朱志宇　著

粒子滤波算法及其应用

科学出版社
www.sciencep.com

(TP-4736.0101)

粒子滤波
算法及其应用

科学出版社
电　话：010-64006909
E-mail：gcjs@mail.sciencep.com

销售分类建议：电子信息、自动化

ISBN 978-7-03-027611-7
9 787030 276117 >
定价：48.00 元

粒子滤波算法及其应用

朱志宇　著

科 学 出 版 社

北　京

内 容 简 介

本书系统介绍粒子滤波算法的基本原理和关键技术，针对标准粒子滤波算法存在的粒子退化、计算量大的缺点介绍了多种改进的粒子滤波算法，包括基于重要性密度函数选择的粒子滤波算法、基于重采样技术的粒子滤波算法、基于智能优化思想的粒子滤波算法、自适应粒子滤波算法、流形粒子滤波算法等，并将粒子滤波算法应用于机动目标跟踪、语音增强、传感器故障诊断、人脸跟踪等领域，最后探讨了粒子滤波算法的硬件实现问题，给出了基于 DSP 和 FPGA 的粒子滤波算法实现方法。

本书可供高等院校电子信息、自动化、计算机应用、应用数学等有关专业高年级本科生和研究生，以及从事控制科学与工程、信号与信息处理领域的工程技术人员和研究人员参考阅读。

图书在版编目(CIP)数据

粒子滤波算法及其应用/朱志宇著. —北京：科学出版社，2010.6

ISBN 978-7-03-027611-7

Ⅰ.①粒… Ⅱ.①朱… Ⅲ.①非线性控制系统 Ⅳ.①O231.2

中国版本图书馆 CIP 数据核字(2010)第 088221 号

责任编辑：孙 芳 王志欣 / 责任校对：陈玉凤
责任印制：赵 博 / 封面设计：耕者设计工作室

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
丽源印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2010 年 6 月第 一 版 开本：B5(720×1000)
2010 年 6 月第一次印刷 印张：16 3/4
印数：1—3 000 字数：324 000

定价：48.00 元

(如有印装质量问题，我社负责调换)

前　言

粒子滤波又称序贯蒙特卡罗方法，是一种基于蒙特卡罗方法和递推贝叶斯估计的统计滤波方法，它依据大数定理，采用蒙特卡罗方法来求解贝叶斯估计中的积分运算。粒子滤波算法首先依据系统状态向量的经验条件分布在状态空间产生一组随机样本的集合，然后根据观测量不断地调整粒子的权重和位置，通过调整后粒子的信息修正最初的经验条件分布。当样本容量很大时，这种蒙特卡罗描述就近似于状态变量真实的后验概率密度函数。粒子滤波适用于任何能用状态空间模型表示的非高斯背景的非线性随机系统，它完全突破了传统的 Kalman 滤波理论框架，对系统的过程噪声和量测噪声没有任何限制，可适用于任何非线性系统，精度可以逼近最优估计，是一种很有效的非线性滤波技术，可广泛应用于数字通信、金融领域数据分析、统计学、图像处理、计算机视觉、自适应估计、语音信号处理、机器学习等方面。粒子滤波算法是现代信号与信息处理学科和统计模拟理论之间的交叉学科，其研究有着重要的理论意义和现实价值，随着计算机性能的迅速提高，这一方法日益受到人们的关注。

近年来，从解决粒子退化和粒子多样性丧失、提高算法实时性和鲁棒性、降低计算复杂度等角度考虑，国内外学者广泛开展了粒子滤波研究。本书系统总结了近年来粒子滤波的研究成果，针对粒子滤波算法的缺点提出了若干种改进算法，包括基于微分流形的粒子滤波算法、基于人工鱼群的粒子滤波算法、基于神经网络的粒子滤波算法、自适应粒子滤波算法等；广泛探讨了粒子滤波算法的各种应用，给出了粒子滤波算法的硬件实现方法。

在本书编撰过程中，作者研读了大量文献，参考融合了国内外专家、学者们在相关领域的研究成果，在此，对他们表示衷心谢意！

王建华教授、姜长生教授、张冰教授对本书的编写工作提供了很多宝贵意见，杨官校、李冀、皇丰辉、刘炜、薄超等同学编制了书中的仿真程序，赵成、苏岭东、姜威威等同学绘制了书中的部分图表。在此，向参与和关心本书编写工作的各位同事和同学表示真诚的感谢！

本书的出版得到了江苏省高校自然科学基金(项目编号：06KJB510030)和中国船舶行业预研基金(项目编号：3.1.5)的资助。

由于作者学术水平有限，书中难免存在不妥之处，殷切期望广大读者批评指正。

作　者

2010 年 3 月

目　　录

第二篇　粒子滤波算法的应用

第一篇　粒子滤波算法

第1章 绪 论

非线性、非高斯随机系统的数据分析和处理在统计学、语音和图像处理、数字通信、计算机视觉、自适应估计、机器学习及自动控制等领域有着广泛应用。以往由于实时处理的要求和计算存储量的限制，通常采用递推滤波算法来求解此类问题，即所谓的Kalman滤波理论，包括扩展Kalman滤波(EKF)、修正增益的扩展Kalman滤波(MGEKF)、多模型算法(MM)等，其基本思想是通过参数化的解析式对系统的非线性进行近似以得到满意的估计精度；但Kalman滤波理论仅仅采用均值和方差表征状态概率分布，当系统为线性系统，且过程和量测噪声为高斯分布时，Kalman滤波可以获得最优状态估计，对于非线性、非高斯分布的状态模型，其滤波和预测精度很难保证。EKF仅适用于滤波误差和预测误差很小的情况，否则滤波初期估计协方差下降太快会导致滤波不稳定甚至发散；MGEKF虽然通过改进增益矩阵而相应改善了状态协方差的估计性能，但其对量测误差有一定的限制，若量测误差较大，算法在收敛精度、收敛时间和稳定性等方面的表现均不太理想。而且，两者仅仅利用了非线性函数泰勒展开式的一阶偏导部分(忽略高阶项)，常常导致在估计状态后验分布时产生较大的误差，影响滤波算法的性能。总之，从理论上来说，当系统为线性系统，且过程和量测噪声为高斯分布时，Kalman滤波可以获得最优状态估计；当不符合上述条件时，其滤波和预测精度将很难保证。

近几年发展起来的粒子滤波是一种基于蒙特卡罗思想的非线性、非高斯系统滤波方法，完全突破了Kalman滤波理论框架，它对系统的过程噪声和量测噪声没有任何限制。粒子滤波通过预测和更新来自于系统概率密度函数的采样集，来近似非线性系统的贝叶斯估计，在处理非高斯、非线性时变系统的参数估计和状态滤波问题方面有独特的优势和广阔的前景。粒子滤波是现代信号与信息处理学科和统计模拟理论之间的交叉学科，其研究有着重要的理论意义和现实价值。

尽管粒子滤波已成为解决非线性、非高斯动态系统的参数估计和状态滤波问题的主流方法，但粒子滤波在当前还不够成熟，仍有许多亟待克服的问题。例如，如何在具体应用中给出逼近最优的建议分布；如何用严格的数学方法给出粒子滤波的收敛性结论；研究在高维数条件下对维数真正不敏感的粒子滤波算法；如何设计更加有效的重采样算法；如何有效克服算法中的权值退化及样本枯竭等问题；各种形式粒子滤波算法的高效、实时实现问题等。因此，有必要进一步深入研究粒子滤波，完善其理论体系，拓展其应用领域。

1.1 粒子滤波的发展和应用

早在20世纪50年代，统计学和理论物理领域就引入了粒子滤波方法，并于60年代末在自动控制领域得到了应用，70年代得到了一定的发展。但是，由于始终未能解决粒子退化现象和计算量制约等问题，因而并未引起足够重视，直到1993年，Gordon等提出了自举粒子滤波(bootstrap particle filter)算法[1]，该算法在递推过程中引入了重新采样思想以克服退化问题；同时，计算机运算能力的急剧增长也为粒子滤波的物理实现提供了客观条件。因此，随后粒子滤波得到了长足发展，而后又有许多改进算法相继被提出，掀起了一股研究粒子滤波的热潮。

