主元分析与偏最小二乘法

王桂增 叶昊 编著

清华大学出版社

主元分析与偏最小二乘法

王桂增 叶昊 编著

清华大学出版社
北京

内容简介

主元分析与偏最小二乘法能较好地解决自变量之间存在的相关性问题，最大限度地概括自变量空间的数据变化信息与自变量对因变量的解释作用，因而被广泛用于解决科学计算、工业控制和信号处理中的特征提取、数据拟合、系统辨识和参数估计等问题。

本书从系统自变量的相关性、系统特性的非线性和时变性等实际问题出发，介绍线性与非线性主元分析方法、线性主元回归及其递推算法、线性与非线性偏最小二乘法及其递推算法、核主元分析与核偏最小二乘法等；最后还介绍了主元分析和偏最小二乘法在数据处理、软测量建模和过程监控等方面的应用案例。

本书可作为高等学校自动化类专业的高年级本科生和研究生的教学参考书，所述内容对从事统计数据处理、软测量建模与过程监控的科研人员和工程技术人员也具有参考价值。

图书在版编目(CIP)数据

主元分析与偏最小二乘法/王桂增，叶昊编著. —北京：清华大学出版社，2012.8
ISBN 978-7-302-27942-6

Ⅰ.①主… Ⅱ.①王… ②叶… Ⅲ.①主元分析 ②最小二乘法 Ⅳ.①TB114 ②O241.5

中国版本图书馆 CIP 数据核字(2012)第 011038 号

责任编辑：王一玲　王丽娜
封面设计：常雪影
责任校对：梁　毅
责任印制：宋　林

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　　邮　　编：100084
社 总 机：010-62770175　　邮　　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质 量 反 馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课 件 下 载：http://www.tup.com.cn，010-62795954

印 装 者：北京嘉实印刷有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：9.5　　**字　　数**：239 千字
版　　次：2012 年 8 月第 1 版　　**印　　次**：2012 年 8 月第1次印刷
印　　数：1～2000
定　　价：29.00 元

产品编号：032463-01

前言
foreword

最小二乘(least squares,LS)方法最早是由 Karl Gauss 为进行行星轨迹预测的研究而提出来的一种数据处理方法[1]。随着计算机的普遍使用,最小二乘算法被广泛用于解决科学计算、工业控制和信号处理中的特征提取、数据拟合、系统辨识和参数估计等问题,成为数据拟合、系统建模和过程监控的一个重要数学工具。

在实际应用中,用于参数估计的自变量之间往往存在一定程度的相关性(线性相关性也称为共线性),采用普通最小二乘方法估计参数将会出现病态解。1901 年,K. Pearson 首先提出了主元分析(principal component analysis,PCA)和主元回归(principal component regression,PCR)方法[2],其基本思想是先对存在相关性的自变量进行主元分析,提取自变量空间中相互正交的主元,然后建立因变量和主元之间的回归模型(称为主元回归模型),较好地解决了共线性的问题。

但在主元回归中,主元的选取以最大限度地概括自变量空间数据的变化信息为准则,并没有考虑主元对于因变量的解释作用,而未被选取的次要主元有可能包含对回归有益的信息,已被选取的主元也有可能包含对于回归无益的噪声。针对这一问题,H. Wold 于 20 世纪 60 年代提出了偏最小二乘 (partial least squares,PLS)方法[3],并应用于计量经济学。20 世纪 70 年代,H. Wold 的儿子 S. Wold 和挪威化学计量学家 H. Martens 将偏最小二乘法应用于化学计量学和化工领域。

PLS 方法综合考虑如何最大限度地概括自变量空间的数据变化信息和自变量对应变量的解释作用,将具有线性相关的数据从高维空间投影到低维特征空间,得到正交的特征向量,再建立特征向量间的线性回归关系。与主元回归相比,PLS 不仅有效地克服了普通最小二乘回归的共线性问题,同时将多元回归问题转化为若干个一元回归问题,适用于样本数较少而变量数较多的过程建模,具有较好的鲁棒性和预测稳定性,广泛地用于过程建模和过程监控领域。

