

参考链接：

[维基百科](#)

[Some tutorials, references, and research related to the Kalman filter](#)

[说说卡尔曼滤波](#)

[Kalman滤波器从原理到实现](#)

历史

Filtering is weighting

滤波的作用就是给不同的信号分量不同的权重

Kalman Filter 是一个最优线性估计器(Estimator)

Kalman滤波器的历史，最早要追溯到17世纪，Roger Cotes开始研究最小均方问题。但由于缺少实际案例的支撑，Cotes的研究让人看着显得很模糊，因此在估计理论的发展中影响很小。17世纪中叶，最小均方估计（Least squares Estimation）理论逐步完善，1777年，77岁的Daniel Bernoulli 发明了最大似然估计算法。递归的最小均方估计理论是由Karl Gauss建立在1809年，当时还有Adrien Legendre在1805年完成了这项工作，Robert Adrain在1808年完成的。

在1880年，丹麦的天文学家Thorvald Nicolai Thiele在之前最小均方估计的基础上开发了一个递归算法，与Kalman滤波非常相似。在某些标量的情况下，Thiele的滤波器与Kalman滤波器时等价的，Thiele提出了估计过程噪声和测量噪声中方差的方法（过程噪声和测量噪声是Kalman滤波器中关键的概念）。

在19世纪40年代，Wiener设计了Wiener滤波器，然而，Wiener滤波器不是在状态空间进行的，Wiener是稳态过程，它假设测量是通过过去无限多个值估计得到的。Wiener滤波器比Kalman滤波器具有更高的自然统计特性。这些也限制其只是更接近理想的模型，要直接用于实际工程中需要足够的先验知识（要预知协方差矩阵），美国NASA曾花费多年的时间研究维纳理论，但依然没有在空间导航中看到维纳理论的实际应用。

在1950末期，大部分工作开始对维纳滤波器中协方差的先验知识通过状态空间模型进行描述。通过状态空间表述后的算法就和今天看到的Kalman滤波已经极其相似了。Peter Swerling实际上在1959年已经推导出了无噪声系统动力学的Kalman滤波器，在他的应用中，他还考虑了使用非线性系统动力学和测量方程。可以这样说，Swerling和发明Kalman滤波器是失之交臂。总结其失之交臂的原因，主要是Swerling没有直接在论文中提出Kalman滤波器的理论，而只是在实践中应用。

Rudolph Kalman在1960年发现了离散时间系统的Kalman滤波器，1961年Kalman和Bucy又推导了连续时间的Kalman滤波器。Ruslan Stratonovich也在1960年也从最大似然估计的角度推导出了Kalman滤波器方程。

目前，卡尔曼滤波已经有很多不同的实现。卡尔曼最初提出的形式现在一般称为简单卡尔曼滤波器。除此以外，还有施密特扩展滤波器、信息滤波器以及很多Bierman, Thornton开发的平方根滤波器的变种。也许最常见的卡尔曼滤波器是锁相环，它在收音机、计算机和几乎任何视频或通讯设备中广泛存在。

举例

假设有一辆质量为 m 的小车，受恒定的力 F ，沿着一个方向做匀加速直线运动。已知小车在 $t - \Delta T$ 时刻的位移是 $s(t - 1)$ ，此时的速度为 $v(t - 1)$ 。求： t 时刻的位移是 $s(t)$ ，速度为 $v(t)$ ？

由牛顿第二定律，求得加速度： $a = \frac{F}{m}$

那么就有下面的位移和速度关系：

$$s(t) = s(t - 1) + v(t - 1)\Delta T + \frac{1}{2} \frac{F}{m} (\Delta T)^2$$

$$v(t) = v(t - 1) + \frac{F}{m} \Delta T$$

写成矩阵的形式，就有：

$$\begin{pmatrix} s(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} 1 & \Delta T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s(t-1) \\ v(t-1) \end{pmatrix} + \begin{pmatrix} \frac{(\Delta T)^2}{2} \\ \Delta T \end{pmatrix} \frac{F}{m} \quad \dots\dots (1)$$

卡尔曼滤波器是建立在动态过程之上，由于物理量（位移，速度）的不可突变特性，这样就可以通过 $t - 1$ 时刻估计（预测） t 时刻的状态，其状态空间模型为（加入了噪声）：

$$x(t) = Ax(t - 1) + Bu(t) + w(t) \quad \dots\dots (2)$$

其中：

$$A = \begin{pmatrix} 1 & \Delta T \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} \frac{(\Delta T)^2}{2} \\ \Delta T \end{pmatrix}$$

