随机系统的可估计性——信息理论方法

章　辉　孙优贤
(浙江大学工业控制技术研究所·杭州, 310027)

摘要：本文在信息理论的框架下研究了随机系统的可估计性问题，对于线性高斯系统导出了其可估计性判据，并讨论了可估计性与相应确定性系统可观测性之间的关系.

关键词：随机系统；可估计性；可观测性；信息理论；熵；互信息

1　引　言

随机系统的状态估计(滤波)是现代控制理论的一个重要分支. 由于 Shannon 信息理论的一些问题在很大程度上同状态估计问题非常相似，近年来有不少学者将状态估计问题置于信息理论框架中进行研究，得出一类最小最大误差熵(MMEE)估计器[3,4,10]. 对于线性高斯系统而言，MMEE 估计则退化为(或称等价于)线性最小方差估计[4].

相对来说各种情况下(如线性、非线性，高斯、非高斯等)的滤波算法研究得较多，而对系统状态的可估计性问题研究得很少. Baram 和 Kailath 在最小均方估计的意义下，给出了线性随机系统的可估计性(Estimability)定义[1]，为随机系统的分析提供了新的视角和方法.

滤波问题和信息理论中的信息传输问题之间有很大相同之处. 过程产生的信号由传感器检测，通过滤波器得出对过程状态的估计. 在信息理论意义下，传感器、滤波器分别可看作是通讯系统中的编码器和译码器，而检测和传输过程中因信号受到噪声干扰，则可看作通过一个受噪声干扰的信息传输通道. 本文第二部分运用信息理论讨论了随机系统的可估计性问题，给出在 MMEE 意义下的可估计性定义，并由此导出了线性高斯离散系统可估计性判据. 第三部分从信息论的角度研究了线性离散系统可估计性与相应确定性系统可观测性(Observability)之间的关系.

2　可估计性

考虑如下的离散随机系统

$$\begin{cases} x(k+1) = F(k)x(k) + G(k)w(k), & x(k) \in \mathrm{R}^n, \\ y(k) = C(k)x(k) + v(k), & y(k) \in \mathrm{R}^m, \quad k = 0,1,\cdots. \end{cases} \tag{1}$$

其中，$\mathrm{E}\{x(0)\} = 0$，$\mathrm{E}\{x(0)x^{\mathrm{T}}(0)\} = \pi(0)$，$\{w(k)\}$ 和 $\{v(k)\}$ 皆为零均值的随机向量序列，$\mathrm{E}\{x(0)w^{\mathrm{T}}(k)\} = 0$，$\mathrm{E}(w(k))v^{\mathrm{T}}(j)\} = 0$，$\mathrm{E}\{w(k)w^{\mathrm{T}}(j)\} = Q(k)\delta_{k,j}$，$\mathrm{E}\{v(k)v^{\mathrm{T}}(j)\} = R(k)\delta_{k,j}$. 当 $k = j$ 时，$\delta_{k,j} = 1$ 时；否则 $\delta_{k,j} = 0$. $\mathrm{E}\{\cdot\}$ 表示数学期望. 设 $\hat{x}(k)$ 表示在给定观测序列 $\{y(k), y(k-1), \cdots\}$ 下，关于系统状态 $x(k)$ 的线性最小均方估计，且 $\tilde{x}(k) \triangleq x(k) - \hat{x}(k)$，$\pi(k) \triangleq E\{x(k)x^{\mathrm{T}}(k)\}$，$P(k) \triangleq \mathrm{E}\{\tilde{x}(k)\tilde{x}^{\mathrm{T}}(k)\}$.

对于上述系统，Baram 和 Kailath[1] 给出了在最小均方估计意义下的“可估计性”定义，如下：

本文于 1995 年 4 月 25 日收到. 1995 年 11 月 28 日收到修改稿.

定义 1　线性随机系统(1)为可估计的，如果

$$P(k) < \pi(k), \quad \forall\, k \geqslant n,$$

即，当且仅当对于任意非零向量 $g \in \mathrm{R}^n$，有

$$g^{\mathrm{T}} P(k) g < g^{\mathrm{T}} \pi(k) g, \quad \forall\, k \geqslant n. \tag{2}$$

这一定义的条件可表述为：状态变量线性估计的后验均方误差矩阵严格小于其先验均方误差矩阵. 文献[1]指出，式(2)反映了这样的条件：状态空间中的任何方向都不与所有过去的输出测量值相正交.

