

用一小时、十分钟、一分钟和一句话 阐释滤波（估计）*

何毓琦

哈佛大学与清华大学

卡尔曼滤波称得上是现代系统科学和控制论的皇冠级成果(卡尔曼因此于2008年获得了被誉为工程诺贝尔奖的德雷珀奖)。一般来说,无论从复杂系统控制与优化的哪方面讲,噪声和扰动情形下的估计都是一个关键问题。过去,我的博客及发表过的论文曾涉及和讨论了这类问题的方方面面^[1,2]。然而,这些材料分散在各处,且时日已久。因此,将所有这些资料整合起来,并结合最新的文献呈现给读者,似乎是一件值得去做的事情。同时,作为一次尝试,本文希望向非专业的普通科学公众阐释这一问题。

读者仅需高中代数而无需概率论知识,便可理解这篇文章的绝大部分内容——包括非线性滤波和估计。至于概率论的概念和解释,我此前有关概率论和随机过程的五篇讲

座博文的内容就足够了^[3,4,5,6,7]。了解概率论中的初等贝叶斯准则会有帮助,但是本文中我也将讲解这一问题。

首先,我们不用概率论或其它知识基础,来较为直观地解释卡尔曼滤波器。考虑如下一个简单的估计问题:已有关于一个未知常数 x 的一组测量值 z_1, z_2, \dots, z_k (通常简记为 z_1^k)。假设有如下可加模型:

$$z_i = x + v_i \quad i = 1, 2, \dots, k, \quad (1)$$

其中 v_i 是测量噪声。在给定 k 个测量值 z_1^k 的条件下,如果没有任何别的信息,则对 x 的一种合理估计是:

$$\hat{x}_k = \frac{1}{k} \sum_{i=1}^k z_i. \quad (2)$$

*该文是何毓琦教授发表在科学网上一篇博文的中文翻译稿,原文请见:

<http://blog.sciencenet.cn/home.php?mod=space&uid=1565&do=blog&id=851754>,

译者是中国科学院数学与系统科学研究院万林副研究员。

由于测量噪声的值可正可负，我们通过取其平均值，使得噪声正负相互抵消（注意，如果噪声的平均值是未知的，那么该平均值与未知量 x 无异，并且无法与 x 区分开来，从而在技术上将我们处于一种不可观测的情形。习题 1：你是否可以将上述注释解释给其他读者？当 k 趋向无穷时，公式（2）的结果是什么？）

现在，通过简单的代数运算，我们可以将公式(2)重写为：

$$\begin{aligned}\hat{x}_k &= \frac{1}{k} \sum_{i=1}^{k-1} z_i + \frac{1}{k} z_k = \frac{k-1}{k} \frac{1}{k-1} \sum_{i=1}^{k-1} z_i + \frac{1}{k} z_k \\ &= \hat{x}_{k-1} + \frac{1}{k} (z_k - \hat{x}_{k-1}).\end{aligned}\quad (3)$$

公式（3）尽管只是公式（2）的一种递推形式，但它却能提供一种有趣的解释——即在 k 个观测值后对 x 的估计，等于 $k-1$ 个观测值后对 x 的估计外加一个修正项。修正项指的是以 $k-1$ 个观测值为基础的估计值 \hat{x}_{k-1} 与测量值 z_k 之间的差。该差值归因于以下两个来源：

(i) 当 \hat{x}_{k-1} 与 x 不同时，需要使用修正项校正新的估计值 \hat{x}_k ；

(ii) 当 \hat{x}_{k-1} 与 x 相等时，应忽略仅仅因噪音 v_k 而引起的修正项。

系数 $1/k$ ，称为修正项的权重因子，体现了以上两种可能性之间合理的权衡。在开始阶段，即在小 k 值的情况下，无法对 x 进行准确的估计，需要重点考虑修正项。随着 k 逐