实际系统往往具有非线性的特性,为适应非线性系统的建模和过程监控的需求,Hastie 等于 1989 年提出了基于主元曲线的非线性主元分析方法[32]。主元曲线是一条通过数据团的平滑曲线,它使所有数据点到主元曲线的距离之和最小,这些数据点在主元曲线上的投影则称为非线性主元。1991 年和 1995 年 Kramer 和 Tan 等分别提出了基于自相关神经元网络和基于输入训练神经元网络的非线性主元分析方法[59,60]。1989 年 S. Wold 等提出非线性 PLS 方法[9]。非线性 PLS 方法的基本思路有两种,一种方法是用回归模型中的原始变量的非线性项对输入矩阵进行扩展,然后对扩展的输入矩阵和因变量数据矩阵实施线性 PLS 回归;另一种方法是保留 PLS 方法的线性外部模型,而采用非线性内部模型来描述自变量与因变量数据的特征向量之间的非线性关系[33]。

Scholkopf B[29]和 Rosipal R[30]分别于 1998 年和 2003 年提出了核主元分析(kernel principal component analysis,KPCA)和核偏最小二乘 (kernel partial least squares,KPLS)方法,统称核分析方法[29,30]。核分析方法是一种新的非线性特征提取方法,它通过非线性映射将自变量映射到高维特征空间,并在高维特征空间中进行基于线性运算的特征提取。由于实际系统的非线性映射函数往往未知,因而无法得到自变量在高维空间中的映射及提取特征所需的协方差阵。核分析方法在映射空间构造映射数据的内积函数和以内积函数为元素的内积矩阵,并用原始自变量空间的核函数去代替内积函数,这样的内积矩阵称为核矩阵。在高维空间对核矩阵进行主元提取,称为核主元分析。而以最大化自变量映射与因变量的互相关为目标,对应变量数据矩阵和核矩阵进行综合分析,称为核偏最小二乘方法。

为使模型适应系统的时变特性,以及为了适应在线建模的需要,要求建模的算法能进行递推计算。Weihua Li 等提出了 PCA 的递推算法[48],Martens H.,Helland K. 和 Qin S. J. 提出了 PLS 递推算法[4,5,6]。文献[4]首先构建一个相对于原来的数据集合小得多的集合,利用 PLS 方法建立模型,然后用所得到的结果和新数据对 PLS 模型进行更新。文献[5]利用原模型中的自变量、因变量数据的负荷向量矩阵及回归因子矩阵,与新的数据块组合,构成在线训练的数据矩阵。这类处理方法以压缩的形式保留旧数据的信息,避免在学习新数据的过程中对旧数据重复建模,提高了 PLS 模型的在线更新速度。文献[6]在 Helland 的基础上提出了一种块式递推偏最小二乘算法。

鉴于标准 PLS 算法需要对原始自变量数据矩阵进行缩减,以求取相应的权值矩阵和得分矩阵,而权值矩阵很难从概念上将得分矩阵与原始自变量数据矩阵之间的关系描述清楚,S. de. Jong 提出了一种直接由原始自变量数据矩阵计算得分矩阵的算法[69]。

由于原始自变量数据矩阵中包含一些与因变量数据矩阵无关的信息,Trygg 等提出 O-PLS(orthogonal PLS)算法,该算法首先对原始自变量数据矩阵进行正交预处理,然后建立 PLS 模型[80]。

本书从系统自变量的相关性、系统特性的非线性和时变性等实际系统的特点出发,介绍线性与非线性主元分析方法、线性主元回归及其递推算法、线性与非线性偏最小二乘法及其递推算法、核主元分析与核偏最小二乘法等;最后还介绍了主元分析和偏最小二乘法在数据处理、软测量建模和过程监控等方面的应用案例。

本书可作为高等学校自动化类专业的高年级本科生和研究生的教学参考书,所述内容对从事统计数据处理、软测量建模与过程监控的科研人员和工程技术人员也具有参考价值。

本书的编写得到了清华大学自动化系阳宪惠教授的大力支持,书的第 7 章“非线性主元分析”和第 11 章“应用案例”中的 11.3.2 节均摘自文献[14]。清华大学自动化系已毕业的博士研究生李春富和刘毅,硕士研究生梁林和程龙等也对本书的编写作出过贡献。在此书成书之际,谨向他们表示衷心的感谢!

