卡尔曼滤波入门指南

作者:刘佳 EMail:lau_jia@aliyun.com

前言：

本文旨在帮助小白入门，理想是任何一个小白都能看懂，然后在实践中应用卡尔曼滤波，至少在初级层面（比如单变量）可以做到正确应用。之前看到网上很多入门指南，可是写得也不是很清楚，也许作者本人是理解这门技术的，很多细节解释上不够清楚，故此我在前人的基础上再进一步，虽然不是最好的，但是我尽最大努力让读者能够达到本文的目标，欢迎读者通过邮件形式对本文提出指教。本文主要参考了两个文献，如果英语 ok 的话，可以自己必应一下文献标题，直达原文。

参考文献：

1. Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation
2. The Extended Kalman Filter: An Interactive Tutorial for Non- Experts

卡尔曼滤波器的优点：

1：不需要全部的历史数据（和维纳滤波器相比）

2：不需要定义窗口大小（超参，和平均滤波相比）

本质：

卡尔曼滤波器是利用理论值和观测值来对实际值做最佳估计。

引入：

要理解卡尔曼滤波，首先要搞清楚这个滤波器面对的需要解决的实际问题。举例：一辆小车以 0.75 米/秒的速度在地面上做匀速直线运动（后面的讲解都借助这个例子）。需求是想每隔一秒了解这个小车实际离开出发点的位移。应该怎么做呢？有两种方式可以做到：

1、（理论法）用牛顿定理，直线运动且速度为 0.75 米/秒，那么结合 S=v*t，就能得到 n 秒后小车移动的距离。

2、（测量法）可以用测量的方式，比如在小车上放个滴漏器，每秒滴一滴墨水，然后用尺来量和原点的距离。具体的测量方法还有很多。

上面的两种方法在理论上可行，但是实际操作中会遇到一些问题，比如：

1、实际中，理论法无法应对时刻变化的外部场景，比如对摩擦力的正确计算，因为平板上无法保证每个点的摩擦力是相同的，一般只能用统计特性来近似的估计摩擦力。即使摩擦力是固定不变的，移动过程中风的阻力也会随着气流的流动不停地发生变化，导致小车速度不能按照 0.75 米/秒做匀速运动。所以没有办法只用理论公式就得到实际操作中准确的结果。

2、测量法也有问题，比如测量工具总有固定的精度，普通的尺精确到毫米，就算用激光，电子电路的时间分辨率也会成为瓶颈。考虑小车滴墨的例子，墨水下落时经过的距离也会对测量造成误差，在高速模式下，测不准原理就会表现出来。

卡尔曼滤波的特色就是认识到了上述的问题，通过结合理论法和测量法，得到一个在统计意义上二者最好的折中。通俗的解释就是，对一个模型所关心的变量，理论上得到结果 Y1，测量法得到结果 Y2，那么估计值（统计学里用字母上加点表示对变量的估计，之所以要采用估计，是因为现实中无法明确知道得到的变量值是 100%正确）$\dot{Y}=Y1+\alpha*(Y2-Y1)$。卡尔曼滤波核心算法的目的就是计算这个α。由于 Y1 和 Y2 都是按时间频率采集的，α按采集频率更新就成了很自然的事情，而且可以想象，α作为一个平衡估计的系数，必然依赖于 Y1 和 Y2

（因为在上述的模型里只有 Y1 和 Y2 两个已知变量），否则α将成为无源之水，无本之木。幸运的是，卡尔曼滤波算法只依赖于上一个时刻的Ẏ（即上一时刻的 Y1 和 Y2，通常有这种类似马尔科夫链算法描述的时候，就可以自然想到该算法必然是不停的迭代的）。

卡尔曼滤波的使用范围：

根据公式Ẏ=Y1+α*(Y2-Y1)可知，卡尔曼滤波在计算时需要知道两个值，一个是理论推断，一个是测量。如果面对的实际问题能满足这两个要求，那么使用卡尔曼滤波是非常方便的。任一条件不满足都无法使用。比如，预测某颗遥远行星的运行轨道，可能由于行星被另外一颗更大的星挡住了，我们无法观测，只能通过理论来预测，这种情况就不适合卡尔曼滤波。再比如语音识别，目前我们一般就只有观测采样值，并没有理论指导，同样也无法使用卡尔曼滤波。股票市场中倒是可以使用卡尔曼滤波，投资客有自己的一套定价理论，通过理论值和实际市场价格的结合，可以删除噪声，找到合理的价格。

