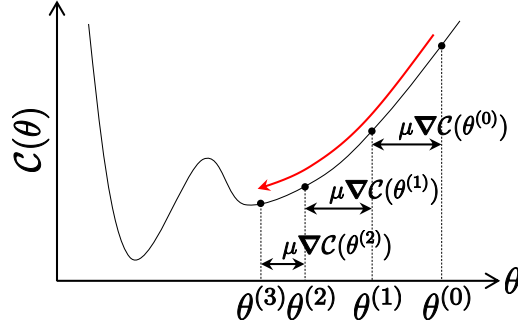


図 7 1 変数の目的関数  $C(\theta)$  に最急降下法を適用したときの変数  $\theta$  の更新の様子．初期値  $\theta^{(0)}$  近傍の局所最小解に向かって最適化が進む．

を用いて関数の停留点に徐々に近づいていく手法である．一階微分ベクトル  $\nabla C(\theta)$  はある点  $\theta$  での「関数値  $C(\theta)$  が大きくなる方向」を示していることから，最小化問題 (32) を解くには「一階微分ベクトル  $\nabla C(\theta)$  の逆方向に変数  $\theta$  を少し移動させる」という処理を反復すれば良いことが分かる． $l$  回反復時の変数  $\theta$  を  $\theta^{(l)}$  と記述するとき，最急降下法による変数の更新は次式で与えられる．

$$\begin{aligned}\theta^{(l+1)} &= \theta^{(l)} - \mu \nabla C(\theta^{(l)}) \\ &= \theta^{(l)} - \mu \begin{pmatrix} \frac{\partial C(\theta^{(l)})}{\partial \theta_1} \\ \frac{\partial C(\theta^{(l)})}{\partial \theta_2} \\ \vdots \\ \frac{\partial C(\theta^{(l)})}{\partial \theta_N} \end{pmatrix}\end{aligned}\quad (35)$$

ここで， $\mu$  はステップサイズと呼ばれるパラメータであり，1 回の反復で  $\theta$  をどの程度移動させるかを調整する値である．1 変数の場合の最急降下法による変数更新の様子を図 7 に示す．この図から分かるように，最急降下法では適当に与えた初期値  $\theta^{(0)}$  を起点として，目的関数の勾配を下る方向 ( $-\nabla C(\theta)$ ) に変数  $\theta$  を更新してゆく．但し，ステップサイズ  $\mu$  が大きすぎると発散して更新値が収束しない場合があるため，アルゴリズムとしての変数の収束性は保証されない点に注意する必要がある．

## 2.7 ICA の最適化アルゴリズム

ICA の目的関数（信号源が 2 つの場合）は式 (31) で表される  $\mathcal{J}(\mathbf{W})$  である．これを最小化する変数  $\mathbf{W}$  を最急降下法で推定するアルゴリズムを考える． $\mathcal{J}(\mathbf{W})$  の第一項は観測信号  $\mathbf{x}$  のみを含んでおり，これは変数  $\mathbf{W}$  を含まない単なる定数である．従って  $\mathcal{J}(\mathbf{W})$  の最小化問題 (20) を解くには，式 (31) の第二項及び第三項を  $\mathbf{W}$  で微分した勾配  $\nabla \mathcal{J}(\mathbf{W})$  を用いて式 (35) の最急降下法を適用すればよい．ここでは詳しい導出を割愛するが，結果として次の更新式が得られる．

$$\mathbf{W}^{(l+1)} = \mathbf{W}^{(l)} - \mu \left( \mathbb{E} \left[ \varphi(\mathbf{y}^{(l)}) \mathbf{y}^{(l)\top} \right] - \mathbf{I} \right) \left( \mathbf{W}^{(l)\top} \right)^{-1} \quad (36)$$

$$\mathbf{y}^{(l)} = \mathbf{W}^{(l)} \mathbf{x} \quad (37)$$

ここで、 $\mathbf{W}^{(l)}$  は  $l$  回反復時の変数  $\mathbf{W}$ ， $\mathbf{I}$  は  $\mathbf{W}$  と同じサイズの単位行列である．また、 $\varphi(\mathbf{y})$  はスコア関数と呼ばれ、次式で定義される．

$$\varphi(\mathbf{y}) = \begin{pmatrix} \varphi_1(y_1) \\ \varphi_2(y_2) \end{pmatrix} \quad (38)$$

$$\varphi_n(y_n) = -\frac{\partial \log p(y_n)}{\partial y_n} \quad (39)$$

このスコア関数を計算するためには、分離音源（即ち信号源）の生成モデル  $p(y_n)$  が必要となる．混合信号しか得られていない状況で、混合前の信号の統計的性質や生成モデルを得ることは通常不可能であるが、ICA においてはこれを深く考えずに定めてしまっても上手くいくことが多い．例えば、図 4 で例示したように、音声信号や音楽信号は優ガウスな分布に従うことが多く、逆に正弦波信号は劣ガウスな分布に従う傾向にある．このように、優ガウスか劣ガウスかという曖昧な判断基準をもって  $p(y_n)$  やスコア関数  $\varphi_n(y_n)$  を定めてしまうことが実用上は有効である．スコア関数に関するより詳細な議論は次節に記述する．