由于本书成书时间匆忙,加之作者水平有限,不足和错误在所难免,诚请读者提出批评、指正。

作 者

2011 年 2 月于清华园

主要符号说明

i,j——序数变量，如变量、权值向量、得分向量和负荷向量的序等

k——采样时刻、数据块的序数

$\bullet^{(k)}$——$\bullet$ 的第 k 次迭代

n,N——采样样本总数

m——自变量的总维数

r——因变量的总维数

h——神经元网络中隐藏层序数

f——高维映射空间的总维数

s——矩阵的秩

$\boldsymbol{X}^o\in\mathbf{R}^{n\times m}$或$\hat{\boldsymbol{X}}\in\mathbf{R}^{n\times m}$——原始数据矩阵

$\boldsymbol{X}\in\mathbf{R}^{n\times m}$——自变量(输入变量)数据矩阵，预测变量数据矩阵

$\boldsymbol{X}(k)\in\mathbf{R}^{k\times m}$(或$\boldsymbol{X}_k\in\mathbf{R}^{k\times m}$)——包含 k 个样本的自变量数据矩阵

$\boldsymbol{E}_i\in\mathbf{R}^{n\times m}$——自变量数据矩阵 $\boldsymbol{X}$ 的第 i 阶缩减矩阵

$\boldsymbol{x}\in\mathbf{R}^{m\times 1}$——自变量向量，$\boldsymbol{x}=[x_1,x_2,\cdots,x_i,\cdots,x_m]^{\mathrm{T}}$

x_i——第 i 个自变量

$\boldsymbol{x}_i\in\mathbf{R}^{n\times 1}$——第 i 个自变量 x_i 的样本向量，$\boldsymbol{x}_i=[x_{i,1},x_{i,2},\cdots,x_{i,k},\cdots,x_{i,n}]^{\mathrm{T}}$

$\boldsymbol{x}(k)\in\mathbf{R}^{m\times 1}$——自变量向量 $\boldsymbol{x}$ 的第 k 个样本，$\boldsymbol{x}(k)=[x_{1,k},x_{2,k},\cdots,x_{i,k},\cdots,x_{m,k}]^{\mathrm{T}}$

$\boldsymbol{Y}\in\mathbf{R}^{n\times r}$——因变量(输出变量)数据矩阵

$\boldsymbol{Y}(k)\in\mathbf{R}^{k\times r}$(或$\boldsymbol{Y}_k\in\mathbf{R}^{k\times r}$)——包含 k 个样本的因变量数据矩阵

$\boldsymbol{F}_i\in\mathbf{R}^{n\times r}$——因变量数据矩阵 $\boldsymbol{Y}$ 的第 i 阶缩减矩阵

$\boldsymbol{y}\in\mathbf{R}^{r\times 1}$——因变量向量，$\boldsymbol{y}=[y_1,y_2,\cdots,y_i,\cdots,y_r]^{\mathrm{T}}$

y_i——第 i 个因变量

$\boldsymbol{y}_i\in\mathbf{R}^{n\times 1}$——第 i 个因变量 y_i 的样本向量，$\boldsymbol{y}_i=[y_{i,1},y_{i,2},\cdots,y_{i,3},\cdots,y_{i,n}]^{\mathrm{T}}$

$\boldsymbol{y}(k)\in\mathbf{R}^{r\times 1}$——因变量向量 $\boldsymbol{y}$ 的第 k 个样本，$\boldsymbol{y}(k)=[y_{1,k},y_{2,k},\cdots,y_{i,k},\cdots,y_{r,k}]^{\mathrm{T}}$