一点点的假设：

1、卡尔曼滤波假设所有涉及变量的精度误差都是正态分布，并且噪声都是加性的（以和的形式施加在模型上，而不是以系数的形式乘在模型上）。

2、卡尔曼滤波器默认模型是线性的。

3、模型当前值仅与其上一状态相关。

用前面的例子，小车实际位移和理论计算结果间的误差（噪声）是正态分布，而工具测量值和小车实际位移值间的误差是正态分布。当然根据实际情况，也可以假设误差属于其它分布，只是在计算联合分布上需要稍微注意，因为正态分布的联合分布还是正态分布，计算上比较简单，别的类型的分布就未必如此了，不过最终的假设还得按照实际情况来设计。根据模型的定义，S=v*t，显然也是线性的。为了充分利用上一次估计的信息，还需要对表达式进行改写 $S_k=1*S_{k-1}+v*\Delta t$。

数学知识：

卡尔曼滤波的数学工具还是比较简单的，如果不考虑多个变量，则只需要用到统计中的联合概率。如果涉及多变量，使用矩阵方式来表达会更方便一点，不过这只是锦上添花的技术。因为线性代数基本上属于工程工具，数学史上很少有这方面的内容。

考虑两个分布，p(x)和 p(y)，如果他们**相互独立**，那么他们的联合分布为

$$p(x,y)=px*py$$

如果这两个分布是正态的，那么联合概率也是正态的。例子中，小车实际位移和理论计算结果间的误差分布与工具测量值和小车实际位移值间的误差分布由于来源于两个不同的体系，所以很显然他们是相互独立的。

再补充一个方差的计算公式：

$$var[A*X]=A*var[X]*A^T$$

如何一一对应：

正态分布的公式如下:

$$f_x(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{[-\frac{(x-\mu)^2}{2\sigma^2}]}$$

在用理论估计小车实际位移时，上述公式中μ对应的是理论计算出的位移值，作为对小车实际位移的最佳估计，而理论值和实际位移的标准差σ则属于一个超参，由于实际环境可能发

生的状况千差万别，想通过实验或者经验评估获得这个σ会存在很大的误差。另外有个概念非常重要，此处讨论的标准差σ仅针对一次评估。在 t 时刻，理论位移值是 s=v*t。但是就好像做了 100 次这个实验，多次获取到实际位移值，然后通过这个多次位移值来计算得到方差。即 t 时刻的实际位移 $f^t(x)\sim N(\mu_t, \sigma_t)$。

在用测量值估计小车实际位移时，上述公式中μ对应的是测量的位移值，作为对小车实际位移的最佳估计。而测量值和实际位移的标准差σ也属于一个超参，不过和上面不同，测量的误差可以通过实验反复抽样得到一个比较好的估计。t 时刻的实际位移也是 $f^t(x)\sim N(\mu_t, \sigma_t)$。

为了区分理论估计和测量估计的记法，使用 $f_1(x)$，μ_1，σ_1和 $f_2(x)$，μ_2，σ_2来区分他们。

如果单独使用上述任一方法都可以作为对实际小车位移的估计。在实际工程中，这个案例得到理论值和测量值都是很方便的，结合两种方式来共同估计实际小车位移似乎是一个更好的选择，因为信息越多，得到的结果越准确。

小车的实际位移既要满足理论估计又要满足测量估计，使用联合分布就变得很自然。通过计算可得

$$f^t_1(x)*f^t_2(x)=f^t(x_1, x_2)\sim N_t\left(\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)$$

那么 t 时刻位移的估计值就是 $\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}$ 。

前面说了，理论估计和测量估计中的方差都是一个超参，测量误差可以通过实验得到，但是理论估计误差由于外部环境千变万化，难以通过实验或者经验得到。利用卡尔曼滤波器的迭代特性（当前状态仅和上一状态相关），就可以得到一个比较好的理论误差估计（卡尔曼滤波器的作用就是在测量噪声统计特性已知的情况下，动态估计系统状态）。

整个卡尔曼滤波器的迭代过程如下：

假设数学模型是 $U_k=a*U_{k-1}+b$ #线性模型

在 t=0 时刻（获取初始状态和计算下一次迭代需要的参数）：

计算理论值阶段

1. μ_1=？ #在 t=0 时刻，理论值的初始值为止
2. σ_1^2=1 #作为初始的超参，理论的方差可以按照经验或者感觉定一个，后续会通过迭代消除误差

计算测量值阶段

1. $\hat{\mu}=\mu_2$ #在 t=0 阶段，由于理论值的μ_1也是通过观察值得到的，所以用观察值作为 t=0 时刻 U 的最佳估计
2. $\hat{\sigma}_1^2=\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ #更新σ_1^2，作为 t=1 时刻理论估算时 t=0 的σ_1^2用(t=1 时需要用)

在 t=1 时刻：

计算理论值阶段

1. μ_1= a*$\hat{\mu}$+b #用上一时刻的$\hat{\mu}$作为理论估计的基础

2. σ_1^2 =a* $\hat{\sigma}_1^2$ *aᵀ #用上一时刻的 $\hat{\sigma}$ 作为上一时刻理论的偏差

计算测量值阶段

1. $\alpha = \frac{\sigma_1^2}{(\sigma_1^2+\sigma_2^2)}$

2. $\hat{\mu} = \mu_1 + \alpha(\mu_2 - \mu_1)$ #Ẏ=Y1+α*(Y2-Y1)