在粒子滤波算法的发展历程中，出现了许多不同的称谓，如自举滤波、凝聚算法、蒙特卡罗滤波、序贯蒙特卡罗方法、适者生存、粒子滤波，现在通称为粒子滤波。

2000年，Doucet等在前人研究的基础上给出了基于序贯重要性采样(sequence importance sampling，SIS)[2]的粒子滤波的通用描述，即采用蒙特卡罗方法求解贝叶斯估计中的多重积分运算，并利用SIS技术在动态状态空间上得到一组粒子，每一个粒子都对应一个重要性权值，最后通过对这些粒子加权求和来获得状态后验概率密度的估计。其后针对粒子滤波不足之处提出的各种改进算法都是在SIS的基础上得到的。另外，Doucet等也证明了当采样粒子足够多时，粒子滤波算法是收敛的，收敛速度不受状态维数的限制[3]。他们为粒子滤波的研究奠定了坚实的基础。

目前，粒子滤波在定位、跟踪领域得到了深入研究[4~6]，主要包括汽车定位、通过地图匹配或地形辅助导航的飞机定位、综合导航及测角、测距或测速跟踪等；国内也有学者将粒子滤波应用于空对海单站无源跟踪、多目标跟踪等领域；视觉跟踪是当前一个热门的研究领域，粒子滤波因其较强的非线性处理能力而被成功地应用于该领域[7]；粒子滤波较好地解决了机器人定位问题；在无线通信中的应用包括盲均衡、平坦衰落信道中的盲检测、盲解卷、信号解调、多用户检测和衰落信道中空时编码的估计与检测等；语音信号是一种典型的非高斯、非平稳信号，因此，应用粒子滤波可以进行语音识别、语音增强与消噪、语音信号盲分离等；粒子滤波还被用于目标识别、系统辨识参数估计、自动控制、动态系统中的故障检测、经济统计等。可以说，凡是需要用到非线性、非高斯递推贝叶斯估计的地方都可以应用粒子滤波，所以，粒子滤波的应用领域极为广泛。

1.2 粒子滤波的缺点和现有的解决方法

1. 重要性函数选择问题

在标准粒子滤波算法中，为了求解方便，一般取重要性概率密度为先验概率密

度，但是，这种方法丢失了当前时刻的量测值，使得当前时刻的状态严重依赖于模型。如果模型不准确，或者量测噪声突然增大，则这种选取方法将不能有效地表示概率密度函数的真实分布。同时，在这种分布下，计算权重时没有考虑系统的模型噪声。因此，从重要性概率密度采样得到的样本与从真实后验概率密度采样得到的样本之间有很大的偏差，尤其当似然函数位于系统状态转移概率密度的尾部或似然函数呈尖峰状态时，这种偏差就更加明显。解决的办法是设法将粒子向似然函数的峰值区移动，如“预编辑”法及类似的“取舍法”采样；或者选用其他更合适的建议分布，如采用先验转移密度的退火形式作为建议分布[8]；用似然函数作为建议分布，用先验转移密度作为权值迭代的比例因子，这就是似然粒子滤波[9]；用于克服尖锐型的似然函数与先验转移密度重叠区过小问题的建议分布还有桥接密度、分割采样、基于梯度的转移密度[10]；此外，还有用 EKF、无味 Kalman 滤波(UKF)或高斯-厄米滤波(GHF)将最近的量测信息计入建议分布的扩展 Kalman 粒子滤波(extended Kalman particle filter，EKPF)、无味粒子滤波(unscented particle filter，UPF)[11]和高斯-厄米粒子滤波(GHPF)等。

上述各种针对实际问题改进的粒子滤波算法对非线性系统的状态模型仍有一定的限制，如要求系统状态的后验分布必须能够用高斯分布近似等，因而对粒子滤波的重要性概率密度函数的通用选择方案的研究缺乏一般借鉴意义；现有的各种粒子滤波算法仍未能很好地解决重采样过程中计算量较大的问题、出现样本枯竭而面临粒子匮乏时导致滤波发散现象。

2. 重采样的样本枯竭问题

粒子退化是标准粒子滤波算法的主要缺陷。粒子退化是指随着迭代次数增加，粒子丧失多样性的现象，Doucet 从理论上证明了粒子退化现象的必然性。解决该问题的最有效方法是选择好的重要性概率密度函数和采用重采样方法。其中，好的重要性概率密度函数的选择准则是使重要性权值的方差最小，重要性概率密度函数应尽可能接近系统状态后验概率密度函数。重采样包括系统重采样、分层重采样和残差重采样等方法，其基本思想是通过对用粒子和相应权值表示的后验概率密度函数重采样产生新的支撑点集。系统重采样由于实现简单、算法复杂度低而得到了广泛运用。但是，重采样带来了新问题：权值越大的粒子的子代越来越多，而权值较小的粒子被剔除，最糟糕的情形是新的粒子集实际都是一个权值最大的粒子的子代，即所谓“样本枯竭”现象，从而导致粒子集的多样性变差，不足以用来近似表征后验密度，难以保证估计精度，尤其是在过程噪声较小时问题就更严重。为保证粒子的多样性，又提出了重采样-移动算法(resampling move algorithm)、正则化方法等。

(1) 重采样-移动算法。在重采样之后加上一步 MCMC(Markov chain Monte

Carlo)移动处理使粒子集趋于平稳分布，减弱粒子间的相关性；而且 Markov 链能使粒子分布更加接近状态概率密度函数分布，使样本分布更加合理。典型的 MCMC 算法有 MH(Metropolis Hastings)算法和 Gibbs 采样。采用 MCMC 移动方法的突出缺陷就是为了保证收敛所需的概率转移次数大，算法增加的运算量大，而且收敛的判断也是个问题。

(2) 正则粒子滤波(regularized particle filter，RPF)。正则化方法通过引入核密度函数和核带宽系数以连续形式计算状态后验概率。较之于标准粒子滤波的离散形式，该算法可以有效缓解重采样过程造成的粒子退化问题，在过程噪声较小时可获得较好的滤波精度，但不能保证样本粒子都能近似表示状态后验概率，而且对非高斯情况核函数和核带宽系数不能达到最优，只是一种次优滤波方法。此外，高维时的正则化难以实现，而且重采样后的粒子集不再是后验密度的渐近无偏估计。

(3) 辅助变量粒子滤波(auxiliary variable particle filter，AVPF)。在重采样前，AVPF 依据似然值的大小对原粒子集中的各个权值进行修正，使得重采样后的粒子向似然函数的高值区移动。该算法从经过平滑的后验密度中重采样，更接近于状态的真值，因此，可以获得更小的权值方差。但是，由于在 AVPF 算法的一次迭代中，对于每个粒子需要计算两次似然函数和权值，所以，其计算量增大了。

(4) 裂变自举粒子滤波(fission bootstrap particle filtering，FBPF)。FBPF 算法在自举粒子滤波算法的基础上改进了重采样过程，引入了权值退化检测和“权值排序(sorting)—裂变繁殖(fission)—权值归一(normalizing)过程”(简称 SFN 预处理过程)。裂变过程是一种随机采样方法，它的采样计划不是始终保持不变的，而是根据实验结果不断进行调整，设计出新的采样计划，从而克服了粒子匮乏问题。

免重采样粒子滤波包括高斯粒子滤波(Gaussian particle filter，GPF)、高斯和粒子滤波(Gaussian sum particle filter，GSPF)等。

(1) GPF[12]。它将高斯滤波和粒子滤波相结合。其前提是用高斯分布来近似后验分布，比其他的高斯滤波方法适用性更强，能处理更多的非线性动态系统问题；而与一般的粒子滤波相比，因为 GPF 用高斯分布近似后验分布，所以，只要所用的高斯分布是正确的，就不会产生粒子退化问题，就不需要对粒子进行重采样，从而降低了算法计算量和复杂度。

(2) GSPF[13]。针对后验分布不能用高斯分布近似的非线性动态空间模型或噪声模型为非加性高斯噪声模型时 GPF 的滤波性能不佳的情形，提出了 GSPF，即用多个高斯分布加权累加来近似滤波和预测分布。

(3) 基于智能优化思想的粒子滤波。在实际应用中，粒子滤波器的样本集合不能太大，否则，将增加算法计算量；但样本集合太小，又会导致样本枯竭现象，影响到粒子的多样性，使得估计精度下降。近些年来，不少学者将模拟退火、遗传算法、模糊[14]、粒子群优化(particle swarm optimization，PSO)算法、人工免疫算法等

