最优估计和滤波

所谓滤波，粗略地说就是要从受到随机噪声干扰的观测信号中，提取所需要的有用信息。通常把有用的信号从干扰(噪声)分离出来的数据处理方法称为滤波或估计。一般估计是根据测量得出的与系统状态 $\boldsymbol{x}(t)$ 有关的数据 $z(t)=\boldsymbol{h}[\boldsymbol{x}(t)]+\boldsymbol{v}(t)$，估计出 $\boldsymbol{x}(t)$ 的值 $\hat{\boldsymbol{x}}(t)$。其中随机向量 $\boldsymbol{v}(t)$ 为测量噪声，$\hat{\boldsymbol{x}}(t)$ 称为 $\boldsymbol{x}(t)$ 的估计。

设在 $[t_0, t_1]$ 时间段内的量测为 $z(t)$，相应的估计为 $\hat{\boldsymbol{x}}(t)$，则

当 $t=t_1$ 时，$\hat{\boldsymbol{x}}(t)$ 称为 $\boldsymbol{x}(t)$ 的估计(滤波)；

当 $t>t_1$ 时，$\hat{\boldsymbol{x}}(t)$ 称为 $\boldsymbol{x}(t)$ 的预测；

当 $t<t_1$ 时，$\hat{\boldsymbol{x}}(t)$ 称为 $\boldsymbol{x}(t)$ 的平滑。

最优估计是指某一指标函数达到最大值或最小值时的估计。若以测量估计偏差的平方和达到最小为指标，即

$$(z - \hat{z})^{T}(z - \hat{z}) = \min$$

则所得估计 $\hat{\boldsymbol{x}}(t)$ 为 $\boldsymbol{x}(t)$ 的最小二乘估计。

若以状态估计 $\hat{\boldsymbol{x}}(t)$ 的均方误差达到最小为指标，即

$$\mathrm{E}[(\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})^{T}] = \min$$

则所得估计 $\hat{\boldsymbol{x}}(t)$ 为 $\boldsymbol{x}(t)$ 的最小方差估计；若 $\hat{\boldsymbol{x}}(t)$ 又是 $z(t)$ 的线性函数，则 $\hat{\boldsymbol{x}}(t)$ 为 $\boldsymbol{x}(t)$ 的线性最小方差估计。

也有用估计值出现的概率作为估计指标的，这样的估计有极大验后估计、贝叶斯估计和极大似然估计等。这里将主要介绍最小二乘估计和线性最小方差估计。

一般最小二乘估计

最小二乘估计是德国科学家高斯(Karl Gauss)在 1795 年为测定行星轨道而提出的参数估计算法。这种估计的特点是算法简单,不需要知道与被估计量及量测量有关的任何统计信息。

设被估计量 $\boldsymbol{x}(t)$是一个未知的 n 维常值向量,即 $\boldsymbol{x}(t)=\boldsymbol{x}$。一般情况下 $\boldsymbol{x}$ 不能直接测量,而只能测量到 $\boldsymbol{x}$ 各分量的线性组合。为了得到 $\boldsymbol{x}$ 的估计,对它进行了 k 次测量,记第 i 次测量 $\boldsymbol{z}_i$ 为

$$\boldsymbol{z}_i = \boldsymbol{H}_i\boldsymbol{x} + \boldsymbol{v}_i \qquad (i = 1,2,\cdots,k) \tag{1}$$

式中,$\boldsymbol{z}_i$ 为 m 维观测向量,$\boldsymbol{H}_i$ 为 $m\times n$ 维测量矩阵,$\boldsymbol{v}_i$ 为随机测量噪声。式(1)可写成如下形式:

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v} \tag{2}$$

式中,$\boldsymbol{z}=[\boldsymbol{z}_1^{\mathrm{T}} \quad \boldsymbol{z}_2^{\mathrm{T}} \quad \cdots \quad \boldsymbol{z}_k^{\mathrm{T}}]^{\mathrm{T}}$,$\boldsymbol{H}=[\boldsymbol{H}_1^{\mathrm{T}} \quad \boldsymbol{H}_2^{\mathrm{T}} \quad \cdots \quad \boldsymbol{H}_k^{\mathrm{T}}]^{\mathrm{T}}$,$\boldsymbol{v}=[\boldsymbol{v}_2^{\mathrm{T}} \quad \boldsymbol{v}_2^{\mathrm{T}} \quad \cdots \quad \boldsymbol{v}_k^{\mathrm{T}}]^{\mathrm{T}}$。

