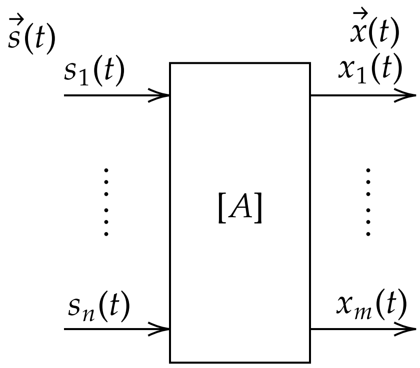
1. 独立元

1.1. 基本概念



$\vec{x}(t) = [x_1(t), x_2(t), \cdots, x_m(t)]^\top \in R^m$

$\vec{s}(t) = [s_1(t), s_2(t), \cdots, s_n(t)]^\top \in R^n$

$\vec{s}(t)$ 中各变量相互独立

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} = [A] \begin{bmatrix} s_1(t) \\ \vdots \\ s_n(t) \end{bmatrix} \text{ 即 } \vec{x}(t) = [A]\vec{s}(t)$$

$[A] \in R^{m\times n}$ 未知

目标：由 $\vec{x}(t)$ 求出 $\vec{s}(t)$

1.2. ICA 求解的可能性和等价性

1.2.1. 求解的可能途径

$\vec{x} = [A]\vec{s}$ $\quad \vec{x} \in R^n \quad \vec{s} \in R^m \quad [A] \in R^{m\times n}$ $\quad m$ 个方程，$m \times n + n$ 个未知数

二阶统计量

$$\{R_x\} = E\left\{\vec{x}(t)\vec{x}(t)^\top\right\} = E\left\{[A]\vec{s}(t)\vec{s}(t)^\top[A]^\top\right\} = [A]E\left\{\vec{s}(t)\vec{s}(t)^\top\right\}[A]^\top = [A][R_s][A]^\top$$

$\dfrac{m+m^2}{2}$ 个方程，$\dfrac{n^2+n}{2} + m\times n$ 个未知数，当 $m \gg n$ 时，方程可能有解

1.2.2. 解的等价性

$\vec{x}(t) = [A]\vec{s}(t)$，令 $[A]' = [A][M]$ $\ \vec{s}'(t) = [M]^{-1}\vec{s}(t)$，则

$$\vec{x}(t) = [A][M][M]^{-1}\vec{s}(t) = [A]'\vec{s}'(t) \text{ 其中 } [M] = [P][A], [P] \text{ 是置换矩阵}$$

1.2.3. 总结

- $s_i(t), i = 1, 2, \cdots, n$ 相互独立
- $p_s(s_1, s_2, \ldots, s_n) = p_1(s_1)\ldots p_n(s_n)$
- $s_i(t), i = 1, 2, \ldots, n$ 是非高斯分布或只有一个元是高斯分布
- 求解结果时，$s_i(t)$ 的次序和幅度是可变的

1.3. 高阶统计量

1.3.1. 矩、中心矩

n 阶矩 $m_n = E\left\{x^n\right\}$ 一阶矩 $m_1 = E\{x\}$

n 阶中心矩 $\mu_n = E\left\{(x-m_x)^2\right\}$ 二阶中心矩 $m_2 = E\left\{(x-m_x)^2\right\} = \sigma^2$

1.3.2. 高阶累积量

$$\phi(\omega) = E\{\exp(j\omega x)\} = E\left\{\sum_{k=0}^{\infty}\frac{x^k(j\omega)^k}{k!}\right\} = \sum_{k=0}^{\infty}E\left\{x^k\right\}\frac{(j\omega)^k}{k!}$$

$$\varphi(\omega) = \ln(\phi(\omega)) = \ln(\exp(j\omega x))$$

$$\varphi(\omega) = \sum_{k=0}^{\infty}K_k\frac{(j\omega)^k}{k!} \quad K_k = (-j)^k\frac{\mathrm{d}^k\phi(\omega)}{\mathrm{d}\omega^k}\bigg|_{\omega=0}$$

一阶累积量 $k=1$

$$\frac{\partial(\varphi(\omega))}{\partial(j\omega)}\bigg|_{\omega=0} = \left(\frac{1}{E(\omega)}E\{x\exp(j\omega x)\}\right)\bigg|_{\omega=0} = E\{x\}$$

二阶累积量 $k=2$

$$\frac{\partial^2(\varphi(\omega))}{\partial(j\omega)^2}\bigg|_{\omega=0} = \left(-\frac{1}{E^2(\omega)}(E\{x\exp(j\omega x)\})^2\right)\bigg|_{\omega=0} + \left(\frac{1}{E(\omega)}E\left\{x^2\exp(j\omega x)^2\right\}\right)\bigg|_{\omega=0}$$
$$= E\left\{x^2\right\} - \left(E\left\{x^2\right\}\right)^2$$