渐变，对 x 的估计也越来越可信，这时只需少量考虑修正项就可以了（如果熟悉概率论，你会联想起大数定律）。当 k 趋向无穷时，情况（ii）适用，我们就可完全忽略修正项。如果我们将 $1/k$ 表示为 P_k ，那么通过简单的代数运算可将 P_k 用以下递推形式表示：

$$P_k = P_{k-1} - P_{k-1}(P_{k-1} + 1)^{-1}P_{k-1}. \quad (4)$$

信不信由你，在这种简单情形下，方程（3-4）就是“卡尔曼滤波”的表达公式。对熟悉“随机逼近”文献的读者，方程（3）就是 $f(x) = x$ 情形下利用梯度下降的思想不断逼近函数 $f(x)$ 的根。公式（1）也可以视为简单的最小二乘拟合问题

$$\text{Min}_{\hat{x}_k} \frac{1}{2} \sum_{i=1}^k (z_i - \hat{x}_k)^2 \quad (5)$$

的解。我们可以将以上最小二乘法（或随机逼近、大数定律、卡尔曼滤波器）的思想推广至更多深刻而有用的情形。下面我们举例说明。

首先，假设 \mathbf{x} 为一个向量，那么根本无需改变公式（3-4）。实际上，我们可以将公式（1）推广为：

$$\mathbf{z}_i = H\mathbf{x} + v_i \quad i = 1, 2, \dots, k. \quad (1')$$

其中 H 是由标量因子构成的矩阵。注意，这里 \mathbf{x} 和 \mathbf{z} 的维数无须相同。例如， \mathbf{x} 可以是飞行器在天空中的位置和速度的六维变量，而 \mathbf{z} 为测量该飞行器与地面某点之间角度关系的一维变量（在本例中 H 为一个 1×6 行向量）。现在公式（3-4）变为：

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + P_k(\mathbf{z}_k - H\hat{\mathbf{x}}_{k-1}), \quad (3')$$

$$P_k = P_{k-1} - P_{k-1}H^T(HP_{k-1}H^T + I)^{-1}HP_{k-1}, \quad (4')$$

其中 P_k 是一个 $n \times n$ 矩阵, 上标 T 表示“转置”。(注意, 通过将全部变量视为标量, 公式 (3-4) 的正确性显而易见)

我们可以将公式 (3-4) 进一步扩展到动态方程的情形。假设未知量 \mathbf{x} 根据如下的方程演化:

$$\mathbf{x}_i = \Phi\mathbf{x}_{i-1}, \quad i=1, 2, \dots, k \quad \text{给定 } \mathbf{x}_0 = \mathbf{x} \quad (5')$$

这时我们想知道如何进一步修改公式 (3'-4')。此时, 虽然代数运算变得有些繁琐, 但基本原则是相同的 (即最小二乘拟合)。我们希望基于已知的全部测量值, 得到关于未知量 \mathbf{x} 的估计 $\hat{\mathbf{x}}_k$ 。为了简化运算, 我们考虑

$$k=2 \text{ 的情况, 即我们希望求解:} \\ \text{Min}_{\mathbf{x}} \frac{1}{2} [(H\Phi\mathbf{x} - \mathbf{z}_2)^2 + (H\mathbf{x} - \mathbf{z}_1)^2]. \quad (6)$$

上述关于 \mathbf{x} 的最优解为

$$\hat{\mathbf{x}}_2 = \hat{\mathbf{x}}_1 + P_2H(\mathbf{z}_2 - H\Phi\hat{\mathbf{x}}_1); \quad \hat{\mathbf{x}}_1 = \mathbf{z}_1. \quad (3'')$$

$$P_2 = M_2 - M_2H^T(HM_2H^T + I)^{-1}HM_2;$$

$$M_2 = \Phi P_1 \Phi^T; \quad P_1 = I. \quad (4'')$$

于是公式 (3'-4') 变为公式 (3''-4'')。

公式 (5) 可进一步推广为:

$$\mathbf{x}_i = \Phi\mathbf{x}_{i-1} + \mathbf{w}_{i-1}, \quad (5'')$$

其中 \mathbf{w}_{i-1} 为上述动态方程中扰动噪声项。最后, 令 \mathbf{x} 为一个 n 维向量, 并令噪声项和扰动项带有不同的权重“ R ”和“ Q ”, 那么我们仅需对其中的矩阵加转置, 将数字 1 变为单位矩阵, 同时插入 Q 和 R 权重项, 最后得到:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + P_k H^T R^{-1} (\mathbf{z}_k - H\Phi\hat{\mathbf{x}}_{k-1}); \quad (3^*) \\ \hat{\mathbf{x}}_1 = \text{给定}.$$

$$P_k = M_k - M_k H^T (H M_k H^T + R)^{-1} H M_k; \\ M_k = \Phi P_{k-1} \Phi^T + Q; \quad P_1 = \text{给定}. \quad (4^*)$$

无疑, 公式 (3*-4*) 展现了卡尔曼滤波公式的全貌。我尤其想强调的是, 以上推导仅需一些矩阵代数运算, 而不涉及矩阵理论。同时, 公式 (3*-4*) 也可以通过求解如下一般的最小二乘问题并施以复杂的矩阵代数运算后得到:

$$\text{Min}_{\hat{\mathbf{x}}_i} \frac{1}{2} \left\{ \sum_{i=1}^k (\hat{\mathbf{x}}_i - \Phi\hat{\mathbf{x}}_{i-1})^T Q (\hat{\mathbf{x}}_i - \Phi\hat{\mathbf{x}}_{i-1}) \right. \\ \left. + \sum_{i=1}^k (\mathbf{z}_i - H\hat{\mathbf{x}}_i)^T R (\mathbf{z}_i - H\hat{\mathbf{x}}_i) \right\}. \quad (6^*)$$

这里矩阵项 (或标量因子) Q 和 R 表示噪声项 \mathbf{w} 和 \mathbf{v} 的相应权重。

通过解决上述最小化问题 (除了使用矩阵代数运算外, 没有涉及其它新的概念), 我们也直接得到了熟知的卡尔曼滤波器公式 (3*-4*)。

图 1 展示了一个关于卡尔曼滤波器实现过程的方框图:



图1 卡尔曼滤波器方框图

因此，可以用一句话来概括卡尔曼滤波器理论的精髓：“**删除所有已知的影响因素，并对余下因素取平均值**”。这是唯一能做的最佳选择。卡尔曼滤波器不过是上述语句的一种数学表述，它是一种绝妙的最小二乘拟合，也是大数定律的一种非常精致的应用。

贝叶斯决策理论也可从概率的角度对卡尔曼滤波器做出非常令人满意的阐释^[8,9]。稍后，我们将最小二乘拟合的概念推广到非线性情形后，再讲解这一内容（参考有关确定性非线性滤波的内容^[2]）。

严格意义上讲，卡尔曼滤波器只适用于带有高斯型扰动和测量噪声的线性系统。由于大量实际问题是非线性和非高斯型的，便自然而然地产生了关于非线性滤波和估计的问题以及与之相关的问题：“在现实中我们该怎么办？”

有关非线性估计，目前最为通行的方法是所谓的**扩展卡尔曼滤波**——即首先对非线

性系统猜测性地估计其轨迹。在估计的轨迹附近进行线性化处理，并且应用通常的卡尔曼滤波器公式于该线性化的系统。你可以通过新的数据和重新线性化不断地递推和更新估计轨迹。一切看起来很有效。唯一的问题是，卡尔曼滤波是基于单峰（高斯型）分布的。如果你的问题涉及和（或）导致多峰分布，你就无法通过不断的线性化逼近这类严重的非线性情形，从而这种方法将无法适用或将得出很坏的估计。

第二种方法是**非线性最小二乘数据拟合**。这种方法可以追溯到我们最初提到的观点，即最优滤波本质上是**最小二乘拟合**。因此，这一方法相当稳健有效。唯一的问题是，它无法实时估计。对每一个新的测量值，都需要解决一个新的两点边界值问题。在测量值较少且观测间隔较长的情形下，这一方法比较有用（例如漫长的太空航行）。否则，除非很少更新，这种方法的计算量极大，在实时情形下缺少实用性。