3. $\hat{\sigma}_1^2$ = $\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ =（1-α）* σ_1^2

后续就以此类推了。

实例应用：

设数学模型：Y_K=1.25*Y_{k-1}-0.1，应用前面的说明，得到下表，其中μ_hat 就是对“模型的现实值”的估计。“模型的现实值”是在“模型理论值”上追加了噪声，而μ2 则是在“模型的现实值”上追加了噪声，模拟测量误差。

时刻	1	2	3	4	5	6	7	8	9	10	11	12
模型的理论值	0.75	0.8375	0.946875	1.08359375	1.254492188	1.468115234	1.735144043	2.068930054	2.486162567	3.007703209	3.659629011	4.474536264
模型的现实值	0.836551468	0.733148167	0.931330218	1.181973055	1.372400653	1.423372643	1.686353494	2.103828604	2.452447319	3.107737925	3.687869691	4.446672917
μ2	1.031620159	0.810100056	0.887146406	1.245783035	1.28996338	1.396262503	1.583262616	1.960264966	2.363609514	3.194615435	3.716859311	4.417749916
μ1		1.189525198	1.150945588	1.214079137	1.430576596	1.635189019	1.8581944	2.126960344	2.50171129	2.980483486	3.697402521	4.528245791
σ1	1	0.015470297	0.009490403	0.007608234	0.006751311	0.006297372	0.006037564	0.005882249	0.005786973	0.005727599	0.005690235	0.005666577
α		0.497512438	0.377871826	0.327472021	0.30171688	0.287257772	0.27870959	0.273500757	0.270268072	0.268238958	0.266956238	0.266141715
μ_hat	1.031620159	1.000756471	1.05126331	1.224461277	1.388151215	1.56655552	1.781568275	2.081369032	2.464386789	3.037922017	3.702596633	4.498838229
σ1_hat	0.00990099	0.006073858	0.00486927	0.004320839	0.004030318	0.003864041	0.003764639	0.003703662	0.003665663	0.00364175	0.003626609	0.003616985

根据上表，再计算一下估计误差

卡尔曼滤波估计的值与“模型的现实值”的误差为：0.164129812

测量值μ2 与“模型的现实值”的误差为：0.105883216

可见，未必一定就是卡尔曼滤波胜过通过观测得到的估计。只能说在理论上，这种估计是最佳的。不过需要注意，如果测量噪声很大，那么卡尔曼滤波的优势是很大的。这个读者可以自行演练（可调高上表中μ2 或者模型现实值的噪声）。另外由于模型采用迭代方式，并且在理论估计时总是使用上一次的最优估计值作为当前理论计算的值，所以卡尔曼滤波具有一定的抗噪声能力，在系统本身噪声很大的时候估计结果会更平滑，使得估计结果和实际值差距较大，所以系统本身噪声很大的时候，直接用观测值可能更好（读者可以调高模型现实值的噪声自行实验）。留一个思考题，如果还是用卡尔曼滤波原理，如果已知系统噪声很大，而测量噪声很小，如何调整卡尔曼滤波，使得估计结果更接近现实值？（提示：Ẏ=Y2+α*(Y1-Y2)）

实践扩展：

需要注意，理论值和测量值的量纲未必总是一致的，所以在更普遍化的理论公式中卡尔曼滤波中有一个 H 变量，其作用就是进行量纲转换，这里不单独再在公式里补充，感兴趣的人可以看一下参考文献。

如果要估计的变量不止一个，那么可以引入矩阵同时计算即可，注意到，在计算理论方差时，采用了

$$var[A*X]=A*var[X]*A^T$$

的形式，就是考虑到了估计量 X 是一个向量，而不是一个标量。

理论扩展：

根据卡尔曼滤波的理论，我们可以发现，除了理论和测量方法，如果能有第三种方法估计实际值，且这种方法与前两种均相互独立，那么就能够得到更加好的估计值。举个例子，比如 A 国向 B 国发射了一颗攻击性导弹，B 国为了拦截这颗导弹需要知道这个导弹的攻击目标。B 国根据形式大致判断 A 国会攻击 B 国最大经济城市 C，但是为了保险，B 国还会通过雷达判断导弹的行进路线，来修正自己的判断。有这两个条件就已经可以使用卡尔曼滤波来估计导弹目标了，但是 B 国为了控制风险，他们动用了在 A 国导弹发射中心的间谍来同时提供情报，这样对于 B 国而言，他们就是用 3 种各自独立的信息来进行估计。由此可见卡尔曼滤波是链式的，可以通过增加信息的来源来增加估计的最优度。