引入到粒子滤波的重采样过程中，通过优化搜索保留那些能够反映系统概率密度函数的最好粒子，在一定程度上使先验粒子向高似然域移动，从而改善了粒子分布，增加了粒子的多样性，提高了估计精度，为解决粒子退化问题提供了新的思路，同时还能加速粒子集的收敛。

3. 粒子滤波的实时性问题

与传统的 Kalman 滤波理论相比，粒子滤波的实时性较差，其计算量随着粒子数的增加成级数增加。因此，粒子滤波算法的实时性问题使粒子滤波距离工程应用尚有一定距离。目前，降低粒子滤波计算量的方法主要有自适应粒子滤波(adaptive particle filter，APF)和实时粒子滤波(real-time particle filter，RTPF)。

(1) APF。APF 是指所用的粒子数不再固定，而是随着信号环境的变化而自适应改变，剔除冗余粒子数可以降低算法实现的复杂度和运算量。目前，用于自适应改变粒子数的方法主要有两类：①基于似然函数的 APF(L-APF)，即所需的粒子数应能保证非归一化似然值的和超过某一预定的门限；②基于 Kullback-Leibler (KL)信息数或 KL 距离(KLD)采样的 APF(KLD APF)[15]，即通过粒子数的自适应变化来保证后验密度的真值与估计值之间的误差限，这种误差限用 KLD 表示。这两种方法都是当概率密度集中在状态空间的小范围内(即状态分布不确定性较小)时采用少量的粒子数，反之采用较多的粒子。但是，KLD APF 的缺点是计算负荷过高；L-APF 的优点是实现简单，缺点是权值方差对确定粒子数影响很大，而且还会增强粒子间的相关性，增加了高速并行实现的难度。

(2) RTPF。通常采用三种方法，即减少粒子集中的粒子数、丢弃数据或组合数据。第一种方法可能会因为粒子数的不足而导致滤波发散；第二种方法在状态剧变时会因为丢失有用数据而导致滤波发散；第三种方法需要对传感器数据作特殊的假定。此外，还可以将 APF 和 RTPF 相结合，得到自适应 RTPF。

(3) 边缘化粒子滤波(marginalized particle filter，MPF)[16]。某些非线性系统可以看做是一个含有线性子结构的状态空间模型，在动态地剔除出状态变量的线性部分后，用 Kalman 滤波器处理线性部分，用粒子滤波处理非线性部分。MPF 有两大优点：第一，对于线性变量的状态估计，Kalman 滤波器是最优的；第二，由于减少了状态变量的维数，所以，较好地克服了粒子滤波的退化现象，并且算法的计算复杂度也得到了降低。类似的还有 Rao-Blackwellised 粒子滤波(RBPF)[17]。

4. 粒子滤波的收敛性问题

虽然当前出现了不少关于粒子滤波的文献，但很少涉及对粒子滤波算法的严格证明，未能从数学上解决算法收敛性的证明问题。各种文献的仿真和实验数据从各个侧面验证或证明了粒子滤波算法的有效性，但仍缺乏说明粒子滤波算法性

能的综合指标体系。若能有效解决收敛性问题，则对粒子退化现象的抑制将有很大帮助。但是，由于粒子间的交互作用使得统计独立的假设不再成立，所以，粒子滤波算法收敛性的分析要复杂得多，这是继续研究粒子滤波算法的一大瓶颈，也是未来研究的重要方向。

5. 粒子滤波算法的硬件实现问题

为了提高粒子滤波算法的运算速度和鲁棒性，研究粒子滤波的硬件实现方法尤为关键。粒子滤波硬件实现的基本思想是[18]：将粒子滤波划分为初始采样、重采样、状态更新等不同过程，利用流水线实现分时并行处理。但迄今为止，尚未研究出成功实用的粒子滤波算法器件。

目前，粒子滤波算法在国外发展得很快，并取得了许多研究成果。国内许多学者也开始积极探索粒子滤波的各种改进算法，并成功地将粒子滤波算法应用到人脸识别、语音增强、目标跟踪、故障诊断等诸多领域，促进了国内粒子滤波算法理论和应用研究的发展。

第 2 章　Kalman 滤波理论

2.1　标准 Kalman 滤波算法

所谓最优滤波或最优估计，是指在最小方差意义下的最优滤波或估计，即要求信号或状态的最优估计值应与相应的真实值的误差方差最小。经典最优滤波理论包括 Wiener 滤波理论和 Kalman 滤波理论，前者采用频域方法，后者采用时域状态空间方法。

Wiener 滤波理论在方法论上采用频域法，局限于处理平稳随机过程，它利用谱分解和平稳随机过程的谱展式解决最优滤波问题，所得到的 Wiener 滤波器是物理上不可实现的。Kalman 突破了经典 Wiener 滤波理论和方法的局限性，提出了时域的状态空间方法，将信号视为状态或状态的分量。状态空间方法的基本特征是：①利用状态方程描写动态系统，利用量测方程提供状态的量测信息；②将状态视为抽象空间中的"点"，从而利用 Hilbert 空间中的影射理论解决最优状态估计问题。Kalman 滤波给出了一套在计算机上容易实时实现的最优递推滤波算法，适合处理多变量系统、时变系统和非平稳随机过程。

Kalman 滤波假定在每一时刻后验概率密度是高斯型的，因此，全部参数为均值和协方差。假设线性系统的离散模型为

$$\boldsymbol{x}_k = \boldsymbol{F}_k \cdot \boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1} \tag{2.1}$$

$$\boldsymbol{z}_k = \boldsymbol{H}_k \cdot \boldsymbol{x}_k + \boldsymbol{v}_k \tag{2.2}$$

式中，$\boldsymbol{F}_k$ 和 $\boldsymbol{H}_k$ 是已知的系统矩阵和测量矩阵；$\boldsymbol{w}_{k-1}$ 和 $\boldsymbol{v}_k$ 分别是均值为零、方差为 $\boldsymbol{Q}_{k-1}$ 和 $\boldsymbol{R}_k$ 且相互独立的过程噪声和量测噪声。初值 $\hat{\boldsymbol{X}}(0)$ 即 $\hat{\boldsymbol{X}}_{0|0}$，一般取 $\hat{\boldsymbol{X}}_{0|0}=E(\boldsymbol{X}_0)=0$，$\boldsymbol{P}(0)$ 可任意假定。

Kalman 滤波是典型的最小方差(MMSE)估计方法。最优 Kalman 滤波问题就是给定量测序列 $z_1, z_2, \cdots, z_{k+1}$，要求找出状态 $\boldsymbol{x}_{k+1}$ 的最优线性估计 $\hat{\boldsymbol{x}}_{k+1|k+1}$，使得估计误差 $\tilde{\boldsymbol{x}}_{k+1|k+1} = \boldsymbol{x}_{k+1} - \hat{\boldsymbol{x}}_{k+1|k+1}$ 的方差最小。

Kalman 滤波算法阐述如下：Kalman 滤波是一个带回馈的估计方法，滤波器先做出相应的估计，然后以含有噪声的量测的形式获得反馈。因此，Kalman 滤波分为两个阶段：时间更新(time update，又称预测)和量测更新(measurement update，又称修正)。预测阶段利用当前的状态及其协方差估计得出下一个时刻这两

个量的先验估计;修正阶段负责处理反馈,将最新获得的量测和先验估计合并以获得改善的后验估计。

Kalman 滤波的时间更新方程如下:

$$\hat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{F}_{k+1}\hat{\boldsymbol{x}}_{k|k} \tag{2.3}$$

$$\boldsymbol{P}_{k+1|k} = \boldsymbol{F}_k\boldsymbol{P}_{k|k}\boldsymbol{F}_k^{\mathrm{T}} + \boldsymbol{Q}_{k+1} \tag{2.4}$$

量测更新方程如下:

$$\boldsymbol{K}_{k+1} = \boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}}(\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}\boldsymbol{H}_{k+1}^{\mathrm{T}} + \boldsymbol{R}_{k+1})^{-1} \tag{2.5}$$

$$\hat{\boldsymbol{x}}_{k+1|k+1} = \hat{\boldsymbol{x}}_{k+1|k} + \boldsymbol{K}_{k+1}(z_{k+1} - \boldsymbol{H}_{k+1}\boldsymbol{x}_{k+1|k}) \tag{2.6}$$

$$\boldsymbol{P}_{k+1|k+1} = (\boldsymbol{I} - \boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1})\boldsymbol{P}_{k+1|k} \tag{2.7}$$

