卡尔曼滤波

Dezeming Family

2022 年 6 月 14 日

DezemingFamily 系列文章和电子书**全部都有免费公开的电子版**，可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列电子书，可以从我们的网站 [https://dezeming.top/] 找到最新的版本。对文章的内容建议和出现的错误也欢迎在网站留言。

目录

一　本文介绍

在工程控制中，卡尔曼滤波占用非常重要的地位，尤其是在对平衡性要求比较高的应用场景中，如何对传感器得到的数据进行去噪是一件非常重要的事。我在阅读互联网和书本上的资料时，感到现有的资料中，要么理论证明难以读懂，对初学者不友好；要么虽然比较通俗，但符号定义混乱，且没有详细的推导过程；有很多文章作者本身对卡尔曼滤波也不甚了解，因此推导中有很多本质性错误。

基于现有资料并不很好的现状，本文将提供一个全面的、层层递进的卡尔曼滤波算法的详细的介绍，不但有最直观的理解，还会有严格的理论推导过程。

本文不但包含直观的卡尔曼滤波理解。还有原理和公式的推导，关于扩展卡尔曼滤波以及一些具体的示例和应用可以参考 Dezeming Family 的其他文章。

二　通过经验估计来修正传感器测量值

2 1　简单的控制场景

考虑你要养热带鱼，需要一个水温恒温 30 度的系统。这个系统里，包含有下面这些物件：鱼缸、水、加热棒（加热棒里有温度传感器和加热管）。

影响鱼缸水温的因素有哪些？主要是外界环境（空气温度、是否往鱼缸里加了一部分水）和加热棒的工作状态。

一个最简单的思路就是使用温度计测温，然后根据温度计的变化来调整加热棒的工作状态。比如，温度低于 30 度就加热，高于 30 度就停止加热，这样看起来简单又可靠，这也是现在很多加热棒都是这么工作的——水温有轻微的温度波动并不会杀死鱼。

2 2　复杂的控制场景

考虑还是恒温系统，但此时此刻你的工作内容是研究某特定温度下融化的金属，你的加热工具必须持续工作，否则金属会很快凝固。此时，温度传感器不能直接放在金属中去测量，这样会直接烧坏温度传感器，因此，你的温度计只能放在容器边缘，因此会没有那么准确，而且很容易受到外界的干扰。

此时，我们需要控制加热工具的功率，但我们很难完全相信温度传感器得到的温度。我们可以去预测一个温度，比如当前温度是 1000 度，我们根据一些经验公式来预测出下一秒的温度会变为 999 度。而到了下一秒，温度传感器却测出温度是 1002 度。到底应该相信是 999 度还是 1002 度？那么卡尔曼滤波给出了答案：都相信，但每个数值相信的程度不同。

换句话说，如果我们更相信我们的先验知识模型，我们就更倾向于相信当前温度是 999 度；反之，我们更相信传感器，就倾向于更相信当前温度是 1002 度。至于我们应该更相信谁，在卡尔曼滤波中，这是一个用户可以调节的参数。

于是，我们可以简单理解为：滤波输出 = 经验模型估计值 × 权重 1 + 真实传感器测量 × 权重 2；其中，权重 1 + 权重 2 = 1。如何计算这个权重，卡尔曼滤波给出了答案。

2 3　时间间隔

真实的系统一般都是离散控制系统，我们能够感受到，要使得这种“经验分析 + 传感数据共同得到最优估计”的方案要能够很好的工作，则需要保证传感器响应和分析的速度要快。

连续性不足，比如你的温度处理每隔一分钟测一次，可能一分钟之后温度变化会非常大，你的经验模型不再起任何作用（时间已经长到你的经验模型难以估计出当前准确温度），你只能去选择完全相信传感器，因此，最好的估计时间间隔要尽量小，比如 1 秒一次。

不同的系统，这种时间间隔的要求也会有所不同。比如二轮平衡小车，我们希望对陀螺仪的数据进行滤波，但如果我们间隔时间太长，比如隔着一秒才估计一次，那可能小车一秒钟的时候就已经倒了，此时前一秒估计的值将不再有参考价值。

