

离散随机模型降阶的最小条件信息损失方法

章 辉, 孙优贤

(浙江大学 控制科学与工程系 现代控制工程研究所 工业控制技术国家重点实验室, 浙江 杭州 310027)

摘要: 针对随机系统的模型降阶问题, 从分析离散线性随机状态方程模型中的条件信息描述机制入手, 讨论了模型状态集聚过程中系统的平均条件信息损失. 运用在模式识别领域中获得成功应用的最小信息损失准则得出了一种新的模型降阶信息论方法——基于状态集聚的最小条件信息损失方法, 并讨论了降阶模型阶次的选择. 分析表明, 当原系统是渐近稳定时, 由该方法得出的降阶模型也是渐近稳定的. 该方法运用简单, 仿真研究也表明由该方法得出的降阶模型具有良好的近似性能.

关键词: 模型降阶; 状态集聚; 最小信息损失; 熵; 离散线性随机系统

中图分类号: TP13 **文献标识码:** A

Minimum conditional information loss approach to discrete stochastic model reduction

ZHANG Hui, SUN You-xian

(National Laboratory of Industrial Control Technology, Institute of Modern Control Engineering,
Department of Control Science & Engineering, Zhejiang University, Hangzhou Zhejiang 310027, China)

Abstract: Aiming at the problem of stochastic model reduction, the average conditional information loss of the state aggregation is discussed by analyzing the conditional information description of discrete linear state space model. A new information theoretic method for linear stochastic model reduction, the minimum conditional information loss approach based on state aggregation, is deduced by the principle of minimum information loss which was applied successfully in the field of pattern recognition. The selection of the order of reduced-order model is also discussed. It is proved that the derived reduced-order model is asymptotically stable when the full-order model is asymptotically stable. Simulation results show that, this method is easily applicable and the reduced-order models derived by it have good approximation performance.

Key words: model reduction; state aggregation; minimum information loss; entropy; discrete linear stochastic system

1 引言 (Introduction)

动态系统模型降阶是一个长期受到关注的问题. 系统和控制理论中的模型降阶是在一定的误差指标下, 寻求一个较低阶次的模型来逼近原系统模型, 以简化问题复杂度, 满足设计任务的需求. 对于随机系统, 人们研究了很多种模型降阶方法, 如状态集聚 (aggregation) 方法^[1,2]、协方差等价实现方法^[3,4]、随机平衡实现 (stochastic balanced realization) 方法^[5,6]、随机包含方法^[7]、 L_2 方法^[8] 等. 另一类有特色的方法是基于信息论的有关概念和原理上的, 如文献[9]和[10], 通过使全阶和降阶模型输出的 Kullback-Leibler 信息距离^[11] 达到最小来获取降阶模型, 称之为 K-L 信息方法.

本文所提出的是另一种基于状态集聚的模型降

阶信息论方法. 与 K-L 信息方法不同, 这里考虑的原则是, 在获取系统降阶模型过程中, 即在状态集聚过程中使包含在系统状态中的平均条件信息损失达到最小, 因而称之为最小条件信息损失方法. 信息论中以熵函数来描述随机变量所包含的信息量. 对于离散随机向量 $\mu \in \mathbb{R}^n$, $H(\mu) - H(\Gamma\mu) > 0$. 其中: $H(\cdot)$ 表示随机变量的熵, Γ 为具有相应维数的奇异矩阵. 在模式识别和特征提取中得到成功应用的最小信息损失准则^[12] 可表述为, 选择奇异矩阵 Γ 使 $H(\mu) - H(\Gamma\mu)$ 达到最小.

2 问题的描述 (Problem statement)

本文考虑离散线性系统的模型降阶. 设渐近稳定的线性时不变随机系统的全阶模型为

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{w}(k), \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k), \end{aligned} \quad (1)$$

其中: $\mathbf{x}(k) \in \mathbb{R}^n$, $\mathbf{w}(k) \in \mathbb{R}^m$, $\mathbf{y}(k), \mathbf{v}(k) \in \mathbb{R}^p$, $\mathbf{A}, \mathbf{B}, \mathbf{C}$ 为具有相应维数的定常矩阵, $\mathbf{w}(k)$ 和 $\mathbf{v}(k)$ 为相互独立的零均值高斯白噪声向量, 协方差分别为 $\mathbf{Q}$ 和 $\mathbf{R}$. 作者希望寻求一个渐近稳定降阶模型

$$\begin{aligned} \mathbf{x}_r(k+1) &= \mathbf{A}_r\mathbf{x}_r(k) + \mathbf{B}_r\mathbf{w}(k), \\ \mathbf{y}_r(k) &= \mathbf{C}_r\mathbf{x}_r(k) + \mathbf{v}(k) \end{aligned} \quad (2)$$

来近似原系统. 其中: $\mathbf{x}_r(k) \in \mathbb{R}^l$, $l < n$, $\mathbf{y}_r(k) \in \mathbb{R}^p$, $\mathbf{A}_r, \mathbf{B}_r, \mathbf{C}_r$ 为具有相应维数的常阵.