最小二乘估计的指标是:使各次量测 z_i 与由估计 $\hat{x}$ 确定的 $\hat{z}_i = H_i\hat{x}$ 之差的平方和最小,即

$$J = (z - \hat{z})^{\mathrm{T}}(z - \hat{z}) = (z - H\hat{x})^{\mathrm{T}}(z - H\hat{x}) = \min \tag{3}$$

式(3)对 $\hat{x}$ 取偏导数,并令其等于零,得

$$\frac{\partial J}{\partial \hat{x}} = -2(H^{\mathrm{T}}z - H^{\mathrm{T}}H\hat{x}) = 0 \tag{4}$$

求得 x 的估计值

$$\hat{x} = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}z \tag{5}$$

式(5)有解的条件是 $(H^{\mathrm{T}}H)^{1}$ 存在。

最小二乘具有下列性质:

若测量噪声 v 是均值为零、方差为 R 的随机向量,则

① 估计是无偏的,即

$$\mathrm{E}[x - \hat{x}] = 0$$

② 估计的均方误差阵为

$$\mathrm{E}[\tilde{x}\tilde{x}^{\mathrm{T}}] = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}RH(H^{\mathrm{T}}H)^{-1}$$

例 10-1 用两台仪器对未知标量 X 各测量一次，量测量分别为 Z_1 和 Z_2，仪器的测量误差是均值为零、方差分别为 r 和 $4r$ 的随机量，求 X 的最小二乘估计，并计算估计的均方误差。

解 由题意得测量方程为

$$\boldsymbol{z} = \boldsymbol{h}X + \boldsymbol{v}$$

式中

$$\boldsymbol{z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \qquad \boldsymbol{h} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{R} = \begin{bmatrix} r & 0 \\ 0 & 4r \end{bmatrix}$$

由式(10-5)得

$$\hat{X} = (\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h})^{-1}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{z} = \frac{1}{2}(Z_1 + Z_2)$$

由式(10-7)得

$$\mathrm{E}[\tilde{x}\tilde{x}^{\mathrm{T}}] = (\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h})^{-1}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{h}(\boldsymbol{h}^{\mathrm{T}}\boldsymbol{h})^{-1} =$$

$$\frac{1}{2}\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} r & 0 \\ 0 & 4r \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\frac{1}{2} = \frac{5}{4}r$$

上式说明，使用精度差一倍的两台仪器同时进行测量，最小二乘估计精度还不如只用一台精度高的仪器效果好。

2. 加权最小二乘估计

从例 10 - 1 看出，一般最小二乘估计精度不高的原因之一是不分优劣地使用了测量值。如果对不同精度的测量值分别给予不同的权值，即对高精度的测量值权重取得大些，低精度的测量值权重取得小些。为此在最小二乘估计的性能指标 $J=(z-H\hat{x})^{\mathrm{T}}(z-H\hat{x})$ 中，引入加权阵 W，使性能指标变为

$$J=(z-H\hat{x})^{\mathrm{T}}W(z-H\hat{x})=\min \tag{10-8}$$

式中，W 为对称正定加权阵。当 $W=I$ 时，式(10 - 8)就是一般最小二乘准则。

将 J 对 $\hat{x}$ 求偏导数，令其等于零，即

$$\frac{\partial J}{\partial \hat{x}}=-2H^{\mathrm{T}}Wz+2H^{\mathrm{T}}WH\hat{x}=0$$

则

$$\hat{x}=(H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}Wz \tag{10-9}$$

由于 J 对 $\hat{x}$ 的二次偏导数为

$$\frac{\partial^2 J}{\partial \hat{x}^2} = 2H^{\mathrm{T}}WH > 0$$

所以 $\hat{x}$ 必使 J 取极小值。

估计误差为

$$\tilde{x} = x - \hat{x} = (H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}WHx - (H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}Wz =$$
$$(H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}W(Hx - z) = -(H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}Wv \qquad (10-10)$$

如果测量误差 v 的均值为零,方差阵为 R,则加权最小二乘估计也是无偏估计。估计的均方误差为

$$\mathrm{E}[\tilde{x}\tilde{x}^{\mathrm{T}}] = (H^{\mathrm{T}}WH)^{-1}H^{\mathrm{T}}WRWH(H^{\mathrm{T}}WH)^{-1} \qquad (10-11)$$