三阶累积量 $k=3$

$$\frac{\partial^3(\varphi(\omega))}{\partial(j\omega)^3}\bigg|_{\omega=0} = E\left\{x^3\right\} - 3E\left\{x^2\right\}E\{x\} + 2(E\{x\})^3$$

四阶累积量 $k=4$

$$\frac{\partial^4(\varphi(\omega))}{\partial(j\omega)^4}\bigg|_{\omega=0} = E\left\{x^4\right\} - 3(E\{x\})^2 - 4E\left\{x^2\right\}E\{x\} + 12E\left\{x^4\right\}(E\{x\})^2 - 6(E\{x\})^4$$

Kurtosis 性质：$kurt(x+y) = kurt(x) + kurt(y) \quad kurt(\alpha x) = \alpha^4 kurt(x)$

1.4. 高斯分布、亚高斯分布、超高斯分布

1.4.1. 高斯分布

$$p_x(x) = \frac{1}{\sqrt{2\pi}\delta} \exp\left[-\frac{(x-E\{x\})^2}{2\sigma^2}\right]$$

矢量形式 $p_x(\vec{x}) = \dfrac{1}{(2\pi)^{\frac{n}{2}} det([C_x])} \exp\left[-\dfrac{1}{2}(\vec{x}-E(\vec{x}))^\top [C_x]^{-1}(\vec{x}-E(\vec{x}))\right]$

1.4.2. 亚高斯分布

比高斯分布更为均匀的分布

1.4.3. 超高斯分布

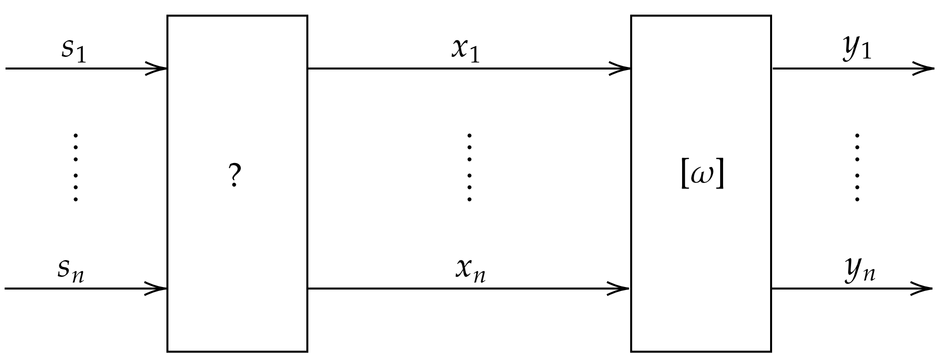
比高斯分布更尖锐的分布

- $kurt = 0 \Rightarrow$ 高斯分布
- $kurt > 0 \Rightarrow$ 超高斯分布
- $kurt < 0 \Rightarrow$ 亚高斯分布

例：均匀分布 $x \sim U(-a, a)$，

$$kurt = E\left\{x^4\right\} - 3(E\{x\})^2 = \int_{-a}^{a} x^4 \frac{1}{2a} \mathrm{d}x - 3\int_{-a}^{a} x^2 \frac{1}{2a} \mathrm{d}x < 0$$

2. 非目标函数 H-J 算法



$m = n$，噪声为 0。$\vec{y} = \vec{x} - [\omega]\vec{y}$，其中 $\omega_{ii} = 1$ $\omega_{ji} : y_j$ 与 x_i 之间的联接权

$([I] + [\omega])\vec{y} = \vec{x} \Rightarrow \vec{y} = ([I] + [\omega])^{-1}\vec{x}$

权重的调整方式

$$\begin{cases} \dfrac{\mathrm{d}\omega_{ij}(t)}{\mathrm{d}t} = \eta E\{f(y_i)g(y_i)\} = \eta E\{f(y_i)\}E\{g(y_i) \approx \eta f(y_i)g(y_i) & i \neq j \\ \omega_{ii} = 1 & i = j \end{cases}$$

其中 y_i 与 y_j 相互独立 $\Leftrightarrow$ $E\{f(y_i)g(y_i) = E\{f(y_i)\}E\{g(y_i)\}$

$f(y_i)$、$g(y_i)$ 必须是奇函数 （一般概率密度函数假设为偶函数）

y_i 与 y_j 相互独立时，$\dfrac{\mathrm{d}\omega_{ij}}{\mathrm{d}t} = 0$

缺点：源信号幅度相差较大或（$[I] + [\omega]$) 接近奇异时，收敛慢

改进形式：$\dfrac{\mathrm{d}\vec{\omega}}{\mathrm{d}t} = \eta\Big([I] - \vec{f}(\vec{y})\vec{g}^\top(\vec{y})\Big)$