匀加速直线运动过程就是卡尔曼滤波状态空间模型的一个典型应用。我们将（2）式表示成离散形式：

$$x(n) = Ax(n - 1) + Bu(n) + w(n) \quad \dots\dots (3)$$

各个量的含义是：

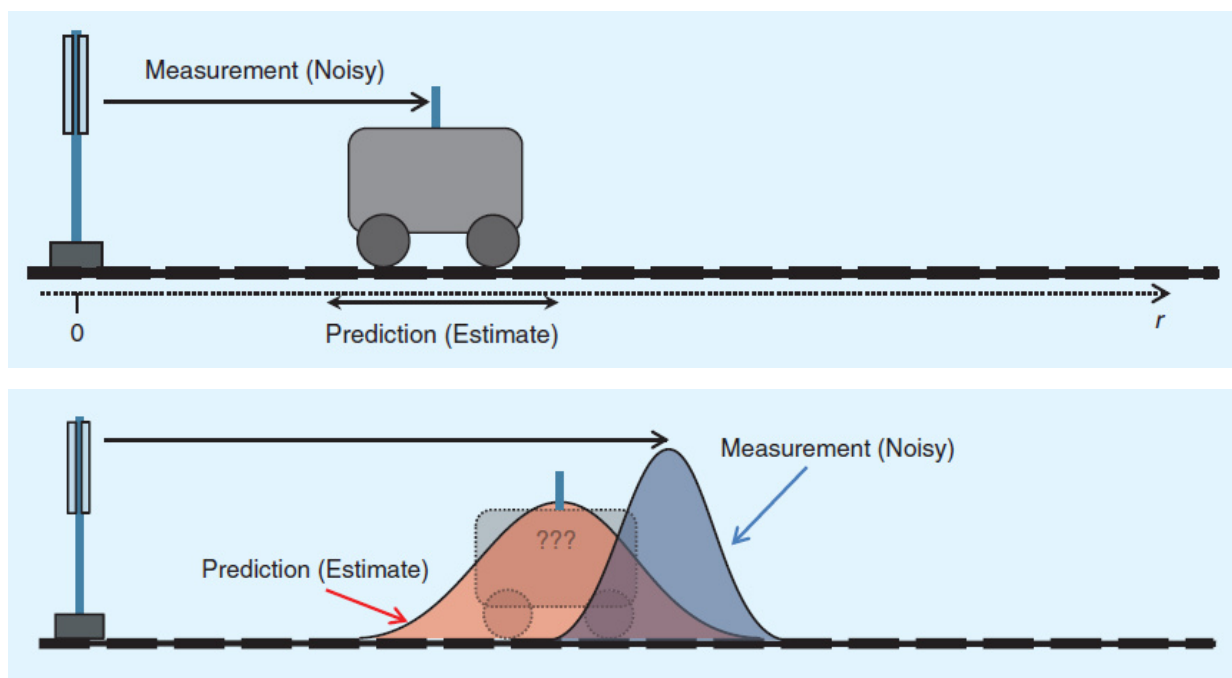
- $x(n)$ 是状态向量，包含了观测的目标（如：位移、速度）
- $u(n)$ 是驱动输入向量，如上面运动过程是通过受力驱动产生加速度，所以 $u(n)$ 和受力有关
- A 是状态转移矩阵，其隐含指示了“ $n - 1$ 时刻的状态会影响到 n 时刻的状态”
- B 是控制输入矩阵，其隐含指示了“ n 时刻给的驱动如何影响 n 时刻的状态”
- 从运动的角度，很容易理解：小车当前 n 时刻的位移和速度一部分来自于 $n - 1$ 时刻的惯性作用，这通过 $Ax(n)$ 来度量，另一部分来自于现在 n 时刻小车新增加的外部受力，通过 $Bu(n)$ 来度量。
- $w(n)$ 是过程噪声， $w(n) \sim N(0, Q)$ 的高斯分布，过程噪声是使用卡尔曼滤波器时一个重要的量，后面会进行分析。

计算 n 时刻的位移，还有一种方法：拿一把长的卷尺，从起点一拉，直接就出来了，设测量值为 $z(n)$ 。计算速度呢？速度传感器往那一用就出来了。

然而，初中物理就告诉我们，测量存在误差，我们暂且将这个误差记为 $v(n)$ 。这种通过直接测量的方式获得所需物理量的值构成观测空间：

$$z(n) = H(n)x(n) + v(n) \quad \dots\dots (4)$$

$z(n)$ 就是测量结果， $H(n)$ 是观测矢量， $x(n)$ 就是要求的物理量（位移、速度）， $v(n) \sim N(0, R)$ 为测量噪声。大部分情况下，如果物理量能直接通过传感器测量， $H(n) = I$ 。