$\boldsymbol{\lambda}$——矩阵的特征值

$\boldsymbol{t}\in\mathbf{R}^{m\times 1}$——$\boldsymbol{X}$ 的主元向量，得分向量

$\boldsymbol{t}=[t_1,t_2,\cdots,t_i,\cdots,t_m]^{\mathrm{T}}$

t_i——第 i 个主元

$\boldsymbol{t}_i\in\mathbf{R}^{n\times 1}$——$\boldsymbol{t}$ 的第 i 个主元(主成分)的样本向量，$\boldsymbol{t}_i=[t_{i,1},t_{i,2},\cdots,t_{i,n}]^{\mathrm{T}}$

$\boldsymbol{t}(k)\in\mathbf{R}^{m\times 1}$——主元向量(主成分向量)$\boldsymbol{t}$ 的第 k 个观测样本，$\boldsymbol{t}(k)=[t_{1,k},t_{2,k},\cdots,t_{m,k}]^{\mathrm{T}}$

$\boldsymbol{t}_i^{\cdot(k)}\in\mathbf{R}^{n\times 1}$——第 i 个主元样本向量 $\boldsymbol{t}_i$ 的第 k 次迭代值，$\boldsymbol{t}_i^{\cdot(k)}=[t_{i,1}^{\cdot(k)},t_{i,2}^{\cdot(k)},\cdots,t_{i,n}^{\cdot(k)}]^{\mathrm{T}}$

$t_{i,k}$——第 i 个主元在 k 时刻的样本值

$\boldsymbol{w}\in\mathbf{R}^{m\times 1}$——$\boldsymbol{X}$ 的权值向量

$\boldsymbol{p}\in\mathbf{R}^{m\times 1}$——$t$ 在 $\boldsymbol{X}$ 上的负荷(负载)向量

a——所取 $\boldsymbol{X}$ 的主元(主成分)的总数,即所取得分向量的维数

$\boldsymbol{u}\in\mathbf{R}^{r\times 1}$——$\boldsymbol{Y}$ 的主元向量

$\boldsymbol{c}\in\mathbf{R}^{r\times 1}$——$\boldsymbol{Y}$ 的权值向量

$\boldsymbol{q}\in\mathbf{R}^{r\times 1}$——$\boldsymbol{u}$ 在 $\boldsymbol{Y}$ 上的负荷向量

$\boldsymbol{b}\in\mathbf{R}^{r\times 1}$——$\boldsymbol{Y}$ 对 t 的回归系数向量

$\mathrm{R}_x\in\mathbf{R}^{m\times m}$——$\boldsymbol{x}$ 的自相关函数

$\boldsymbol{\psi}_x\in\mathbf{R}^{m\times m}$——$\boldsymbol{x}$ 的自协方差矩阵

$\boldsymbol{\Theta}_x\in\mathbf{R}^{m\times m}$——$\boldsymbol{x}$ 的方差阵,$\boldsymbol{\Theta}_x=\mathrm{diag}[\sigma_{x_1}^2,\sigma_{x_2}^2,\cdots,\sigma_{x_m}^2]$

$\mathrm{tr}\boldsymbol{\Theta}_x$——$\boldsymbol{\Theta}_x$ 的迹,标量,$\mathrm{tr}\boldsymbol{\Theta}_x=\sum_{i=1}^{m}\sigma_{x_i}^2=\frac{1}{n}\sum_{k=1}^{n}\boldsymbol{x}^{\mathrm{T}}(k)\boldsymbol{x}(k)$

$\boldsymbol{\psi}_{xy}\in\mathbf{R}^{m\times r}$——$\boldsymbol{x},\boldsymbol{y}$ 的互协方差矩阵

$\mathrm{R}_{xy}\in\mathbf{R}^{m\times r}$——$\boldsymbol{x},\boldsymbol{y}$ 的互相关函数

$\boldsymbol{X}^{\mathrm{T}}$——矩阵 $\boldsymbol{X}$ 的转置

$|\cdot|$——矩阵的行列式,或标量的绝对值

$\|\boldsymbol{x}\|$——向量 $\boldsymbol{x}$ 的模(长度)