在状态集聚方法^[1]中, $\mathbf{x}_r(k)$ 称为原模型状态 $\mathbf{x}(k)$ 的集聚状态向量 (aggregated state vector), 且

$$\mathbf{x}_r(k) = \mathbf{\Lambda}\mathbf{x}(k). \quad (3)$$

其中 $\mathbf{\Lambda} \in \mathbb{R}^{l \times n}$ 称为集聚矩阵 (aggregation matrix), 为行满秩的, 即 $\text{rank } \mathbf{\Lambda} = l$, 且 $(\mathbf{\Lambda}\mathbf{\Lambda}^T)^{-1}$ 的每个元素都是非负的^[1].

由式(1)~(3)可以得到 $\mathbf{\Lambda}\mathbf{A} = \mathbf{A}_r\mathbf{\Lambda}$, $\mathbf{B}_r = \mathbf{\Lambda}\mathbf{B}$. 降阶模型是对原模型的近似, 即 $\mathbf{y}_r(k) \approx \mathbf{y}(k)$, 于是可得到 $\mathbf{C}_r\mathbf{\Lambda} \approx \mathbf{C}$. 从近似的目的(越接近越好)出发, 为获得降阶模型参数可以设 $\mathbf{y}_r(k) = \mathbf{y}(k)$, 则 $\mathbf{C}_r\mathbf{\Lambda} = \mathbf{C}$ ^[20]. 设 $\mathbf{\Lambda}^+$ 为 $\mathbf{\Lambda}$ 的 Moore-Penrose 伪逆, 当 $\text{rank } \mathbf{\Lambda} = l$ 时, 有^[19] $\mathbf{\Lambda}^+ = \mathbf{\Lambda}^T(\mathbf{\Lambda}\mathbf{\Lambda}^T)^{-1}$. 这样, 在状态集聚方法中可以利用下面的关系来获得降阶模型的参数矩阵^[19,20]:

$$\mathbf{A}_r = \mathbf{\Lambda}\mathbf{A}\mathbf{\Lambda}^+, \mathbf{B}_r = \mathbf{\Lambda}\mathbf{B}, \mathbf{C}_r = \mathbf{C}\mathbf{\Lambda}^+. \quad (4)$$

下文将阐述获得集聚矩阵 $\mathbf{\Lambda}$ 的最小信息损失方法.

3 最小信息损失方法 (Method of minimizing information loss)

模型的动态特性由其结构和参数决定, 而关于动态特性的“信息”则反映在系统状态上. 模型(1)是对真实系统的“全阶”描述, 它所描述的动态信息反映在状态向量 $\mathbf{x}(k)$ 上; 同样, 模型(2)是对真实系统的“降阶”描述, 它所描述的动态信息则反映在状态向量 $\mathbf{x}_r(k)$ 上. 设 $\mathbf{Y}^k = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(k)\}$ 为真实系统输出的一组观测. 条件概率 $p(\mathbf{x}(k+1) | \mathbf{Y}^k)$ 反映了全阶模型在已知观测序列 $\mathbf{Y}^k$ ($\mathbf{Y}^k$ 固定) 条件下对于真实系统的似然度, $p(\mathbf{x}_r(k+1) | \mathbf{Y}^k)$ 则反映了降阶模型对于真实系统的似然度. 而两个模型关于真实系统似然度的差异可采用对数似然率来描述:

$$\ln \frac{p(\mathbf{x}_r(k+1) | \mathbf{Y}^k)}{p(\mathbf{x}(k+1) | \mathbf{Y}^k)} =$$

$$\ln p(\mathbf{x}_r(k+1) | \mathbf{Y}^k) - \ln p(\mathbf{x}(k+1) | \mathbf{Y}^k). \quad (5)$$

若对函数(5)关于 $p(\mathbf{x}_r(k+1) | \mathbf{Y}^k)$ 取数学期望, 则得到关于模型(1)和(2)状态预测的 Kullback-Leibler 信息^[11]

$$KL(\mathbf{x}_r(k+1); \mathbf{x}(k+1) | \mathbf{Y}^k) =$$

$$\sum_{\mathbf{x}(k+1)} p(\mathbf{x}_r(k+1) | \mathbf{Y}^k) \ln \frac{p(\mathbf{x}_r(k+1) | \mathbf{Y}^k)}{p(\mathbf{x}(k+1) | \mathbf{Y}^k)}. \quad (6)$$

当选择集聚矩阵使指标函数(6)达到最小时, 得到的降阶模型在 K-L 信息意义上与全阶模型的距离最小. 这也是文献[9]和[10]所采用的方法(所不同的是文献[9]和[10]所采用的是关于输出的 K-L 信息, 而这里关注系统的状态).