式 (37) 中の  $E[\cdot]$  は期待値であり、次式で定義される．

$$E[F(z)] = \int_{-\infty}^{\infty} F(z)p(z)dz \quad (40)$$

ここで、 $F(\cdot)$  は任意の関数である．例えば、 $F(z) = z$ （恒等関数）では式 (40) は確率変数  $z$  の平均であり、 $F(z) = (z - E[z])^2$  では  $z$  の分散となる．また、 $F(z) = z^n$  のとき、式 (40) は  $z$  の  $n$  次統計量（モーメント）に対応する．実際には観測信号  $\mathbf{x}(t)$  が有限のサンプル数であること考慮し、各時間サンプルが独立同一分布に従う（independent and identically distributed: i.i.d.）ことを仮定して、次式のように経験期待値（標本期待値）で近似できる．

$$\begin{aligned} E[\varphi(\mathbf{y})\mathbf{y}^T] &\approx \frac{1}{T} \sum_{t=1}^T \varphi(\mathbf{y}(t))\mathbf{y}(t)^T \\ &= \frac{1}{T} \sum_{t=1}^T \begin{pmatrix} \varphi_1(y_1(t)) \\ \varphi_2(y_2(t)) \end{pmatrix} \begin{pmatrix} y_1(t) & y_2(t) \end{pmatrix} \\ &= \frac{1}{T} \sum_{t=1}^T \mathbf{R}(t) \end{aligned} \quad (41)$$

$$\mathbf{R}(t) = \begin{pmatrix} \varphi_1(y_1(t))y_1(t) & \varphi_1(y_1(t))y_2(t) \\ \varphi_2(y_2(t))y_1(t) & \varphi_2(y_2(t))y_2(t) \end{pmatrix} \quad (42)$$

最急降下法に基づく式 (36) はベル＝セイノフスキーのアルゴリズム<sup>\*6</sup>と呼ばれる．しかしながら、式 (36) のアルゴリズムは行列  $\mathbf{W}^T$  の逆行列演算を含むため、安定性に欠ける不安がある．より高速な収束と逆行列演算の問題解決を同時に達成したアルゴリズムとして、自然勾配法（natural gradient algorithm）<sup>\*7</sup>が後に提案された．自然勾配法は、最適化変数  $\mathbf{W}$  が張る空間の幾何学的な構造を加味<sup>\*8</sup>して、より最適な方向へと変数を更新する工夫が導入されている．自然勾配法は、

<sup>\*6</sup> A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Computation*, no. 7, pp. 1129–1159, 1995.

<sup>\*7</sup> S. Douglas and S. Amari, “Natural-gradient adaptation,” in *Unsupervised adaptive filtering*, Ed. S. Haykin, vol. I, pp. 13–61, Wiley, 2000.

<sup>\*8</sup> 最適化変数空間の歪みを補正するリーマン計量を掛けて、歪んだ空間での最急勾配を計算している．

ICA の最急降下法に基づく更新式 (36) の右辺に対して、右から  $\mathbf{W}^T \mathbf{W}$  を乗じた形となっており、 $\mathbf{W}^T$  の逆行列が打ち消されて次式のようになる。

$$\mathbf{W}^{(l+1)} = \mathbf{W}^{(l)} - \mu \left( \mathbf{E} \left[ \varphi(\mathbf{y}^{(l)}) \mathbf{y}^{(l)T} \right] - \mathbf{I} \right) \mathbf{W}^{(l)} \quad (43)$$

これにより、自然勾配法は最急降下法における逆行列計算の不安定性を解消しながら、より高速な収束が期待できるアルゴリズムとなっている。

以上をまとめると、ICA による BSS は以下のような手順となる。

1. 乱数や単位行列等で適当に初期分離行列  $\mathbf{W}^{(0)}$  を定める
2. 式 (43) を規定回数  $L$  回反復計算し、最終的な分離行列  $\mathbf{W}^{(L)}$  を得る
3. 最終的な分離行列  $\mathbf{W}^{(L)}$  から分離信号  $\mathbf{y}(t)^{(L)}$  を式 (37) で計算する