若$W=R^{-1}$,则加权最小二乘估计

$$\hat{x}=(H^TR^{-1}H)^{-1}H^TR^{-1}z \tag{10-12}$$

又称马尔柯夫估计。

马尔柯夫估计的均方误差为

$$\mathrm{E}[\tilde{x}\tilde{x}^T]=(H^TR^{-1}H)^{-1} \tag{10-13}$$

马尔柯夫估计的均方误差比任何其他加权最小二乘估计的均方误差都要小,是加权最小二乘中的最优者。

例 10-2 对例 10-1 采用马尔柯夫估计，并计算估计的均方误差。

解 取

$$\boldsymbol{W}=\boldsymbol{R}^{-1}=\begin{bmatrix}1/r & 0\\ 0 & 1/4r\end{bmatrix}$$

由式(10-12)和式(10-13)得

$$\hat{X}=\frac{4}{5}Z_1+\frac{1}{5}Z_2$$

$$\mathrm{E}[\tilde{X}\tilde{X}^{\mathrm{T}}]=\frac{4}{5}r<r$$

这说明，对高精度的测量值取权系数$\frac{4}{5}$，对低精度的测量值取权系数$\frac{1}{5}$，估计精度高于仅用高精度量测结果所得值。所以增加量测值，并根据精度高低，区别加以利用，可以有效地提高估计精度。

3. 递推最小二乘

对于加权最小二乘，量测值越多，估计的均方误差就越小。采用批处理实现的最小二乘算法，须存储所有的量测值。若量测值非常庞大，则所需计算机的存储量太大，这显然是不经济的。而递推最小二乘则可克服上述缺点，仅从每次获得的测量值中提取出被估计量信息，用于修正上一步所得的估计。

在任意时刻的最小二乘估计。$\hat{x}$ 和 P_0 的选取是任意的，一般可取 $\hat{x}_0=0$，$P_0=\rho I$。其中 ρ 为很大的正数。由于初值选取盲目，所以在递推过程中，刚开始计算时，估计误差跳跃剧烈，随着测量次数的增加，初值影响逐渐消失，估计值逐渐趋于稳定而逼近被估量。

设 x 为确定性常值向量，假定已进行了 k 次测量，量测方程为

$$\bar{z}_k = \bar{H}_k x + \bar{v}_k$$

式中

$$\bar{z}_k = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \end{bmatrix}, \quad \bar{H}_k = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_k \end{bmatrix}, \quad \bar{v}_k = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix}$$

z_i 为第 i 次量测，量测方程为

$$z_i = H_i x + v_i \qquad (i = 1,2,\cdots,k)$$

则前 $k+1$ 次量测为

$$\bar{z}_{k+1} = \bar{H}_{k+1} x + \bar{v}_{k+1}$$

式中

$$\bar{z}_{k+1} = \begin{bmatrix} \bar{z}_k \\ z_{k+1} \end{bmatrix}, \quad \bar{H}_{k+1} = \begin{bmatrix} \bar{H}_k \\ H_{k+1} \end{bmatrix}, \quad \bar{v}_{k+1} = \begin{bmatrix} \bar{v}_k \\ v_{k+1} \end{bmatrix}$$

线性最小方差估计

假设被估计量 x 是一个 n 维随机量，它和 m 维观测量 z 的一、二阶矩已知。如果限定估计 $\hat{x}$ 是观测量 z 的线性函数，即 $\hat{x}=a+Bz$，则线性最小方差估计的性能指标为

$$J=\mathrm{E}[\tilde{x}^{\mathrm{T}}\tilde{x}]=\mathrm{E}[(x-a-Bz)^{\mathrm{T}}(x-a-Bz)]=\min \qquad (10-20)$$

线性最小方差估计的性质

① 线性最小方差估计是一种无偏估计。

对线性最小方差估计式(10-26)两边求均值，即得

$$\mathrm{E}[\hat{x}]=\mathrm{E}[x]+\mathrm{cov}\ (x,z)[\mathrm{var}(z)]^{-1}\mathrm{E}[z-\mathrm{E}(z)]=\mathrm{E}[x]$$

这表明估计量的均值等于被估计量的均值，因此线性最小方差估计是一种无偏估计。这时估计误差 $\hat{x}$ 的均值等于零，即

$$\mathrm{E}[\hat{x}]=\mathrm{E}[x-\hat{x}]=\mathrm{E}[x]-\mathrm{E}[\hat{x}]=0$$

② 线性最小方差估计是线性无偏估计中误差方差阵最小的一种估计。

③ 线性最小方差估计是被估计量在观测矢量上的正交投影。

维纳滤波

1. 维纳滤波问题的提法

设系统的观测方程为

$$Z(t) = X(t) + V(t) \tag{10-31}$$

式中,$X(t)$为有用信号,$Z(t)$为观测信号,$V(t)$为观测误差(干扰)。

设 $X(t)$,$Z(t)$和 $V(t)$都是均值为零,具有各态历经的平稳随机过程。根据观测值 $Z(t)$来估计 $X(t)$,使估值 $\hat{X}(t)$接近 $X(t)$。