2 4 简单的总结

在通过公式来介绍卡尔曼滤波之前，我们进行一下总结。我们当前是一个不断工作的系统，我们需要不断测量系统的某些状态。这个测量过程其实是递归的（参考《递归最小二乘法》），即当有新的观测数据进来后，需要更新当前状态值。

卡尔曼滤波的解释：当最优前估计值 $\Leftarrow$ 上一次估计值 + 系数 $\times$ (当前测量值 - 根据经验估计的当前状态值)。

三　卡尔曼滤波的直观理解

我们希望构建一个小车匀速直线运动系统，此时系统中有一个小车，有小车的测速装置，有检测小车移动距离的

小车的位置估计可以写为：

$$x_t = x_{t-1} + v_{t-1}\Delta t + \frac{1}{2}a_t(\Delta t)^2 \tag{三.1}$$

其中，x_t 表示当前时刻小车总共前进的距离，x_{t-1} 表示小车上一个时刻前进的距离，a_t 表示当前时刻的加速度（与施加的力有关），v_{t-1} 表示上一个时刻的速度。小车的速度可以表示为：

$$v_t = v_{t-1} + a_t\Delta t \tag{三.2}$$

3 1 状态方程

在这个系统中，小车的状态量有两个值，一是小车前进的距离 x，二是小车的速度 v：

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ v_t \end{bmatrix} \tag{三.3}$$

我们可以把状态写成状态方程：

$$\begin{bmatrix} x_t^- \\ v_t^- \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \widehat{x}_{t-1} \\ \widehat{v}_{t-1} \end{bmatrix} + \begin{bmatrix} \frac{1}{2}(\Delta t)^2 \\ \Delta t \end{bmatrix} a_t \tag{三.4}$$

注意上标 $-$ 表示根据先验模型预测的结果，上标 ^ 表示上一次的最优估计结果。上式可以简写为：

$$\mathbf{x}_t^- = \mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t \tag{三.5}$$

注意这里的状态矩阵 $\mathbf{F}_t$ 代表了我们的先验知识。我们当前有两个状态（距离和速度），所以是 2×2 的矩阵；如果有 N 个状态，那么这个状态矩阵就是 $N \times N$ 大小的矩阵。

$\mathbf{B}_t$ 是控制矩阵，a_t 相当于我们施加的控制效果，比如加大油门，小车就会加速。施加的作用 a 乘以控制矩阵，作用到状态上，就会改变状态。我们可以想象，如果没有控制矩阵 $\mathbf{B}_t$ 和控制项 a_t 的作用，小车就是按照匀速在前进：

$$\begin{aligned} x_t^- &= \widehat{x}_{t-1} + \widehat{v}_{t-1}\Delta t \\ v_t^- &= \widehat{v}_{t-1} \end{aligned} \tag{三.6}$$

注意此时的系统假定是无摩擦阻力的匀速运动系统；也可以是该系统有一个稳定的功率，使得小车在没有控制 $\mathbf{B}_t$ 和 a_t 的时可以保持匀速运动（当然也不是绝对的匀速，速度大小也有一定的噪声，噪声符合高斯分布，距离和速度都是状态量）。

这个状态方程是我们根据先验知识得到的，用于给定上一个时刻的值 $\widehat{\mathbf{x}}_{t-1}$ 来预测下一时刻的值。我们知道，我们的系统本身也是符合这个特征的，但是会包含误差，所以系统真实值 $\mathbf{x}_t$ 相对于上一次系统真实值 $\mathbf{x}_{t-1}$ 而言，除了控制和状态转移矩阵的作用外，还需要加一个误差变量 $\mathbf{w}_k$（过程噪声）：

$$\mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{B}_t a_t + \mathbf{w}_t \tag{三.7}$$

我们需要再次强调这一点。系统自身的这种性质和我们的先验估计看着有点相似：

$$\mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{B}_t a_t + \mathbf{w}_t \tag{三.8}$$

$$\mathbf{x}_t^- = \mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t \tag{三.9}$$