在信息理论的概念中，两随机变量间的"互信息"(mutual information)描述了一个随机变量包含另一个随机变量的信息的多少，当两个变量相互独立时，互信息为零. 由此出发，这里运用信息理论的观点来研究可估计性问题.

考虑随机系统

$$\begin{cases} x(k+1) = f(k, x(k), w(k)), & x(k) \in \mathrm{R}^n, \\ y(k) = h(k, x(k), v(k)), & y(k) \in \mathrm{R}^m, \quad k = 0, 1, 2, \cdots. \end{cases} \tag{3}$$

式中 $w(k)$，$v(k)$ 为零均值的随机噪声. 设 $\hat{x}(k)$ 为在给定观测序列 $\{y(k), y(k-1), \cdots\}$ 下的对 $x(k)$ 的最小最大误差熵估计(注：当估计误差的最大熵达到最小时，此估计为 MMEE 意义下的最优估计；而对线性高斯系统，这一最优估计即等价于线性最小均方意义下的最优估计).

对系统(3)，给出 MMEE 意义下的可估计性定义如下：

定义 2　系统(3)为可估计的，当且仅当

$$I(x(k); \hat{x}(k)) > 0, \quad \forall\, k \geqslant n. \tag{4}$$

式中 $I(\cdot\,;\,\cdot)$ 为互信息.

定义 2 是针对较一般的系统给的，从这个意义上说，它较定义 1 具有更广泛的意义. 对系统(1)，若假设其概率特性为高斯的，则可以证明这两个定义是一致的. 有：

定理 1　对线性离散系统(1)，若 $x(k)$，$w(k)$，$v(k)$ 皆服从高斯分布，则

$$I(x(k); \hat{x}(k)) > 0 \Leftrightarrow g^{\tau} \pi(k) g > g^{\mathrm{T}} P(k) g, \quad \forall\, g \in \mathrm{R}^n, \quad g \neq 0; \quad \forall\, k \geqslant n, \tag{5}$$

即定义 2 与定义 1 是等价的.

对于正态随机向量的协方差矩阵 $(\Pi(k), P(k))$，我们知其是正定对称的[9]. 并且，Bellman[7] 指出它们具有这样的特性：对于任意非零向量 x，有

$$(\det P(k))^{-1/2} = \pi^{-n/2} \int^{\mathrm{R}_n} \exp(-x^{\mathrm{T}} P(k) x) \mathrm{d}x. \tag{6}$$

正是基于这样一个特性，人们才得出在线性高斯系统情况下最小最大误差熵估计与线性最小方差估计的等价关系[4]. 这里，我们先由式(6)得出这样一个引理：

引理 1　若正态分布随机向量的方差矩阵 A，B 对任意非零向量 x 恒有 $x^{\mathrm{T}} A x \geqslant x^{\mathrm{T}} B x$，则 $\det A \geqslant \det B$. 反之亦成立.

此引理的证明可参阅文献[9]第 57～62 页.

定理 1 的证明.

由信息理论知识及有关定义可知

$$I(x(k); \hat{x}(k)) = \mathrm{H}(x(k)) - \mathrm{H}(x(k) \mid \hat{x}(k))$$

$$=\frac{1}{2}\ln(2\pi e)^n\det\Pi(k)-\frac{1}{2}\ln(2\pi e)^n\det P(k),$$

式中 H(·) 代表熵. 因此

$$I(x(k);\hat{x}(k))>0\Leftrightarrow\det\Pi(k)>\det P(k).$$

再由引理 1 有

$$I(x(k);\hat{x}(k))>0\Leftrightarrow g^{\mathrm{T}}\Pi(k)g>g^{\mathrm{T}}P(k)g,\quad\forall\, g\in\mathrm{R}^n,\quad g\neq 0.$$

证毕.