维纳滤波的任务就是考虑如何设计出一个线性定常系统 L(如图 10-2 所示),使得系统的输出 $Y(t)$与 $X(t)$具有最小的方差,即

$$J = \mathrm{E}\{[X(t) - Y(t)]^2\} = \min \tag{10-32}$$

这样的 $Y(t)$就作为 $X(t)$的估值 $\hat{X}(t)$。

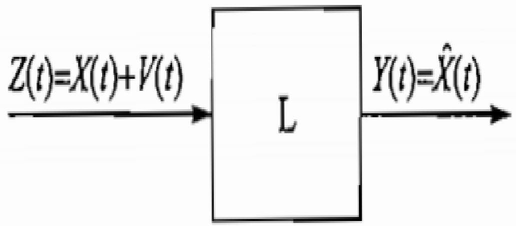

图 10-2　维纳滤波

如果系统 L 的脉冲过渡函数为 $h(t)$，则

$$Y(t)=\int_{0}^{\infty}h(\tau)Z(t-\tau)\mathrm{d}\tau$$

式中，$Y(t)$是系统 L 根据输入信号 $Z(t)$在$(-\infty,t)$上的全部过去值所给出的实际输出。$Y(t)$是 $Z(t-\tau)$的线性函数($\tau>0$)。

根据问题的性质，可以看出维纳滤波要有下列三个条件：

① 信号与噪声都是零均值的，具有各态历经性的平稳随机过程。

② 滤波器是一个物理可实现的线性定常系统。当 $\lambda<0$ 时，$h(\lambda)=0$。

③ 最优准则是滤波的方差为最小。

这些条件使维纳滤波受到很大限制。

最小二乘估计法实用于对常值向量或随机向量的估计。由于使用的最优指标是使量测估计的精度达到最佳，估计过程中不必使用与被估计量有关的动态信息与统计信息，甚至连量测误差的统计特性也可不必使用，所以估计精度不高。这种方法的最大优点是算法简单，在对被估计量缺乏了解的情况下，仍可使用。

线性最小方差估计是所有线性估计过程中的最优者。线性最小方差估计可适用于随机过程的估计，估计过程中只需知道被估计量和量测量的一、二阶矩。对于平稳过程，这些一、二阶矩都为常值；但对于非平稳随机过程，一、二阶矩随时间变化，必须确切知道每一估计时刻的一、二阶矩，才能给出估计值，这种要求很多时候难以满足。所以线性最小方差估计适用于平稳过程，而难以适用于非平稳过程。

维纳滤波是线性最小方差估计的一种。维纳滤波器是一种线性定常系统，适用于对有用信号和干扰信号都是零均值的平稳随机过程。但由于维纳-霍夫方程求解困难，故在工程中应用较少。

Kalman滤波

前面介绍的线性最小方差估计具有一定的实用性，不同时刻的量测信息利用得越多，估计的精度就越高；但这种算法使用了被估计量与量测量的一、二阶矩，对于非平稳过程，必须确切知道一、二阶矩的变化规律，这种要求是十分苛刻的，一般无法满足。此外，算法中采用对不同时刻的量测值作集中处理的办法，这使计算随着估计过程的推移而逐渐加重，所以线性最小方差估计并不是一种实用的估计算法，特别是对非平稳过程的处理更是困难重重。人们试图找到一种算法，适用于非平稳过程，并且算法是递推的，不需要大量的存储空间。卡尔曼滤波就具有这样的特点。

1960 年由卡尔曼(R. E. Kalman)首次提出的卡尔曼滤波是一种线性最小方差估计。其主要特点如下：

① 算法是递推的，且使用状态空间法在时域内设计滤波器，所以卡尔曼滤波适用于对多维随机过程的估计。

② 用动力学方程即状态方程描述被估计量的动态变化规律，被估计量的动态统计信息由激励白噪声的统计信息和动力学方程确定。由于激励白噪声是平稳过程，动力学方程已知，所以被估计量既可以是平稳的，也可以是非平稳的，即卡尔曼滤波也适用于非平稳过程。

