盲信号处理基础知识

- 随机矢量及其统计量
- 线性方程求解

随机矢量及其统计量

- 对盲信号处理中所涉及到的随机矢量及其统计量基本概念的介绍

－需要有概率论的基本知识

－强调一些盲信号处理中常用的一些概念，如相关性，独立性，非高斯性以及各阶统计量

概率分布函数

给定随机变量ξ，它的取值不超过实数x的事件概率$P(\xi \le x)$是x的函数，称为ξ的概率分布函数，记作

$$F_\xi(x) = P(\xi \le x) \qquad (-\infty < x < \infty)$$

概率分布函数的基本性质：

1、$0 \le F_\xi(x) \le 1$

2、若$x_1 < x_2$，则$F_\xi(x_1) < F_\xi(x_2)$

概率密度函数

如果随机变量ξ的概率分布函数$F_\xi(x)$可以表示为：

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(x)dx$$

就称ξ是连续型随机变量，$p_\xi(x)$是随机变量ξ的概率密度函数。

概率密度函数的基本性质：

1、$p_\xi(x) = \frac{d}{dx}F_\xi(x) \ge 0$

2、$\int_{-\infty}^{\infty} p_\xi(x)dx = 1$

随机变量的函数的概率分布函数

如果随机变量η是随机变量ξ的函数

$$\eta = f(\xi)$$

则

$$F_\eta(x) = \int_{f(y)<x} dF_\xi(y)$$

$$F_\eta(x) = \int_{f(y)<x} p_\xi(y)dy$$

随机矢量的联合概率分布函数

给定随机矢量$\xi = (\xi_1, \xi_2, \cdots \xi_n)$，它的取值不超过实数矢量$x = (x_1, x_2, \cdots x_n)$的事件概率$P(\xi \le x)$是$x$的函数，称为$\xi$的联合概率分布函数，记作

$$F_\xi(x) = P(\xi \le x) \qquad (-\infty < x < \infty)$$

边缘分布函数

记随机矢量 $\xi=\left(\xi_1,\xi_2,\cdots\xi_n\right)$ 中任意 $m(m<n)$ 个随机变量构成的随机矢量为 $\xi^c=\left(\xi_{i_1},\xi_{i_2},\cdots\xi_{i_m}\right)$，则称随机矢量 ξ^c 的联合概率分布函数为随机矢量 ξ 的边缘分布函数

$$F_{\xi^c}(x_{i_1},x_{i_2},\cdots x_{i_m})=F_\xi(\infty,\cdots,x_{i_1},\cdots\infty,\cdots x_{i_m},\cdots\infty)$$

统计独立性

随机矢量 $\xi=\left(\xi_1,\xi_2,\cdots\xi_n\right)$ 的各个随机变量相互独立的条件是：

$$F_\xi(x)=\prod_{i=1}^{n}F_{\xi_i}(x_i)$$

随机变量的数字特征

数学期望(Expectation)

如果随机矢量 η 是随机矢量 ξ 的函数

$$\eta=f(\xi)$$

则 η 的数学期望定义为：

$$E\{\eta\}=\int_{-\infty}^{\infty}f(x)p_\xi(x)dx$$

数学期望可以由信号进行估计得到，而不需要知道其概率密度函数

矩（Moment）

若 $\eta=f(\xi)$ 中函数为随机矢量 ξ 中各个变量的乘积，则得到的 η 的数学期望被称为矩（原点矩）

一阶矩： $m_\xi=E\{\xi\}=\int_{-\infty}^{\infty}xp_\xi(x)dx$

相关（Correlation，二阶矩）

随机矢量 $\xi=\left(\xi_1,\xi_2,\cdots\xi_n\right)$ 中任意一对变量的互相关定义为：

$$r_{ij}=E\left\{\xi_i\xi_j\right\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_ix_jp_{\xi_i,\xi_j}(x_i,x_j)dx_idx_j$$