第三种方法称为粒子滤波。我们将其放在本教程的概率方法部分进行讨论。

到这里，在无需概率论知识的情形下，关于实际滤波问题，我想你们已经了解了全部值得了解的内容。

(第一部分结束)

现在，我们从概率的角度将以上的公式重新推导：

当我们讲 \mathbf{x} 是一个随机变量时，意味着被采样之前，该变量的实际值是未知的。我们使用概率密度函数 (pdf, $p(\mathbf{x})$)，即采样值介于 \mathbf{x} 和 $\mathbf{x} + d\mathbf{x}$ 之间的概率，来表示这种可变性。可以说，**概率是用来精确地处理不确定性问题的方法**。作为确定性函数，概率密度函数 (pdf, $p(\mathbf{x})$) 正是此类精确方法。目前，即便可以对一个多变量函数进行计算，也是非常麻烦的^[10]（这是应用数学论文和教科书中鲜少提及的难言困惑）。为了简化随机变量的描述，我们通常使用其均值和方差来刻画一个概率密度函数的粗略特征（即平均值和以平均值为中心的扩展程度）。我在此前关于概率论的教程中已经解释了上述基础概念（请阅读第一部分提到的五篇教程）。

在系统研究中，我们对所谓的输入输出关系甚感兴趣。如果一个系统的输入是随机的，那么输出也将是随机的。基于输入的概率密度函数，我们希望了解输出的概率密度函数。特别地，我们希望知道当 $\mathbf{y} = f(\mathbf{x})$ 且

给定 $p(\mathbf{x})$ 时 $p(\mathbf{y})$ 是什么样子。下面对公式 (3'-4') 和 (5) 或 (5') 在概率情形下进行阐述。

(i) **传播方程**。假设 $\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{w}_{k-1})$ ，这是公式 (5) 的非线性输入输出的推广。那么，给定 $p(\mathbf{x}_{k-1})$ 和 $p(\mathbf{w}_{k-1})$ ，我们得到：

$$p(\mathbf{x}_k) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}, \quad (5'')$$

这里，原则上在给定 $p(\mathbf{w}_k)$ 及关系式 $\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{w}_{k-1})$ 的情况下， $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ 是可知的。

(ii) **更新方程**。假设 $\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k)$ ，这是公式 (1') 的非线性推广。于是根据贝叶斯准则，公式(4) (或 4' 或 4*) 的概率表述为：

$$p(\mathbf{x} | \mathbf{z}) = \frac{p(\mathbf{z} | \mathbf{x})}{p(\mathbf{z})} p(\mathbf{x}), \quad (4'')$$

其中 $p(\mathbf{x})$ 是 \mathbf{x} 的所谓先验概率密度函数， $p(\mathbf{x} | \mathbf{z})$ 是 \mathbf{x} 的后验概率密度函数，即在测量 \mathbf{z} 条件下的 \mathbf{x} 的概率。同样，已知 $p(\mathbf{v}_k)$ 和关系式 $\mathbf{z}_k = h(\mathbf{x}_k, \mathbf{v}_k)$ ，我们可以原则上给出 $p(\mathbf{z} | \mathbf{x})$ 和 $p(\mathbf{z}) = \int p(\mathbf{z} | \mathbf{x}) d\mathbf{x}$ 。

注意，我们讲的是“原则上”与公式(4'') 和 (5'') 的关联。这里值得再一次强调应用数学的难言之困惑^[10]，即在计算上基本不可能处理一般的多变量方程。如果 $\mathbf{x}, \mathbf{z}, \mathbf{w}$ 及 \mathbf{v} 都是 n 维向量的话，公式 (4'') 和 (5'') 看似简单，实则难以计算。因此，我们寻求对