③ 卡尔曼滤波具有连续型和离散型两类算法，离散型算法可直接在数字计算机上实现。

Kalman滤波的由来

滤波估计经历的三个阶段：

1最小二乘法：计算简单 ，但没考虑被估参数和观测数据的统计特性，不具最优性。

2 Wiener滤波：频域中的统计最优滤波器，但运算复杂，存储空间大，应用范围有限。

3Kalman滤波：时域滤波，采用状态空间描述系统，运用递推形式使计算简单，数据存储量小，应用广泛。

Kalman滤波理论的作用

Kalman滤波定义：

Kalman滤波是一种实时递推算法，它所处理的是随机信号，利用系统噪声和观测噪声的统计特性，以系统的观测量作为滤波器的输入，以所要估计值（状态或参数）作为滤波器的输出，滤波器输入与输出是由时间更新和观测更新算法联系在一起的，根据系统方程和观测方程估计出所需要处理的信号——实质是一种最优估计方法。

问题分解：运用白噪声

噪声信号 $\omega(t)$ 满足：

$$\begin{cases} E[\omega(t)]=0 \\ E[\omega(t)\omega^{T}(\tau)]=q\delta(t-\tau) \end{cases}$$

则称 $\omega(t)$ 为白噪声，式中 q 为 $\omega(t)$ 方差强度

白噪声序列特点：

1零均值且具有独立性

2与时间无关，与时间间隔有关，所以它具有平稳性

卡尔曼滤波与最优估计

卡尔曼滤波是一种最优估计技术！

它能将仅与部分状态有关的测量值进行处理，得出从某种统计意义上讲估计误差最小的更多的状态的估计值。

估计误差最小的标准称为估计准则。

根据不同的估计准则和估计计算方法，有各种不同的最优估计。

卡尔曼滤波是一种递推线性最小方差估计。

最小方差估计 → 线性最小方差估计 → 递推线性最小方差估计

- 最小方差估计具有无偏性质，即它的估计误差用 $\tilde{X}$ 表示）的均值为零。即：

$$E\{[X-\hat{X}(Z)]\}=E\{\tilde{X}\}=0$$

- 估计的均方误差就是估计误差的方差，即：

$$E\{\tilde{X}\tilde{X}^{T}\}=E\{[\tilde{X}-E(\tilde{X})][\tilde{X}-E(\tilde{X})]^{T}\}$$

- 因此，最小方差估计不但使估值 $\hat{X}(Z)$ 的均方误差最小，而且这种最小的均方误差就是估计的误差方差

离散卡尔曼最优预测

设离散型系统的方程为

$$\boldsymbol{x}(k+1)=\boldsymbol{\phi}(k+1,k)\boldsymbol{x}(k)+\boldsymbol{G}(k+1,k)\boldsymbol{u}(k)+\boldsymbol{\Gamma}(k+1,k)\boldsymbol{w}(k) \tag{11-16a}$$

$$\boldsymbol{z}(k)=\boldsymbol{H}(k)\boldsymbol{x}(k)+\boldsymbol{v}(k) \tag{11-16b}$$

为简单起见，状态方程中暂不考虑系统的控制信号 $\boldsymbol{u}(k)$，因为它是一个确定性时间序列，故系统状态方程为

$$\boldsymbol{x}(k+1)=\boldsymbol{\phi}(k+1,k)\boldsymbol{x}(k)+\boldsymbol{\Gamma}(k+1,k)\boldsymbol{w}(k) \tag{11-17}$$

观测方程为

$$\boldsymbol{z}(k)=\boldsymbol{H}(k)\boldsymbol{z}(k)+\boldsymbol{v}(k) \tag{11-18}$$

且有

$$\left.\begin{aligned}
&\mathrm{E}[\boldsymbol{w}(k)]=\mathrm{E}[\boldsymbol{v}(k)]=\boldsymbol{0}\\
&\mathrm{E}[\boldsymbol{w}(k)\boldsymbol{w}^{\mathrm{T}}(j)]=\boldsymbol{Q}_k\delta_{kj}\\
&\mathrm{E}[\boldsymbol{v}(k)\boldsymbol{v}^{\mathrm{T}}(j)]=\boldsymbol{R}_k\delta_{kj}\\
&\mathrm{E}[\boldsymbol{w}(k)\boldsymbol{v}^{\mathrm{T}}(j)]=\boldsymbol{0}
\end{aligned}\right\} \tag{11-19}$$

式中

$$\delta_{kj}=\begin{cases}1, & k=j\\ 0, & k\neq j\end{cases}$$

初始状态 $\boldsymbol{x}(0)$ 的统计特性为

$$\left.\begin{aligned} &\mathrm{E}[\boldsymbol{x}(0)] = m_0 \\ &\mathrm{E}\{[\boldsymbol{x}(0)] - m_0][\boldsymbol{x}(0) - m_0]^{\mathrm{T}}\} = \boldsymbol{P}_0 \end{aligned}\right\} \tag{11-20}$$