独立时 $\vec{f}(\vec{y})\vec{g}^\top(\vec{y}) = \begin{bmatrix} f(y_1)g(y_1) & 0 & \cdots & 0 \\ 0 & f(y_2)g(y_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(y_n)g(y_n) \end{bmatrix}$

3. 四阶统计量 ICA

先白化预处理，应用 4 阶矩方法求解。

$\vec{x}(t) \to \vec{Z}(t)$，且 $\delta_{ij}^2 = E\{Z_i, Z_j\} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \vec{x}(t) \in R^m, \vec{s}(t) \in R^n, m \geqslant n$

3.1. 白化预处理

无噪声时，$\vec{x}(t) = [A]\vec{s}(t)$，其中 $\vec{x}(t)$ 已去除均值。

设协方差矩阵 $[C_x]$ 的特征向量为 $\vec{u}_1, \cdots, \vec{u}_m$，对应的特征值为 $\lambda_1, \cdots, \lambda_m$，对 $[C_x]$ 进行奇异值分解，则

$$\begin{aligned} [C_x] &= [\vec{\mu}_1, \cdots, \vec{\mu}_m] diag(\lambda_1, \cdots, \lambda_m)[\vec{\mu}_1, \cdots, \vec{\mu}_m]^\top \\ &= [\vec{\mu}_1, \cdots, \vec{\mu}_m] diag\Big(\underbrace{\lambda_1, \cdots, \lambda_m}_{\text{源的个数}}, \underbrace{0, \cdots, 0}_{m-n}\Big)[\vec{\mu}_1, \cdots, \vec{\mu}_m]^\top \end{aligned}$$

有噪声时 $\vec{x}(t)=[A]\vec{s}(t)+n(t)$，对 $[C_x]$ 进行奇异值分解，则

$$
\begin{aligned}
[C_x] &= E\left\{([A]\vec{s}(t)+\vec{n}(t))([A]\vec{s}(t)+\vec{n}(t))^\top\right\} \\
&= [A]E\{\left(\vec{s}(t)\vec{s}(t)^\top\right\}[A]^\top + E\left\{\vec{n}(t)\vec{n}(t)^\top\right\} \\
&= [A][R_s][A]^\top + \delta^2[I] \\
&= (\vec{\mu}_1,\vec{\mu}_2,\cdots,\vec{\mu}_m)diag\Big(\underbrace{\lambda_1+\delta^2,\cdots,\lambda_n+\delta^2}_{\text{源的个数}},\underbrace{\delta^2,\cdots,\delta^2}_{m-n}\Big)(\vec{\mu}_1,\cdots,\vec{\mu}_m)^\top
\end{aligned}
$$

令 $[M]=diag\left(\frac{1}{\sqrt{\lambda_1}},\frac{1}{\sqrt{\lambda_2}},\cdots,\frac{1}{\sqrt{\lambda_n}}\right)(\vec{\mu}_1,\vec{\mu}_2,\cdots,\vec{\mu}_n)^\top\in R^{n\times m}$，为白化矩阵

则可知 $\vec{Z}(t)=[M]\vec{x}(t)=[M][A]\vec{s}(t)\in R^{n\times 1}$，由此可以推出

$$
\begin{aligned}
E\left\{\vec{Z}(t)\vec{Z}(t)^\top\right\} &= [M]\left(E\left\{\vec{x}(t)\vec{x}(t)^\top\right\}\right)[M]^\top \\
&= diag\left(\frac{1}{\sqrt{\lambda_1}},\cdots,\frac{1}{\sqrt{\lambda_m}}\right)[\vec{\mu}_1,\cdots,\vec{\mu}_m]^\top[C_x][\vec{\mu}_1,\cdots,\vec{\mu}_m]diag\left(\frac{1}{\sqrt{\lambda_1}},\cdots,\frac{1}{\sqrt{\lambda_m}}\right)^\top \\
&= diag\left(\frac{1}{\sqrt{\lambda_1}},\cdots,\frac{1}{\sqrt{\lambda_m}}\right)[\vec{\mu}_1,\cdots,\vec{\mu}_m]^\top[\vec{\mu}_1,\cdots,\vec{\mu}_m]diag(\lambda_1,\cdots,\lambda_m)\ldots \\
&\ldots\ [\vec{\mu}_1,\cdots,\vec{\mu}_m]^\top[\vec{\mu}_1,\cdots,\vec{\mu}_m]diag\left(\frac{1}{\sqrt{\lambda_1}},\cdots,\frac{1}{\sqrt{\lambda_m}}\right)^\top \\
&= diag\left(\frac{1}{\sqrt{\lambda_1}},\cdots,\frac{1}{\sqrt{\lambda_m}}\right)diag(\lambda_1,\cdots,\lambda_m)diag\left(\frac{1}{\sqrt{\lambda_1}},\cdots,\frac{1}{\sqrt{\lambda_m}}\right)^\top \\
&= [\mathrm{I}]
\end{aligned}
$$