若不是对函数(5)关于 $p(\mathbf{x}_r(k+1) | \mathbf{Y}^k)$ 取数学期望, 而是对式(5)等号右边的第一项和第二项分别关于 $p(\mathbf{x}_r(k+1) | \mathbf{Y}^k)$ 和 $p(\mathbf{x}(k+1) | \mathbf{Y}^k)$ 取数学期望, 则得到

$$IL(\mathbf{x}(k+1), \mathbf{x}_r(k+1) | \mathbf{Y}^k) =$$

$$\sum_{\mathbf{x}_r(k+1)} p(\mathbf{x}_r(k+1) | \mathbf{Y}^k) \ln p(\mathbf{x}_r(k+1) | \mathbf{Y}^k) -$$

$$\sum_{\mathbf{x}(k+1)} p(\mathbf{x}(k+1) | \mathbf{Y}^k) \ln p(\mathbf{x}(k+1) | \mathbf{Y}^k) =$$

$$H(\mathbf{x}(k+1) | \mathbf{Y}^k) - H(\mathbf{x}_r(k+1) | \mathbf{Y}^k). \quad (7)$$

由于已知观测序列 $\mathbf{Y}^k$ 为固定的, 所以上式第二个等号成立. 式中 $H(\cdot | \cdot)$ 表示条件熵. 该式反映的是模型降阶后, 包含在状态向量中的系统动态信息的损失.

设 $\hat{\mathbf{x}}(k+1) = E[\mathbf{x}(k+1) | \mathbf{Y}^k]$ 为已知观测序列 $\mathbf{Y}^k$ ($\mathbf{Y}^k$ 固定, k 足够大) 条件下对于全阶模型(1)的系统状态的一步向前 Kalman 估计, 为条件均值估计. 设 $\bar{\mathbf{x}}(k+1) = \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1)$, $\Sigma(k+1) = E[\bar{\mathbf{x}}(k+1)\bar{\mathbf{x}}^T(k+1)]$. 设 $\bar{\mathbf{y}}(i) = \mathbf{y}(i) - \mathbf{C}\hat{\mathbf{x}}(i)$, $i = 1, 2, \dots, k$. 由于 $\mathbf{x}(i)$ 和 $\mathbf{y}(i)$ 为联合高斯的, 新息序列 $\bar{\mathbf{Y}}^k = \{\bar{\mathbf{y}}(1), \bar{\mathbf{y}}(2), \dots, \bar{\mathbf{y}}(k)\}$ 为一个高斯白噪声过程, 且 $p(\mathbf{x}(k+1) | \mathbf{Y}^k) = p(\mathbf{x}(k+1) | \bar{\mathbf{Y}}^k)$ ^[13]. 对于独立过程 $\bar{\mathbf{Y}}^k$, 由估计问题的有关知识^[13] 可知, $\Sigma(k+1) = \Sigma_{\mathbf{x} | \bar{\mathbf{Y}}^k(k+1)}$, 其中 $\Sigma_{\mathbf{x} | \bar{\mathbf{Y}}^k(k+1)} = \text{cov}\{\mathbf{x}(k+1), \mathbf{x}(k+1) | \bar{\mathbf{Y}}^k\}$ 为已知 $\bar{\mathbf{Y}}^k$ 条件下 $\mathbf{x}(k+1)$ 的协方差. 由于已知观测序列 $\mathbf{Y}^k$ 为固定的, 由信息论知识可得^[14,15]

$$H(\mathbf{x}(k+1) | \mathbf{Y}^k) =$$

$$\sum_{\mathbf{x}(k+1)} p(\mathbf{x}(k+1) | \mathbf{Y}^k) \ln p(\mathbf{x}(k+1) | \mathbf{Y}^k) =$$

$$H(\bar{\mathbf{x}}(k+1)) = \frac{n}{2} \ln(2\pi e) + \frac{1}{2} \ln \det \Sigma(k+1).$$

同样,设 $\hat{x}_k(k+1) = E[x_r(k+1) | Y^k]$ 为已知观测序列 Y^k 条件下对于降阶模型(2)的系统状态向量的一步向前 Kalman 估计,且 $\bar{x}_r(k+1) = x_r(k+1) - \hat{x}_r(k+1)$, $\Sigma_r(k+1) = E[\bar{x}_r(k+1)\bar{x}_r^T(k+1)]$. 则

$$H(x_r(k+1) | Y^k) = H(\bar{x}_r(k+1)) = \frac{l}{2} \ln(2\pi e) + \frac{1}{2} \ln \det \Sigma_r(k+1),$$

因而有

$$IL(x(k+1), x_r(k+1) | Y^k) = H(\bar{x}(k+1)) - H(\bar{x}_r(k+1)). \quad (8)$$

由式(3)可知,状态集聚矩阵 Λ 显然是奇异的,因而在本文所讨论的状态集聚问题中总存在状态信息的损失.

