**Kalman filter卡尔曼滤波**

参考：An Introduction to the Kalman Filter

B站作者UP：逆风引弓

**1，明确问题**

卡尔曼滤波处理的是离散时间控制过程，目的是对维的状态矢量进行估计。适合用在能算但算的不准，能测但测的不准的情况下，在诸多领域有着广泛的应用。

**状态转移方程**为：

**测量方程**为：

* 和分别代表第个时间步和第个时间步的维状态矢量。
* 代表第个时间步的可选的外部控制输入，是个维矢量。
* 代表转移过程中的高斯噪声，是维随机矢量。
* 代表第个时间步的维测量值矢量。为测量时的高斯噪声，是维随机矢量。
* 是行列的矩阵，是行列的矩阵，是行列的矩阵。
* 现实中和可能每个时间步都不一样，但我们这里假设它是恒定的。
* 假定、分别满足（多元）**正态分布**，，,是各自的协方差矩阵。，，和是通过**离线校准**、**微调**得到的（后文详谈）。
* 不同时间步的彼此独立同分布，即彼此之间独立，不同时间步的彼此独立同分布，即彼此之间独立。任意和之间彼此独立，比如和彼此独立。

为了分析问题我们定义一些变量：

* 先验状态估计priori state estimate：。它表示根据第步之前的测量值预测出的。
* 后验状态估计posteriori state estimate：。它代表根据包含了第步测量结果的得出的。
* 先验估计误差priori estimate errors：
* 后验估计误差posteriori estimate errors：
* 先验误差协方差矩阵为：
* 后验误差协方差矩阵为：。

**注意**：、是维随机矢量。对矩阵求期望值表示对矩阵的每个元素求期望值。

我们希望得到的后验估计的后验误差协方差矩阵的迹**越小越好**。

卡尔曼滤波是一种递推算法，利用上一时间步的2个参数：后验估计&后验误差协方差矩阵，就可以推导出当前时间步的后验估计&后验误差协方差矩阵。其余定义的变量需要用于中间过程的计算。

1. **先验估计&根据测量结果(feedback)对先验估计做修正**

用上一时间步的后验估计代入转移方程**得到当前时间步的先验估计**（当然，我们不可能知道噪声是多少，所以没有）：

先验状态估计并**未用到**的测量结果，且**忽略了过程噪声**，与真实的肯定有所偏差，我们试着刻画这种偏差。状态值和测量值之间通过测量方程联系起来，如果和真实的很接近，那么测量值和之间应该也很接近。于是可以用两者之差描述这种偏差，我们将偏差乘以**权重矩阵**，再加到先验估计上，就能得到根据测量结果**修正的后验估计**，即融入了测量信息的估计：

* 权重矩阵是行列矩阵，叫做**混合因子blending factor/卡尔曼增益因子**。
* 被称为预期测量值measurement prediction。
* 被称为测量革新measurement innovation，或者叫残差residual。

**3，确定混合因子矩阵**

混合因子要如何选择？必须选择合适的，使得后验误差协方差矩阵的迹最小，。我们先推导与的关系：

对上式求均值，不含随机变量的矩阵不参与求均值，可以直接提出来：

把（和互相独立）代入：

要求出合适的使得最小，等价于求：

先回顾一下对矩阵求导的相关定义。

=====================================================================

参考：https://zhuanlan.zhihu.com/p/273729929

复习一下标量函数对矩阵（假设是行列）求导的规则：

=====================================================================

注意到协方差矩阵都是对称矩阵：

是**已知的**测量矩阵，是**已知的**测量噪声的协方差，是先验估计误差的协方差矩阵，所以我们**要先求出**。指数上的-1要理解成矩阵求逆。

**注意到**每个时间步的都不同（因为**不同**），所以加上下标变成：

倘若每个时间步测量矩阵和测量噪声协方差都有所不同，也必须用带下标的和来描述：。当测量噪声趋于0时~0，，代表完全采信测量值，后验估计值的趋势会紧密跟随测量值。当很大时，会很小，代表不愿采信测量值，根据测量值做的修正很小。而这很显然是正确的。