假定 $E\left\{\vec{s}(t)\vec{s}(t)^\top\right\}=[\mathrm{I}]_{\min}$（即已经归一化），则可知

$$
E\left\{\vec{Z}(t)\vec{Z}(t)^\top\right\}=E\left\{[M][A]\vec{s}(t)([M][A]\vec{s}(t))^\top\right\}\xlongequal{\text{令 }[B]=[M][A]}E\left\{[B]\vec{s}(t)\vec{s}(t)^\top[B]^\top\right\}
$$

则 原式 $=[B][\mathrm{I}][B]^\top=[B][B\}^\top=[\mathrm{I}]$

3.2. 应用 4 阶矩求解

白化后样本 $\vec{z}^1, \vec{z}^2, \cdots, \vec{z}^N$，其四阶矩为

$$[R_{z4}] = E\left\{\vec{Z}(t)\vec{Z}(t)^\top\left(\vec{Z}(t)\vec{Z}(t)^\top\right)^\top\right\} = E\left\{\vec{Z}(t)\vec{Z}(t)^\top\vec{Z}(t)\vec{Z}(t)^\top\right\} = E\left\{||\vec{Z}(t)||^2\vec{Z}(t)\vec{Z}(t)^\top\right\}$$

其中 $||\vec{Z}(t)||^2 = ([B]\vec{s}(t))^\top[B]\vec{s}(t) = \vec{s}(t)^\top[B]^\top[B]\vec{s}(t) = \vec{s}(t)^\top\vec{s}(t) = ||\vec{s}(t)||^2 = \sum_{k=1}^{n} s_n^2(t)$

$\vec{Z}(t)\vec{Z}(t)^\top = [B]\vec{s}(t)\vec{s}(t)^\top[B]^\top = \sum_{i=1}^{n}\sum_{j=1}^{n} s_i(t)s_j(t)\vec{\beta}_i\vec{\beta}_j^\top$，$\vec{\beta}_i, \vec{\beta}_j$ 是 [B] 第 i 和 j 行的行向量

由此可推出四阶矩为

$$[R_{z4}] = E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} s_i(t)s_j(t)s_k^2(t)\vec{\beta}_i\vec{\beta}_j^\top\right\}$$

当 $i \neq j$ 时，讨论如下：

- $i \neq j \neq k \;\Rightarrow\; E\left\{s_i(t)s_j(t)s_k^2(t)\right\} = E\{s_i(t)\}E\{s_j(t)\}E\left\{s_k^2(t)\right\} = 0$
- $i \neq j$ 且 $i = k \;\Rightarrow\; E\left\{s_i(t)s_j(t)s_k^2(t)\right\} = E\{s_j(t)\}E\left\{s_k^3(t)\right\} = 0$
- $i \neq j$ 且 $j = k \;\Rightarrow\; E\left\{s_i(t)s_j(t)s_k^2(t)\right\} = E\{s_i(t)\}E\left\{s_j^3(t)\right\} = 0$

当 $i = j$ 时，讨论如下：

- $i = j = k \;\Rightarrow\; E\left\{s_i(t)s_j(t)s_k^2(t)\right\} = E\left\{s_i^4(t)\right\} = \mu_i$
- $i = j \neq k \;\Rightarrow\; E\left\{s_i(t)s_j(t)s_k^2(t)\right\} = E\left\{s_i^2(t)\right\}E\left\{s_k^2(t)\right\} = 1$

则 $[R_{z4}] = \sum_{i=1}^{n}(\mu_i + n - 1)\vec{\beta}_i\vec{\beta}_j^\top \;\Rightarrow\; [R_{z4}]\vec{\beta}_i = (\mu_i + n - 1)\vec{\beta}_i \quad \vec{\beta}_i^\top\vec{\beta}_j = \begin{cases}1 & i = j\\ 0 & i \neq j\end{cases}$

正交的 $\vec{\beta}_i(i = 1, \cdots, n)$ 是 $[R_{z4}]$ 的特征向量

则可以由 $[R_{z4}] \;\Rightarrow\; [B]$ （$[R_{z4}]$ 特征向量是 $[B]$ 的行向量）

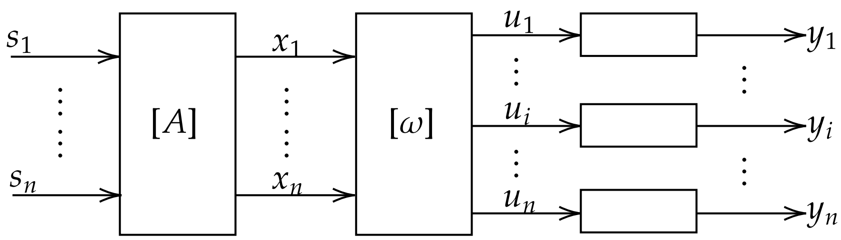
$\therefore \vec{s}(t) = [B]^\top\vec{Z}(t)$