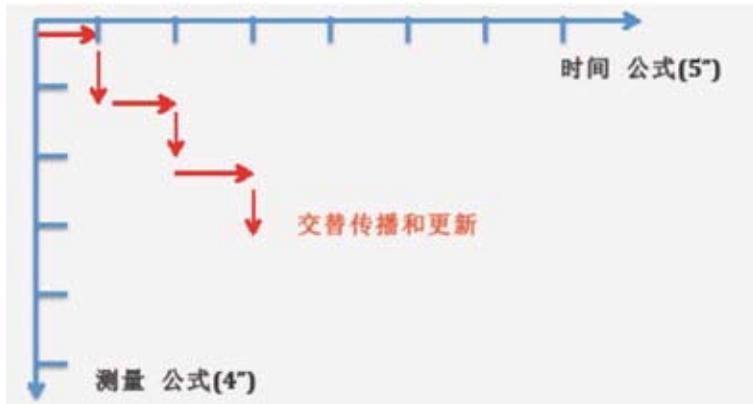


图2 动态系统传播和更新概率密度函数示意图

于概率密度函数的参数化描述,以解决公式(4"-5")的计算上的困境。为此,我们引入共轭概率密度函数或再生概率密度函数的概念。例如,如果公式(5")中的 $p(\mathbf{x}_{k-1})$ 和 $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ 满足高斯分布,那么公式(5")中的 $p(\mathbf{x}_k)$ 同样为高斯分布。进而,我们可以用一个有限维方程表达公式(5"),使后验概率密度函数的参数与先验概率密度函数相关联。实际上,如果 $p(\mathbf{x}_{k-1})$ 为高斯分布,均值为 $\hat{\mathbf{x}}_{k-1}$ 及方差为 P_{k-1} ,且 \mathbf{x}_k 和 \mathbf{x}_{k-1} 通过公式(5')线性相关,那么对公式(3*-4*)进行复杂的代数运算便可证明 $p(\mathbf{x}_k)$ 同样为高斯分布,并可知其均值和方差。同理,我们可以证明公式(5)中的 $p(\mathbf{x})$ 和 $p(\mathbf{x} | \mathbf{z})$ 亦为再生或共轭配对。如果 $p(\mathbf{x})$ 满足高斯分布,则 $p(\mathbf{x} | \mathbf{z})$ 也是高斯分布,其先验均值及方差与后验均值及方差通过公式(3*-4*)相关联。换句话讲,卡尔曼滤波器仅仅是高斯再生情形下对公式(4"-5")的传播和更新形式。在我们的心目中,可以构建一个关于动态系统传播和更新的概率密度函数,如图2所示。

在我45年前著的《应用最优化控制》一书^[9]第12章中的第322-328页,可以找到一个比上图更为形象化的示意图。

总而言之,我们考虑的动态系统和观测方程无法满足卡尔曼滤波严格的线性要求。那么,如何处理公式(4"-5")呢?一个显而易见的方法是使用离散直方图(即抽样的离散化)来逼近概率密度函数,进而实现公式(4"-5")。这就是所谓的粒子滤波——这一方法的领军人物之一是雷神(Raytheon)公司的Fred Daum博士,他拥有超过20年的实践经验和众多成功案例。那么,什么是粒子滤波呢?用一句话来说——试图通过抽取样本(粒子)来传播和更新关于系统状态的多维密度函数。当然,你是通过由状态向量的实际样本组成的直方图来逼近多维密度函数的。传播直方图(即组成直方图的粒子)计算起来并不是那么复杂(想想看,只是针对粒子运行蒙特卡洛计算)。但是,使用新的测量来更新直方图就不那么容易了。首先,如果状态维数很高,为了使直方图具有充分的代表性,你需要指数量级的粒子数。这是应用数学家不愿谈及的普遍性困惑^[10]。更棘手