上式第二行是我们根据上一次估计得到的最优估计值 $\widehat{\mathbf{x}}_{t-1}$ 与本次先验估计 $\mathbf{x}_t^-$ 之间的关系，上式第一行则是上次的真实值 $\mathbf{x}_{t-1}$ 与本次真实值 $\mathbf{x}_t$ 之间的关系。有些文章里会错误的把这两个内容混用，这里我们强调一下，第一行与系统本身有关，与预测无关；第二行与预测有关，但是你无法预测当前状态的噪声，所以没有噪声项。

3 2 观测方程

我们不能完全依靠系统状态的方程来得到下一时刻的信息，因为总是存在外在因素来干扰，我们又不知道 $\mathbf{w}_t$ 是怎么干扰系统的，所以需要借助传感器测量。观测值肯定与当前状态存在某种关系，我们考虑的只是线性系统，所以关系符合线性关系：

$$\mathbf{y}_t^{\cdot} = \mathbf{H}\mathbf{x}_t^- + \mathbf{v}_t \tag{三.10}$$

注意，我们观测到的值其实 $\mathbf{y}_t$，我们假设这个 $\mathbf{y}_t^{\cdot}$ 跟状态 $\mathbf{x}_t^-$ 要符合上述关系，这里的上标 $\cdot$ 表示观测值。$\mathbf{H}$ 可以称为观测矩阵，$\mathbf{v}_t$ 表示观测噪声。举个例子，如果我们使用超声波测距，且使用电磁编码器测速度，那么测量得到的 $\mathbf{y}_t$ 和 $\mathbf{x}_t$ 一样，都包含了距离和速度，也就是说：

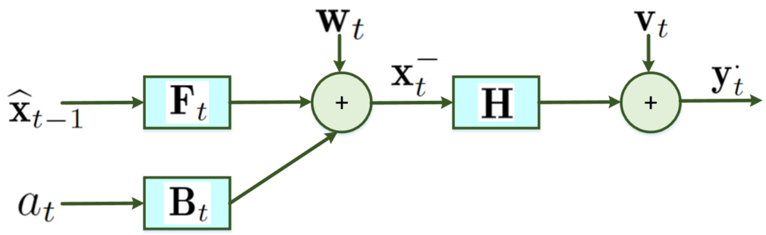
$$\mathbf{y}_t^{\cdot} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}_t^- + \mathbf{v}_t \tag{三.11}$$

当然，或许我们只观测了距离，因此：

$$\mathbf{y}_t^{\cdot} = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}_t^- + \mathbf{v}_t \tag{三.12}$$

此时噪声量 $\mathbf{v}_t$ 是一个标量。

总的关系式可以用图示来表示（注意，这并不是一个流程，而是一个关系，用来表示预测值和观测值的内在联系）：



状态方程 + 观测方程称为“状态空间表达式”。

另外需要注意的是，观测值 $\mathbf{y}_t^{\cdot}$ 和真实值 $\mathbf{x}_t$ 之间也应该符合这么一个关系：

$$\mathbf{y}_t^{\cdot} = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \tag{三.13}$$

这里的 $\mathbf{v}_t$ 同样表示高斯分布噪声。这里的 $\mathbf{v}_t$ 和前面的 $\mathbf{v}_t$ 未必是同一种高斯分布，只是都用符号 $\mathbf{v}_t$ 来表示观测噪声。

3 3 协方差矩阵

我们现在有了当前系统的估计值 $\mathbf{x}_t^-$ 和观测值 $\mathbf{y}_t^{\cdot}$，我们如何利用这两个值来得到当前系统的最优估计值呢？经过一系列比较复杂但是并不难的推导（我们会在后面的章节进行描述），可以得到最优估计关系式，在此之前，从直观理解的角度，我们先思考一下各个量之间的误差关系。

首先我们知道状态方程的估计是有误差的，而且对于 N 个状态（比如前面同时包含了速度和距离的两个状态），不但每个状态自己都有噪声方差，它们之间也有一些关系（比如速度和距离其实存在某种线