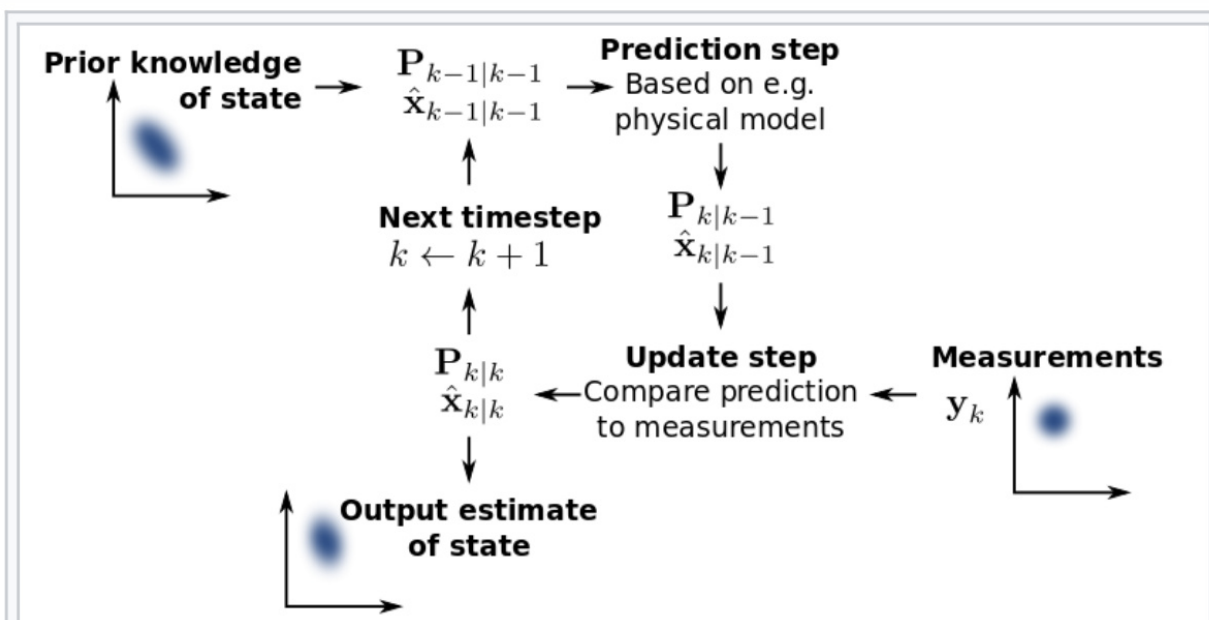
现在有了两种方法（如上）可以得到 n 时刻的位移和速度：一种就是通过（3）式的状态空间递推计算（Prediction），另一种就是通过（4）式直接测量得到（Measurement）。然而这两种方式没一个是精确无误的，都存在 0 均值高斯分布的误差 $w(n)$ 和 $v(n)$ 。

那么我们最终的结果应该取哪个比较好呢？

Kalman 滤波器的具体过程

图片来源：维基百科

Kalman 滤波器不停的迭代状态矢量以及误差的协方差阵（也就是状态矢量预测值的不确定性）*



The Kalman filter keeps track of the estimated state of the system and the variance or uncertainty of the estimate. The estimate is updated using a state transition model and measurements. $\hat{x}_{k|k-1}$ denotes the estimate of the system's state at time step k before the k -th measurement y_k has been taken into account; $P_{k|k-1}$ is the corresponding uncertainty.

为充分利用测量值（Measurement）和预测值（Prediction），Kalman 滤波并不是简单的取其中一个作为输出，也不是求平均。

设预测过程噪声 $w(n) \sim N(0, Q)$ ，测量噪声 $v(n) \sim N(0, R)$ 。Kalman 计算输出分为预测过程和修正过程如下：

预测

预测值：

$$\hat{x}(n|n-1) = A\hat{x}(n-1|n-1) + Bu(n) \quad \dots\dots(5)$$

误差协方差阵：

$$P(n|n-1) = AP(n-1|n-1)A^T + Q \quad \dots\dots(6)$$

修正：

误差增益：

$$K(n) = P(n|n-1)H^T(n)[R(n) + H(n)P(n|n-1)H^T(n)]^{-1} \quad \dots\dots(7)$$

修正值：

$$\hat{x}(n|n) = \hat{x}(n|n-1) + K(n)[z(n) - H(n)\hat{x}(n|n-1)] \quad \dots\dots(8)$$