在 k 时刻的预测过程计算系统状态及其协方差矩阵的先验值,在修正阶段计算 Kalman 增益矩阵 $\boldsymbol{K}_k$,然后获得系统的状态及其协方差的后验值。式(2.3)~式(2.7)构成了 Kalman 滤波的核心方程。式(2.7)另一种常见的形式为

$$\boldsymbol{P}_{k+1|k+1} = \boldsymbol{P}_{k+1|k} - \boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1}\boldsymbol{K}_{k+1}^{\mathrm{T}} \tag{2.8}$$

可将计算 $\boldsymbol{P}_{k+1|k}$ 的公式展开为如下的形式:

$$\boldsymbol{P}_{k+1|k} = \boldsymbol{F}_k[\boldsymbol{P}_{k|k-1} - \boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}(\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k)^{-1}\boldsymbol{H}_k\boldsymbol{P}_{k|k-1}]\boldsymbol{F}_k^{\mathrm{T}} + \boldsymbol{Q}_{k+1} \tag{2.9}$$

该方程被称为 Ricatti 方程。当系统是时不变的($\boldsymbol{F}$、$\boldsymbol{H}$ 为常矩阵),且噪声是平稳的($\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 是常矩阵)时,则随着 $k \to +\infty$,该方程的解收敛到一个正定矩阵 $\bar{\boldsymbol{P}}$,这个解可以作为判断 Kalman 滤波是否稳态的一个标准。

对于线性的、正态的系统状态估计问题,期望的概率密度函数仍然是正态分布,它的分布特性可以用均值和方差来描述,Kalman 滤波很好地解决了这类估计问题。但是,Kalman 方法也有其固有的缺点,它要求系统的过程与量测噪声为高斯白噪声,且相互独立,否则,Kalman 滤波可能会出现发散。针对 Kalman 滤波的不足,学者们提出了许多改进方法,如采用自适应算法调整 Kalman 滤波器中的过程噪声协方差矩阵 $\boldsymbol{Q}_k$;基于进化程序求解最优 Kalman 滤波器;考虑状态之间的约束关系,通过将非约束的 Kalman 解投影到状态约束表面,从而极大地提高了滤波器的预测精度;基于 H_2 和 H_∞ 思想的 Kalman 滤波器的鲁棒性十分适合解决不确定问题。值得一提的是,这些改进的 Kalman 滤波算法一般都会增加计算量,降低了算法的效率。

2.2　α-β-γ 滤波器

α-β-γ 滤波器的最大优点在于其增益矩阵可以离线计算,而且计算量相对于 Kalman 滤波器来说非常小。α-β-γ 滤波器实际上是 Kalman 滤波器的稳态解形

式，参数 α、β、γ 与 Kalman 滤波的稳态增益矩阵、稳态预测协方差矩阵、状态噪声方差和量测噪声方差之间存在着某种对应关系。

α-β-γ 滤波器的状态方程和量测方程与 Kalman 滤波器相同，其滤波方程为

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k-1}) \tag{2.10}$$

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}_{k|k-1}\hat{\boldsymbol{x}}_{k-1|k-1} \tag{2.11}$$

$$\boldsymbol{K}_k = \begin{bmatrix} \alpha & \dfrac{\beta}{T} & \dfrac{\gamma}{T^2} \end{bmatrix} \tag{2.12}$$

Kalman 滤波的稳态增益矩阵 $\boldsymbol{K}=[k_1 \quad k_2 \quad k_3]$与 α、β、γ 的对应关系为

$$k_1 = \alpha, \quad k_2 = \frac{\beta}{T}, \quad k_3 = \frac{\gamma}{T^2} \tag{2.13}$$

假设 Kalman 滤波的稳态对称预测协方差矩阵 $\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{23} & p_{33} \end{bmatrix}$，则 $\boldsymbol{P}$ 与 α、β、γ 的对应关系为

$$\alpha = \frac{p_{11}}{p_{11}+r}, \quad \beta = \frac{Tp_{12}}{p_{11}+r}, \quad \gamma = \frac{T^2 p_{13}}{p_{11}+r} \tag{2.14}$$

Kalman 滤波中的状态噪声方差 q 和量测噪声方差 r 与 α、β、γ 的对应关系为

$$\alpha = 2\left(\frac{q}{r}\right), \quad \beta = 2T\left(\frac{q}{r}\right)^{1/2}, \quad \gamma = T^2\left(\frac{q}{r}\right)^{1/2} \tag{2.15}$$

需要特别指出的是，α-β-γ 滤波器是常增益滤波器，适用于稳态且增益值比较小的情况；而在滤波开始阶段属瞬态过程，滤波性能较差，估计误差方差比较大。通过在滤波过程中修正 α、β、γ 的数值，可以提高滤波性能。

2.3 EKF 滤波算法

由于 Kalman 滤波是在线性高斯情况下利用最小均方误差准则获得系统状态的动态估计，而在实际应用过程中，许多情况下量测数据与系统动态参数间的关系是非线性的。对于非线性滤波问题，至今尚未得到完善的解法。通常的处理方法是利用线性化技巧将非线性问题转化为一个近似的线性滤波问题，套用线性滤波理论得到求解原非线性滤波问题的次优滤波算法。

2.3.1 非线性贝叶斯滤波

考虑混合系统估计问题，非线性系统的离散时间状态系统方程和量测方程分

别为

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}, \boldsymbol{w}_{k-1}) \tag{2.16}$$

$$\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{v}_k) \tag{2.17}$$

式中，$\boldsymbol{x}_k \in R^n$ 表示 k 时刻的系统状态向量；$\boldsymbol{z}_k \in R^n$ 是 k 时刻含有加性噪声的量测矢量；$\boldsymbol{w}_{k-1}$是过程噪声；$\boldsymbol{v}_k$ 为量测噪声；$\boldsymbol{w}_{k-1}$和$\boldsymbol{v}_k$ 是相互独立、协方差分别为 $\boldsymbol{Q}_{k-1}$ 和 $\boldsymbol{R}_k$ 的零均值加性噪声序列；$\boldsymbol{u}_k$ 表示已知输入量。

在式(2.16)和式(2.17)描述的系统中，假设系统的状态服从一阶 Markov 过程，即 $p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, \boldsymbol{x}_{k-2}, \cdots, \boldsymbol{x}_0) = p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1})$，且状态独立给出量测值 $p(\boldsymbol{z}_k | \boldsymbol{x}_k)$。根据贝叶斯学派的观点，状态估计就是在给定量测数据集合 $\boldsymbol{z}_{1:k} = \{\boldsymbol{z}_0, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_k\}$ 的条件下估算状态向量 $\boldsymbol{x}_k$ 的值，即估计后验概率密度函数 $p(\boldsymbol{x}_k | \boldsymbol{z}_{1:k})$。若假定初始先验概率密度函数 $p(\boldsymbol{x}_0 | \boldsymbol{z}_0) \equiv p(\boldsymbol{x}_0)$是已知的($\boldsymbol{x}_0$ 表示初始状态向量，$\boldsymbol{z}_0$ 表示没有量测值)，则从原则上讲，通过预测和更新就可以采用递归的方式估计出后验概率密度函数 $p(\boldsymbol{x}_k | \boldsymbol{z}_{1:k})$。

假定在 $k-1$ 时刻概率密度函数 $p(\boldsymbol{x}_{k-1} | \boldsymbol{z}_{1:k-1})$已知，由于系统(2.16)满足一阶 Markov 过程，即 $p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1}, \boldsymbol{z}_{1:k-1}) = p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1})$，因此，就可以预测 k 时刻的先验概率密度为

$$p(\boldsymbol{x}_k \mid \boldsymbol{z}_{1:k-1}) = \int p(\boldsymbol{x}_k \mid \boldsymbol{x}_{k-1}) \cdot p(\boldsymbol{x}_{k-1} \mid \boldsymbol{z}_{1:k-1}) \mathrm{d}\boldsymbol{x}_{k-1} \tag{2.18}$$