相关矩阵（大小为 $n\times n$）

$$R_\xi=E\left\{\xi\xi^T\right\}=\left[r_{ij}\right]$$

对称的，具有非负的特征值和相互正交的特征向量。

中心矩（Central moment）

中心矩和原点矩的定义一样，只是在求数学期望前从随机矢量中减去其均值矢量。

协方差(Covariance)：和相关 r_{ij} 对应的中心矩

$$c_{ij}=E\left\{(\xi_i-m_{\xi_i})(\xi_j-m_{\xi_j})\right\}$$

协方差矩阵

$$C_\xi=E\left\{(\xi-m_\xi)(\xi-m_\xi)^T\right\}=\left[c_{ij}\right]$$

对称的，具有非负的特征值和相互正交的特征向量。

$$C_\xi=R_\xi-m_\xi m_\xi{}^T$$

考虑单个随机变量，则一阶矩为均值，二阶原点矩为平均能量，二阶中心矩为方差。

对于两个随机矢量ξ,η，他们的联合数学期望可以利用他们的联合概率密度得到。两个重要的数字特征是：

互相关矩阵（Cross-correlation）

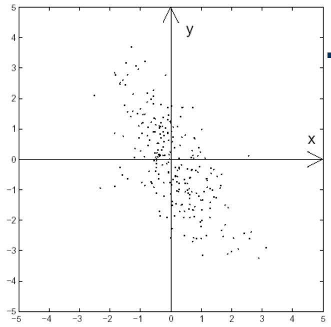
$$R_{\xi\eta}=E\left\{\xi\eta^{T}\right\}$$

互协方差矩阵（Cross-covariance）

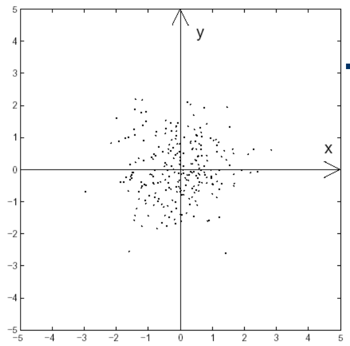
$$C_{\xi\eta}=E\left\{(\xi-m_{\xi})(\eta-m_{\eta})^{T}\right\}$$

度量两个随机矢量ξ,η之间相关性，是二阶统计量。

x, y
负相关



x, y
不相关



考虑离散数字信号$x_i, i=1,\cdots N$，其数字特征可以由信号本身估计得到，而不需要知道其概率密度函数，如

$$m_x=E\{x\}=\frac{1}{N}\sum_{i=1}^{N}x_i$$

高阶统计量（Higher order statistics, HOS）

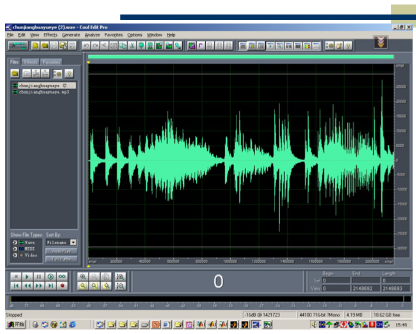
基于二阶统计量有许多标准的信号处理技术。使用如下假设

1、 信号高斯分布

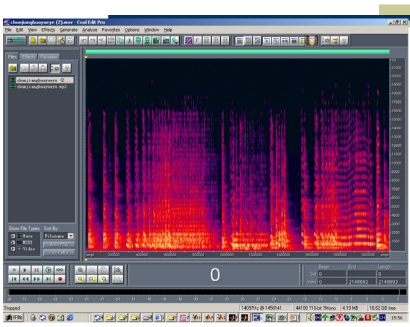
2、 信号是平稳的

当信号是非高斯分布时，高阶统计量携带大量有用的信息。盲反褶积，盲信号分离，独立成分分析都利用了高阶统计量。

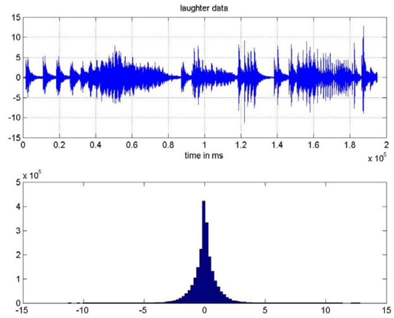
春江花月夜—统计分析



短时谱（振幅谱）



直方图



统计结果

均值Mean	-4.54256e-005
方差Variance	0.00418846
歪斜度Skewness 三阶统计量	0.0442806
峰度Kurtosis 四阶统计量	5.46975

高阶矩(Higher order moment)

随机矢量 ξ 的 j 阶矩的定义为：

$$m_{\xi}^{j}=E\left\{\xi^{j}\right\}=\int_{-\infty}^{\infty}x^{j}p_{\xi}(x)dx，\quad j=1,2,\cdots$$