误差协方差阵：

$$P(n|n) = [I - K(n)H(n)]P(n|n-1) \quad \dots\dots(9)$$

(5) - (9)式为 Kalman 滤波的关键的五个公式，其中：

- $x(n) : N \times 1$ 的状态矢量
- $z(n) : M \times 1$ 的观测矢量, Kalman 滤波器的输入
- $\hat{x}(n|n-1)$: 用 n 时刻以前的数据进行对 n 时刻的估计结果
- $\hat{x}(n|n)$: 用 n 时刻及 n 时刻以前的数据对 n 时刻的估计结果, 这也是 Kalman 滤波器的输出
- $P(n|n-1) : N \times N$, 误差协方差阵, 预测 (先验) 协方差

$$P(n|n-1) = Cov(x(n) - \hat{x}(n|n-1))$$
- $K(n) : N \times M$, Kalman 增益, 从表达式看相当于“预测误差”占“测量误差与预测误差之和”的权重。(???猜测 $K(n)$ 中的 $H(n)$ 是否由于后续【(8)式中】会被约去, 故而消去了)
- $P(n|n) : N \times N$, 修正后的误差协方差阵

$$P(n|n) = Cov(x(n) - \hat{x}(n|n))$$

测量值与估计值的巧妙融合

我们知道 Kalman 滤波器是充分结合了估计值和测量值得到 n 时刻更接近真值的估计方法, 那么它是如何结合的呢?

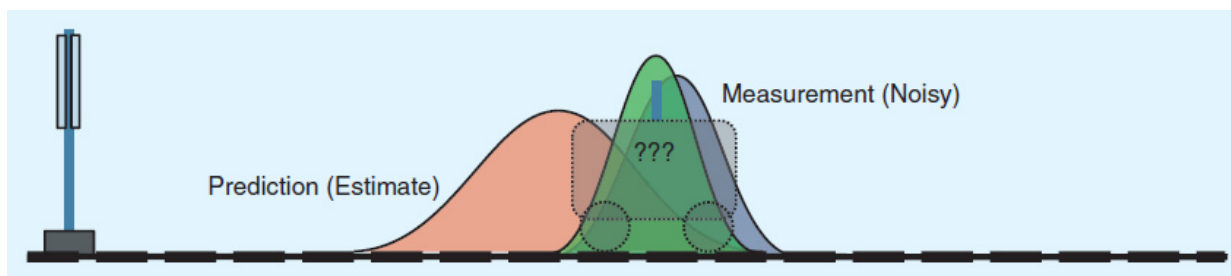
Kalman 滤波器有个假设是“独立高斯分布”。

(对一维的情形):

也就是假设测量值 $z(n) \sim N(\mu_z, \sigma_z^2)$, 估计值 $x(n) \sim N(\mu_x, \sigma_x^2)$ 。

Kalman 滤波器巧妙的使用“独立高斯分布的乘积”将这测量值与估计值进行了融合。

如下图:



$$y = \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{(r-\mu_z)^2}{2\sigma_z^2}} * \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(r-\mu_x)^2}{2\sigma_x^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r-\mu)^2}{2\sigma^2}}$$

计算之后可得:

$$\mu = \frac{\mu_x \sigma_z^2 + \mu_z \sigma_x^2}{\sigma_z^2 + \sigma_x^2} = \mu_x + \frac{\sigma_x^2}{\sigma_z^2 + \sigma_x^2} (\mu_z - \mu_x) \quad \dots\dots (10)$$

$$\sigma^2 = \frac{\sigma_z^2 \sigma_x^2}{\sigma_z^2 + \sigma_x^2} = \sigma_x^2 (1 - \frac{\sigma_x^2}{\sigma_z^2 + \sigma_x^2}) \quad \dots\dots (11)$$

$$\text{令 } K = \frac{\sigma_x^2}{\sigma_z^2 + \sigma_x^2}$$

则 $\mu = \mu_x + K(\mu_z - \mu_x), \sigma^2 = \sigma_x^2 (1 - K)$ 。

与 Kalman 公式中的修正步骤相同。

部分公式推导

推导 $P(n|n)$

根据定义

$$P(n|n) = \text{Cov}(x(n) - \hat{x}(n|n))$$

代入 $\hat{x}(n|n)$ 的定义

$$P(n|n) = \text{Cov}[x(n) - (\hat{x}(n|n-1) + K(n)[H(n)x(n) + v(n) - H(n)\hat{x}(n|n-1)])]$$