性关系），这些方差关系可以用协方差矩阵来描述（描述不同分布之间的关系）。

$$cov(\mathbf{x}_t, \mathbf{x}_t) = \begin{bmatrix} cov(x_t^-, x_t^-) & cov(x_t^-, v_t^-) \\ cov(v_t^-, x_t^-) & cov(v_t^-, v_t^-) \end{bmatrix} \tag{三.14}$$

自身与自身的协方差其实就是方差：

$$cov(x_t^-, x_t^-) = var(x_t^-, x_t^-) \tag{三.15}$$

如果我们有上一次状态估计的协方差矩阵 $\mathbf{P}_{t-1}$，根据状态转移矩阵 $\mathbf{F}_t$ 我们就可以计算出本次估计的协方差矩阵 $\mathbf{P}_t^-$，注意本次估计的协方差矩阵 $\mathbf{P}_t^-$ 也是有估计误差的。

协方差矩阵如何作为最优估计的参数，以及具体怎么根据上一次估计的协方差矩阵 $\mathbf{P}_{t-1}$ 求出本次估计的协方差矩阵 $\mathbf{P}_t^-$ 将会放在后面的原理推导部分。

为什么需要协方差矩阵？从结果来说，协方差矩阵是用于进行最优估计的手段，进行最优估计需要协方差矩阵作为参数（因为我们知道估计和观测都是有噪声的，所以最优估计肯定需要考虑这些噪声，也就需要一个参数来衡量噪声对最优估计的作用，这个参数就是协方差矩阵）。

3 4 通过状态和观测来估计

现在，我们需要根据估计值 $\mathbf{x}_t^-$ 和观测值 $\mathbf{y}_t$ 来得到最优估计。先给出最优估计的式子：

$$\hat{\mathbf{x}}_t = \mathbf{x}_t^- + \mathbf{K}_t\left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^-\right) \tag{三.16}$$

注意前面说过 $\mathbf{H}$ 是观测矩阵，$\mathbf{K}_t$ 叫做“卡尔曼增益”，相当于估计值和观测值的混合权重。经过一系列推导，$\mathbf{K}_t$ 的表示如下：

$$\mathbf{K}_t = \frac{\mathbf{P}_t^-\mathbf{H}^T}{\mathbf{H}\mathbf{P}_t^-\mathbf{H}^T + \mathbf{R}} \tag{三.17}$$

其中 $\mathbf{R}$ 表示观测噪声的协方差矩阵。

根据上面计算最优估计 $\hat{\mathbf{x}}_t$ 的式子中，我们可以看出，最优估计与状态方程协方差矩阵、当前估计值和当前观测值都有关系。括号里的部分 $\left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^-\right)$ 表示估计值和观测值的差值。当观察噪声的协方差矩阵 $\mathbf{R}$ 的值越大（传感器噪声越大），则 $\mathbf{K}_t$ 就越小，表示此时观察值和估计值的差的作用性就越小，最优估计更接近于估计值。

另外注意，$\mathbf{H}$ 矩阵一定要是可逆的，因为如果不可逆，说明我们选择的观测量中的某些量之间呈倍数的关系，则属于重复观测量。比如你的观测量中的两个量都是距离，一个量是速度，则 $\mathbf{H}$ 矩阵就是：

$$\mathbf{y}_t = \begin{bmatrix} x_t \\ x_t \\ v_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t^- \\ x_t^- \\ v_t^- \end{bmatrix} + \mathbf{v}_t \tag{三.18}$$

此时的 $\mathbf{H}$ 矩阵就不是一个可逆矩阵了。

3 5 卡尔曼滤波的适用条件

前面说了这么多，我想大家应该能感受到卡尔曼滤波应该主要应用于线性系统，且系统的噪声（波动）需要符合高斯分布；注意这里的噪声不但包括传感器测量噪声，还包括估计噪声。

所谓线性系统，需要满足叠加性和齐次性。也就是该系统需要满足：

$$\begin{aligned} & input\ x_1 \rightarrow output\ y_1 \\ & input\ x_2 \rightarrow output\ y_2 \\ \Longrightarrow\ & input\ a \times x_1 + b \times x_2 \rightarrow output\ a \times y_1 + b \times y_2 \end{aligned} \tag{三.19}$$