均值：$j=1$

平均能量：$j=2$

$j>2$：高阶矩

高阶中心矩

和高阶矩定义一样，只是在求高阶矩以前，减去随机矢量 ξ 的均值。

零均值随机矢量的高阶中心矩和高阶矩是一致的

后面的分析假设随机矢量为零均值，高阶中心矩和高阶矩是一致的

偏斜度（Skewness）

三阶中心矩被称为 Skewness，

$$m_{\xi}^{3}=E\left\{\xi^{3}\right\}$$

可以用来度量随机变量的概率密度函数对称性。

$m_{\xi}^{3}=0$ 时，对称的

高阶累计量（Higher order Cumulants）

随机变量ξ的第一特征函数（矩生成函数）

$$\Phi(\omega)=E\{\exp(j\omega\xi)\}=\int_{-\infty}^{\infty}\exp(j\omega x)p_{\xi}(x)dx$$

写成 Laplace 算子形式：$\Phi(s)=E\left\{e^{s\xi}\right\}$

对上式求k阶导数，得到: $\Phi^{k}(s)=E\left\{\xi^{k}e^{s\xi}\right\}$

则：$\Phi^{k}(0)=E\left\{\xi^{k}\right\}=m_{\xi}^{k}$

随机变量的第二特征函数（累计量生成函数）

$\Psi(s)=\ln(\Phi(s))$

随机变量ξ的k阶累计量定义为上式的k阶导数在原点的取值

$$c_{\xi}^{k}=\frac{d^{k}(\Psi(s))}{ds^{k}}|_{s=0}$$

对于零均值随机变量

$c_{\xi}^{1}=0=m_{\xi}^{1}$

$c_{\xi}^{2}=m_{\xi}^{2}$

$c_{\xi}^{3}=m_{\xi}^{3}$

$c_{\xi}^{4}=m_{\xi}^{4}-3\left(m_{\xi}^{2}\right)^{2}$

高阶累计量和高阶矩一样，可以推广到随机矢量。即可以得到互高阶累计量和互高阶矩。

峰度（Kurtosis）

$kurt(\xi)=c_{\xi}^{4}=m_{\xi}^{4}-3\left(m_{\xi}^{2}\right)^{2}$ 四阶累计量

归一化峰度（Normalized kurtosis）

$k(\xi)=m_{\xi}^{4}/\left(m_{\xi}^{2}\right)^{2}-3$

具有如下性质：

$kurt(x+y)=kurt(x)+kurt(y)$

$kurt(\alpha x)=\alpha^{4}kurt(x)$

Kurtosis：一个最简单的非高斯性度量

如果随机变量ξ是高斯分布的，则

$kurt(\xi)=0$，$k(\xi)=0$

四阶矩没有以上的性质，这是 Kurtosis 比四阶矩应用广泛的原因。

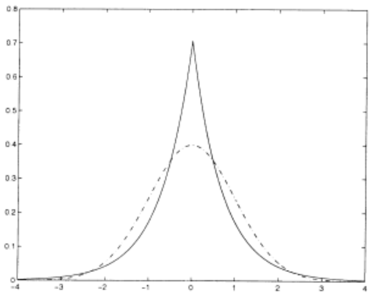
广义高斯密度函数

$$p_{x}(x)=\beta e^{\left(-\frac{|x|^{\gamma}}{\gamma E\{|x|^{\gamma}\}}\right)}$$