由系统(2.16)和统计值$\boldsymbol{v}_{k-1}$可以确定状态演化的概率密度 $p(\boldsymbol{x}_k | \boldsymbol{x}_{k-1})$。在时刻 k 获得量测值 $\boldsymbol{z}_k$ 后，利用贝叶斯规则更新后验概率，得到

$$p(\boldsymbol{x}_k \mid \boldsymbol{z}_{1:k}) = \frac{p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) \cdot p(\boldsymbol{x}_k \mid \boldsymbol{z}_{1:k-1})}{p(\boldsymbol{z}_k \mid \boldsymbol{z}_{1:k-1})} \tag{2.19}$$

常数 $p(\boldsymbol{z}_k \mid \boldsymbol{z}_{1:k-1}) = \int p(\boldsymbol{z}_k \mid \boldsymbol{x}_k) \cdot p(\boldsymbol{x}_k \mid \boldsymbol{z}_{1:k-1}) \mathrm{d}\boldsymbol{x}_k$ 取决于由量测模型(2.17)和统计值 $\boldsymbol{w}_k$ 所定义的似然函数 $p(\boldsymbol{z}_k | \boldsymbol{x}_k)$。在更新公式(2.19)中，测量值 $\boldsymbol{z}_k$ 被用来修正先验概率密度，以获取当前状态的后验概率密度函数。式(2.18)和式(2.19)是最优贝叶斯估计的一般概念表达式，通常不可能对其进行精确解析。在满足一定条件时，可以得到最优贝叶斯解。但如果条件不满足，可以利用 EKF 等滤波方法获得次优贝叶斯解。

2.3.2　离散型 EKF 滤波算法

考虑如下离散随机非线性系统的数学模型：

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1) + \boldsymbol{\Gamma}_{k-1}\boldsymbol{W}_{k-1} \\ \boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_k, k) + \boldsymbol{V}_k \end{cases} \tag{2.20}$$

假设在 k 时刻得到系统的滤波值为 $\boldsymbol{x}_{k-1|k-1}$ 和 $\boldsymbol{x}_{k|k-1}$，将离散系统的状态方程在 $\hat{\boldsymbol{x}}_{k-1|k-1}$ 附近按泰勒级数展开，取其前两项，可得

$$\begin{aligned}\boldsymbol{x}_k &= \boldsymbol{f}(\hat{\boldsymbol{x}}_{k-1|k-1}, k-1) \\ &\quad + \frac{\partial \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1)}{\partial \boldsymbol{x}_{k-1}}\bigg|_{\boldsymbol{x}_{k-1}=\boldsymbol{x}_{k-1|k-1}} (\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}) \\ &\quad + \boldsymbol{\Gamma}(\hat{\boldsymbol{x}}_{k-1|k-1}, k-1)\boldsymbol{W}_{k-1}\end{aligned} \tag{2.21}$$

$$\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_{k|k-1}, k) + \frac{\partial \boldsymbol{h}(\boldsymbol{x}_k, k-1)}{\partial \boldsymbol{x}_k}\bigg|_{\boldsymbol{x}_k=\boldsymbol{x}_{k|k-1}} (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{V}_k \tag{2.22}$$

令

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1)}{\partial \boldsymbol{x}_{k-1}}\bigg|_{\boldsymbol{x}_{k-1}=\boldsymbol{x}_{k-1|k-1}} = \boldsymbol{F}_k$$

$$\frac{\partial \boldsymbol{h}(\boldsymbol{x}_k, k-1)}{\partial \boldsymbol{x}_k}\bigg|_{\boldsymbol{x}_k=\boldsymbol{x}_{k|k-1}} = \boldsymbol{H}_k$$

$$\boldsymbol{f}(\hat{\boldsymbol{x}}_{k-1|k-1}, k-1) - \frac{\partial \boldsymbol{f}(\boldsymbol{x}_{k-1}, k-1)}{\partial \boldsymbol{x}_{k-1}}\bigg|_{\boldsymbol{x}_{k-1}=\boldsymbol{x}_{k-1|k-1}} \hat{\boldsymbol{x}}_{k-1|k-1} = \boldsymbol{U}_{k-1}$$

$$\boldsymbol{h}(\boldsymbol{x}_{k|k-1}, k) - \frac{\partial \boldsymbol{h}(\boldsymbol{x}_k, k-1)}{\partial \boldsymbol{x}_k}\bigg|_{\boldsymbol{x}_k=\boldsymbol{x}_{k|k-1}} \hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{Y}_k$$

则得到原非线性系统的线性化模型为

$$\begin{cases}\boldsymbol{x}_k = \boldsymbol{F}_k \boldsymbol{x}_{k-1} + \boldsymbol{U}_{k-1} + \boldsymbol{\Gamma}_{k-1}\boldsymbol{W}_{k-1} \\ \boldsymbol{z}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{Y}_k + \boldsymbol{V}_k\end{cases} \tag{2.23}$$

当 $\boldsymbol{W}_k$ 与 $\boldsymbol{V}_k$ 互不相关，且一步预测误差 $\boldsymbol{x}_{k-1} - \hat{\boldsymbol{x}}_{k-1|k-1}$ 和滤波误差 $\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}$ 都较小时，k 时刻非线性离散系统的 EKF 滤波方程如下所示。

预测方程：

$$\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{F}_k \hat{\boldsymbol{x}}_{k-1} \tag{2.24}$$

预测协方程：

$$\boldsymbol{Q}_{k|k-1} = \boldsymbol{F}_k \boldsymbol{Q}_{k|k-1} \boldsymbol{F}_k^{\mathrm{T}} \tag{2.25}$$

Kalman 增益：

$$\boldsymbol{K}_k = \boldsymbol{Q}_{k|k-1} \boldsymbol{H}_k^{\mathrm{T}} [\boldsymbol{H}_k \boldsymbol{Q}_{k|k-1} \boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k]^{-1} \tag{2.26}$$

滤波方程：

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k [\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k-1}] \tag{2.27}$$

滤波协方差：

$$\boldsymbol{Q}_k = [\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k] \boldsymbol{Q}_{k|k-1} [\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k]^{\mathrm{T}} + \boldsymbol{K}_k \boldsymbol{R}_k \boldsymbol{K}_k^{\mathrm{T}} \tag{2.28}$$

只要通过量测或者先验信息就可以确定初始估计值 $\boldsymbol{x}_0$、$\boldsymbol{Q}_0$，通过上述方程即可递推求出系统的状态解。

由于 EKF 中的预测值可能存在一定误差，当这种误差经非线性函数放大后可能会导致 EKF 滤波的结果变坏甚至发散，因此，也可以采用多次迭代运算的 IEKF 方法。IEKF 方法是在 EKF 滤波过程中，假定已经获得第 k 时刻的预测值 $\hat{\boldsymbol{x}}_{k|k-1}$ 和预测协方差 $\boldsymbol{Q}_{k|k-1}$ 以后，算法公式及流程如下：

$$\hat{\boldsymbol{x}}_{k|k}^{0} = \hat{\boldsymbol{x}}_{k|k-1}, \quad \boldsymbol{Q}_{k|k}^{0} = \boldsymbol{Q}_{k|k-1}$$

For $m=0$ to $m-1$

$$\boldsymbol{H}_k^m = \left.\frac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k|k}^m}$$

$$\boldsymbol{K}_k^m = \boldsymbol{Q}_{k|k-1}(\boldsymbol{H}_k^m)^{\mathrm{T}}[\boldsymbol{H}_k^m\boldsymbol{Q}_{k|k-1}(\boldsymbol{H}_k^m)^{\mathrm{T}} + \boldsymbol{R}_k]^{-1}$$

$$\boldsymbol{x}_{k|k}^{m+1} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k^m[\boldsymbol{z}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_{k|k}^m) - \boldsymbol{H}_k^m(\hat{\boldsymbol{x}}_{k|k-1} - \hat{\boldsymbol{x}}_{k|k}^m)]$$

$$\boldsymbol{Q}_{k|k}^m = (\boldsymbol{I} - \boldsymbol{K}_k^m\boldsymbol{H}_k^m)\boldsymbol{Q}_{k|k-1}(\boldsymbol{I} - \boldsymbol{K}_k^m\boldsymbol{H}_k^m)^{\mathrm{T}} + \boldsymbol{K}_k^m\boldsymbol{R}_k(\boldsymbol{K}_k^m)^{\mathrm{T}}$$