由状态估计的信息理论方法可知,对于线性高斯系统,MMEE 估计 $\hat{x}(k)$ 等同于线性最小方差估计(Kalman 估计),即 $\hat{x}(k)$ 为无偏最优估计. 因而有

引理 2　对于线性高斯系统(1)有

$$I(x(k);\hat{x}(k))=I(x(k);Y_k),\quad\forall\, k\geqslant n.\tag{7}$$

式中

$$Y_k=[y^{\mathrm{T}}(0)\ y^{\mathrm{T}}(1)\ \cdots\ y^{\mathrm{T}}(k)]^{\mathrm{T}}.$$

由定义 2 及引理 2 可直接导出:

定理 2　线性离散高斯系统(1)(即 $x(0),w(k),v(k)$ 皆为高斯分布)如可估计的,当且仅当

$$I(x(k);Y_k(k))>0,\quad\forall\, k\geqslant n.\tag{8}$$

由信息理论知识: $I(x(k);Y_k)\geqslant 0$ 当且仅当 $x(k)$ 与 Y_k 相互独立时等号成立,即此时 Y_k 中不包含任何关于 $x(k)$ 的信息. 只要 Y_k 中任一分量与 $x(k)$ 不独立, $I(x(k);Y_k)$ 即大于零. 由此我们可以看出,式(8) 与式(2) 所反映的条件是一致的:状态空间任何方向都不与所有的过去观测值相正交.

对于线性高斯系统(1),设 $Z_k\triangleq[x^{\mathrm{T}}(k)\quad Y_k^{\mathrm{T}}]^{\mathrm{T}}, R_{Y_k}\triangleq E\{Y_kY_k^{\mathrm{T}}\}, M_k\triangleq E\{Y_kx^{\mathrm{T}}(k)\}$. 则有

$$R_{Z_k}\triangleq E\{Z_kZ_k^{\mathrm{T}}\}=\left[\begin{array}{c|c}\pi(k) & M_k^{\mathrm{T}}\\\hline M_k & R_{Y_k}\end{array}\right],$$

$$\det R_{Z_k}=\det Y_{Y_k}\det[\pi(k)-M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k].$$

所以有

$$\begin{aligned}I(x(k);Y_k)&=\frac{1}{2}\ln(2\pi e)^n\det\pi(k)+\frac{1}{2}\ln(2\pi e)^{m(k+1)}\det R_{Y_k}\\&\quad-\frac{1}{2}\ln(2\pi e)^{[n+m(k+1)]}\det R_{Z_k}\\&=\frac{1}{2}\ln\det\pi(k)-\frac{1}{2}\ln\det[\pi(k)-M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k].\end{aligned}\tag{9}$$

由于最小最大误差熵估计 $\hat{x}(k)$ 在线性高斯系统情况下亦为 $x(k)$ 的线性无偏最小方差估计,因而由最小方差估计理论有[8]

$$P(k)=\pi(k)-M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k,\tag{10}$$

即 $\pi(k)-M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k$ 即是估计误差的方差阵,且为正定对称的. 所以,可以由引理 1 得:条件(8) 等价于:

$$g^{\mathrm{T}}\pi(k)g>g^{\mathrm{T}}[\pi(k)-M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k]g,\quad\forall\, g\in\mathrm{R}^n,\quad g\neq 0,$$

即

$$g^{\mathrm{T}}(M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k)g^{\mathrm{T}}>0\Leftrightarrow I(x(k);Y_k)>0.\tag{11}$$

由矩阵理论知识可得:

$$g^{\mathrm{T}}(M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k)g > 0,\quad \forall\, g\in \mathrm{R}^n,\quad g\neq 0,\quad \forall\, k\geqslant n.$$

等价于

$$\mathrm{rank}[M_k^{\mathrm{T}}R_{Y_k}^{-1}M_k] = n,\quad \forall\, k\geqslant n,$$

即

$$\mathrm{rank}[M_k^{\mathrm{T}}M_k] = n,\quad \forall\, k\geqslant n. \tag{12}$$

而

$$M_k^{\mathrm{T}}M_k = \sum_{i=0}^{k}\Phi(k,i)\pi(i)C^{\mathrm{T}}(i)C(i)\pi(i)\Phi^{\mathrm{T}}(k,i)$$

$$= \sum_{i=0}^{k}\Phi(k,i)N(i)N^{\mathrm{T}}(i)\Phi^{\mathrm{T}}(k,i). \tag{13}$$

其中 $\Phi(k,i) = F(k-1)F(k-2)\cdots F(i),\quad N(i) = \pi(i)C^{\mathrm{T}}(i).$

至此，由定理 2，式(11)，(12)，(13)可得出如下结论：

推论 1(可估计性判定定理) 线性高斯系统(1)为可估计的充分必要条件是可估计性 Grame 矩阵 W_k：

$$W_k \triangleq \sum_{i=0}^{k}\Phi(k,i)N(i)N^{\mathrm{T}}(i)\Phi^{\mathrm{T}}(k,i),\quad \forall\, k\geqslant n \tag{14}$$

满秩，即 $\mathrm{rank}\ W_k = n$.