$\cdot^{\Phi}$——Φ 为“$\cdot$”所在非线性映射空间的标识

$\Phi(\cdot)$——非线性映射函数

$\boldsymbol{\Phi}(\boldsymbol{X})$——$\boldsymbol{X}$ 的非线性映射数据矩阵

$\boldsymbol{\varphi}(\boldsymbol{x})$——$\boldsymbol{x}$ 的非线性映射向量

目录

contents

chapter 1 第1章

随机过程的基本知识

自然界的事物变化万千，但从其变化过程可以分为两大类：确定性过程与随机过程。本书所涉及的信号处理方法均与随机信号有关，为此，本章简要地介绍一些随机过程的基本概念和随机信号的数字特征。

1.1 基本概念

1.1.1 事物变化过程的分类

事物变化过程可以分为两大类。

一类是确定性过程(deterministic process)，这类过程的变化具有确定的形式，或者说具有必然的规律。从数学的角度来讲，其变化过程可以用时间上的确定函数来描述。这类过程称为确定性过程。如物体以初速度 0 开始自由落体，则其离初始点的距离为 $d(t)=\frac{1}{2}gt^2$，g 为重力加速度。这个关系是确定的，在条件相同的情况下，在任何时间、任何地方做这样的实验，结果都是一样的。

另一类是随机过程(stochastic process)，这类过程没有确定的变化规律，即过程的变化不能用时间 t 的确定函数来描述。对同一事物的变化过程独立进行的数次观测所得到的时间函数各不相同。如同一台电子放大器的热噪声电压，在不同时刻所观测到的值是不同的；多台同样的电子放大器在同样的条件下，同一时刻所观测到的热噪声电压也是不同的，而且相互之间没有“确定”的关系，是一个随机变量 $X(t)$。这里说没有“确定的关系”，不是说所观测到的热噪声电压值之间“没有关系”。

对上述不确定过程在不同时刻进行观测所得到的一簇无穷多个、相互“有关”的随机变量 $X(t_1),X(t_2),\cdots,X(t_n)$，记为 $\{X(t),t\in T\}$，构成随机过程。

这里所说的“有关”强调的是对同一过程所进行的观测所得到的随机变量的集合才构成随机过程，而互不相干的随机变量的集合是不能构成随机过程的。

上述时间集合 T 有 3 种情况：

(1) $T=\{0,1,2,\cdots,\infty\}$

(2) $T=\{\cdots,-2,-1,0,1,2,\cdots\}$

(3) $T=\{a,b\}$，a 和 b 可以是$\pm\infty$区间上的任意值(不一定是整数)。当 T 为(1)和(2)的情况时，上述一簇随机变量构成的随机过程称为随机序列(random sequence)，或称离散时间(参数)随机过程。$X(t)$中的 t 一般表示时间，广义地讲，也可以表示一些随机事件。

1.1.2 随机过程的样本与状态

1. 随机过程的样本函数

随机过程的任一个时间函数 $x_1(t),x_2(t),\cdots,x_n(t)$称为随机过程的样本函数(sample function)，它们是随机过程的不同实现(realization)。图 1-1 所示为随机过程的 m 个样本函数。

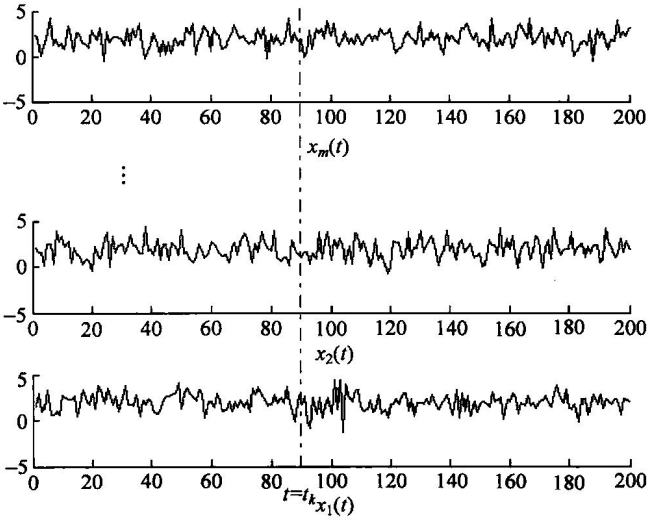

图 1-1 随机过程的样本函数与状态

2. 随机过程的状态

随机过程$\{X(t),t\in T\}$在 $t=t_k$ 时刻的取值 $X(t_k)$称为随机过程的一个状态，$X(t_k)$也是一个随机变量。

1.1.3 集合(总体)平均

随机过程在特定时刻的状态 $X(t_i)$是一个随机变量，其取值为 $x_j(t_i),j=1,2,\cdots,m$(如图 1-1 所示)，则称

$$\mu_X(t_i)=\mathrm{E}[X(t_i)]=\lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}x_j(t_i)$$

