

高等学校教材

卡尔曼滤波 与组合导航原理

秦永元 张洪钺 汪叔华 编著



西北工业大学出版社

TN713

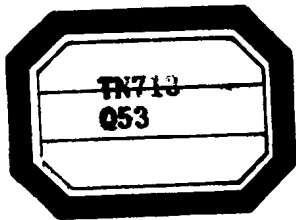
447292

Q53

高等学校教材

卡尔曼滤波与组合导航原理

秦永元 张洪钺 汪叔华 编著



西北工业大学出版社

1998年11月 西安

(陕)新登字 009 号

【内容摘要】 本书着重阐述了卡尔曼滤波基本理论,近十年发展起来的有关卡尔曼滤波的新理论和新方法,容错组合导航设计理论和方法,以及作者的部分科研成果。内容安排上力求循序渐进,由浅入深,确保知识连贯。为便于读者理解概念内涵,公式和定理一般都附有详细推导和证明。

本书是航空工业总公司确定的控制、导航专业研究生通用教科书,也可作为相关专业高年级学生及研究工作者的参考书。

13月17日

高等学校教材
卡尔曼滤波与组合导航原理
秦永元 张洪钺 汪叔华 编著
责任编辑 郑家梅
责任校对 齐随印

©1998 西北工业大学出版社出版发行
(邮编:710072 西安市友谊西路127号 电话:8493844)

全国各地新华书店经销
陕西省富平印刷有限责任公司印装
ISBN 7-5612-1034-5/TN·63(课)

开本: 787毫米×1092毫米 1/16 印张: 21.875 字数: 537千字
1998年11月第1版 1998年11月第1次印刷
印数: 1-2000册 定价: 23.50元

购买本社出版的图书,如有缺页、错页的,本社发行部负责调换。

前 言

卡尔曼滤波是对随机信号作估计的算法之一。与最小二乘、维纳滤波等诸多估计算法相比,卡尔曼滤波具有显著的优点:采用状态空间法在时域内设计滤波器,用状态方程描述任何复杂多维信号的动力学特性,避开了在频域内对信号功率谱作分解带来的麻烦,滤波器设计简单易行;采用递推算法,实时量测信息经提炼被浓缩在估计值中,而不必存储时间过程中的量测量。所以,卡尔曼滤波能适用于白噪声激励的任何平稳或非平稳随机向量过程的估计,所得估计在线性估计中精度最佳。正由于其独特的优点,卡尔曼滤波在60年代初一经提出,立即受到工程界,特别是空间技术和航空界的高度重视,阿波罗登月计划中的导航系统设计和C-5A飞机的多模式导航系统的设计是卡尔曼滤波早期应用中最为成功的实例。随着计算机技术的发展,目前卡尔曼滤波的应用几乎涉及通讯、导航、遥感、地震测量、石油勘探、经济和社会学研究等所有领域。

本书是航空工业总公司组织编写的研究生用统编教材。全书着重阐述卡尔曼滤波基本理论,近十年发展起来的有关卡尔曼滤波的新理论和新方法。此外,还专门介绍了与组合导航系统设计理论和方法有关的内容以及作者的部分研究成果。既使读者能全面、系统地了解卡尔曼滤波理论,又为组合导航系统设计提供完整的设计理论和方法,也为卡尔曼滤波理论在其他领域中的应用提供研究方法上的参考和借鉴。

本书共分八章。第一至第七章系统介绍卡尔曼滤波的基本理论,包括:滤波基本方程,滤波中的技术处理,滤波稳定性分析,广义滤波理论,滤波器的容错设计理论等;第八章介绍滤波理论在组合导航系统设计中的应用,包括:各种常用导航子系统的介绍,卡尔曼滤波在惯性导航系统对准和容错组合导航系统设计中的应用。在内容阐述上注意到知识的连贯性,力求做到循序渐进,由浅入深。为了使读者能从本质上理解概念的内涵,公式和定理的提出一般都经过详细的推导和证明。

本书是在西北工业大学俞济祥教授主编的《卡尔曼滤波及其在惯性导航中的应用》^[1]一书的基础上撰写的,由秦永元主编。各章编写分工:第一至第五章、第八章由西北工业大学秦永元编写,第六章由南京航空航天大学汪叔华编写,第七章由北京航空航天大学张洪钺编写。本书承蒙空军工程学院张宗麟教授详细审阅提出了许多宝贵意见,在此谨致深切谢意。

编著者

1998年2月

目 录

第一章 绪论	1
1.1 卡尔曼滤波所要解决的问题	1
1.2 卡尔曼滤波理论的发展和工程应用	2
1.3 组合导航简介	3
第二章 几种最优估计和卡尔曼滤波基本方程	5
2.1 几种最优估计	5
2.2 离散型卡尔曼滤波	33
2.3 连续型卡尔曼滤波	56
习题	66
第三章 卡尔曼滤波中的技术处理	69
3.1 有色噪声的白化	69
3.2 有色噪声条件下的卡尔曼滤波	77
3.3 序贯处理	82
3.4 信息滤波处理	84
3.5 卡尔曼滤波发散的抑制	86
3.6 平方根滤波	92
3.7 UDU^T 分解滤波	102
3.8 自适应滤波	117
3.9 次优滤波	120
3.10 卡尔曼滤波误差分析	128
习题	138
第四章 卡尔曼滤波稳定性介绍	141
4.1 稳定性定义	141
4.2 判别卡尔曼滤波稳定的充分条件	142
4.3 适用于惯导系统的滤波稳定判别条件	152
习题	154
第五章 滤波系统的校正	155
5.1 概述	155