End

当 $m=1$ 时，IEKF 就退化为 EKF 方法。IEKF 方法的物理意义在于其利用多次迭代方法来更新状态估计以逼近非线性量测 $\boldsymbol{z}_k$，使得状态估计误差进一步减小，从而达到优化滤波估计的目的。注意 IEKF 方法虽然提高了估计精度，但是，它是以增加滤波的计算量为代价的。

EKF 是一种比较常用的非线性滤波方法。在这种滤波方法中，非线性因子的存在对滤波稳定性和状态估计精度都有很大的影响，其滤波结果的好坏与过程噪声和量测噪声的统计特性也有很大关系。由于 EKF 中预先估计的过程噪声协方差 $\boldsymbol{Q}_k$ 和量测噪声协方差 $\boldsymbol{R}_k$ 在滤波过程中一直保持不变，如果这两个噪声的协方差不太准确的话，在滤波过程中就容易产生误差积累，导致滤波发散；而且对于维数较大的非线性系统，估计的过程噪声协方差矩阵和量测噪声协方差矩阵容易出现异常现象，即 $\boldsymbol{Q}_k$ 失去半正定性，$\boldsymbol{R}_k$ 失去正定性，也容易导致滤波发散。只有当系统的动态模型和量测模型都接近线性时，也就是线性化模型误差较小时，EKF 的滤波结果才有可能接近真实值。此外，EKF 还有一个缺点，就是状态的初始值很难确定，如果假设的状态初始值和初始协方差误差较大的话，也容易导致滤波发散。

2.4 MVEKF 算法

根据前面所述的 EKF 方法的优缺点，通过修改滤波的协方差可以达到改善 EKF 滤波性能的目的，因而出现了一种修正协方差的 MVEKF 方法。

对于高斯白噪声条件下线性系统的 Kalman 滤波，根据正交投影定理可以得到

$$\boldsymbol{H}_k\boldsymbol{Q}_k = 0 \tag{2.29}$$

在实际的 EKF 中，由于量测方程往往也是非线性的，且预测值 $\hat{\boldsymbol{x}}_{k|k-1}$ 存在一定的偏差，导致在 $\hat{\boldsymbol{x}}_{k|k-1}$ 处所求得的 Jacobi 矩阵也存在偏差，使得

$$\boldsymbol{H}_k\boldsymbol{Q}_k \neq 0 \tag{2.30}$$

MVEKF 的基本思想是在 EKF 方程中按照式(2.24)～式(2.28)计算出状态滤波值 $\hat{\boldsymbol{x}}_{k|k}$ 之后，采用状态滤波值 $\hat{\boldsymbol{x}}_{k|k}$ 重新计算 Jacobi 矩阵，即

$$\boldsymbol{H}_k^+ = \left.\frac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{x=\hat{x}_{k|k}} \tag{2.31}$$

对协方差矩阵进行更新，从而得到更加准确的协方差矩阵，使得

$$\boldsymbol{H}_k^+\boldsymbol{Q}_k \approx 0 \tag{2.32}$$

实际上，相当于直接用滤波值计算 $\boldsymbol{H}_k^+ = \left.\frac{\partial \boldsymbol{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{x=\hat{x}_{k|k}}$ 后，重新计算一次增益和协方差矩阵。由于实际上对下一次迭代有贡献的是按照这种方法修正的协方差矩阵 $\boldsymbol{Q}_k$，这也就是将这种方法称之为“修正协方差”EKF 的原因。

$$\boldsymbol{K}_k = \boldsymbol{Q}_{k|k-1}\boldsymbol{H}_k^{+\mathrm{T}}[\boldsymbol{H}_k^+\boldsymbol{Q}_{k|k-1}\boldsymbol{H}_k^{+\mathrm{T}} + \boldsymbol{R}_k]^{-1} \tag{2.33}$$

$$\boldsymbol{Q}_k = [\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k^+]\boldsymbol{Q}_{k|k-1}[\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{H}_k^+]^{\mathrm{T}} + \boldsymbol{K}_k\boldsymbol{R}_k\boldsymbol{K}_k^{\mathrm{T}} \tag{2.34}$$

将式(2.34)计算的 $\boldsymbol{Q}_k$ 和按照式(2.33)计算的滤波协方差再进行迭代滤波，由此得到的协方差矩阵隐含了本次测量值 z_i 的信息，从而改善了估计的性能。

仔细比较迭代次数 $k=2$ 时的 IEKF 算法和 MVEKF 算法可以看出，实际上 MVEKF 比前者只是少了一次更新状态估计。

2.5 UKF 算 法

另外一种比较常用的次优非线性贝叶斯滤波方法是 UKF，该方法是在无味变换的基础上发展起来的。无味变换的基本思想是由 Juiler 等首先提出来的。无味变换是用于计算经过非线性变换的随机变量统计的一种新方法，其不需要对非线性状态和量测模型进行线性化，而是对状态向量的概率密度函数进行近似化，近似化后的概率密度函数仍然是高斯的，但它表现为一系列选取好的采样点。

2.5.1 无味变换

无味变换用固定数量的参数去近似一个高斯分布，这样做比近似非线性函数的线性变换更容易。其实现原理为：在原先状态分布中按某一规则取一些点，使这

些点的均值和协方差等于原状态分布的均值和协方差；将这些点代入非线性函数中，相应得到非线性函数点集，通过该点集求取变换后的均值和协方差。由于这样得到的函数值没有经过线性化，没有忽略其高阶项，因而由此得到的均值和协方差的估计比 EKF 方法要精确。

假设 n_x 维状态向量 $\boldsymbol{x}$ 的统计特性为：均值为 $\bar{\boldsymbol{x}}$，方差为 $\boldsymbol{P}_x$，$\boldsymbol{x}$ 通过任意一个非线性函数 $\boldsymbol{f}:R^{n_x}\to R^{n_y}$ 变换得到 n_y 维变量 $\boldsymbol{y}$：$\boldsymbol{y}=\boldsymbol{f}(\boldsymbol{x})$，$\boldsymbol{x}$ 的统计特性通过非线性函数 $\boldsymbol{f}(\cdot)$ 进行传播，得到 $\boldsymbol{y}$ 的统计特性 $\bar{\boldsymbol{y}}$ 和 $\boldsymbol{P}_y$。

无味变换的基本思想是：根据 $\boldsymbol{x}$ 的均值 $\bar{\boldsymbol{x}}$ 和方差 $\boldsymbol{P}_x$，选择 $2n_x+1$ 个加权样点 $\boldsymbol{S}_i=\{\boldsymbol{W}_i,\boldsymbol{x}_i\}(i=1,2,\cdots,2n_x+1)$ 来近似随机变量 $\boldsymbol{x}$ 的分布，称 $\boldsymbol{x}_i$ 为 σ 点（粒子）；基于设定的粒子 $\boldsymbol{x}_i$ 计算其经过 $\boldsymbol{f}(\cdot)$ 的传播 γ_i；然后基于 γ_i 计算 $\bar{\boldsymbol{y}}$ 和 $\boldsymbol{P}_y$。

无味变换的具体过程描述如下：

$$\boldsymbol{x}_0=\bar{\boldsymbol{x}},\quad W_0=\frac{\lambda}{n_x+\lambda},\quad i=0 \tag{2.35}$$

$$\boldsymbol{x}_i=\bar{\boldsymbol{x}}+\left(\sqrt{(n_x+\lambda)\boldsymbol{P}_x}\right)_i,\quad i=1,2,\cdots,n_x \tag{2.36}$$

$$\boldsymbol{x}_i=\bar{\boldsymbol{x}}-\left(\sqrt{(n_x+\lambda)\boldsymbol{P}_x}\right)_i,\quad i=n_x+1,n_x+2,\cdots,2n_x \tag{2.37}$$

$$W_0^{(m)}=\frac{\lambda}{n_x+\lambda},\quad W_0^{(c)}=\frac{\lambda}{n_x+\lambda}+(1-\alpha^2+\beta),\quad \lambda=\alpha^2(n+k)-n \tag{2.38}$$

$$W_i^{(m)}=W_i^{(c)}=\frac{1}{2(n_x+k)},\quad i=1,2,\cdots,2n_x \tag{2.39}$$

式中，$\alpha>0$ 是一个比例因子，它可以调节粒子的分布距离，降低高阶矩的影响，减小预测误差，一般取小的正值，如 0.01；$\beta\geqslant 0$ 的作用是改变 $W_0^{(c)}$，调节 β 的数值可以提高方差的精度，控制估计状态的峰值误差。α 和 β 的值随 $\boldsymbol{x}$ 分布的不同而不同，如果 $\boldsymbol{x}$ 服从正态分布，则一般取 $n_x+k=3$。$\left(\sqrt{(n_x+k)\boldsymbol{P}_x}\right)_i$ 是矩阵 $(n_x+k)\boldsymbol{P}_x$ 的均方根的第 i 行（列），可以利用 QR 分解或 Cholesky 分解得到矩阵 $(n_x+k)\boldsymbol{P}_x$ 的均方根。$W_i^{(m)}$ 是求一阶统计特性时的权系数，$W_i^{(c)}$ 是求二阶统计特性时的权系数。