或者通俗点，输入增加几倍大小，输出也就增加几倍大小。

关于高斯分布其实有很多值得掌握和了解的地方，我们并不展开介绍。下一章节我们会推导卡尔曼滤波的所有公式。

四 原理推导

我们这里先给出卡尔曼滤波的五个公式，见 table 1。

表 1: 卡尔曼滤波公式组

预测公式	校正公式
先验估计 (1)：$\mathbf{x}_t^- = \mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t$	卡尔曼增益 (3)：$\mathbf{K}_t = \frac{\mathbf{P}_t^-\mathbf{H}^T}{\mathbf{H}\mathbf{P}_t^-\mathbf{H}^T+\mathbf{R}}$
先验误差协方差 (2)：$\mathbf{P}_t^- = \mathbf{F}_t\mathbf{P}_{t-1}\mathbf{F}_t + \mathbf{Q}$	后验估计 (4)：$\widehat{\mathbf{x}}_t = \mathbf{x}_t^- + \mathbf{K}_t\big(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^-\big)$
	更新误差协方差 (5)：$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_t^-$

注意先验估计 (1) 中不含有噪声变量（因为我们并不知道噪声具体是多少，我们估计的时候也不会把随机噪声加到估计值中），加上噪声的先验估计其实相当于是真实值：

$$\mathbf{x}_t = \mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t + \mathbf{w}_t \tag{四.1}$$

我们在前面已经了解到了先验估计公式和后验估计公式，以及观测状态关系式 (6)：

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \tag{四.2}$$

目前还剩三个公式需要推导：先验误差协方差公式、卡尔曼增益公式以及更新误差协方差公式。

4 1 先验误差协方差估计公式

4 1.1 方法一

在推导如何根据上一次的协方差矩阵得到当前协方差矩阵之前，先给出一个关于协方差性质的公式：

$$cov(\mathbf{A}\mathbf{x} + \mathbf{c}, \mathbf{A}\mathbf{x} + \mathbf{c}) = \mathbf{A}cov(\mathbf{x}, \mathbf{x})\mathbf{A}^T \tag{四.3}$$

这里的 $\mathbf{A}$ 表示作用于一个矩阵，$\mathbf{c}$ 表示某个常量。设：

$$\mathbf{P}_{t-1} = cov(\widehat{\mathbf{x}}_{t-1}, \widehat{\mathbf{x}}_{t-1}) \tag{四.4}$$

当 $\mathbf{x}$ 表示 $\mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t + \mathbf{w}_t$ 时，由于 $\mathbf{B}_t a_t$ 是一个无噪声的常量，$\mathbf{w}_t$ 是符合高斯分布的噪声量，所以 $\mathbf{P}_t^-$ 可以写为：

$$\begin{aligned}\mathbf{P}_t^- &= cov(\mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t + \mathbf{w}_t, \mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t + \mathbf{w}_t)\\ &= \mathbf{F}_t cov(\widehat{\mathbf{x}}_{t-1}, \widehat{\mathbf{x}}_{t-1})\mathbf{F}_t^T + cov(\mathbf{w}_t, \mathbf{w}_t)\\ &= \mathbf{F}_t\mathbf{P}_{t-1}\mathbf{F}_t + \mathbf{Q}\end{aligned} \tag{四.5}$$

4 1.2 方法二

我们也可以定义误差符号 $\mathbf{e}_t^- = \mathbf{x}_t - \mathbf{x}_t^-$ 表示先验概率误差，先验估计协方差矩阵 $\mathbf{P}_t^-$ 可以写为 $E(\mathbf{e}_t^-(\mathbf{e}_t^-)^T)$ 的形式，也就是 $\mathbf{e}_t^-(\mathbf{e}_t^-)^T$ 矩阵中每个元素的期望。

当我们把 $\mathbf{e}_t^-$ 展开时，真实值 $\mathbf{x}_t$ 要写为这种形式：

$$\mathbf{x}_t = \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{B}_t a_t + \mathbf{w}_t \tag{四.6}$$

$\mathbf{x}_t^-$ 要写为这种形式：

$$\mathbf{x}_t^- = \mathbf{F}_t\widehat{\mathbf{x}}_{t-1} + \mathbf{B}_t a_t \tag{四.7}$$