**4，获取先验误差&后验误差的协方差矩阵&**

的计算需要，如何得知？，，先验估计，我们把该式子和转移方程相减得到：

上式LHS正是第时间步的先验估计误差，RHS正是作用于时间步的后验估计误差加噪声项：

就得到了先验误差的协方差矩阵。有了就能利用算出最优的。

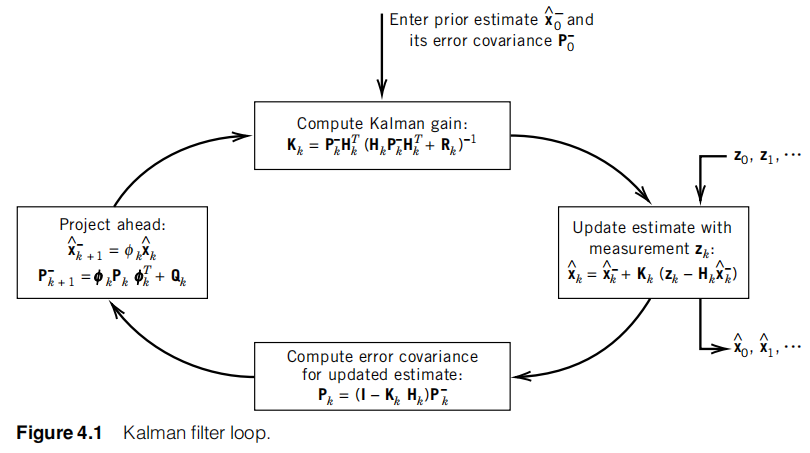
得到最优的后，就得到了后验估计：，我们还需要求出后验估计误差的协方差矩阵。我们**只需要再把此时的代回表达式**：

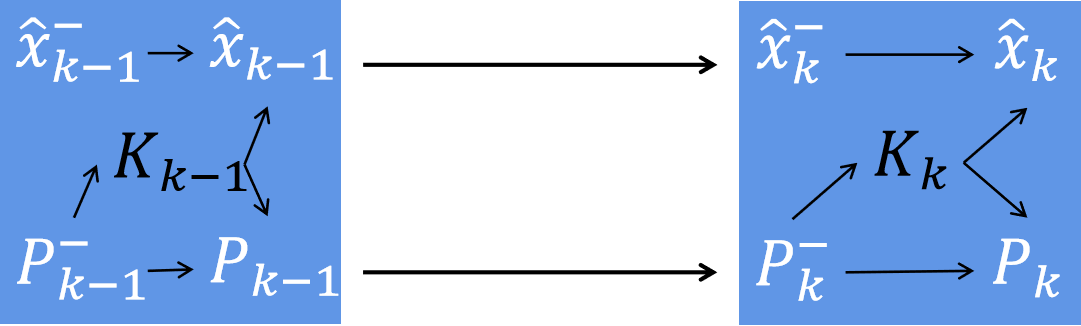
**5，综合所有5个关键方程&参数的离线校准、微调**

时间更新方程time update equations/predictor equations，分别获得先验估计，先验误差协方差矩阵：

测量更新方程measurement update equations/corrector equations，分别获得混合因子，后验估计，后验误差协方差矩阵：

依次计算这5个公式，就可以得到第时间步的，，就完成了卡尔曼滤波的1次递推。不过**迭代过程总得有个初始值**，你得有最**初始的和**才能开始往下递推。初始值不需要多精确，只要、、时间步长参数设置没问题，经过若干轮迭代都会收敛到真实值（如果参数有问题...就会发散或产生偏差），就和牛顿迭代法解方程的初始值选取类似。





**问题在于、参数的设置、调节相当麻烦**。一开始我们说了，转移过程和测量过程的噪声协方差矩阵和其实是未知的，我们得预先通过测量校准、微调这两个参数。是测量过程的协方差矩阵，取决于测量系统自身，的校准很方便，只要离线做几次测量即可（嫌麻烦就直接查传感器手册）。可以通过换更高端的传感器、优化测量方式等等手段缩小，越小说明测量越准。的校准十分麻烦，因为转移过程的噪声是不可见的，**噪声来自于实际系统复杂的、被忽略的物理机制**，我们又无法直接观测。很多时候只能从小到大一点点微调，然后做实验验证效果，直到得到一个合适的。

**理想情况下，和都是恒定的**，此时经过若干次迭代很快会稳定下来：

因为，是只依赖的一个固定的函数。是的函数，因此也是只依赖的固定函数，于是可写成：

那么（一般来说）会收敛、稳定在该迭代式的不动点，、也会稳定下来，于是我们完全可以**预先计算好**稳定后的，，，计算量可以大大减小。

实际案例中，测量噪声协方差矩阵**可能会随时间/位置/速度等等其他参数变动**，比如，对某台手机的GPS定位，手机进入信号不好的地区，就会增大。类似的，过程噪声协方差矩阵**可能会随某些参数变动**，例如我们跟踪机器人时，选定为时间步的机器人坐标，若该时间步机器人移动速度很快，我们会认为较大。若速度较慢，我们会认为较小。