3.3. 算法步骤

$\vec{x}(t) \xrightarrow{\text{白化}}$ 确定源的个数 $\rightarrow$ 白化矩阵 $[M] \rightarrow \vec{Z}(t) = [M]\vec{x}(t) \rightarrow$ 求四阶矩 $[R_{z4}] \rightarrow$ 求 $[R_{z4}]$ 的特征向量 $\rightarrow [B]$ ($[R_{z4}$ 的特征向量为 $[B]$ 的行向量 $\rightarrow \vec{s}(t) = [B]^\top\vec{Z}(t)$ $\left([A] = [M]^{-1}[B]\right)$

4. 最大熵 ICA （有目标函数）

4.1. 概述



$$y_i = \frac{1}{1+e^{-u_i}}$$

$$u_i = \sum_{k=1}^{n} \omega_{kn} x_n$$

$p_y(\vec{y}):\vec{y}$ 的概率密度函数

$$H(\vec{y}) = -\int p_y(\vec{y})\log p_y(\vec{y})\ \mathrm{d}\vec{y} = -E\{\log p_y(\vec{y})\}$$

目标：使得 $\vec{y}$ 中各元相互独立，此时 $H(\vec{y})$ 最大

4.2. 梯度法 $\Delta[\omega] = \eta\dfrac{\partial(H(\vec{y}))}{\partial[\omega]}$

设 $p_x(\vec{x})$ 为 $\vec{x}$ 的概率密度函数， $\vec{y} = f(\vec{x}) = (y_1, y_2, \cdots, y_n)^\top$，则 $p_y(\vec{y}) = \dfrac{p_x(\vec{x})}{|J(x)|}$

其中 J 是 Jacobi 矩阵，$J(\vec{x}) = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y_n}{\partial x_1} & \cdots & \dfrac{\partial y_n}{\partial x_n} \end{bmatrix}$ $\quad \dfrac{\partial y_i}{\partial x_j} = \dfrac{\partial y_i}{\partial u_i}\dfrac{\partial u_i}{\partial x_j} \xlongequal{u_i=\sum_{k=1}^{n}\omega_{ik}x_k} \omega_{ij}\dfrac{\partial y_i}{\partial u_i}$

则可知 $|J(\vec{x})| = \begin{bmatrix} \omega_{11}\dfrac{\partial y_1}{\partial u_1} & \cdots & \omega_{1n}\dfrac{\partial y_1}{\partial u_1} \\ \vdots & \ddots & \vdots \\ \omega_{n1}\dfrac{\partial y_n}{\partial u_n} & \cdots & \omega_{nn}\dfrac{\partial y_n}{\partial u_n} \end{bmatrix} = det[\omega]\prod_i \dfrac{\partial y_i}{\partial u_i}$

$$H(\vec{y}) = -\int p_y(\vec{y})\log p_y(\vec{y})\ \mathrm{d}\vec{y} = -\int p_y(\vec{y})[\log p_x(\vec{x}) - \log|J(x)|]\ \mathrm{d}\vec{y} = -E\{\log p_x(\vec{x}) - \log|J(\vec{x})|\}$$

则 $\dfrac{\partial H(\vec{y})}{\partial[\omega]} = \dfrac{\partial E\{\log|J(x)|\}}{\partial[\omega]} = \dfrac{\partial E\left\{\log\left(det[\omega]\prod_i \frac{\partial y_i}{\partial u_i}\right)\right\}}{\partial[\omega]} = \dfrac{\partial \log det[\omega]}{\partial[\omega]} + \dfrac{\partial\left(\log \prod_i \frac{\partial y_i}{\partial u_i}\right)}{\partial[\omega]}$

$$\Rightarrow \frac{\partial H(\vec{y})}{\partial[\omega]} = \left([\omega]^\top\right)^{-1} + (\vec{I} - 2\vec{y})\vec{x}^\top$$

由此，可知 $\Delta[\omega]=\eta\dfrac{\partial H(\vec{y})}{\partial[\omega]}=\eta\left[\left([\omega]^{\top}\right)^{-1}+(\vec{I}-2\vec{y})\vec{x}^{\top}\right]\ \ \eta>0\quad(*)$

4.3. 求解步骤

- 初始化，随机选取 $[\omega]$
- 求 $\vec{u},\vec{y}$
- 用公式 $(*)$ 更新 $[\omega]$，直到 $\Delta[\omega]<\Delta[\varepsilon]$