$\gamma=2$，标准高斯分布

$\gamma>2$，SubGaussian，例如：均匀分布，$\gamma\to\infty$。$kurt(\xi)<0$

$\gamma<2$，SupperGaussian，例如：Laplace 分布时，$\gamma=1$。$kurt(\xi)>0$

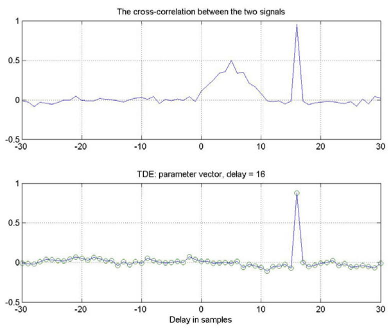


实线：Laplace分布； 虚线：Gaussian分布

高阶统计量的优点：
对非高斯信号，提供更多的信息
如盲信道估计，高阶统计量可以提供相位信息。而二阶统计量只提供振幅信息。
对加性高斯噪声不敏感，时延估计。

高阶统计量的缺点：
1、 估计需要大样本。
2、 对大野值敏感。

时延估计



不相关性、独立性和非高斯性
Uncorrelatedness, Independence and NonGaussianity

两个随机矢量 x, y 的互协方差矩阵为零，即 $C_{xy}=0$，则 x, y 不相关。

等价为：$R_{xy}=E\left\{xy^T\right\}=E\left\{x\right\}E\left\{y^T\right\}=m_x m_y^T$

一个随机矢量 x 的协方差矩阵为对角阵，即 $C_{xx}=D$，则 x 的各个成分变量不相关。

统计独立性：
两个随机变量 x, y，如果知道 y 值，不能给 x 的取值提供任何信息，则 x, y 是独立的，他们具有如下性质：

$$p_{x,y}(x,y)=p_x(x)p_y(y)$$

$$E\left\{g(x)h(y)\right\}=E\left\{g(x)\right\}E\left\{h(y)\right\}$$

其中，$g(), h()$ 是任意函数。可以看到，不相关性是 $g(), h()$ 为线性函数时的特例。

在 ICA 中，非高斯性＝独立性
中心极限定理：
如果相互独立的具有相同分布的随机变量 $\xi_1, \xi_2, \cdots \xi_n$ 的均值和方差均存在，则他们的和比单个随机变量更趋向高斯分布。
在盲信号处理时，最大化输出信号的非高斯性等价于最大化输出信号的独立性

线性方程组求解

信号处理中的许多问题与线性方程组求解有关

1、 反褶积问题
2、 信号分离问题
3、 盲反褶积问题
4、 盲信号分离问题

问题描述

$$Ax = b$$

其中：A 是 $m\times n$ 维矩阵，x 是 $n\times 1$ 向量，b 是 $m\times 1$ 向量。A 和 b 已知。
很多信号处理问题中，$m>n$，称为超定方程组。
$m<n$ ，称为欠定方程组。

$Rank(A)<n$ 时，上述方程组是病态的，需要增加别的约束条件，得到唯一解。

本节内容：

- 从信号处理角度，分别介绍超定方程组的如下解法的特点：

1. 最小二乘解
2. 加权最小二乘解
3. 总体最小二乘解
4. 最小1-范数解
5. 最小∞-范数解

目标函数（能量函数）

$$E = \left\| Ax - b \right\|_p$$

方程组求解是通过最小化上述目标函数实现的。
使用一个假设：噪声只存在于观测信号 b 中，A 是不含噪声的
$e = b - Ax$
选用不同阶范数来构造目标函数，得到不同的解。

最小二乘解（Least-Squares,LS）

目标函数：$E = \left\| Ax - b \right\|_2$

$$p = 2$$

最小二乘解为：

$$x_* = \left(A^T A\right)^{-1} A^T b$$

最小二乘解为最优解的条件为：
当观测信号 b 中的噪声是高斯分布

最小1-范数解

目标函数：$E = \left\| Ax - b \right\|_1$

$$p = 1$$

当观测信号 b 中的噪声是超高斯分布，最小 1-范数解优于最小二乘解。
换一句话来说，当观测信号中只有一小部分信号被大的噪声污染，最小 1-范数解更合理。
最小 1-范数准则被用于稀疏信号表达中。

最小1-范数解

最小化 $E=\|Ax-b\|_1$ 问题，可以利用多阶 LS 来求解

1、 求解最小二乘解：$x_*=\left(A^TA\right)^{-1}A^Tb$

2、 计算每个方程对应的偏差：$e(b)=\left|b-Ax_*\right|$

并判断是否满足终止条件，如不满足，到第 3 步

3、 去掉偏差大的方程，利用剩下的 r 个方程构成新的方程组：

$A_r x=b_r$

返回第 1 步。

最小 ∞ -范数解

目标函数：$E=\|Ax-b\|_\infty$

$p=\infty$

当观测信号 b 中的噪声是亚高斯分布，最小 ∞-范数解可能优于最小二乘解。

最小 ∞-范数解又被称为最小最大解，即使误差信号 e 的绝对最大值最小。

加权最小二乘解

不同方程在求解方程组时所起的作用不同，需要采用加权最小二乘技术。

目标函数：

$E(x)=(b-Ax)^T\sum_e(b-Ax)$

式中，$\sum_e$ 是一个正的 $m\times m$ 的对角阵

$\sum_e=diag\{\sigma_1,\sigma_2,\cdots,\sigma_m\}$

σ_i 越大，对应的方程所带来的误差在目标函数中所起的作用就越大，即该方程越重要。

加权最小二乘解

$x_*=x_{wls}=\left(A^T\sum_e A\right)^{-1}A^T\sum_e b$

当噪声 e 是零均值，其协方差矩阵 R_e 是正定的，$\sum_e$ 用噪声 e 的协方差矩阵 R_e^{-1} 代替，可以得到最优线性无偏估计（Best Linear Unbiased Estimator, BLUE）：