为随机过程在 t_i 时刻的集合平均均值(ensemble average)，式中 $\mathrm{E}[X(t_i)]$为状态的数学期望。一般来说，集合平均均值 $\mu_X(t_i)$是时间的函数，但对平稳随机过程来说，集合平均均值不随时间变化，故可用 μ_X 表示。

实际中，集合平均均值不易得到，因为理论上它要求有 m 个($m\to\infty$)信号源。

1.1.4 时间平均

对随机过程的一个样本函数 $x_j(t)$，$t=1,2,\cdots,n$(如图 1-1 所示)，称

$$\bar{\mu}_X(j) = \mathrm{E}[x_j(t)] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_j(t_i)$$

为随机过程的第 j 个样本函数的时间平均均值(time average)。一般来说，样本函数的时间平均均值是样本 j 的函数，而不是时间的函数。

1.2 随机过程的数字特征

在数学上，随机过程一般用概率密度函数和概率分布函数来描述，但在实际中要确定概率密度函数和概率分布函数并加以分析是很困难的，甚至是不可能的。而随机过程的数字特征不仅能刻画随机过程的主要特征，又便于在实际中进行测量和运算，在各个领域得到广泛的应用。

1.2.1 数学期望

数学期望(expectation)是一个统计概念，理论上要求样本函数中的 $T\to\infty$。而在实际中，由于 T 是有限值，一般称为均值。

设随机过程在特定时刻的状态为 $X(t_i)$，$X(t_i)$是一个随机变量。

若 $X(t_i)$是连续型随机变量，则其集合平均均值为

$$\begin{aligned}\mu_X(t_i) &= \mathrm{E}[X(t_i)]\\ &= \int_{-\infty}^{\infty}xf(x,t_i)\mathrm{d}x\end{aligned}\tag{1-1}$$

式中，$f(x,t_i)$为 $X(t_i)$的概率密度函数。

若 $X(t_i)$为离散型随机变量，其集合平均均值为

$$\mu_X(t_i) = \lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}x_j(t_i)\tag{1-2}$$

m 为 $X(t_i)$的 m 个取值。

对连续时间参数随机过程，样本函数的时间平均均值可用式(1-3)求得，即

$$\bar{\mu}_X = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)\mathrm{d}t\tag{1-3}$$

对离散时间参数随机过程，样本序列的时间平均均值可表示为

$$\bar{\mu}_X = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x(i)\tag{1-4}$$

式中，i 为离散时间参数。

1.2.2 方差

设随机过程在特定时刻的状态为 $X(t_i)$，$X(t_i)$是一个随机变量。若 $X(t_i)$是连续型随

机变量，则其集合平均方差(variance)为

$$\sigma^2(t_i)=\mathrm{E}\{[x(t_i)-\mu_X(t_i)]^2\}=\int_{-\infty}^{\infty}[x(t_i)-\mu_X(t_i)]^2 f(x,t_i)\mathrm{d}x \tag{1-5}$$

式中，$x(t_i)$为 $X(t_i)$的取值，是一个连续函数，$f(x,t_i)$为 $X(t_i)$的概率密度函数。

对离散型随机变量 $X(t_i)$，其集合平均方差为

$$\sigma_X^2(t_i)=\lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}[x_j(t_i)-\mu_X(t_i)]^2 \tag{1-6}$$

式中，$x_j(t_i)$为 $X(t_i)$的第 j 个离散取值。

对连续时间参数随机过程，任一样本函数的方差可用式(1-7)求得，即

$$\bar{\sigma}_X^2=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[x(t)-\bar{\mu}_X]^2\mathrm{d}t \tag{1-7}$$