上述判据是从信息理论的角度由式(4)导出的，它完全等同于文献[1]在线性最小方差估计意义下由式(2)导出的结论. 这也进一步说明了定义 1 和定义 2 的等价性.

3 可估计性与可观测性的关系

确定性系统的可观测性问题在现代控制理论中已很成熟. Kam 等[1]从信息理论的角度对它进行了研究.

考虑一个单输出线性定常系统，即设式(1)中 $x(k)\in \mathrm{R}^n, y(k)\in R^1, k = 0,1,2,\cdots, F, G, C$ 为定常矩阵，$w(k), v(k)$ 是相互独立且与 $x(0)$ 独立的零均值随机噪声，$x(0) = [x_1(0)\quad x_2(0)\quad \cdots\quad x_n(0)]^{\mathrm{T}}$，所有可能的初始状态向量的集合为 X_0.

设上述系统具有离散初始状态，即每个 $x_i(0), i = 1,2,\cdots,n$ 都取值于一个有限的集合 Ω_i. 为了应用信息理论有关概念进行研究，在每个 Ω_i“人为”地定义一个概率测度，$x_i(0)$ 的概率密度由“最客观”的 Jaynes 最大熵准则[5]确定. 设

$$Y = [y(0)\quad y(1)\quad \cdots\quad y(n-1)]^{\mathrm{T}}. \tag{15}$$

引理 3[2] Y 包含了所有可能反映在无限长观测序列 $y(i), i = 0,1,\cdots,n-1,n,\cdots$ 中的 $x(0)$ 的信息，即

$$H\{[y(0)\quad y(1)\quad \cdots\quad y(n-1)\quad \cdots]^{\mathrm{T}}\} = H[Y]. \tag{16}$$

另由信息理论知识，有：

引理 4 对一个离散随机变量 z 和其函数 $f(z)$，有

$$H(z)\geqslant H(f(z)), \tag{17}$$

当且仅当 $f(\cdot)$ 具有单值逆时等号成立.

对上述随机系统，Kam 等给出了关于其相应确定性系统的可观测性问题的如下定理：

定理 3[2] 具有取值于 X_0 的离散初始状态的确定性系统(即在上述系统中 $w(k) = 0$, $v(k) = 0$)，当且仅当

$$H(X_0) = H(Y) \tag{18}$$

时，是可观测的.

定理3的证明可从引理3和引理4得出(详见文献[2]).

虽然式(18)不能作为一个实用的判定条件，然而在下面的分析中我们可以利用它，在信息理论框架下讨论随机系统可估计性和相应确定性系统可观测性之间的关系.

先给出定理：

定理4　若 $\operatorname{rank} F = n$，则上述离散初始状态系统为可估计的，当且仅当其相应的确定性系统(即 $w(k) = 0, v(k) = 0$) 为可观测的.

证　上述随机系统的状态向量为 $x(k)$，输出测量序列为 $Y_k = [y(0) \quad y(1) \quad \cdots \quad y(k)]^{\mathrm{T}}$；相应的确定性系统状态向量为 $x'(k)$，输出序列为 $Y_k' = [y'(0) \quad y'(1) \quad \cdots \quad y'(k)]^{\mathrm{T}}$. 由引理3可知，$H(Y_k') = H(Y'_{k-1})$. 另为方便起见，设 $w(k) = 0$. 设 $V_k = [v(0) \quad v(1) \quad \cdots \quad v(k)]^{\mathrm{T}}$.

充分性. 反证法. 假设当确定性系统为可观测(即 $H(X_0) = H(Y_k'), \forall k \geqslant n$) 时，随机系统是不可估计的，即 $I(x(k);Y_k) = 0, \forall k \geqslant n$.