$x_*=x_{BLUE}=\left(A^TR_e^{-1}A\right)^{-1}A^TR_e^{-1}b$

总体最小二乘解（Total least-squares, TLS）

前面分析的方程组的解在实际信号处理中得到广泛的应用，但请记住他们的隐含条件：

噪声只存在于观测信号 b 中，A 是不含噪声的或者其噪声可以忽略不计。

考虑噪声不只存在于观测信号 b 中，而且存在在 A 中，导出 TLS

$(A+N)x=b+e$

式中，N 是数据矩阵 A 中的噪声

总体最小二乘解

$(A+N)x_{TLS}=b+e$

TLS 试图估计噪声矩阵 N 和噪声向量 e，满足以上的方程组，x_{TLS} 是精确解。一般情况下，存在许多噪声矩阵 N 和噪声向量 e，满足以上的方程组。利用以下准则进行选择：

$\|[N,-e]\|_F^2=\sum_{i=1}^m e_i^2+\sum_{i=1}^m\sum_{j=1}^n n_{ij}^2$

总体最小二乘解

重写

$$(A+N)x_{TLS}=b+e$$

为

$$(B+D)z=0$$

式中：$B=[-b\,|\,A], D=[-e\,|\,N], z=\begin{bmatrix}1\\ x_{TLS}\end{bmatrix}$

需要估计 D，使 $B+D$ 非满秩。

利用奇异值分解：

$$B=\sum_{i=1}^{n+1}\sigma_i u_i v_i^T$$

其秩为 n 的近似矩阵为: $\tilde{B}=\sum_{i=1}^{n}\sigma_i u_i v_i^T$

则使 $\tilde{B}z=0$ 的解为：$z=\frac{v_{n+1}}{v_{n+1,1}}=\begin{bmatrix}1\\ x_{TLS}\end{bmatrix}$

数据最小二乘解（data least-squares,DLS）

$e=0$　时，TLS 被称为 DLS

一个简单的直线拟合例子

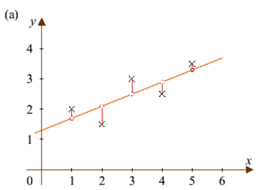
利用五个点：

$(x_k,y_k)=(1,2),(2,1.5),(3,3),(4,2.5),(5,3.5)$

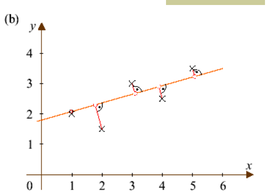
决定一条最优直线：$y=mx+c$

写成方程组：$Ax=b$

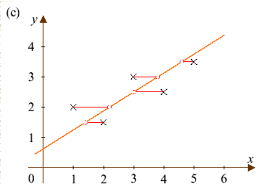
其中：$A=\begin{bmatrix}1&2&3&4&5\\1&1&1&1&1\end{bmatrix}^T$，$x=[m,c]^T$，$b=[2\quad 1.5\quad 3\quad 2.5\quad 3.5]^T$



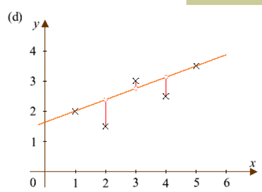
LS



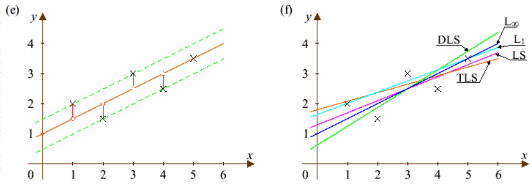
TLS



DLS



L_1



L_∞

资料来源和进一步阅读文献

- Profs. Juha Karhunen and Erkki Oja at the Helsinki University of Technology

  http://www.cis.hut.fi/projects/ica/book/
- Andrzej Cichocki, and Shun-ichi Amari, Adaptive blind signal and image processing: learning algorithms and applications, John Wiley & Sons, Ltd. 2002. 第二章
- 数学手册，高等教育出版社，1979，概率统计部分