5.2	离散系统的分离定理	156
5.3	连续系统的分离定理	160
5.4	离散系统的估计直接反馈控制	161
5.5	连续系统的估计直接反馈控制	165
	习题	165
第六章	卡尔曼滤波的推广	167
6.1	最优线性平滑	167
6.2	非线性系统的卡尔曼滤波	175
	习题	188
第七章	容错组合导航的设计理论	190
7.1	概述	190
7.2	联邦滤波器算法原理	191
7.3	系统级故障检测与隔离的原理与方法	205
7.4	惯性器件的故障检测与隔离原理	212
7.5	故障的统计检测原理与风险分析	227
第八章	卡尔曼滤波理论在组合导航系统设计中的应用	238
8.1	概述	238
8.2	组合导航系统的设计模式	239
8.3	组合导航系统设计中一些常用导航子系统的误差模型	243
8.4	惯性导航系统的误差模型	267
8.5	卡尔曼滤波理论在惯导系统初始对准中的应用	286
8.6	应用基本滤波理论设计组合导航系统	311
8.7	应用联邦滤波理论设计容错组合导航系统	326
	附录	337
	参考文献	340

第一章 绪 论

1.1 卡尔曼滤波所要解决的问题

所谓滤波就是从混合在一起的诸多信号中提取出所需要的信号。

信号是传递和运载信息的时间或空间函数。有一类信号的变化规律是既定的,如调幅广播中的载波信号、阶跃信号、脉宽固定的矩形脉冲信号等,它们都具有确定的频谱,这类信号称为确定性信号。另一类信号没有既定的变化规律,在相同的初始条件和环境条件下,信号的每次实现都不一样,如陀螺漂移、海浪、作水平飞行的飞机飞越山区时无线电高度表的输出信号、惯导系统的导航输出误差、GPS的SA误差等,它们没有确定的频谱,这类信号称为随机信号。

由于确定性信号具有确定的频谱,所以可根据各信号频带的不同,设置具有相应频率特性的滤波器,如低通、高通、带通、带阻滤波器,使有用信号无衰减地通过,使干扰信号受到抑制。这类滤波器可用物理方法实现,此即模拟滤波器,也可用计算机通过算法实现,此即数字滤波器。对确定性信号的滤波处理也称常规滤波。

随机信号没有确定的频谱,无法用常规滤波提取或抑制信号,但随机信号具有确定的功率谱,所以可根据有用信号和干扰信号的功率谱设计滤波器。维纳滤波是解决此类问题的方法之一。但设计维纳滤波器须作功率谱分解,只有当被处理信号为平稳的,干扰信号和有用信号均为一维,且功率谱为有理分式时,维纳滤波器的传递函数才可用伯特-香农设计法较容易地求解出。否则设计维纳滤波器存在着诸多困难。维纳滤波除设计思想与常规滤波不同外,对信号作抑制和选通这一点是相似的。

卡尔曼滤波从与被提取信号有关的量测量中通过算法估计出所需信号。其中被估计信号是由白噪声激励引起的随机响应,激励源与响应之间的传递结构(系统方程)已知,量测量与被估测量之间的函数关系(量测方程)也已知。估计过程中利用了如下信息:系统方程、量测方程、白噪声激励的统计特性、量测误差的统计特性。由于所用信息都是时域内的量,所以卡尔曼滤波器是在时域内设计的,且适用于多维情况,这就完全避免了维纳滤波器在频域内设计遇到的限制和障碍,适用范围远比维纳滤波器广。

从以上简述中可看出卡尔曼滤波有如下特点:

- (1) 卡尔曼滤波处理的对象是随机信号;
- (2) 被处理信号无有用和干扰之分,滤波的目的是要估计出所有被处理信号;
- (3) 系统的白噪声激励和量测噪声并不是需要滤除的对象,它们的统计特性正是估计过程中需要利用的信息。

所以确切地说,卡尔曼滤波应称作最优估计理论,此处所谓的滤波与常规滤波具有完全不同的概念和含意。

就实现形式而言,卡尔曼滤波器实质上是一套由数字计算机实现的递推算法,每个递推周

期中包含对被估计量的时间更新和量测更新两个过程。时间更新由上一步的量测更新结果和设计卡尔曼滤波器时的先验信息确定,量测更新则在时间更新的基础上根据实时获得的量测值确定。因此,量测量可看做卡尔曼滤波器的输入,估计值可看做输出。输入与输出之间由时间更新和量测更新算法联系,这与数字信号处理概念是类似的,所以有些书上称卡尔曼滤波为广义数字信号处理似有一定道理。

1.2 卡尔曼滤波理论的发展和工程应用

随机信号没有既定的变化规律,对它们的估计也不可能完全准确,所谓最优估计也仅仅是指在某一准则下的最优。根据不同的最优准则,可获得随机信号的不同最优估计。使贝叶斯风险达到最小的估计为贝叶斯估计;使关于条件概率密度的似然函数达到极大的估计为极大似然估计;使验后概率密度达到极大的估计为极大验后估计;使估计误差的均方误差达到最小的估计为最小方差估计,若估计具有线性形式,则估计为线性最小方差估计,卡尔曼滤波即属此类估计。