このとき、ステップサイズ  $\mu$  や反復回数  $L$  を事前に決めておく必要があるが、これは反復毎の目的関数値 (31) の減少を観察しながら試行により適切な値に定めれば良い。但し、目的関数値 (31) は確率変数に対する計算であるため、無限個のサンプルに対する期待値計算（積分）を含んでいる。実際には、 $\mathbf{x}(t)$  が有限のサンプル数であることを考慮し、各時間サンプルが i.i.d. であることを仮定して、経験期待値で近似できる。

$$\begin{aligned} \mathcal{J}(\mathbf{W}) &\propto -\log |\det \mathbf{W}| - \left[ \int_{-\infty}^{\infty} p(y_1) \log p(y_1) dy_1 + \int_{-\infty}^{\infty} p(y_2) \log p(y_2) dy_2 \right] \\ &= -\log |\det \mathbf{W}| - \sum_{n=1}^2 \mathbf{E} [\log p(y_n)] \\ &\approx -\log |\det \mathbf{W}| - \frac{1}{T} \sum_{t=1}^T \sum_{n=1}^2 \log p(y_n(t)) \end{aligned} \quad (44)$$

ここで、最適化変数の含まない定数項（式 (31) の第 1 項）は省いている。

## 2.8 頻繁に用いられる生成モデルとスコア関数

式 (39) のスコア関数には、以下の仮定が良く用いられる\*9。

- ラプラス分布（優ガウス分布）

$$p(y) = \frac{1}{2\sigma} \exp \left( -\frac{|y|}{\sigma} \right) \quad (45)$$

$$\varphi(y) = \frac{1}{\sigma} \frac{y}{|y|} \quad (46)$$

- 双曲線正割分布（優ガウス分布）

$$p(y) = \frac{1}{\pi\sigma^2} \operatorname{sech} \left( \frac{y}{\sigma^2} \right) \quad (47)$$

$$\varphi(y) = \tanh \left( \frac{y}{\sigma^2} \right) \quad (48)$$

---

\*9 ここで示したもの以外の生成モデルとスコア関数については、次の書籍に詳しい: A. Cichocki and S. Amari, “Adaptive blind signal and image processing,” Wiley, 2002.

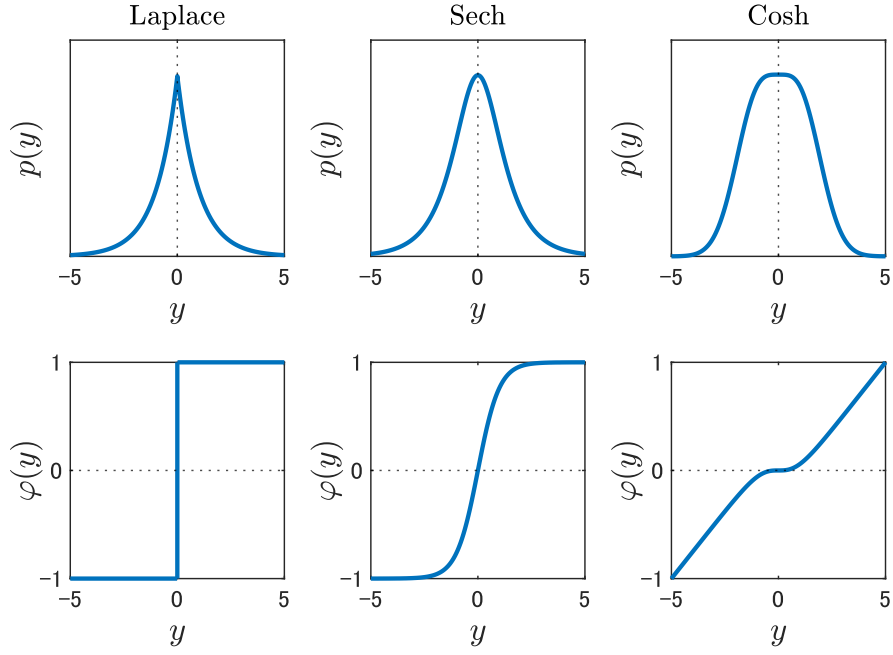


図8 様々な生成モデル  $p(y)$  とスコア関数  $\varphi(y)$  のグラフ．上段が分布の概形，下段がスコア関数をそれぞれ表している．ラプラス分布及び双曲線正割分布は優ガウスな分布であり，双曲線余弦分布は劣ガウスな分布である．