由于 $\boldsymbol{w}$ 与 $\boldsymbol{v}$ 噪声的污染，从观测值 $\boldsymbol{z}$ 中必须利用最佳估计原理对真实状态进行估计。卡尔曼基于正交投影的性质推出的递推估计算法，对每一瞬时的状态进行预测。下面介绍一种推导预测基本方程的方法，采用数学归纳法进行。

一步最优预测问题的提法如下：

对于由式(11-17)～式(11-20)所描述的线性系统，当已得到观测向量 $\boldsymbol{z}(0),\boldsymbol{z}(1),\cdots,\boldsymbol{z}(k)$ 之后，要求确定$(k+1)$步的状态向量 $\boldsymbol{x}(k+1)$ 的最优估计值 $\hat{\boldsymbol{x}}(k+1/k)$，使得估计误差的方差

$$\mathrm{E}[\tilde{\boldsymbol{x}}(k+1/k)\tilde{\boldsymbol{x}}^{\mathrm{T}}(k+1/k)] = \min \tag{11-21}$$

式中，$\tilde{\boldsymbol{x}}(k+1/k)=\boldsymbol{x}(k+1)-\hat{\boldsymbol{x}}(k+1/k)$，且 $\hat{\boldsymbol{x}}(k+1/k)$ 是观测值 $\boldsymbol{z}(0),\boldsymbol{z}(1),\cdots,\boldsymbol{z}(k)$ 的线性函数，估计要求是无偏的，即

$$\mathrm{E}[\tilde{\boldsymbol{x}}(k+1/k)] = \boldsymbol{0}$$

下面分几步进行推导。

(1) 最优两步预测

设对于获得的观测值 $z(0), z(1), \cdots, z(k-1)$ 已得到了 k 时刻状态的最优预测估计为 $\hat{\boldsymbol{x}}(k/k-1)$。这一符号可解释为根据第 $k-1$ 步及其以前的观测值对第 k 步状态 $\boldsymbol{x}$ 所作的线性最小方差估计。据此可推得第 $k+1$ 步的状态最优预测为

$$\hat{\boldsymbol{x}}(k+1/k-1)=\boldsymbol{\phi}(k+1,k)\hat{\boldsymbol{x}}(k/k-1) \tag{11-22}$$

式(11-22)表示了根据到第 $k-1$ 步所获得的观测值对第 $k+1$ 步状态的估计。$\boldsymbol{w}(k)$ 将不在式(11-22)中出现。同理，这时还不曾得到观测值 $\boldsymbol{z}(k)$，只能由第 k 步的状态最优估计预测第 k 步的可能观测值 $\hat{\boldsymbol{z}}(k/k-1)$，因此有

$$\hat{\boldsymbol{z}}(k/k-1)=\boldsymbol{H}(k)\hat{\boldsymbol{x}}(k/k-1) \tag{11-23}$$

这一预测中也不可能包含 $\boldsymbol{v}(k)$。

可以证明，如果 $\hat{\boldsymbol{x}}(k/k-1)$ 确实为最优估计，则依据第 $k-1$ 步的观测值对第 $k+1$ 步状态所作的估计 $\hat{\boldsymbol{x}}(k+1/k-1)$ 也一定是最优的，则第 $k+1$ 步的状态估计误差为

$$\begin{aligned}\tilde{\boldsymbol{x}}(k+1/k-1)=&\boldsymbol{x}(k+1)-\hat{\boldsymbol{x}}(k+1/k-1)=\\&\boldsymbol{\phi}(k+1,k)\boldsymbol{x}(k)-\boldsymbol{\phi}(k+1,k)\hat{\boldsymbol{x}}(k/k-1)+\boldsymbol{\Gamma}(k+1,k)\boldsymbol{w}(k)=\\&\boldsymbol{\phi}(k+1,k)\tilde{\boldsymbol{x}}(k/k-1)+\boldsymbol{\Gamma}(k+1,k)\boldsymbol{w}(k)\end{aligned}$$

式中，$\tilde{\boldsymbol{x}}(k/k-1)=\boldsymbol{x}(k)-\hat{\boldsymbol{x}}(k/k-1)$。

因为 $\hat{\boldsymbol{x}}(k/k-1)$已设为最优估计，因此必满足

$$\mathrm{E}[\tilde{\boldsymbol{x}}(k/k-1)\boldsymbol{z}^{\mathrm{T}}(j)]=\boldsymbol{0} \qquad (j=0,1,2,3,\cdots,k-1)$$