该方法特点：

- $m=n$（若 $m\neq n$，先白化预处理）
- y_i 与 u_i 之间的关系：$y_i=\dfrac{1}{1+e^{-u_i}}$

5. 最小互信息 ICA

5.1. 概述

互信息：$I(\vec{y})=\displaystyle\int p(\vec{y})\log\frac{p(\vec{y})}{\prod_i p(y_i)}\,\mathrm{d}\vec{y}=\int p(\vec{y})\log p(\vec{y})\,\mathrm{d}\vec{y}-\int p(\vec{y})\log\prod_i p(y_i)\,\mathrm{d}\vec{y}$

由此可得，$I(\vec{y})=-H(\vec{y})-\displaystyle\sum_{i=1}^{n}H(y_i)=-H(\vec{y})-E\left\{\log\prod_i p(y_i)\right\}$

当 $p(\vec{y})=p(y_1)p(y_2)\cdots p(y_n)$ 时，$y_1,y_2,\cdots,y_n$ 相互独立，$I(\vec{y})$ 最小

5.2. 梯度法

5.2.1. 推导

$$\Delta[\omega]=-\eta\frac{\partial I(\vec{y})}{\partial[\omega]}=-\eta\left(-\frac{\partial H(\vec{y})}{\partial[\omega]}-\frac{\partial E\left\{\log\prod_i p(y_i)\right\}}{\partial[\omega]}\right)$$

$$\because\ p(y_i)=\frac{p(u_i)}{\partial y_i/\partial u_i}\quad\therefore\ \frac{\partial E\left\{\log\prod_i p(y_i)\right\}}{\partial[\omega]}=\frac{\partial E\left\{\log\prod_i\frac{p(u_i)}{\partial y_i/\partial u_i}\right\}}{\partial[\omega]}\approx\frac{\partial\log\prod_i\frac{p(u_i)}{\partial y_i/\partial u_i}}{\partial[\omega]}$$

假设 $p(u_i)=\dfrac{\partial y_i}{\partial u_i}$，在此假设下可以知道

$$\Delta[\omega]=-\eta\frac{\partial I(\vec{y})}{\partial[\omega]}=\eta\left\{\frac{\partial\log det[\omega]}{\partial[\omega]}+\frac{\partial}{\partial[\omega]}\log\prod_{i=1}^{n}\frac{\partial y_i}{\partial u_i}\right\}$$

$$\frac{\partial\left(\log\prod_{i=1}^{n}\frac{\partial y_i}{\partial u_i}\right)}{\partial[\omega]}\xlongequal{p(u_i)=\frac{\partial y_i}{\partial u_i}}\frac{\partial\left(\log\prod_{i=1}^{n}p(u_i)\right)}{\partial[\omega]}\xlongequal{\text{设 } p(\vec{u})=\prod_{i=1}^{n}p(u_i)}\frac{1}{p(\vec{u})}\frac{\partial p(\vec{u})}{\partial\vec{u}}\frac{\partial\vec{u}}{\partial[\omega]}\xlongequal{\vec{u}=[\omega]\vec{x}}\frac{1}{p(\vec{u})}\frac{\partial p(\vec{u})}{\partial\vec{u}}\vec{x}^\top$$

令 $\vec{\phi}(\vec{u})=-\dfrac{1}{p(\vec{u})}\dfrac{\partial p(\vec{u})}{\partial\vec{u}}$，则可知

$$\Delta[\omega]=\eta\left\{\frac{\partial\log det[\omega]}{\partial[\omega]}+\frac{\partial}{\partial[\omega]}\log\prod_{i=1}^{n}\frac{\partial y_i}{\partial u_i}\right\}=\eta\left\{\left([\omega]^\top\right)^{-1}-\vec{\phi}(\vec{u})\vec{x}^\top\right\}$$

5.2.2. 自然梯度法

$$\Delta[\omega]=-\eta\frac{\partial I(\vec{y})}{\partial[u]}\cdot[\omega]^\top[\omega]$$

$$[\omega(t+1)]=[\omega(t)]+\eta\Big([I]-\vec{\phi}(\vec{u})\vec{u}^\top(t)\Big)[\omega(t)]$$

5.2.3. 迭代公式中 $\vec{\phi}(\vec{u})$ 计算

$$\begin{aligned}
\vec{\phi}(\vec{u})&=-\frac{1}{p(\vec{u})}\frac{\partial p(\vec{u})}{\partial\vec{u}}=-\frac{1}{\prod_i p(u_i)}\frac{\partial\left(\prod_i p(u_i)\right)}{\partial\vec{u}}\\
&=-\frac{1}{\prod_i p(u_i)}\left[\frac{\partial p(u_1)}{\partial u_1}\prod_{i\neq 1}p(u_i),\ \cdots,\frac{\partial p(u_n)}{\partial u_n}\prod_{i\neq n}p(u_i)\right]^\top\\
&=-\left[\frac{1}{p(u_1)}\frac{\partial p(u_1)}{\partial u_1},\ \cdots,\frac{1}{p(u_n)}\frac{\partial p(u_n)}{\partial u_n}\right]^\top
\end{aligned}$$