卡尔曼滤波理论的创立是科学技术和社会需要发展到一定程度的必然结果。早在 1795 年,高斯(Karl Gauss)为测定行星运动轨道而提出了最小二乘估计法。本世纪 40 年代,为了解决火力控制系统精确跟踪问题,维纳(N. Wiener)于 1942 年提出了维纳滤波理论^[2]。维纳根据有用信号和干扰信号的功率谱确定出线性滤波器的频率特性,首次将数理统计理论与线性系统理论有机地联系在一起,形成了对随机信号作平滑、估计或预测的最优估计新理论。比维纳稍早,前苏联科学家戈尔莫克洛夫(A. N. Kolmogorov)于 1941 年也曾提出过类似理论^[3]。维纳给出了由功率谱求解维纳滤波器频率特性闭合解的一般方法,包括对功率谱的上、下平面分解及傅里叶变换和反变换,运算繁杂,解析求解十分困难。1950 年,伯特和香农给出了在功率谱为有理谱这一特殊条件下,由功率谱直接求取维纳滤波器传递函数的设计方法^[4],这一方法简单易行,具有一定的工程实用价值。维纳滤波的最大缺点是适用范围极其有限,它要求被处理信号必须是平稳的,且是一维的。人们试图将维纳滤波推广到非平稳和多维的情况^[5,6],都因无法突破计算上的困难而难以推广和应用。

采用频域设计法是造成维纳滤波器设计困难的根本原因。因此人们逐渐转向寻求在时域内直接设计最优滤波器的新方法^[7-12],其中卡尔曼的研究最具有代表性,他提出的递推最优估计理论也因此而被称为卡尔曼滤波。由于采用了状态空间法描述系统,算法采用递推形式,所以卡尔曼滤波能处理多维和非平稳的随机过程。

卡尔曼滤波理论一经提出,立即受到了工程界的重视,而工程应用中遇到的实际问题又使卡尔曼滤波的研究更深入更完善。1959 年起美国太空署即 NASA(National Aeronautics and Space Administration)开始研究载人太空船登月方案,当时提出了两个主要问题:① 中途导航和制导;② 液体燃料助推器大挠度条件下的自动驾驶问题。因这两项研究的工作量都很庞大,无力同时进行,所以选择前者作为重点,即宇宙飞船的测轨问题。导航问题中主要解决对太空船运动状态的估计。量测信息来自三个子系统:飞船装备的惯性测量装置和天文观测仪,地面测轨系统,测轨数据经数据链传送至太空船。估计方法曾试图采用递推加权最小二乘和维纳滤波,均因精度满足不了要求和计算过于繁杂而不得不放弃。1960 年秋,卡尔曼访问了 NASA,提出了卡尔曼滤波算法,立即引起重视并投入研究。由于最初提出的卡尔曼滤波仅适

用于线性系统,而实际系统是非线性系统,滤波初值应如何取才合理,这些都迫使卡尔曼作进一步的思考,广义卡尔曼滤波就是在此情况下提出来的,阿波罗计划中的导航系统后由麻省理工学院研制完成^[13]。卡尔曼滤波早期应用中的另一成功实例为 C-5A 飞机的多模式导航系统^[14]。

卡尔曼滤波比维纳滤波的应用范围广,设计方法也简单易行得多,但它必须在计算机上执行,而 60 年代初,无论是速度、字长、容量,计算机还处于低水平阶段。为了适应当时的技术水平,避免由于字长不够产生的舍入误差引起卡尔曼滤波的计算发散,Bierman,Carlson 和 Schmidt 等人提出了平方根滤波算法和 UDU^T 分解滤波算法,以确保卡尔曼滤波增益回路中的滤波方差阵始终正定^[15~23]。

卡尔曼最初提出的滤波基本理论只适用于线性系统,并且要求量测也必须是线性的。在之后的 10 多年时间内,Bucy,Sunahara 等人致力于研究卡尔曼滤波理论在非线性系统和非线性量测情况下的推广^[24~33],拓宽了卡尔曼滤波理论的适用范围。

卡尔曼滤波最成功的工程应用是设计运载体的高精度组合导航系统。80 年代起,可供运载体装备的导航系统越来越多,非相似导航子系统的增加使量测信息增多,这对提高组合导航系统的精度十分有利。但是,如果采用集中式卡尔曼滤波器实现组合,则存在两个致命问题:① 滤波器计算量以状态维数的三次方剧增,无法满足导航的实时性要求;② 导航子系统的增加使故障率也随之增加,只要有一个子系统发生故障又没有及时检测出并隔离掉,则整个导航系统都会被污染。为了解决这一矛盾,1979 年至 1985 年间,Speyer,Bierman 和 Kerr 等人先后提出了分散滤波思想^[34~40]。并行计算技术的成熟为分散滤波的发展创造了有利条件。1988 年起,Carlson 提出了联邦滤波理论(Federated Filtering)^[41~46],旨在为容错组合导航系统提供设计理论。Carlson 在装备运载体的诸多非相似导航子系统中选择导航信息全面、输出速率高、可靠性绝对保证的子系统作为公共参考系统,与其余子系统两两组合,形成若干个子滤波器。各子滤波器并行运行,获得建立在子滤波器局部量测基础上的局部最优估计,这些局部最优估计在第二级滤波器即主滤波器内按融合算法合成,获得建立在所有量测基础上的全局估计。全局估计再按信息守恒原则反馈给各子滤波器。实际设计的联邦滤波器是全局次优的,但是对于自主性要求特别重要的重要运载体来说,导航系统的可靠性比精度更为重要。采用联邦滤波结构设计组合导航系统,虽然相对最优损失了少许精度,但换来的却是组合导航系统的高容错能力。目前美国空军已将联邦滤波器确定为新一代导航系统的通用滤波器^[12,16]。