又由于 $\boldsymbol{w}(k)$与 $\boldsymbol{z}(j)$均相互独立，因此有

$$\mathrm{E}[\boldsymbol{w}(k)\boldsymbol{z}^{\mathrm{T}}(j)]=\boldsymbol{0} \qquad (j=0,1,2,3,\cdots,k-1)$$

由此可得

$$\begin{aligned}&\mathrm{E}[\tilde{\boldsymbol{x}}(k+1/k-1)\boldsymbol{z}^{\mathrm{T}}(j)]=\\&\qquad \mathrm{E}[\boldsymbol{\phi}(k+1,k)\tilde{\boldsymbol{x}}(k/k-1)\boldsymbol{z}^{\mathrm{T}}(j)]+\mathrm{E}[\boldsymbol{\Gamma}(k+1,k)\boldsymbol{w}(k)\boldsymbol{z}^{\mathrm{T}}(j)]=\boldsymbol{0}\\&\qquad\qquad j=0,1,2,3,\cdots,k-1\end{aligned}$$

这就证明了 $\hat{\boldsymbol{x}}(k+1/k-1)$确实是最优估计。

(2) 最优一步预测

当得到了第 k 步的观测 $z(k)$之后，应当重新估计第 $k+1$ 步的状态 $x(k+1)$。下面就$z(k)$的两种情形分别讨论。

① 如果新得到的观测值 $z(k)$正好等于其预测值 $\hat{z}(k/k-1)$，则说明实际的 $z(k)$值并没有比原来的预测值增加新的信息，这时依据一组新的预测值 $z(j)$，$j=0,1,2,3,\cdots,k$ 重新作出的状态估计必然仍然是 $\hat{x}(k/k-1)$。由此可断定第 $k+1$ 步的最佳估计也仍然是 $\hat{x}(k+1/k-1)=\boldsymbol{\phi}(k+1,k)\hat{x}(k/k-1)$，可证明其估计误差

$$\tilde{x}(k+1/k-1)=x(k+1)-\hat{x}(k+1/k-1) \tag{11-24}$$

与新的观测值 $z(k)=\hat{z}(k/k-1)$是正交的。事实上，

$$\begin{aligned}
\mathrm{E}[\tilde{x}(k+1/k-1)z^{\mathrm{T}}(k/k-1)] &= \mathrm{E}\{[\tilde{x}(k+1/k-1)][\boldsymbol{H}(k)\hat{x}(k/k-1)]^{\mathrm{T}}\} = \\
&\mathrm{E}\{[\boldsymbol{\phi}(k+1,k)x(k)-\boldsymbol{\phi}(k+1,k)\hat{x}(k/k-1)+ \\
&\boldsymbol{\Gamma}(k+1,k)w(k)][\boldsymbol{H}(k)\hat{x}(k/k-1)]^{\mathrm{T}}\} = \\
&\boldsymbol{\phi}(k+1,k)\mathrm{E}[\tilde{x}(k/k-1)\hat{x}^{\mathrm{T}}(k/k-1)]\boldsymbol{H}^{\mathrm{T}}(k)+ \\
&\boldsymbol{\Gamma}(k+1,k)\mathrm{E}[w(k)\hat{x}^{\mathrm{T}}(k/k-1)]\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{0}
\end{aligned}$$

② 实际上 $z(k)$不可能完全等于 $\hat{z}(k/k-1)$，在 k 时刻必然有噪声 $v(k)$，因此应有

$$z(k)=\boldsymbol{H}(k)\boldsymbol{x}(k)+\boldsymbol{v}(k)$$

式中，$\boldsymbol{x}(k)$也不会完全等于 $\hat{\boldsymbol{x}}(k/k-1)$。观测误差为

$$\begin{aligned}\tilde{z}(k/k-1)=z(k)-\hat{z}(k/k-1)=\\ \boldsymbol{H}(k)\boldsymbol{x}(k)-\boldsymbol{H}(k)\hat{\boldsymbol{x}}(k/k-1)+\boldsymbol{v}(k)=\\ \boldsymbol{H}(k)\tilde{\boldsymbol{x}}(k/k-1)+\boldsymbol{v}(k)\end{aligned} \tag{11-25}$$

通常称 $\tilde{z}(k/k-1)$为新息。在得到新息之后，应当用来对原先的估计 $\hat{\boldsymbol{x}}(k+1/k-1)$进行修正。由于 $z(k)$不等于 $\hat{z}(k/k-1)$，故 $\hat{\boldsymbol{x}}(k+1/k-1)$也不再是最优线性预测了。修正后的第 $k+1$ 步的最优线性预测用 $\hat{\boldsymbol{x}}(k+1/k)$表示。根据线性最小方差准则要求，修正后的最优线性预测应当是 $z(k)$的线性函数，故设

$$\hat{\boldsymbol{x}}(k+1/k)=\boldsymbol{\phi}(k+1,k)\hat{\boldsymbol{x}}(k/k-1)+\boldsymbol{K}(k)\tilde{z}(k/k-1)$$