- 双曲線余弦分布（劣ガウス分布）

$$p(y) = \frac{1}{\alpha} \exp\left(-\frac{y^2}{2}\right) \cosh(y) \quad (49)$$

$$\varphi(y) = y - \tanh(y) \quad (50)$$

ここで， $\sigma$  は分布のスケールに関するパラメータ， $\alpha$  は正規化係数である．式 (45)–(50) のグラフを図8に示している．混合前の信号源が優ガウスに従う信号と予想されるならば， $p(y_n)$  をラプラス分布や双曲線正弦関数に基づく分布と仮定し，式 (46) 又は (48) を用いる．また，劣ガウスに従う信号と予想されるならば， $p(y_n)$  を双曲線余弦分布と仮定し，式 (50) を用いる．なお，本資料では割愛したが，ICA による BSS は中心極限定理（central limit theorem）に基づいている．その関係上，ガウス分布に従う複数の信号源の混合を分離することは原理的に不可能である．従って，ICA では常に「ガウス分布に従う信号は高々1つ以下である」という状況を想定しており，基本的には非ガウス（優ガウス又は劣ガウス）な分布に従う信号源が混合されていると仮定する点に注意が必要である．

## 2.9 ICA に基づく BSS の後処理

ICA は信号源間の独立性と，信号源そのものの統計的性質（非ガウス性）に基づいているため，分離信号には以下の不確実性が存在する．

1. 分離信号の順番は推定できない
2. 分離信号のスケールは推定できない

これらの不確実性は分離信号に対して図9のように現れる．上記の不確実性1より，元々の信号源

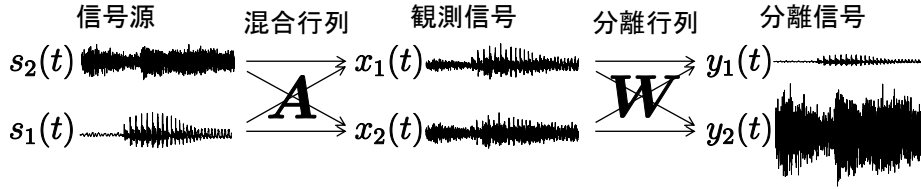


図9 ICAにおける不確定性. ICAは分離信号の順番とスケール（大きさ）を推定することはできない.

の順番が入れ替わる可能性がある. これは最適化における局所解の違いであり, 分離行列の初期値  $\mathbf{W}^{(0)}$  やステップサイズ  $\mu$  によって,  $y_1(t)$  が  $s_1(t)$  と  $s_2(t)$  のどちらの推定信号となるかは変わってしまう ( $y_2$  は残りの信号源の推定信号となる), という現象である. また, 不確定性 2 より, 分離信号のスケール (信号の大きさ) が混合前の源信号のスケールから変化してしまう可能性がある. これは, 変数  $\mathbf{W}$  に適当な係数  $b$  を乗じて  $b\mathbf{W}$  としても, あるいは分離行列の行ベクトル毎に異なる係数  $b_n$  を乗じて  $b_n \mathbf{w}_n$  としても, ICA の目的関数値 (31) は影響を受けないことに起因している. 確かに, 信号の統計的性質や信号間の独立性は信号の順番や定数倍に対して不変であることから, ICA による BSS に上記の不確定性が存在することは直感的に理解できる.

ICA を混合信号に適用する際に, 事前に信号の順番を望み通りに固定することは少々困難である. しかしながら, 信号の順番に興味がある状況は (時間信号に対する ICA においては) 稀であり, 「分離後に信号を確認して判断する」という対処法でも問題がない場合が多い. 一方で, 信号のスケールに関しては補正する必要がある場合が多い. 例えば, 音声の混合信号を分離した結果, 音量が非常に大きな, あるいは零に等しい分離信号が得られても, そのままでは有用性に欠ける. 特定の振幅値に正規化を施す後処理を適用しても良いが, より自然な後処理としては「観測信号  $\mathbf{x}(t)$  中の  $s_1(t)$  及び  $s_2(t)$  のスケールに揃える」という方法がある. これはプロジェクションバック法と呼ばれ, 信号源数と観測信号数がそれぞれ 2 つの場合は次式の後処理となる.