变换过程如下。

(1) 选定参数 k、α 和 β 的数值。

(2) 按照式(2.35)～式(2.39)计算得到 $2n_x+1$ 个调整后的粒子及其权值。

(3) 对每个粒子点进行非线性变换，形成变换后的点集 $\boldsymbol{y}_i=\boldsymbol{f}(\boldsymbol{x}_i)$，$i=1,2,\cdots,2n_x$。

(4) 变换后的点集的均值 $\bar{\boldsymbol{y}}$ 和方差 $\boldsymbol{P}_y$ 由下式计算：

$$\bar{\boldsymbol{y}}=\sum_{i=0}^{2n_x}W_i^{(m)}\boldsymbol{y}_i,\quad \boldsymbol{P}_y=\sum_{i=0}^{2n_x}W_i^{(c)}(\boldsymbol{y}_i-\bar{\boldsymbol{y}})(\boldsymbol{y}_i-\bar{\boldsymbol{y}})^{\mathrm{T}} \tag{2.40}$$

由于无味变换得到的函数值没有经过线性化，没有忽略其高阶项，同时因为避

免了 Jacobi 矩阵(线性化)的计算,因而由此得到的均值和协方差的估计比 EKF 方法要精确。该算法对于 $\boldsymbol{x}$ 均值和协方差的计算精确到真实后验分布的二阶矩,而且误差可以通过 k 来调节,而 EKF 只是非线性函数的一阶近似,因此,该算法具有比 EKF 更高的精度。如果已知 $\boldsymbol{x}$ 概率密度分布的形状,可以通过将 β 设为一个非零值来减小 4 阶项以上带来的误差。

2.5.2 算法描述

在无味变换基础上建立起来的 UKF 是 1996 年由剑桥大学的 Julier 首次提出来的,UKF 是无味变换和标准 Kalman 滤波体系的结合。与 EKF 不同,UKF 是通过上述无味变换使非线性系统方程适用于线性假设下的标准 Kalman 滤波体系,而不是像 EKF 那样通过线性化非线性函数实现递推滤波。由于 UKF 不需要求导,它比 EKF 能更好地逼近状态方程的非线性特性,具有更高的估计精度,计算量却与 EKF 同阶,因而获得了广泛关注。

假设系统如式(2.16)和式(2.17)表示,过程噪声和量测噪声均为不相关的零均值高斯白噪声,其协方差分别为 $\boldsymbol{Q}_k$ 和 $\boldsymbol{R}_k$,则 UKF 滤波算法如下所述。

(1) 初始化。初始状态 $\boldsymbol{x}_0$ 的统计特性为:$E[\boldsymbol{x}_0]=\bar{\boldsymbol{x}}_0$,$\mathrm{var}\,[\boldsymbol{x}_0]=E[(\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0)\cdot(\boldsymbol{x}_0-\bar{\boldsymbol{x}}_0)^{\mathrm{T}}]=\boldsymbol{P}_0$,且 $E[\boldsymbol{x}_0,\boldsymbol{w}_k]=0$,$E[\boldsymbol{x}_0,\boldsymbol{v}_k]=0$。考虑噪声,扩展后的初始状态向量及其方差为

$$\boldsymbol{x}_0^a=[\bar{\boldsymbol{x}}^{\mathrm{T}}\quad 0\quad 0]^{\mathrm{T}},\quad \boldsymbol{P}_0^a=E[(\boldsymbol{x}_0^a-\bar{\boldsymbol{x}}_0^a)(\boldsymbol{x}_0^a-\bar{\boldsymbol{x}}_0^a)^{\mathrm{T}}]=\begin{bmatrix}\boldsymbol{P}_0 & 0 & 0\\ 0 & \boldsymbol{Q} & 0\\ 0 & 0 & \boldsymbol{R}\end{bmatrix}$$

(2) 系统的扩展状态向量表示为

$$\boldsymbol{x}_k^a=[\boldsymbol{x}_k^{\mathrm{T}}\quad \boldsymbol{w}_k^{\mathrm{T}}\quad \boldsymbol{v}_k^{\mathrm{T}}]^{\mathrm{T}},\quad \boldsymbol{P}_k^a=E[(\boldsymbol{x}_k^a-\bar{\boldsymbol{x}}_k^a)(\boldsymbol{x}_k^a-\bar{\boldsymbol{x}}_k^a)^{\mathrm{T}}]=\begin{bmatrix}\boldsymbol{P}_k & 0 & 0\\ 0 & \boldsymbol{Q}_k & 0\\ 0 & 0 & \boldsymbol{R}_k\end{bmatrix} \tag{2.41}$$

选取粒子

$$\boldsymbol{x}_{k-1}^a=[\bar{\boldsymbol{x}}_{k-1}^a\quad \bar{\boldsymbol{x}}_{k-1}^a+\sqrt{(n_a+\lambda)\boldsymbol{P}_{k-1}^a}\quad \bar{\boldsymbol{x}}_{k-1}^a-\sqrt{(n_a+\lambda)\boldsymbol{P}_{k-1}^a}] \tag{2.42}$$

且

$$\boldsymbol{x}_{k-1}^a=[\boldsymbol{x}_{k-1}^x,\boldsymbol{x}_{k-1}^w,\boldsymbol{x}_{k-1}^v]^{\mathrm{T}} \tag{2.43}$$

式中,$n_a=n_x+n_w+n_v$,n_x、n_w 和 n_v 分别为系统状态向量、过程噪声和量测噪声的维数;$\boldsymbol{x}_{k-1}^x$、$\boldsymbol{x}_{k-1}^w$ 和 $\boldsymbol{x}_{k-1}^v$ 分别为 $\boldsymbol{x}_{k-1}^a$ 中对应于状态向量、过程噪声和量测噪声的分量。

(3) 时间更新。若不考虑有输入作用,由式(2.38)、式(2.39)可计算得权值

W_i，则有

$$\boldsymbol{x}_{k|k-1}^{x} = \boldsymbol{f}(\boldsymbol{x}_{k-1}^{x}, \boldsymbol{u}_{k-1}, \boldsymbol{x}_{k-1}^{w}) \tag{2.44}$$

$$\bar{\boldsymbol{x}}_{k|k-1} = \sum_{i=0}^{2n_a} W_i^{(m)} \boldsymbol{x}_{i,k|k-1}^{x} \tag{2.45}$$

$$\boldsymbol{P}_{k|k-1} = \sum_{i=0}^{2n_a} W_i^{(c)} [\boldsymbol{x}_{i,k|k-1}^{x} - \bar{\boldsymbol{x}}_{k|k-1}][\boldsymbol{x}_{i,k|k-1}^{x} - \bar{\boldsymbol{x}}_{k|k-1}]^{\mathrm{T}} \tag{2.46}$$

$$\boldsymbol{z}_{k|k-1} = \boldsymbol{h}(\boldsymbol{x}_{k|k-1}^{x}, \boldsymbol{x}_{k|k-1}^{v}) \tag{2.47}$$

$$\bar{\boldsymbol{z}}_{k|k-1} = \sum_{i=0}^{2n_a} W_i^{(m)} \boldsymbol{z}_{i,k|k-1} \tag{2.48}$$