式中，$\boldsymbol{K}(k)$称为最优增益矩阵，又称加权阵。上式还可表示为

$$\hat{\boldsymbol{x}}(k+1/k)=\boldsymbol{\phi}(k+1,k)\hat{\boldsymbol{x}}(k/k-1)+\boldsymbol{K}(k)[z(k)-\boldsymbol{H}(k)\hat{\boldsymbol{x}}(k/k-1)] \tag{11-26}$$

完整的卡尔曼最优一步预测方程为

① 最优预测估计方程：

$$\hat{x}(k+1/k)=\phi(k+1,k)\hat{x}(k/k-1)+K(k)[z(k)-H(k)\hat{x}(k/k-1)]$$

② 最优增益方程：

$$K(k)=\phi(k+1,k)P(k/k-1)H^{T}(k)[H(k)P(k/k-1)H^{T}(k)+R_{k}]^{-1}$$

③ 最优误差方差阵估计方程：

$$P(k+1/k)=\phi(k+1,k)P(k/k-1)\phi^{T}(k+1,k)-$$
$$K(k)H(k)P(k/k-1)\phi^{T}(k+1,k)+\Gamma(k+1,k)Q_{k}\Gamma^{T}(k+1,k)$$

最优预测估计的性质：在推导最优预测的过程中已规定了 $\hat{x}(k+1/k)$是 $z(k)$的线性函数，同时在推导 $K(k)$时也保证了 $\tilde{x}(k+1/k)$与观测序列的正交。因此，只需证明估计是无偏的。

例 11-1 设系统的状态方程和观测方程为

$$x(k+1)=0.5x(k)+w(k)$$

$$z(k)=x(k)+v(k)$$

$w(k)$，$v(k)$为零均值互不相关的白噪声序列，且有 $\mathrm{E}[w(k)w(j)]=1\cdot\delta_{kj}$，$\mathrm{E}[v(k)v(j)]=2\cdot\delta_{kj}$。初始状态 $x(0)$的统计特性为 $\mathrm{E}[x(0)]=m_0=0$，$P(0/0_-)=1$；观测值为 $z(0)=0$，$z(1)=4$，$z(2)=2$。试求 $x(k)$的一步最优预测估计。

解 由题意可知 $\phi(k+1,k)=0.5$，$\Gamma(k+1,k)=1$，$H(k)=1$，$Q_k=1$，$R_k=2$,则

$$\hat{x}(k+1/k)=0.5\hat{x}(k/k-1)+K(k)[z(k)-\hat{x}(k/k-1)]$$

$$K(x)=0.5P(k/k-1)[P(k/k-1)+2]^{-1}$$

$$P(k+1/k)=0.25P(k/k-1)-0.25[P(k/k-1)]^2[P(k/k-1)+2]^{-1}+1$$

根据 $P(0/0_-)=1$,取 $k=0,1,2$,计算 $P(k+1/k)$和 $K(k)$,得

$$P(1/0)=0.25-0.25\times\frac{1}{3}+1=1.166$$

$$P(2/1)=0.25\times 1.166-0.25\times 1.166^2\times\frac{1}{3.166}+1=1.184$$

$$P(3/2)=0.25\times 1.184-0.25\times 1.184^2\times\frac{1}{3.184}+1=1.186$$

$$K(0)=0.5\times\frac{1}{3}=0.166$$

$$K(1)=0.5\times 1.166\times\frac{1}{3.166}=0.184$$

$$K(2)=0.5\times 1.184\times\frac{1}{3.184}=0.186$$

根据 $\hat{x}(0/0_-)=0$ 及 $P(k+1/k)$和 $K(k)$求得 x 的预测估计为

$$\hat{x}(1/0)=0,\qquad \hat{x}(2/1)=0.184\times 4=0.736$$

$$\hat{x}(3/2)=0.5\times 0.736+0.186(2-0.736)=0.603$$

由于初始状态的估计值正好等于初始状态的均值,故估计是无偏的。

如果估计误差的初值 $P(0/0_-)$事先不知道,则只能任意假设一个 $P(0/0_-)$值。如果所设的值小于实际的误差方差,则以后每一步的估计误差方差也将小于实际的误差方差;反之将大于实际值。当系统满足一定条件时,随着 k 的增大,误差方差阵 $P(k+1/k)$将趋于常数。