1.3 组合导航简介

将运载体从起始点引导到目的地的技术或方法称为导航。导航系统测量并解算出运载体的瞬时运动状态和位置,提供给驾驶员或自动驾驶仪实现对运载体的正确操纵或控制。随着科学技术的发展,可资利用的导航信息源越来越多,导航系统的种类也越来越多。以航空导航为例,目前可供装备的机载导航系统有惯性导航系统、GPS 导航系统、多卜勒导航系统、罗兰 C 导航系统等,这些导航系统各有特色,优缺点并存。比如,惯性导航(以下简称惯导)系统的优点是:不需要任何外来信息也不向外辐射任何信息,可在任何介质和任何环境条件下实现导航,且能输出飞机的位置、速度、方位和姿态等多种导航参数,系统的频带宽,能跟踪运载体的任何机动运动,导航输出数据平稳,短期稳定性好。但惯导系统具有固有的缺点:导航精度随时间而

发散,即长期稳定性差。GPS 导航系统导航精度高,在美国国防部加入 SA(Selective Availability) 误差后,使用 C/A 码信号的水平 and 垂直定位精度仍分别可达 100 m 和 157 m(2σ),且不随时间发散,这种高精度和长期稳定性是惯导系统望尘莫及的。但 GPS 导航系统也有其致命弱点:频带窄,当运载体作较高机动运动时,接收机的码环和载波环极易失锁而丢失信号,从而完全丧失导航能力;完全依赖于 GPS 卫星发射的导航信息,受制于他人,且易受人为干扰和电子欺骗。其余导航系统也有各自的优缺点。

各种导航系统单独使用时是很难满足导航性能要求的,提高导航系统整体性能的有效途径是采用组合导航技术,即用两种或两种以上的非相似导航系统对同一导航信息作测量并解算以形成量测量,从这些量测量中计算出各导航系统的误差并校正之。采用组合导航技术的系统称组合导航系统,参与组合的各导航系统称子系统。

实现组合导航有两种基本方法:

(1) 回路反馈法,即采用经典的回路控制方法,抑制系统误差,并使各子系统间实现性能互补。第二次世界大战期间,英国为解决海上巡逻机长时间连续飞行的导航精度问题,率先研究了该型系统。

(2) 最优估计法,即采用卡尔曼滤波或维纳滤波,从概率统计最优的角度估计出系统误差并消除之。两种方法都使各子系统内的信息互相渗透,有机结合,起到性能互补的功效。但由于各子系统的误差源和量测误差都是随机的,所以第二种方法远优于第一种方法。设计组合导航系统时一般都采用卡尔曼滤波。由于惯导和 GPS 在性能上正好形成互补,所以采用该两种系统作为组合导航设计中的子系统是世界公认的最佳方案,如图 1.3.1 所示。

组合导航系统一般具有以下 1~3 种功能:

(1) 协合超越功能。组合导航系统能充分利用各子系统的导航信息,形成单个子系统不具备的功能和精度。

(2) 互补功能。由于组合导航系统综合利用了各子系统的信息,所以各子系统能取长补短,扩大使用范围。

(3) 余度功能。各子系统感测同一信息源,使测量值冗余,提高整个系统的可靠性。

组合导航系统的发展方向是容错组合导航系统和导航专家系统,这些系统具有故障检测、诊断、隔离和系统重构的功能。目前,很多国家都在发展以激光陀螺捷联惯导系统和 GPS 导航系统为主要子系统并辅以其余子系统的组合导航系统,如 Litton 公司的 FLAGSHIP 系统和 Honeywell 公司的 GAINS 系统,平均故障时间可达 12 000 小时,可根据不同的使用要求和系统故障情况选择不同的工作模式,都已成功地应用于诸多先进飞机上。

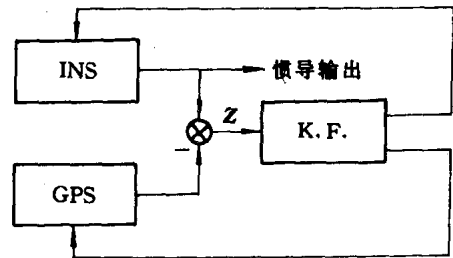


图 1.3.1 惯导/GPS 组合导航系统原理图

第二章 几种最优估计和卡尔曼滤波基本方程

2.1 几种最优估计

所谓估计就是根据测量得出的与状态 $X(t)$ 有关的数据 $Z(t) = h[X(t)] + V(t)$ 解算出 $X(t)$ 的计算值 $\hat{X}(t)$, 其中随机向量 $V(t)$ 为量测误差, \hat{X} 称为 X 的估计, Z 称为 X 的量测。因为 $\hat{X}(t)$ 是根据 $Z(t)$ 确定的, 所以 $\hat{X}(t)$ 是 $Z(t)$ 的函数。若 \hat{X} 是 Z 的线性函数, 则 \hat{X} 称作 X 的线性估计。

设在 $[t_0, t_1]$ 时间段内的量测为 Z , 相应的估计为 $\hat{X}(t)$, 则

当 $t = t_1$ 时, $\hat{X}(t)$ 称为 $X(t)$ 的估计;

当 $t > t_1$ 时, $\hat{X}(t)$ 称为 $X(t)$ 的预测;