的是，更新方程中用到的贝叶斯准则要求先验密度与似然函数相乘，而后者在实际测量值附近会极速地下降。因此，只有一小部分粒子会得到显著的更新。这种计算相当低效。Daum 博士的专长和贡献在于，他发展了一些技巧用于改善这一问题（他近期与 Huang 在一篇发表于 SPIE 文集的文章中详细阐述了他们最为成功的技术手段^[11]）。这一方法仍然面临维度灾难这一局限。幸运的是，许多实际问题只有 6 个维度的状态向量（例如航天制导和应用控制中的 3 个位置和 3 个速度），这就减轻了计算负担。实际上，事后想来，在我半个世纪前出版的那本教科书的第 381 页上的非线性滤波的例子，就是一个极其简单（只有两个可能状态）的粒子滤波问题。只不过，在大约 50 年前，粒子滤波这一名词还没出现，所以书中没有标示。

综上所述，以下是就我所知道的有关滤波或估计的全部概念的阐释：

1. 它只不过是以复杂精妙形式展现的简单的最小二乘拟合，

2. 它的基础是有关概率密度函数的传播和更新的两个简单的概率公式，

3. 卡尔曼滤波是上述 1 或 2 在线性系统和(或)再生高斯概率密度函数中的具体运用。

根据听众和时间长短的不同，你可以任选本文的某个部分来阐释卡尔曼滤波——50 多年来系统控制领域的皇冠级成就和德雷珀奖工作。

理论上讲，若想严格推导连续时间下的非线性滤波方程，需要谨慎运用测度论这一工程领域大多数人难以理解的工具。但是，在实际应用中，由于使用数字计算机进行离散时间的计算，测度论的知识变得不再必不可少。我们虽然没有提及这些知识，但不能掩盖了他们的理论贡献。

参考文献

- [1] Ho Y C. Proximity to Fame (#2) – the 2008 Draper Prize. <http://blog.sciencenet.cn/blog-1565-14253.html>, 2008.
- [2] Ho Y C. A Short Course on Nonlinear Filtering. <http://blog.sciencenet.cn/blog-1565-426323.html>, 2008.
- [3] [3] Ho Y C. Probability and Stochastic Process Tutorial (1). <http://blog.sciencenet.cn/blog-1565-664051.html>, 2013.
- [4] Ho Y C. Probability and Stochastic Process Tutorial (2). <http://blog.sciencenet.cn/blog-1565-665359.html>, 2013.
- [5] Ho Y C. Probability and Stochastic Process Tutorial (3). <http://blog.sciencenet.cn/blog-1565-666599.html>, 2013.
- [6] Ho Y C. Probability and Stochastic Process Tutorial (4). <http://blog.sciencenet.cn/blog-1565-669532.html>, 2013.
- [7] Ho Y C. Probability and Stochastic Process Tutorial (5). <http://blog.sciencenet.cn/blog-1565-671318.html>, 2013.
- [8] Ho Y C & Lee R C K A. Bayesian approach to problems in stochastic estimation and control. IEEE Trans. on Automatic Control, 1964, 9(Oct.): 333–

339.

- [9] Bryson A E & Ho Y C. Applied Optimal Control: Optimization, Estimation and Control. Taylor & Francis, 1969.
- [10] Ho Y C. Some Dirty Secrets of Applied Mathematics. <http://blog.sciencenet.cn/blog-1565-26889.html>, 2008.
- [11] Daum F & Huang J. How to avoid normalization of particle flow for nonlinear filters: Bayesian decisions and transport. Proceedings of SPIE 2014.
- [12] Ho Y C. On stochastic approximation methods and optimal filtering. J. of Math. Anal. and Application , 1963, 6(1): 152-154.
- [13] Kalman R E , Ho Y C, & Narendra K.S. Controllability of linear dynamical systems. Contributions to Diff. Equation , 1963, 1(Interscience) : 189-213.
- [14] Kalman R E. A new approach to linear filtering and prediction problems. Transactions of the ASME- Journal of Basic Engineering, 1960, 82 (Series D): 35-45.
- [15] Kalman R E. Discovery and invention: the newtonian revolution in system technology. Journal of Guidance and Dynamics, 2003, 26(6): 1-7.