对离散时间参数随机过程，任一样本序列的均值可表示为

$$\bar{\sigma}_X^2=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}[x(i)-\bar{\mu}_X]^2 \tag{1-8}$$

1.2.3 相关函数

1. 自相关函数

设随机过程在两个特定时刻的状态为 $X(t_i)$和 $X(t_i+\tau)$，$X(t_i)$和 $X(t_i+\tau)$均为随机变量。为了描述同一随机过程的两个不同状态之间的统计联系，引入自相关函数(auto-correlation function)的概念。

对连续型随机变量 $X(t_i)$和 $X(t_i+\tau)$，其集合平均自相关函数为

$$R_{XX}(t_i,t_i+\tau)=\mathrm{E}[x(t_i)x(t_i+\tau)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_{t_i}x_{t_i+\tau}f(x_{t_i},x_{t_i+\tau})\mathrm{d}x_{t_i}\mathrm{d}x_{t_i+\tau} \tag{1-9}$$

式中，$f(x_{t_i},x_{t_i+\tau})$为 $X(t_i)$和 $X(t_i+\tau)$的二维联合概率密度函数。

对离散型随机变量 $X(t_i)$和 $X(t_i+\tau)$，其集合平均自相关函数为

$$R_{XX}(\tau)=E[x_{j,t_i}x_{j,(t_i+\tau)}]=\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}x_{j,t_i}x_{j,(t_i+\tau)} \tag{1-10}$$

对连续时间参数随机过程，任一样本函数的时间平均自相关函数定义为

$$\bar{R}_{XX}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)x(t+\tau)\mathrm{d}t \tag{1-11}$$

对离散时间参数随机过程，任一样本序列的时间平均自相关函数可用式(1-12)求得，即

$$\bar{R}_{XX}(i)=\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}x_jx_{j+i} \tag{1-12}$$

时间平均自相关函数 $\bar{R}_{XX}(\tau)$和 $\bar{R}_{XX}(i)$又可简写为 $\bar{R}_X(\tau)$和 $\bar{R}_X(i)$。

对平稳随机过程来说，自相关函数具有下列性质：

(1) $R_X(0)>0$

(2) 自相关函数是 τ 的偶函数，即

$$R_X(\tau)=R_X(-\tau)$$

或

$$R_X(i)=R_X(-i)$$

(3) $R_X(0)$在自相关函数中取最大值，即

$$|R_X(\tau)|\leqslant R_X(0)$$

或

$$|R_X(i)|\leqslant R_X(0)$$

这意味着，$R_X(\tau)$可取负值。

(4) 如果随机过程的样本函数满足 $x(t)=x(t+T)$，T 为周期，则自相关函数也是周期函数，且周期也为 T。

自相关函数反映给定的随机过程的变化快慢，或者说其样本函数的变化快慢。如样本函数的变化缓慢，则自相关函数也变化缓慢，反之亦然。

2. 互相关函数

为了描述不同随机过程之间的统计联系，引入互相关函数(cross-correlation function)的概念。

设有两个平稳随机过程$\{X(t)\}$和$\{Y(t)\}$，它们在两个特定时刻的状态分别为 $X(t_i)$和 $Y(t_i+\tau)$，$X(t_i)$和 $Y(t_i+\tau)$均为随机变量。

对连续型随机变量 $X(t_i)$和 $Y(t_i+\tau)$，定义集合平均互相关函数

$$\begin{aligned}R_{XY}(\tau)&=\mathrm{E}[X(t_i)Y(t_i+\tau)]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,t_i;\ y,t_i+\tau)\mathrm{d}x\mathrm{d}y\end{aligned}\tag{1-13}$$

式中，$f(x,t_i;\ y,t_i+\tau)$为 $X(t_i)$和 $Y(t_i+\tau)$的联合概率密度函数。

对离散型随机变量 $X(t_i)$和 $Y(t_i+\tau)$，其集合平均互相关函数为

$$R_{XY}(\tau)=\lim_{m\to\infty}\frac{1}{m}\sum_{j=1}^{m}X_{j,t_i}Y_{j,(t_i+\tau)}\tag{1-14}$$