式中，$\bar{\boldsymbol{x}}_{k|k-1}$为所有粒子点的一步预测加权和。

（4）测量更新。

$$\boldsymbol{P}_{\boldsymbol{z}_{k|k-1}\boldsymbol{z}_{k|k-1}} = \sum_{i=0}^{2n_a} W_i^{(c)} [\boldsymbol{z}_{i,k|k-1} - \bar{\boldsymbol{z}}_{k|k-1}][\boldsymbol{z}_{i,k|k-1} - \bar{\boldsymbol{z}}_{k|k-1}]^{\mathrm{T}} \tag{2.49}$$

$$\boldsymbol{P}_{\boldsymbol{x}_{k|k-1}\boldsymbol{z}_{k|k-1}} = \sum_{i=0}^{2n_a} W_i^{(c)} [\boldsymbol{x}_{i,k|k-1}^{x} - \bar{\boldsymbol{x}}_{k|k-1}][\boldsymbol{z}_{i,k|k-1} - \bar{\boldsymbol{z}}_{k|k-1}]^{\mathrm{T}} \tag{2.50}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{\boldsymbol{x}_{k|k-1}\boldsymbol{z}_{k|k-1}} \boldsymbol{P}_{\boldsymbol{z}_{k|k-1}\boldsymbol{z}_{k|k-1}}^{-1} \tag{2.51}$$

$$\bar{\boldsymbol{x}}_k = \bar{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \bar{\boldsymbol{z}}_{k|k-1}) \tag{2.52}$$

$$\hat{\boldsymbol{P}}_k = \boldsymbol{P}_{k|k-1} + \boldsymbol{K}_k \boldsymbol{P}_{\boldsymbol{z}_{k|k-1}\boldsymbol{z}_{k|k-1}} \boldsymbol{K}_k^{\mathrm{T}} \tag{2.53}$$

至此，得到了 UKF 在 k 时刻的滤波状态和方差。

如果系统的状态噪声和量测噪声为加性噪声，则无需像式(2.43)那样增广状态向量，算法中的时间更新和状态更新方程得以简化。当量测方程和状态方程均为线性方程时，由 UKF 得到的滤波结果和标准的线性 Kalman 滤波得到的结果相同。

尽管 UKF 的估计精度比 EKF 高，但 UKF 不能应用于非高斯分布系统，而粒子滤波可以解决这一问题。

第 3 章　从贝叶斯理论到粒子滤波

贝叶斯方法为动态系统的估计问题提供了一类严谨的解决框架，可以得到对系统状态估计的最优解。对于高斯噪声的线性系统，Kalman 滤波很好地解决了这类估计问题；对于非线性系统的估计问题，得到广泛应用的方法以 EKF 为代表，但 EKF 仅仅利用了非线性函数泰勒展开式的一阶偏导部分（忽略高阶项），常常导致在状态的后验分布的估计上产生较大的误差，影响滤波算法的性能；且此类算法要求状态估计的概率密度函数满足高斯分布，但实际系统往往是非高斯、非线性的（即系统状态/量测模型是非线性的，模型/量测噪声不服从高斯分布），这就迫切要求研究针对非线性、非高斯系统估计的新算法。

对于非线性、非高斯系统的状态估计问题，比较有发展前途的是基于序贯蒙特卡罗方法和 SIS 的粒子滤波方法，有些文献也称之为自举滤波，该方法对系统噪声没有任何限制，它通过预测和更新来自于系统概率密度函数的采样集来近似非线性系统的随机贝叶斯估计。其基本思想是：假设在 k 时刻得到了一组描述系统状态后验概率分布 $P[\boldsymbol{x}(k)|z_k]$ 的采样值，记为 $\{\boldsymbol{x}(k,i), i=1,\cdots,N\}$，然后以此为基础，根据粒子滤波算法，同时利用系统状态方程和量测方程寻找一组采样值 $\{\boldsymbol{x}(k+1,i), i=1,\cdots,N\}$，使其近似于 $P[\boldsymbol{x}(k+1)|z_{k+1}]$。这种滤波方法的优点是：采样集中在高概率的区域，采样计算的过程也很简单，是一种非线性、非高斯系统递推贝叶斯估计。粒子滤波采用具有不同概率密度分布的采样近似，同时还突破了 Kalman 滤波算法固有的缺点——只能用于噪声为高斯密度分布的线性系统。

本章从介绍贝叶斯估计理论开始，引出适用于非线性、非高斯系统估计的粒子滤波算法。

3.1　动态空间模型

用某种适当的模型来描述一个实际物理系统，对分析、研究该物理系统非常重要。在信号处理、通信、雷达、声呐等许多实际工程应用中，经常采用动态空间模型来描述其中的许多问题。动态空间模型是一个很重要的统计分析工具，如 Kalman 滤波采用的高斯-马尔可夫线性模型就是一个很好的例子，它用系统方程来描述状态随时间演变的过程，而用量测方程来描述与状态有关的噪声变量。同样的，只要将高斯-马尔可夫线性模型写成一般的数学映射，就可以用这两个方程来描述更一般的动态系统。

时变问题分析的一般动态状态空间模型分为状态转移模型 $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$和量测模型 $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$，其中，$\boldsymbol{x}_k \in R^{n_x}$代表系统在时间 k 的状态变量(隐含变量或参数)，$\boldsymbol{z}_k \in R^{n_z}$为系统在时间 k 的量测值。对于非线性、非高斯过程，其模型可表示为

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, \boldsymbol{v}_{k-1}) \tag{3.1}$$

$$\boldsymbol{z}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{n}_k) \tag{3.2}$$

式中，$\boldsymbol{v}_k \in R^{n_v}$和 $\boldsymbol{n}_k \in R^{n_n}$ 分别为过程噪声和量测噪声，并且是相互独立、协方差分别为 $\boldsymbol{Q}_k$ 和 $\boldsymbol{R}_k$ 的零均值加性噪声序列；$\boldsymbol{f}: R^{n_x} \to R^{n_x}$，$\boldsymbol{h}: R^{n_x} \to R^{n_z}$ 为有界非线性映射。

3.2 贝叶斯估计理论

总体来说，整个统计学界大体分为两大学派：经典学派[19,20]和贝叶斯学派[21,22]。贝叶斯学派发展起步较晚，1736 年，贝叶斯提出了著名的贝叶斯定理，但一直不为人重视，直到 20 世纪 30 年代在 Ramsey 等的努力下才得到了长足发展，逐渐形成了贝叶斯学派，与经典学派共同成为现代两大统计学派。与经典学派相比，贝叶斯学派有两个特点：一是将未知参数作为随机变量；二是建立先验分布。

3.2.1 贝叶斯定理

贝叶斯定理的主要内容为：假设 $\boldsymbol{z}=[z_1, z_2, \cdots, z_m]$为独立同分布的可量测随机变量，每一变量有相对于未知参数 $\boldsymbol{x}$ 的条件概率密度 $p(\boldsymbol{x}|\boldsymbol{z})$，则未知参数 $\boldsymbol{x}$ 的后验概率密度为

$$p(\boldsymbol{x} \mid \boldsymbol{z}) = \frac{p(\boldsymbol{z} \mid \boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{z})} = \frac{p(\boldsymbol{z} \mid \boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{z} \mid \boldsymbol{x})p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}} \tag{3.3}$$

对贝叶斯定理有几点需要说明。

(1) 式中的 $p(\boldsymbol{z}|\boldsymbol{x})$表示给定参数 $\boldsymbol{x}$ 的数据后的似然函数，其中，各数据是相互独立的，并且似然函数相对于变量 $\boldsymbol{x}$ 而言是唯一的。

(2) 式中的 $p(\boldsymbol{x})$是未知参数 $\boldsymbol{x}$ 的概率密度，称之为先验分布密度或先验分布，它是根据以往的实践经验和认识在量测实验数据之前确定的。

(3) $p(\boldsymbol{x}|\boldsymbol{z})$为后验分布密度，也被称为后验分布，是在量测了实验数据之后确定的。

(4) 由式(3.3)可以看出，分母实际上只依赖于 $\boldsymbol{z}$，与 $\boldsymbol{x}$ 无关。因此，贝叶斯定理可等价描述为：后验分布密度∝似然函数×先验分布密度，即

$$p(\boldsymbol{x} \mid \boldsymbol{z}) \propto p(\boldsymbol{z} \mid \boldsymbol{x})p(\boldsymbol{x}) \tag{3.4}$$