**6，卡尔曼滤波的背后是贝叶斯定理**

参考：Introduction to Random Signals and Applied Kalman Filtering 4ed 4.7节

===============================================================================

参考：苏淳 马群强 《概率论》5.5节 定理5.5.1 定理5.5.6

首先回顾一些多维高斯分布的性质。已知维随机向量，满足均值为，协方差矩阵为的多维正态分布，多元高斯概率密度函数为：

**性质1：**若，是一个常数向量，则维随机变量满足正态分布。

**性质2**：若，是一个行列的常数矩阵，则维随机变量满足正态分布。

**性质3**：若，和都是随机变量，分别满足正态分布和。则满足正态分布。

===============================================================================

我们将利用贝叶斯定理证明，在高斯分布假设下，卡尔曼滤波是最优的。我们将用贝叶斯定理推导出卡尔曼滤波的公式。

**状态转移方程**为：

**测量方程**为：

**高斯分布假设**：

卡尔曼滤波的高斯分布假设是指，转移过程噪声、测量噪声、状态值后验分布都满足高斯分布。对初始值，我们已知真实的初始状态值满足高斯分布。之所以这么假设是因为后续证明将用到高斯分布的3点性质。

**命题**：

运用数学归纳法的思想，我们接下来通过**贝叶斯定理**证明2点。

1，若k时间步的kalman filter先验估计是最优的：

则k时间步的后验估计也是最优的：

其中，，与卡尔曼滤波中的迭代公式一致。代表的测量结果序列。

2，若k时间步的后验估计是最优的：

则k+1时间步的先验估计也是最优的：

**Proof-1**：

由贝叶斯定理可知：

：代表的先验分布，由命题的前提条件给定：

：分布同时取决于的分布和测量噪声，通过测量方程，再利用上述多维高斯分布的性质2&3，有

也可以表述为。

：它的含义是在固定状态变量的前提下，测量值的条件概率分布。是个只跟测量噪声有关的随机变量：。部分此时看做常数，因此是一个常数加一个高斯分布噪声：

在第个时间步，之前的测量值都是既定事实，是前提，所以此时的

将3个高斯分布的函数形式代入：

接下来就是代入多维概率密度函数，以及冗长的化简过程：

相关化简计算十分冗长，但也不用完全化简，只需明白

然后通过对比二次项、一次项的系数可以看出：

这还不是我们熟悉的卡尔曼滤波中所用的形式，接着化简要用到**矩阵求逆引理matrix inversion lemma**：*截图参考自《Gaussian Processes for Machine Learning》附录A公式(A.9)*



请注意蓝色部分恰好是卡尔曼增益，于是得到我们熟悉的形式：

同样通过比较一次项的系数，有：

注意：。证明如下：

相同的结论参考论文：Introduction to Random Signals and Applied Kalman Filtering 4ed (4.7.10)(4.7.11)公式。

**总之**，化简后仍然得到一个高斯分布：

均值为

协方差矩阵为

**这与卡尔曼滤波的迭代公式一致，后验估计正是后验分布的均值，很明显此时也是最小**。我们从贝叶斯定理出发，再度得到kalman filter的公式，这充分说明kalman filter的迭代公式背后是贝叶斯定理。

**Proof-2**：

接着proof-1的结论，证明：若k时间步的后验估计是最优的：，则下一时间步的先验估计就是最优的。这是显然的，因为先验估计无非是把上一轮的结果代入转移方程：

这意味着在高斯分布的前提下，只要初始有了一个最优的先验 or 后验估计，之后卡尔曼滤波所有的后验估计都是严格的最优估计。

**END**

参考：Introduction to Random Signals and Applied Kalman Filtering 4ed chapter6有smoothing、prediction相关内容。普通的卡尔曼滤波叫做real-time filtering。

===============================================================================

参考：苏淳 马群强 《概率论》5.5节 定理5.5.1 定理5.5.6

首先回顾一些多维高斯分布的性质。已知维随机向量，满足均值为，协方差矩阵为的多维正态分布，多元高斯概率密度函数为（https://zhuanlan.zhihu.com/p/58987388）：

**性质1**：若，是一个常数向量，则维随机变量满足正态分布。

**性质2**：若，是一个行列的常数矩阵，则维随机变量的概率密度函数为：

换元需要用到雅克比行列式。矩阵转置不改变行列式。由此可知满足正态分布。若是一个行列的矩阵，变换后有是维随机矢量，当时无法保证反函数/逆矩阵存在，此时互相不独立，其中必定有个可以表达成其余分量的线性组合，可以舍弃这个分量。当时总可以补充个方程使得，此时结论依然成立。

**性质3**：若，和都是随机变量，分别满足正态分布和。则满足正态分布。

===============================================================================

===============================================================================

参考：苏淳 马群强 《概率论》5.5节 定理5.5.1 定理5.5.6

首先回顾一些多维高斯分布的性质。已知维随机向量，满足均值为，协方差矩阵为的多维正态分布，多元高斯概率密度函数为：

**性质1：**若，是一个常数向量，则维随机变量满足正态分布。

**性质2**：若，是一个行列的常数矩阵，则维随机变量满足正态分布。

**性质3**：若，和都是随机变量，分别满足正态分布和。则满足正态分布。

===============================================================================