这是因为一个与系统自身有关，一个与我们的估计有关（在第三章第一节的最后“状态方程”部分我进行了强调）。

之后再简单推导就能得到结论了。

4.2 卡尔曼增益公式

卡尔曼滤波中，推导最麻烦的部分是卡尔曼增益，这里面涉及比较多的概率分布、矩阵微积分的内容，但基本流程并不难。

t 时刻的估计状态为 $\hat{\mathbf{x}}_t$，设实际真实状态是 $\mathbf{x}_t$，估计存在一定的噪声，设估计值与真实值之间的差值为 $\mathbf{e}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t$，该误差值符合高斯分布，均值为 0（差值是一个向量，每个元素是各个状态的估计值和真实值的差）。我们按照前面的小车前进的例子：

$$\mathbf{e}_t = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} x_t \\ v_t \end{bmatrix} - \begin{bmatrix} \hat{x}_t \\ \hat{v}_t \end{bmatrix} \tag{四.8}$$

根据误差的分布，得到误差的协方差矩阵为：

$$\mathbf{P}_t = E(\mathbf{e}_t\mathbf{e}_t^T) = \begin{bmatrix} cov(e_1,e_1) & cov(e_1,e_2) \\ cov(e_2,e_1) & cov(e_2,e_2) \end{bmatrix} \tag{四.9}$$

我们希望寻找合适的 $\mathbf{K}_t$，使得方差最小，也就是协方差矩阵的迹 $cov(e_1,e_1)+cov(e_2,e_2)$ 最小。换言之，希望矩阵 $\mathbf{P}_t$ 的迹（对角线元素之和）最小。

我们把 $\hat{\mathbf{x}}_t$ 的估计式（后验估计 (4) 式）和观测状态关系式 (6) 代入到 $\mathbf{x}_t - \hat{\mathbf{x}}_t$ 中，就能得到：

$$\begin{aligned} \mathbf{e}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t &= (\mathbf{I} - \mathbf{K}_t\mathbf{H})(\mathbf{x}_t - \mathbf{x}_t^-) - \mathbf{K}_t\mathbf{v}_t \\ &= (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{e}_t^- - \mathbf{K}_t\mathbf{v}_t \end{aligned} \tag{四.10}$$

注意这里的 $(\mathbf{x}_t - \mathbf{x}_t^-)$ 表示为先验估计误差 $\mathbf{e}_t^-$。因此：

$$\begin{aligned} \mathbf{P}_t &= E(\mathbf{e}_t\mathbf{e}_t^T) \\ &= E\Big[(\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{e}_t^-(\mathbf{e}_t^-)^T(\mathbf{I} - \mathbf{K}_t\mathbf{H})^T\Big] - E\Big[(\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{e}_t^-\mathbf{v}_t^T\mathbf{K}_t^T\Big] \\ &\quad - E\Big[\mathbf{K}_t\mathbf{v}_t(\mathbf{e}_t^-)^T(\mathbf{I} - \mathbf{K}_t\mathbf{H})^T\Big] + E\Big[\mathbf{K}_t\mathbf{v}_t\mathbf{v}_t^T\mathbf{K}_t^T\Big] \end{aligned} \tag{四.11}$$

注意，当两个变量 A 和 B 互相独立，则 $E(AB) = E(A)E(B)$。而观测噪声 $\mathbf{v}_t$ 和状态估计噪声 $\mathbf{e}_t^-$ 是相互独立的，所以：

$$E\Big[(\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{e}_t^-\mathbf{v}_t^T\mathbf{K}_t^T\Big] = (\mathbf{I} - \mathbf{K}_t\mathbf{H})E\big[\mathbf{e}_t^-\big]E\big[\mathbf{v}_t^T\big]\mathbf{K}_t^T = 0 \tag{四.12}$$