亚高斯分布

$p(u_i)=\frac{1}{2}\left(N_i\left(u,\sigma^2\right)+N_i\left(-\mu,\sigma^2\right)\right)$，其中 $N\left(\mu,\sigma^2\right)=\frac{1}{\sqrt{2\pi\delta}}\exp\left[-\frac{(\mu_i-\mu)^2}{2\delta^2}\right]$

则可知 $\frac{\partial p(u_i)}{\partial u_i}=\frac{1}{\sqrt{2\pi\delta}}\left[-\frac{\mu_i+\mu}{\delta^2}\exp\left(-\frac{(\mu_i+\mu)^2}{2\delta^2}\right)-\frac{\mu_i-\mu}{\delta^2}\exp\left(-\frac{(\mu_i-\mu)^2}{2\delta^2}\right)\right]$

$\frac{1}{p(\mu_i)}\frac{\partial p(\mu_i)}{\partial \mu_i}=-\frac{\mu_i}{\delta^2}+\frac{\mu_i}{\delta^2}\tanh\frac{\mu_i\mu}{\delta^2}$，其中 $\tanh\frac{\mu_i\mu}{\delta^2}=\frac{\exp\left(\frac{\mu_i\mu}{\delta^2}\right)-\exp\left(-\frac{\mu_i\mu}{\delta^2}\right)}{\exp\left(\frac{\mu_i\mu}{\delta^2}\right)+\exp\left(-\frac{\mu_i\mu}{\delta^2}\right)}$

如令 $\mu=1,\sigma^2=1$，则 $\frac{1}{p(\mu_i)}\cdot\frac{\partial p(\mu_i)}{\partial \mu_i}=-\mu_i+\tanh(\mu_i)$

亚高斯分布时 $\vec{\phi}(\vec{\mu})=(\mu_1-\tanh(\mu_1),\ \cdots,\mu_n-\tanh(\mu_n))=\vec{\mu}-\tanh(\vec{\mu})$

超高斯分布

$p(\mu_i)=p_G(\mu_i)\mathrm{sech}(\mu_i),\ \mathrm{sech}(\mu_i)=\frac{2}{e^{-\mu_i}+e^{\mu_i}}$， $p_G(\mu_i)$：高斯分布（$\delta^2=1,\mu=0$）

则 $\frac{\partial p(\mu_i)}{\partial \mu_i}=\frac{\partial p_G(\mu_i)}{\partial \mu_i}\mathrm{sech}(\mu_i)+p_G(\mu_i)\frac{\partial\,\mathrm{sech}(\mu_i)}{\partial \mu_i}=-\mu_i p(\mu_i)-p(\mu_i)\tanh(\mu_i)$

则 $\frac{1}{p(\mu_i)}\frac{\partial p(\mu_i)}{\partial \mu_i}=-\mu_i-\tanh(\mu_i)\ \Rightarrow\ \vec{\phi}(\vec{\mu})=\vec{\mu}+\tanh(\vec{\mu})$

5.2.4. 迭代公式

SubGaussion：$[\omega(t+1)]=[\omega(t)]+\eta\left([I]+\tanh(\vec{\mu})\vec{\mu}^{\top}-\vec{\mu}\vec{\mu}^{\top}\right)[\omega(t)]$

SuperGaussion：$[\omega(t+1)]=[\omega(t)]+\eta\left([I]-\tanh(\vec{\mu})\vec{\mu}^{\top}-\vec{\mu}\vec{\mu}^{\top}\right)[\omega(t)]$

根据 $kurt(u_i)=E\left\{u_i^4\right\}-3\left(E\left\{u_i^2\right\}\right)^2=\begin{cases}>0 & Super\\ <0 & Sub\end{cases}$

注意：$\vec{u}$ 中若既有超高斯的信号，又有亚高斯的信号，迭代需注意相应信号 tanh 前正负号

6. Fast ICA

目标函数 $F=E\left\{G_2\left(\vec{\omega}_i^{\top}\vec{z}\right)\right\}+\frac{1}{2}\beta||\vec{\omega}_i||^2$ $\vec{\omega}_i$：已归一化

在目标点 $\vec{\omega}_i^*$ 附近对 F 进行二阶 Taylor 展开，如下

$$F(\vec{\omega}_i) = F\left(\vec{\omega}_i^*\right) + \left.\frac{\partial F(\vec{\omega}_i)}{\partial \vec{\omega}_i}\right|_{\vec{\omega}_i=\vec{\omega}_i^*} \Delta\vec{\omega}_i + \frac{1}{2}\Delta\vec{\omega}_i^\top \left.\frac{\partial^2 F(\vec{\omega}_i)}{\partial \vec{\omega}_i}\right|_{\vec{\omega}_i=\vec{\omega}_i^*} \Delta\vec{\omega}_i$$