化简

$$P(n|n) = \text{Cov}[(I - K(n)H(n))(x(n) - \hat{x}(n|n-1)) - K(n)v(n)]$$

由于测量误差 $v(k)$ 与其他项不相关，因此

$$P(n|n) = \text{Cov}[(I - K(n)H(n))(x(n) - \hat{x}(n|n-1))] - \text{Cov}[K(n)v(n)]$$

展开后可得到

$$P(n|n) = (I - K(n)H(n))\text{Cov}[x(n) - \hat{x}(n|n-1)](I - K(n)H(n))^T - K(n)\text{Cov}[v(n)]K(n)^T$$

代入 $P(n|n-1), R(n)$ 的定义

$$P(n|n) = (I - K(n)H(n))P(n|n-1)(I - K(n)H(n))^T - K(n)R(n)K(n)^T$$

当 $K(n)$ 为卡尔曼增益时，上述式子还可继续化简

推导 Kalman 增益

Kalman 滤波器是一个最小均方误差估计法，后验估计状态矢量的误差为

$$x(n) - \hat{x}(n|n)$$

我们寻求找到该向量的最小均方

$$\mathbb{E} [\|x(n) - \hat{x}(n|n)\|^2]$$

这等价于最小化误差协方差阵 $P(n|n)$ 的迹 (trace)，我们展开上一小节中 $P(n|n)$ 的公式：

$$\begin{aligned} P(n|n) &= P(n|n-1) - K(n)H(n)P(n|n-1) - P(n|n-1)H(n)^T K(n)^T \\ &\quad + K(n)[H(n)P(n|n-1)H(n)^T + R(n)]K(n)^T \end{aligned}$$

$$\text{令 } S(n) = H(n)P(n|n-1)H(n)^T + R(n),$$

$P(n|n)$ 的迹 (trace) 的最小值在它对 $K(n)$ 的矩阵导数为 0 时取到。对上式求导：

$$\frac{\partial \text{tr}(P(n|n))}{\partial K(n)} = -2(H(n)P(n|n-1))^T + 2K(n)S(n) = 0$$

求解上式可得：

$$K(n)S(n) = (H(n)P(n|n-1))^T = P(n|n-1)H(n)^T$$

$$K(n) = P(n|n-1)H(n)^T S(n)^{-1}$$

这个增益 $K(n)$ 便是最优 Kalman 增益。

简化后验误差协方差阵 $P(n|n)$

对 $K(n)$ 两边同时乘以 $S(n)K(n)^T$:

$$K(n)S(n)K(n)^T = P(n|n-1)H(n)^T K(n)^T$$

由于

$$P(n|n) = P(n|n-1) - K(n)H(n)P(n|n-1) - P(n|n-1)H(n)^T K(n)^T + K(n)S(n)K(n)^T$$

我们发现后两项消去了，因此

$$P(n|n) = P(n|n-1) - K(n)H(n)P(n|n-1) = [I - K(n)H(n)]P(n|n-1)$$

Kalman 滤波器的扩展

Extended Kalman filter (EKF)

在 EKF 中，状态空间模型以及观测模型可能是非线性的方程，如下：

$$x(n) = f(x(n-1), u(n)) + w(n)$$

$$z(n) = h(x(n)) + v(n)$$

在这种情况下， f, h 不能直接用来计算协方差，而是计算一个偏导数矩阵。这个过程相当于是将非线性函数在当前估计值附近线性化。

EKF 的不足：

- 当非线性函数 Taylor 展开的高阶项无法忽略时，系统会产生巨大误差
- 许多实际问题中很难得到非线性函数的雅克比矩阵求导
- EKF 需要求导，所以必须清楚了解非线性函数的具体形式，无法作到黑盒封装，从而难以模块化应用

Unscented Kalman filter (UKF)

当 f, h 高度非线性时，EKF 的表现有可能会很差。

由于近似非线性函数的概率密度分布比近似非线性函数更容易，因此使用采样方法近似非线性分布来解决非线性问题受到了广泛关注。

UT变换（无迹变换）采用确定性采样策略，用多个粒子点逼近函数的概率密度分布，从而获得更高阶次的均值与方差。

UKF 的特点：

- 对非线性函数的概率密度分布进行近似,而不是对非线性函数进行近似
- 可以处理不可导的非线性函数
- 不要求导计算雅克比矩阵
- 对高斯输入的非线性函数近似时,使均值精确到三阶,方差精确到二阶
- 计算量与EKF相当