也就是前面的式子中第二三项都是 0，另外，$E\big[\mathbf{e}_t^-(\mathbf{e}_t^-)^T\big] = \mathbf{P}_t^-$，$E\big[\mathbf{v}_t\mathbf{v}_t^T\big] = \mathbf{R}$：

$$\begin{aligned} \mathbf{P}_t &= E\Big[(\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{e}_t^-(\mathbf{e}_t^-)^T(\mathbf{I} - \mathbf{K}_t\mathbf{H})^T\Big] + E\Big[\mathbf{K}_t\mathbf{v}_t\mathbf{v}_t^T\mathbf{K}_t^T\Big] \\ &= \mathbf{P}_t^- - \mathbf{K}_t\mathbf{H}\mathbf{P}_t^- - \mathbf{P}_t^-\mathbf{H}^T\mathbf{K}_t^T + \mathbf{K}_t\mathbf{H}\mathbf{P}_t^-\mathbf{H}^T\mathbf{K}_t^T + \mathbf{K}_t\mathbf{R}\mathbf{K}_t^T \end{aligned} \tag{四.13}$$

该式的第二项 $\mathbf{K}_t\mathbf{H}\mathbf{P}_t^-$ 和第三项 $\mathbf{P}_t^-\mathbf{H}^T\mathbf{K}_t^T$ 互为转置（$\mathbf{P}_t^-$ 是协方差矩阵，是一个对角线对称矩阵），所以迹相同：

$$tr(\mathbf{P}_t) = tr(\mathbf{P}_t^-) - 2tr(\mathbf{K}_t\mathbf{H}\mathbf{P}_t^-) + tr(\mathbf{K}_t\mathbf{H}\mathbf{P}_t^-\mathbf{H}^T\mathbf{K}_t^T) + tr(\mathbf{K}_t\mathbf{R}\mathbf{K}_t^T) \tag{四.14}$$

对迹求导，原理见矩阵微积分相关内容，这里给出结论：

$$\frac{d\ tr(\mathbf{AB})}{d\ \mathbf{A}} = \mathbf{B}^T$$
$$\frac{d\ tr(\mathbf{ABA}^T)}{d\ \mathbf{A}} = 2\mathbf{A}(\mathbf{B} + \mathbf{B}^T)$$

可以得到：

$$\frac{d\ tr(\mathbf{P}_t)}{d\ \mathbf{K}_t} = 0 - 2(\mathbf{H}\mathbf{P}_k^-)^T + 2\mathbf{K}_t\mathbf{H}\mathbf{P}_t^-\mathbf{H}^T + 2\mathbf{K}_t\mathbf{R} = 0 \tag{四.15}$$

因此：

$$\mathbf{K}_t = \frac{\mathbf{P}_t^-\mathbf{H}^T}{\mathbf{H}\mathbf{P}_t^-\mathbf{H}^T + \mathbf{R}} \tag{四.16}$$

4 3 更新误差协方差公式

我们上一节已经得到了 $\mathbf{P}_t$ 的公式，以及 $\mathbf{K}_t$ 的表达式，因此，将 $\mathbf{K}_t$ 代入到 $\mathbf{P}_t$ 中，化简得到：

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_t^- \tag{四.17}$$

五 总结

至此，本文已经将卡尔曼滤波的很多方面都进行了解释和阐述。然而，这里的卡尔曼滤波仅仅是广义卡尔曼滤波的最简单的情况，对于非线性系统的滤波会更复杂，我们留到后面再慢慢介绍。

由于文章公式确实比较繁琐，出错之处也或许很难避免，还请读者批评指正。

参考文献

[1] Pizzinga A. Restricted Kalman filtering: Theory, methods, and application[M]. Springer Science & Business Media, 2012.

[2] Leach B W. An Introduction to Kalman Filtering[J]. National Aeronautical Establishment, NAE-MISC 57, 1984.

[3] Chui C K, Chen G. Kalman filtering[M]. Berlin, Germany: Springer International Publishing, 2017.

[4] https://space.bilibili.com/1337726246

[5] https://space.bilibili.com/352976834

[6] Zarchan P, Musoff H. Fundamentals of Kalman filtering: a practical approach, vol. 190[J]. Progress in Astronautics and Aeronautics, American Institute of Aeronautics and Astronautics (AIAA), 2000.