因为在目标函数附近 $F(\vec{\omega}_i) - F\left(\vec{\omega}_i^*\right) \approx 0$，故可以得出

$$0 = \left.\frac{\partial F(\vec{\omega}_i)}{\partial \vec{\omega}_i}\right|_{\vec{\omega}_i=\vec{\omega}_i^*} \Delta\vec{\omega}_i + \frac{1}{2}\Delta\vec{\omega}_i^\top \left.\frac{\partial^2 F(\vec{\omega}_i)}{\partial \vec{\omega}_i}\right|_{\vec{\omega}_i=\vec{\omega}_i^*} \Delta\vec{\omega}_i$$

牛顿法 $\Delta\vec{\omega}_i = -\left[\frac{\partial F(\vec{\omega}_i)}{\partial \vec{\omega}_i^*}\right]^{-1} \frac{\partial F(\vec{\omega}_i)}{\partial \vec{\omega}_i}$

其中 $\frac{\partial F(\vec{\omega}_i)}{\partial \vec{\omega}_i} = E\left\{g_2\left(\vec{\omega}_i^\top \vec{z}\right)\vec{z}\right\} + \beta\vec{\omega}_i$

$$\frac{\partial^2 F(\vec{\omega}_i)}{\partial \vec{\omega}_i^2} = E\left\{g_2\left(\vec{\omega}_i^\top \vec{z}\right)\vec{z}\vec{z}^\top\right\} + \beta[I] = E\left\{g_2'\left(\vec{\omega}_i^\top \vec{z}\right)\right\}E\left\{\vec{z}\vec{z}^\top\right\} + \beta[I] = \left(E\left\{g_2'\left(\vec{\omega}_i^\top \vec{z}\right)\right\} + \beta\right)[I]$$

$$\Delta\vec{\omega}_i = -\left[\left(E\left\{g_2'\left(\vec{\omega}^\top \vec{z}\right)\right\} + \beta\right)[I]\right]^{-1}\left[E\left\{g_2\left(\vec{\omega}_i^\top \vec{z}\right)\vec{z}\right\} + \beta\vec{\omega}_i\right] = -\frac{E\left\{g_2\left(\vec{\omega}_i^\top \vec{z}\right)\vec{z} + \beta\vec{\omega}_i\right\}}{E\left\{g_2'\left(\vec{\omega}^\top \vec{z}\right) + \beta\right\}}$$

则可知 $\Delta\vec{\omega}_i(t) = \Delta\vec{\omega}_i(t-1) - \frac{E\left\{g_2\left(\vec{\omega}_i^\top (t-1)\vec{z}\right)\vec{z} + \beta\vec{\omega}_i(t-1)\right\}}{E\left\{g_2\left(\vec{\omega}_i^\top (t-1)\vec{z}\right) + \beta\right\}}$

整理上式，$\vec{\omega}_i$ 需归一化

$$\vec{\omega}_i(t) = E\left\{g_2\left(\vec{\omega}_i^\top (t-1)\vec{z}\right)\vec{z}\right\} - E\left\{g_2'\left(\vec{\omega}_i^\top (t-1)\vec{z}\right)\vec{\omega}_i(t-1)\right\}$$

其中 $G_2(y) = \frac{1}{a}\log(\cosh(a,y)) \quad 1 \leqslant a \leqslant 2 \quad or \quad G_2(y) = -\exp\left(-\frac{y^2}{2}\right)$

$$g_2(y) = \frac{\partial G_2}{\partial y} = \begin{cases} \tan(a,y) & 1 \leqslant a \leqslant 2 \\ y\exp\left(-\frac{y^2}{2}\right) & \end{cases} \qquad g_2'(y) = \begin{cases} a\left(1-\tanh^2(a,y)\right) & 1 \leqslant a \leqslant 2 \\ -\left(1-y^2\right)\exp\left(-\frac{y^2}{2}\right) & \end{cases}$$

Fast ICA 算法步骤

- $\vec{x} \rightarrow \vec{z}$
- 随机初始化 $\vec{\omega}_i(0)$ $\rightarrow$ 归一化 $\vec{\omega}_i(t) = \dfrac{\vec{\omega}_i(t)}{||\vec{\omega}_i(0)||}$
- 计算 $g_i, g_i{'} \rightarrow$ 计算 $\vec{\omega}_i$ 并归一化
- 正交化所有 $\vec{\omega}_i$ $(i = 1, 2, 3, \cdots, n)$

  $$\vec{\omega}_p(t) = \vec{\omega}_p(t-1) - \sum_{k=1}^{p-1} \vec{\omega}_p^{\top}(t-1)\vec{\omega}_k(t)\vec{\omega}_k(t) \quad p = 1, 2, \cdots, n$$

  $\rightarrow$ 再归一化正交化 $\vec{\omega}_p$
- 如果 $\vec{\omega}_p$ 收敛，回到第三步