当 $t < t_1$ 时, $\hat{X}(t)$ 称为 $X(t)$ 的平滑。

最优估计是指某一指标函数达到最值时的估计。

若以量测估计 \hat{Z} 的偏差的平方和达到最小为指标, 即

$$(Z - \hat{Z})^T(Z - \hat{Z}) = \min$$

则所得估计 \hat{X} 为 X 的最小二乘估计。

若以状态估计 \hat{X} 的均方误差集平均达到最小为指标, 即

$$E[(X - \hat{X})^T(X - \hat{X})] = \min$$

则所得估计 \hat{X} 为 X 的最小方差估计; 若 \hat{X} 又是 X 的线性估计, 则 \hat{X} 为 X 的线性最小方差估计。

也可用估计值出现的概率作为估计指标, 这样的估计有极大验后估计、贝叶斯估计和极大似然估计。本节将对上述各种估计作详细介绍。

2.1.1 最小二乘估计

1. 一般最小二乘估计

最小二乘估计是高斯(Karl Gauss)在1795年为测定行星轨道而提出的参数估计算法。这种估计的特点是算法简单, 不必知道与被估计量及量测量有关的任何统计信息。

设 X 为某一确定性常值向量, 维数为 n 。一般情况下对 X 不能直接测量, 而只能测量到 X 各分量的线性组合。记第 i 次量测 Z_i 为

$$Z_i = H_i X + V_i \quad (2.1.1)$$

式中 Z_i 为 m 维向量, H_i 和 V_i 为第 i 次测量的量测矩阵和随机量测噪声。

若共测量 r 次, 即

$$\left. \begin{aligned} Z_1 &= H_1 X + V_1 \\ Z_2 &= H_2 X + V_2 \\ &\dots\dots \\ Z_r &= H_r X + V_r \end{aligned} \right\} \quad (2.1.2)$$

则由上述诸式可得描述 r 次量测的量测方程

$$Z = HX + V \quad (2.1.3)$$

式中 Z, V 为 $\sum_{i=1}^r m_i = m$ 维向量, H 为 $m \times n$ 矩阵。

最小二乘估计的指标是:使各次量测 Z_i 与由估计 \hat{X} 确定的量测的估计 $\hat{Z}_i = H_i \hat{X}$ 之差的平方和最小,即

$$J(\hat{X}) = (Z - H\hat{X})^T (Z - H\hat{X}) = \min \quad (2.1.4)$$

而要使上式达到最小,需满足(详细推导见附录)

$$\left. \frac{\partial J}{\partial X} \right|_{X=\hat{X}} = -2H^T(Z - H\hat{X}) = 0$$

若 H 具有最大秩 n , 即 $H^T H$ 正定, 且 $m = \sum_{i=1}^r m_i > n$, 则 X 的最小二乘估计为

$$\hat{X} = (H^T H)^{-1} H^T Z \quad (2.1.5)$$

从上式可看出最小二乘估计是一种线性估计。

为了说明最小二乘估计最优的含义, 将式(2.1.4)改写成

$$J(\hat{X}) = [(Z_1 - H_1 \hat{X})^T (Z_2 - H_2 \hat{X})^T \dots (Z_r - H_r \hat{X})^T] \begin{bmatrix} Z_1 - H_1 \hat{X} \\ Z_2 - H_2 \hat{X} \\ \vdots \\ Z_r - H_r \hat{X} \end{bmatrix} =$$

$$\sum_{i=1}^r (Z_i - H_i \hat{X})^T (Z_i - H_i \hat{X}) = \min$$

这说明, 最小二乘估计虽然不能满足式(2.1.2)中的每一个方程, 使每个方程都有偏差, 但它使所有方程偏差的平方和达到最小, 这实际上兼顾了所有方程的近似程度, 使整体误差达到最小, 这对抑制测量误差 V_1, V_2, \dots, V_r 的影响是有益的。

最小二乘估计具有如下性质:

若量测噪声 V 是均值为零、方差为 R 的随机向量, 则

(1) 最小二乘估计是无偏估计, 即

$$E[\hat{X}] = X \quad (2.1.6)$$

或

$$E[\tilde{X}] = 0 \quad (2.1.7)$$

式中 $\tilde{X} = X - \hat{X}$ 为 \hat{X} 的估计误差。

(2) 最小二乘估计的均方误差阵为

$$E[\tilde{X}\tilde{X}^T] = (H^T H)^{-1} H^T R H (H^T H)^{-1} \quad (2.1.8)$$

证明

(1) 由式(2.1.5), \hat{X} 的估计误差为

$$\tilde{X} = X - \hat{X} = X - (H^T H)^{-1} H^T Z$$

由于 $E[V] = 0$, 所以

$$E[\tilde{X}] = E[(H^T H)^{-1}(H^T H)X - (H^T H)^{-1}H^T Z] = (H^T H)^{-1}H^T E[HX - Z] = - (H^T H)^{-1}H^T E[V] = 0$$

$$(2) E[\tilde{X}\tilde{X}^T] = (H^T H)^{-1}H^T E[VV^T]H(H^T H)^{-1} = (H^T H)^{-1}H^T R H(H^T H)^{-1}$$

例 2-1 用一台仪器对未知确定性标量 X 作 r 次直接测量, 量测值分别为 Z_1, Z_2, \dots, Z_r , 测量误差的均值为零, 方差为 R , 求 X 的最小二乘估计 \hat{X} , 并计算估计的均方误差。