因为 $\operatorname{rank} F = n$，由信息理论知识有

$$\begin{aligned} I(x(k);Y_k) &= I(X_0;Y_k) = I(Y_k';Y_k') \\ &= H(Y_k) - H(Y_k|Y_k') \\ &= H(Y_k) - H(Y_k' + V_k|Y_k') \\ &= H(Y_k) - H(V_k). \end{aligned} \tag{19}$$

设 $\overline{N}_Y, \overline{N}_{Y'}, \overline{N}_V$ 分别为随机向量 Y_k, Y_k', V_k 的熵功率. 若 $I(x(k);Y_k) = 0$，则由上式及熵功率的定义有

$$\overline{N}_Y = \overline{N}_V. \tag{20}$$

而另一方面，据熵功率不等式有

$$\overline{N}_Y \geqslant \overline{N}_{Y'} + \overline{N}_V. \tag{21}$$

因而由式(20)和(21)：$\overline{N}_{Y'} \leqslant 0$. 这显然与熵功率的定义相矛盾，则假设不成立. 所以当 $H(X_0) = H(Y_k')$ 时，有 $I(x(k);Y_k) > 0$.

必要性. 当 $I(x(k);Y_k) > 0$ 时，据信息理论知识有

$$I(x'(k);Y_k') \geqslant I(x(k);Y_k) > 0. \tag{22}$$

由定理2可知，上式表明：对无噪声干扰的确定性系统，存在 $x'(k)$ 的在给定 Y_k' 下的最优估计. 因没有量测噪声，这一估计值精确地等于状态本身[10]，即 $\hat{x}'(k) = x'(k)$，所以

$$I(x'(k);Y_k') = I(x'(k);\hat{x}'(0)) = I(x'(k);x'(k)) = H(x'(k)).$$

又因为 $\operatorname{rank} F = n$，由上式及引理4有：

$$I(X_0;Y_k') = I(x'(k);Y_k') = H(x'(k)) = H(X_0), \tag{23}$$

而

$$I(X_0;Y_k') = H(X_0) - H(X_0|Y_k'). \tag{24}$$

由式(23)、(24)可得

$$H(X_0|Y_k') = H(Y_k'|X_0) = 0,$$

则

$$I(X_0;Y_k') = H(Y_k') - H(Y_k'|X_0) = H(Y_k'). \tag{25}$$

至此,由(23)、(25)式可得:当 $I(x(k);Y_k) > 0$ 时,有 $H(X_0) = (Y'_k)$. 证毕.

定理 4 可推广到时变系统,其中 rank $F = n$ 的条件相应变为 rank $\Phi(k,0) = n$,$\Phi(k,0) = F(k-1)F(k-2)\cdots F(0)$.

4 结束语

本文运用 Shannon 信息理论方法讨论了随机系统可估计性问题,并研究了线性单输出随机系统可估计性与其相应确定性系统可观测性之间的关系,旨在为随机系统的分析探索一条新的途径.

参考文献

[1] Baram, Y. and Kailath, T.. Estimability and Regulability of Linear Systems. IEEE Trans. Automat. Contr., AC-33 (12):1116—1121

[2] Kam, M., Chenh, R. and Kalata, P.. An Information-Theoretic Interpretation of Stability and Observability. Proc. of the 1987 ACC, 3:1957—1962

[3] Weidemann, H. L.. Entropy Analysis of Feedback Control Systems. Adv. Contr. Syst., 1969, 7:225—255

[4] Tomita, Y., Ohmatsu, S. and Soeda, T.. An Application of the Information Theory to Estimation Problems. Information & Control, 1976, 32:101—111

[5] Jaynes, E. T.. Information Theory and Statistical Mechanics. Physical Review, 1957, 106:620—630

[6] Wolfowitz, J.. Coding Theorems of Information Theory. Second Eition, New York:Springer, 1964

[7] Bellman, R. E.. Introduction to Matrix Analysis. Mcgraw-Hill, New York, 1960

[8] Astrom, K.. 随机控制理论导论. 北京:科学出版社, 1983

[9] 王宏禹. 随机数字信号处理. 北京:科学出版社, 1988

[10] 李树英, 许茂增. 随机系统的滤波与控制. 北京:国防工业出版社, 1991

Estimability of Stochastic Systems——An Information Theory Approach

ZHANG Hui and SUN Youxian

(Institute of Industrial Control Technology, Zhejiang University · Hangzhou, 310027, PRC)

Abstract: This paper is concered with the estimability of stochastic systems by means of information-theoretic concepts, and the connection between the estimablility and the observability of certain systems is also presented.

Key words: stochastic system; estimability; observability; information theory; entropy; mutual information

本文作者简介

章 辉 1967 年生. 浙江大学工业控制技术研究所硕士生. 研究方向为控制系统的信息理论方法, 工业过程模型化及控制.

孙优贤 见本刊 1996 年第 2 期第 258 页.