式中，X_{j,t_i}为 $X(t_i)$的第 j 个取值，而 $Y_{j,(t_i+\tau)}$为 $Y(t_i+\tau)$的第 j 个取值。

设两个平稳随机过程$\{X(t)\}$和$\{Y(t)\}$的任一样本函数分别为 $x(t)$和 $y(t)$，则定义其时间平均互相关函数为 $\bar{R}_{XY}(\tau)$或 $\bar{R}_{XY}(i)$。

对连续时间参数随机过程

$$\bar{R}_{XY}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)y(t+\tau)\mathrm{d}t\tag{1-15}$$

对离散时间参数随机过程(随机序列)

$$\bar{R}_{XY}(\tau)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_iy_{i+\tau}\tag{1-16}$$

对平稳随机过程，互相关函数具有下列性质：

(1) $R_{XY}(0)$并不一定取得最大值；

(2) $R_{XY}(\tau)$既不是偶函数也不是奇函数，而有 $R_{XY}(\tau)=R_{YX}(-\tau)$，因为

$$R_{XY}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x(t)y(t+\tau)\mathrm{d}t$$

$$= \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} y(t+\tau)x(t)\mathrm{d}t$$
$$= R_{YX}(-\tau)$$

(3) 如果两个随机过程$\{X(t)\}$和$\{Y(t)\}$相互统计独立，且一个均值为0，则其互相关函数为0。

证：因$\{X(t)\}$和$\{Y(t)\}$相互统计独立，则有

$$R_{XY}(\tau) = \mathrm{E}[x(t)y(t+\tau)]$$
$$= \mathrm{E}[x(t)]\mathrm{E}[y(t+\tau)]$$
$$= 0$$

注意逆定理不一定成立，即如果两个随机过程$\{X(t)\}$和$\{Y(t)\}$的互相关函数为0，$\{X(t)\}$与$\{Y(t)\}$不一定相互统计独立。

1.2.4 功率谱密度函数

1. 自相关谱密度函数(自谱密度函数)

我们用电路中功率的例子来引出功率谱密度(power density spectrum)的概念。如果将一个随时间变化的电压函数$x(t)$加于1Ω的电阻上，则在电阻上的瞬时功率损耗函数为$x^2(t)$，其平均功率损耗函数为

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T x^2(t)\mathrm{d}t$$

若此平均功率函数为非周期函数，则通过傅里叶展开可得到其连续的平均功率谱密度。

假如$x(t)$是随机过程的一个样本函数，其自相关函数

$$R_X(0) = \lim_{T\to\infty}\frac{1}{T}\int_0^T x^2(t)\mathrm{d}t$$

在数值上，$x(t)$的自相关函数$R_X(0)$恰与$x(t)$的平均功率相等。这就将一个函数的平均功率与此函数的自相关函数建立起联系。因而，对一个随机过程的样本函数$x(t)$的自相关函数$R_X(0)$进行傅里叶展开得到该函数的自功率谱密度就不难理解了。将此概念推广到$\tau\neq 0$的情况就得到功率谱密度函数的一般定义：对一个随机过程的自相关函数进行傅里叶变换就得到该随机过程的自相关功率谱密度函数，用$S_X(\omega)$表示，它是自相关函数在频域的表示式，ω为角频率。

$$S_X(\omega) = \mathrm{F}[R_X(\tau)]$$
$$= \int_{-\infty}^{\infty} R_X(\tau)\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau$$
$$= \int_{-\infty}^{\infty} R_X(\tau)[\cos\omega\tau - \mathrm{j}\sin\omega\tau]\mathrm{d}\tau \qquad (1\text{-}17)$$

自相关谱密度函数的性质：由于$R_X(\tau)$和$\cos\omega\tau$是τ的偶函数，而$\sin\omega\tau$是τ的奇函数，因而自相关谱密度函数又可以写成

$$S_X(\omega) = 2\int_0^{\infty} R_X(\tau)\cos\omega\tau\mathrm{d}\tau$$

自相关谱密度函数是ω的正实偶函数。