解 由题意, r 次直接测量所得的量测方程为

$$Z = HX + V$$

式中 $Z = [Z_1 \ Z_2 \ \dots \ Z_r]^T, H = [1 \ 1 \ \dots \ 1]^T, E[VV^T] = RI$

根据式(2.1.5), 得

$$\hat{X} = \frac{1}{r}(Z_1 + Z_2 + \dots + Z_r)$$

估计的均方误差为

$$E[\tilde{X}^2] = \frac{R}{r}$$

例 2-2 设有线性系统满足

$$X_k = \Phi_{k,k-1}X_{k-1}$$

对 X_k 的量测为标量, 满足

$$Z_k = h_k X_k + V_k$$

式中: X_k 为 n 维向量; V_k 为随机误差; k 为时刻 t_k 。求 X_0 基于 $Z_1, Z_2, \dots, Z_m (m \geq n)$ 的最小二乘估计 $\hat{X}_{0/m}$ 。

解 X_0 的第 i 次量测为

$$Z_i = h_i X_i + V_i = h_i \Phi_{i,0} X_0 + V_i \quad i = 1, 2, \dots, m$$

X_0 的前 m 次量测为

$$Z_m = H_{m,0} X_0 + V_m$$

式中

$$Z_m = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_m \end{bmatrix}, H_{m,0} = \begin{bmatrix} h_1 \Phi_{1,0} \\ h_2 \Phi_{2,0} \\ \vdots \\ h_m \Phi_{m,0} \end{bmatrix}, V_m = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_m \end{bmatrix}$$

根据式(2.1.5), 得 X_0 基于 Z_1, Z_2, \dots, Z_m 的最小二乘估计为

$$\hat{X}_{0/m} = (H_{m,0}^T H_{m,0})^{-1} H_{m,0}^T Z_m$$

例 2-3 用两台仪器对未知标量 X 各直接测量一次, 量测量分别为 Z_1 和 Z_2 。仪器的测量误差是均值为零, 方差分别为 r 和 $4r$ 的随机量, 求 X 的最小二乘估计, 并计算估计的均方误差。

解 由题意得量测方程

$$Z = HX + V$$

式中 $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, H = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, R = \begin{bmatrix} r & 0 \\ 0 & 4r \end{bmatrix}$

根据式(2.1.5)和式(2.1.8), 得

$$\hat{X} = \frac{1}{2}(Z_1 + Z_2)$$

$$E[\hat{X}^2] = \frac{1}{2} [1 \ 1] \begin{bmatrix} r & 0 \\ 0 & 4r \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \frac{1}{2} = \frac{5}{4}r > r$$

上式说明,使用精度差一倍的两台仪器同时进行测量,最小二乘估计效果还不如只使用一台精度高的仪器时好。

2. 加权最小二乘估计

从例2-3看出,一般最小二乘估计精度不高的原因之一是不分优劣地使用了量测值。如果对不同量测值的质量有所了解,则可用加权的办法分别对待各量测量,精度质量高的权重取得大些,精度质量差的权重取得小些。这就是加权最小二乘估计的思路。加权最小二乘估计 \hat{X} 的求取准则是

$$J(\hat{X}) = (Z - H\hat{X})^T W (Z - H\hat{X}) = \min \quad (2.1.9)$$

式中 W 是适当取值的正定加权矩阵,当 $W = I$ 时,上式就是一般最小二乘准则。

要使式(2.1.9)成立, \hat{X} 应满足

$$\frac{\partial J(\hat{X})}{\partial \hat{X}} \Big|_{\hat{X}} = -H^T(W + W^T)(Z - H\hat{X}) = 0$$

从中解得

$$\hat{X} = [H^T(W + W^T)H]^{-1}H^T(W + W^T)Z$$

一般情况下,加权阵取成对称阵,即 $W^T = W$,所以加权最小二乘估计为

$$\hat{X} = (H^TWH)^{-1}H^TWZ \quad (2.1.10)$$

估计误差为

$$\begin{aligned} \tilde{X} = X - \hat{X} &= (H^TWH)^{-1}H^TWHX - (H^TWH)^{-1}H^TWZ = \\ &= (H^TWH)^{-1}H^TW(HX - Z) = \\ &= - (H^TWH)^{-1}H^TWV \end{aligned}$$

如果量测误差 V 的均值为零,方差阵为 R ,则加权最小二乘估计也是无偏估计,估计的均方误差为

$$E[\tilde{X}\tilde{X}^T] = (H^TWH)^{-1}H^TWRWH(H^TWH)^{-1} \quad (2.1.11)$$

如果 $W = R^{-1}$,则加权最小二乘估计

$$\hat{X} = (H^TR^{-1}H)^{-1}H^TR^{-1}Z \quad (2.1.12)$$

又称为马尔柯夫估计。

马尔柯夫估计的均方误差为

$$E[\tilde{X}\tilde{X}^T] = (H^TR^{-1}H)^{-1} \quad (2.1.13)$$

马尔柯夫估计的均方误差比任何其他加权最小二乘估计的均方误差都要小,所以是加权最小二乘估计中的最优者。下面证明这一结论。

设 A 和 B 分别为 $n \times m$ 和 $m \times l$ 的矩阵,且 AA^T 满秩,则因

$$\begin{aligned} [B - A^T(AA^T)^{-1}AB]^T [B - A^T(AA^T)^{-1}AB] &= \\ B^TB - 2B^TA^T(AA^T)^{-1}AB + B^TA^T(AA^T)^{-1}AA^T(AA^T)^{-1}AB &= \\ B^TB - (AB)^T(AA^T)^{-1}(AB) &\geq 0 \end{aligned}$$