- 分離信号  $y_1(t)$  に対するスケール補正

$$\begin{pmatrix} \hat{y}_{11}(t) \\ \hat{y}_{12}(t) \end{pmatrix} = \mathbf{W}^{(L)-1} \begin{pmatrix} y_1(t)^{(L)} \\ 0 \end{pmatrix} \quad (51)$$

- 分離信号  $y_2(t)$  に対するスケール補正

$$\begin{pmatrix} \hat{y}_{21}(t) \\ \hat{y}_{22}(t) \end{pmatrix} = \mathbf{W}^{(L)-1} \begin{pmatrix} 0 \\ y_2(t)^{(L)} \end{pmatrix} \quad (52)$$

ここで,  $\hat{y}_{nm}(t)$  は  $m$  番目のセンサの観測信号にスケールを合わせた  $n$  番目の推定信号源である. このプロジェクションバック法は, 分離行列  $\mathbf{W}$  によって生じた信号のスケールの歪みを, その逆行列  $\mathbf{W}^{-1}$  によって復元するという処理である. 当然, 逆行列  $\mathbf{W}^{-1}$  を左から乗じる処理は再び分離信号を混合してしまうが, 式 (51) 及び (52) のように, あらかじめ片方の分離信号を 0 に置き換えておくことで, 分離信号を混合することなくスケールを補正している.

スケール補正後の分離信号はセンサ番号の数だけ得られるが, 1 次元信号を出力したい場合は適当にリファレンスとなるセンサの信号を選択すればよい. 例えばリファレンスとなるセンサを  $n = 1$  とすると,  $\hat{y}_{11}(t)$  及び  $\hat{y}_{21}(t)$  が最終的な分離信号となる.

## 2.10 ICA に基づく BSS の疑似アルゴリズム

本章のまとめとして、ICA に基づく BSS の疑似アルゴリズムを Algorithm 1 に示す。実装における注意点としては、8 行目の  $\mathbf{R}(t)$  の計算が  $2 \times 1$  の縦ベクトルと  $1 \times 2$  の横ベクトルの行列積であり、演算結果が  $2 \times 2$  の行列になる点、5~10 行目及び 13~15 行目のような  $t$  に関する繰り返し演算は for 文を用いずに一気に計算してしまう方が高速になる点（但しこれはプログラム言語に依存する）等が挙げられる。また、音源数と観測信号数を 2 ではなく  $N$  として一般化した形で実装すれば、信号源数とセンサ数が 3 や 4 に増加した場合も全く同じアルゴリズムで分離することができるので挑戦して欲しい。

---

**Algorithm 1:** 自然勾配法に基づく ICA による BSS の疑似アルゴリズム

---

入力: 多次元観測信号  $\mathbf{x}(t)$  ( $t = 1, \dots, T$ ), ステップサイズ  $\mu$ , 反復回数  $L$ , スコア関数  $\varphi(\cdot)$

出力: スケール補正後の分離信号  $\hat{y}_{nm}(t)$  ( $n = 1, 2, m = 1, 2, t = 1, \dots, T$ )

```
1 乱数や単位行列等で分離行列  $\mathbf{W}^{(0)}$  を初期化;
2  $\mathbf{I}$  をサイズ  $2 \times 2$  の単位行列として定義; // 式 (43) で利用
3 for  $l = 0$  to  $L - 1$  do
4    $\mathbf{E}$  をサイズ  $2 \times 2$  の零行列で初期化; // 式 (41) における経験期待値の計算用
5   for  $t = 1$  to  $T$  do
6      $\mathbf{y}(t) = \mathbf{W}^{(l)} \mathbf{x}(t)$ ; // 式 (37) 参照,  $\mathbf{y}(t)$  のサイズは  $2 \times 1$ 
7      $\mathbf{p} = \varphi(\mathbf{y}(t))$ ; // 式 (38) 参照,  $\mathbf{p}$  のサイズは  $2 \times 1$ 
8      $\mathbf{R}(t) = \mathbf{p} \mathbf{y}(t)^T$ ; // 式 (42) 参照,  $\mathbf{R}(t)$  のサイズは  $2 \times 2$ 
9      $\mathbf{E} = \mathbf{E} + (1/T) \mathbf{R}(t)$ ; // 式 (41) における経験期待値の計算
10  end
11  $\mathbf{W}^{(l+1)} = \mathbf{W}^{(l)} - \mu (\mathbf{E} - \mathbf{I}) \mathbf{W}^{(l)}$ ; // 式 (43) 参照
12 end
13 for  $t = 1$  to  $T$  do
14    $\mathbf{y}(t)^{(L)} = \mathbf{W}^{(L)} \mathbf{x}(t)$ ; // 分離信号
15 end
16 for  $t = 1$  to  $T$  do
17   式 (51) 及び (52) より  $\hat{y}_{nm}(t)$  を計算; // プロジェクションバック法によるスケール補正
18 end
```

---

## 3 実験手順

本実験では、Mathworks 社の MATLAB という科学技術用のスクリプト言語を用いて ICA のアルゴリズムを実装する。但し、MATLAB は有償ソフトウェアであるため利用が困難な場合は、MATLAB と互換性の高い GNU Octave でも代用できる。