所以有

$$B^T B \geq (AB)^T (AA^T)^{-1} (AB) \quad (2.1.14)$$

这就是矩阵型的许瓦茨不等式。

由矩阵理论知, 正定阵 R 可表示成 $R = S^T S$, 其中 S 为满秩矩阵。令

$$A = H^T S^{-1} \quad (2.1.15a)$$

$$B = SWH(H^TWH)^{-1} \quad (2.1.15b)$$

则

$$AB = H^T S^{-1} SWH(H^TWH)^{-1} = I \quad (2.1.16)$$

$$B^T B = (H^TWH)^{-1} H^T WRWH (H^TWH)^{-1} \quad (2.1.17)$$

根据许瓦茨不等式和式(2.1.16)及式(2.1.15a), 有

$$B^T B \geq (AB)^T (AA^T)^{-1} (AB) = (AA^T)^{-1} = (H^T R^{-1} H)^{-1}$$

式(2.1.17)式代入上式, 得

$$(H^TWH)^{-1} H^T WRWH (H^TWH)^{-1} \geq (H^T R^{-1} H)^{-1} \quad (2.1.18)$$

上式说明, 只有当 $W = R^{-1}$ 时, 估计的均方误差才达到最小, 最小值为 $(H^T R^{-1} H)^{-1}$ 。

例 2-4 对例 2-3 采用马尔柯夫估计, 并求估计的均方误差。

解

$$\text{取 } W = R^{-1} = \begin{bmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{4r} \end{bmatrix}$$

$$\hat{X} = ([1 \ 1] \begin{bmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{4r} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix})^{-1} [1 \ 1] \begin{bmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{4r} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \frac{4}{5} Z_1 + \frac{1}{5} Z_2$$

$$E[\tilde{X}\tilde{X}] = ([1 \ 1] \begin{bmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{4r} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix})^{-1} = \frac{4}{5} r < r$$

这说明, 对精度高的量测值取权重系数 $\frac{4}{5}$, 对精度低的量测值取权重系数 $\frac{1}{5}$, 估计精度高于一用精度高的量测所得估计的精度。所以增加不同的量测值, 并根据其精度质量区别对待利用之, 能有效提高估计精度。

3. 递推最小二乘估计

从上述诸例可看出, 量测值越多, 只要处理得合适, 最小二乘估计的均方误差就越小。采用批处理实现的最小二乘算法, 须存储所有的量测值。若量测值数量十分庞大, 则计算机必须具备巨大的存储容量, 这显然是不经济的。递推最小二乘估计从每次获得的量测值中提取出被估计量信息, 用于修正上一步所得的估计。获得量测的次数越多, 修正的次数也越多, 估计的精度也越高。下面详细介绍该算法。

设 X 为确定性常值向量, 前 k 次观测积累的量测为 \bar{Z}_k , 量测方程为

$$\bar{Z}_k = \bar{H}_k X + \bar{V}_k$$

式中

$$\bar{Z}_k = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{bmatrix}, \bar{H}_k = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_k \end{bmatrix}, \bar{V}_k = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix}$$

Z_i 为第 i 次量测, 量测方程为

$$Z_i = H_i X + V_i \quad i = 1, 2, \dots, k$$

则前 $k+1$ 次量测为

$$\bar{Z}_{k+1} = \bar{H}_{k+1} X + \bar{V}_{k+1}$$

式中

$$\bar{Z}_{k+1} = \begin{bmatrix} \bar{Z}_k \\ Z_{k+1} \end{bmatrix}, \bar{H}_{k+1} = \begin{bmatrix} \bar{H}_k \\ H_{k+1} \end{bmatrix}, \bar{V}_{k+1} = \begin{bmatrix} \bar{V}_k \\ V_{k+1} \end{bmatrix}$$

Z_{k+1} 为第 $k+1$ 次量测, 量测方程为

$$Z_{k+1} = H_{k+1} X + V_{k+1}$$

根据式(2.1.10), 由前 k 次量测确定的加权最小二乘估计为

$$\hat{X}_k = (\bar{H}_k^T \bar{W}_k \bar{H}_k)^{-1} \bar{H}_k^T \bar{W}_k \bar{Z}_k$$

式中

$$\bar{W}_k = \begin{bmatrix} W_1 & & & \\ & W_2 & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & & & W_k \end{bmatrix}$$

令

$$P_k = (\bar{H}_k^T \bar{W}_k \bar{H}_k)^{-1} \quad (2.1.19)$$

则

$$\hat{X}_k = P_k \bar{H}_k^T \bar{W}_k \bar{Z}_k \quad (2.1.20)$$

由前 $k+1$ 次量测确定的加权最小二乘估计为

$$\begin{aligned} \hat{X}_{k+1} &= (\bar{H}_{k+1}^T \bar{W}_{k+1} \bar{H}_{k+1})^{-1} \bar{H}_{k+1}^T \bar{W}_{k+1} \bar{Z}_{k+1} = P_{k+1} [\bar{H}_k^T \ H_{k+1}^T] \begin{bmatrix} \bar{W}_k & \mathbf{0} \\ \mathbf{0} & W_{k+1} \end{bmatrix} \begin{bmatrix} \bar{Z}_k \\ Z_{k+1} \end{bmatrix} = \\ & P_{k+1} \bar{H}_k^T \bar{W}_k \bar{Z}_k + P_{k+1} H_{k+1}^T W_{k+1} Z_{k+1} \end{aligned} \quad (2.1.21)$$

式中

$$\begin{aligned} P_{k+1} &= (\bar{H}_{k+1}^T \bar{W}_{k+1} \bar{H}_{k+1})^{-1} = ([\bar{H}_k^T \ H_{k+1}^T] \begin{bmatrix} \bar{W}_k & \mathbf{0} \\ \mathbf{0} & W_{k+1} \end{bmatrix} [\bar{H}_k \\ H_{k+1}])^{-1} = \\ & (\bar{H}_k^T \bar{W}_k \bar{H}_k + H_{k+1}^T W_{k+1} H_{k+1})^{-1} = (P_k^{-1} + H_{k+1}^T W_{k+1} H_{k+1})^{-1} \end{aligned} \quad (2.1.22)$$

由矩阵反演公式

$$(A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1} A_{12} (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} A_{21} A_{11}^{-1} \quad (2.1.23)$$

令 $A_{11} = P_k^{-1}$, $A_{12} = -H_{k+1}^T$, $A_{22}^{-1} = W_{k+1}$, $A_{21} = H_{k+1}$, 得

$$P_{k+1} = P_k - P_k H_{k+1}^T (W_{k+1}^{-1} + H_{k+1} P_k H_{k+1}^T)^{-1} H_{k+1} P_k \quad (2.1.24)$$

再考察式(2.1.21)中的第一项, 由式(2.1.20)式和(2.1.22)式, 得

$$\bar{H}_k^T \bar{W}_k \bar{Z}_k = P_k^{-1} \hat{X}_k$$

$$P_k^{-1} = P_{k+1}^{-1} - H_{k+1}^T W_{k+1} H_{k+1}$$

所以该项为

$$P_{k+1} \overline{H}_k^T \overline{W}_k \overline{Z}_k = P_{k+1} P_k^{-1} \hat{X}_k = P_{k+1} (P_{k+1}^{-1} - H_{k+1}^T W_{k+1} H_{k+1}) \hat{X}_k = \hat{X}_k - P_{k+1} H_{k+1}^T W_{k+1} H_{k+1} \hat{X}_k$$

因此式(2.1.21)成

$$\begin{aligned} \hat{X}_{k+1} &= \hat{X}_k - P_{k+1} H_{k+1}^T W_{k+1} H_{k+1} \hat{X}_k + P_{k+1} H_{k+1}^T W_{k+1} Z_{k+1} = \\ & \hat{X}_k + P_{k+1} H_{k+1}^T W_{k+1} (Z_{k+1} - H_{k+1} \hat{X}_k) \end{aligned} \quad (2.1.25)$$

式(2.1.24)和(2.1.25)即为递推最小二乘估计的全套算法。

式(2.1.25)说明, $k+1$ 时刻的估计由对 k 时刻的估计作修正而获得, 修正量由对 $k+1$ 时刻的量测的估计误差 $\tilde{Z}_{k+1} = Z_{k+1} - H_{k+1} \hat{X}_k$ 经增益阵 $K_{k+1} = P_{k+1} H_{k+1}^T W_{k+1}$ 加权后确定, 其中 P_{k+1} 由式(2.1.24)确定。

式(2.1.24)和式(2.1.25)确定的算法是递推的, 只要给定初始值 \hat{X}_0 和 P_0 , 即可获得 X 在任意时刻的最小二乘估计。 \hat{X}_0 和 P_0 的选取可以是任意的, 一般可取 $\hat{X}_0 = \mathbf{0}$, $P_0 = \rho I$, 其中 ρ 为很大的正数。由于初值选取盲目, 所以递推过程中, 刚开始计算时, 估计误差跳跃剧烈, 随着量测次数的增加, 初值影响逐渐消失, 估计值逐渐趋于稳定而逼近被估计量。

例 2-5 在例 2-1 中求出 \hat{X}_k 后, 试求获得新量测值 Z_{k+1} 后的递推最小二乘估计。

解 例 2-1 采用了一般最小二乘算法, 这实际上加权阵取作单位阵。根据式(2.1.19)、式(2.1.24)及式(2.1.25), 得

$$\begin{aligned} P_k &= ([1 \ 1 \ \cdots \ 1] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix})^{-1} = \frac{1}{k} \\ P_{k+1} &= \frac{1}{k} - \frac{1}{k} \left(1 + \frac{1}{k}\right)^{-1} \frac{1}{k} = \frac{1}{k+1} \\ \hat{X}_{k+1} &= \hat{X}_k + \frac{1}{k+1} (Z_{k+1} - \hat{X}_k) \end{aligned}$$

最小二乘估计的最大优点是算法简单, 特别是一般最小二乘估计, 根本不必知道量测误差的统计信息。但正是这种优点又引起了使用上的局限性, 主要体现在如下两点上:

- (1) 最小二乘算法只能估计确定性的常值向量, 而无法估计随机向量的时间过程。
- (2) 最小二乘的最优指标只保证了量测的估计均方误差之和最小, 而并未确保被估计量的估计误差达到最佳, 所以估计精度不高。

2.1.2 最小方差估计

1. 最小方差估计与条件均值

设 X 为随机向量, Z 为 X 的量测向量, 即 $Z = Z(X) + V$, 求 X 的估计 \hat{X} 就是根据 Z 解算出 X , 显然 \hat{X} 是 Z 的函数, 由于 V 是随机误差, 所以 X 无法从 Z 的函数关系式中直接求取, 而必须按统计意义的最优标准求取。

最小方差估计是使下述指标达到最小的估计 $\hat{X}_{MV}(Z)$

$$J = E_{X,Z} \{ [X - \hat{X}(Z)]^T [X - \hat{X}(Z)] \} \Big|_{\hat{X}(Z) = \hat{X}_{MV}(Z)} = \min \quad (2.1.26)$$