由于系统是渐近稳定的,有 $\lim_{k \rightarrow \infty} \Sigma(k+1) = \Sigma$, $\lim_{k \rightarrow \infty} \Sigma_r(k+1) = \Sigma_r$, 皆为定常矩阵.由 Kalman 估计理论可知,稳态估计误差的协方差阵 Σ 满足 Riccati 方程

$$\Sigma = A(\Sigma - \Sigma C^T(C\Sigma C^T + R)^{-1}C\Sigma)A^T + BQB^T. \quad (9)$$

设

$$\begin{aligned} \tilde{X}^k &= [\bar{x}^T(1), \bar{x}^T(2), \dots, \bar{x}^T(k)]^T, \\ \tilde{X}_r^k &= [\bar{x}_r^T(1), \bar{x}_r^T(2), \dots, \bar{x}_r^T(k)]^T. \end{aligned}$$

定义 Toeplitz 阵

$$T_{\tilde{X}^k} = [Q(j-i)]_{j,i=-k,\dots,-1,0,1,\dots,k}.$$

其中 $Q(j-i) \in \mathbb{R}^{n \times n}$ 为 $\tilde{x}(k)$ 的协方差阵. $\tilde{X}^k$ 的嫡为

$$H(\tilde{X}^k) = \frac{1}{2} \ln(2\pi e)^{nk} + \frac{1}{2} \ln \det T_{\tilde{X}^k},$$

因为 $\bar{x}(k)$ 为零均值的平稳高斯序列,由信息论及 Toeplitz 阵的 Szegö 极限定理可得^[16]

$$\bar{H}(\bar{x}) = \lim_{k \rightarrow \infty} \frac{1}{k} H(\tilde{X}^k) = \frac{n}{2} \ln(2\pi e) + \frac{1}{2} \ln \det \Sigma. \quad (10)$$

同理可得

$$\bar{H}(\bar{x}_r) = \lim_{k \rightarrow \infty} \frac{1}{k} H(\tilde{X}_r^k) = \frac{l}{2} \ln(2\pi e) + \frac{1}{2} \ln \det \Sigma_r. \quad (11)$$

式中 $\bar{H}(\cdot)$ 表示随机过程的嫡率^[14],描述了每单位时间的平均信息量.

函数(8)反映了全阶模型和降阶模型之间的瞬态条件信息损失,而模型随时间演化过程中的平均条件信息则由式(10)和(11)描述.因而可以进一步考虑平均信息损失函数,设为

$$\overline{IL}(x, x_r) = \bar{H}(\bar{x}) - \bar{H}(\bar{x}_r). \quad (12)$$

当选择集聚矩阵 Λ 使 $\overline{IL}(x, x_r)$ 达到最小时,得到最小平均条件信息损失的降阶模型.

由于 $\hat{x}_r(k) = E[\Lambda x(k) | Y^{k-1}] = \Lambda \hat{x}(k)$, 所以 $\bar{x}_r(k) = \Lambda \bar{x}(k)$. 由式(10)~(12)可得

$$\begin{aligned} \overline{IL}(x, x_r) &= \frac{(n-l)}{2} \ln(2\pi e) + \frac{1}{2} (\ln \det \Sigma - \ln \det \Sigma_r) = \\ &= \frac{(n-l)}{2} \ln(2\pi e) + \frac{1}{2} (\ln \det \Sigma - \ln \det (\Lambda \Sigma \Lambda^T)). \end{aligned} \quad (13)$$

设 $\rho_1, \rho_2, \dots, \rho_l$ 为 $\Lambda \Sigma \Lambda^T$ 的特征值, 则 $\det(\Lambda \Sigma \Lambda^T) = \rho_1 \rho_2 \dots \rho_l$. 于是

$$\bar{H}(\bar{x}_r) = \frac{l}{2} \ln(2\pi e) + \frac{1}{2} \sum_{i=1}^l \ln \rho_i.$$

式(13)等号右边第一项是由于系统维数改变而造成的直接信息损失,与集聚矩阵元素的选择无关,因而使 $\overline{IL}(x, x_r)$ 达到最小等价于使 $\sum_{i=1}^l \ln \rho_i$ 达到最大.

而使 $\sum_{i=1}^l \ln \rho_i$ 达到最大的集聚矩阵为^[12]

$$\Lambda = [\eta_1 \quad \eta_2 \quad \dots \quad \eta_l]^T. \quad (14)$$

式中 $\eta_1, \eta_2, \dots, \eta_l$ 分别为相应于估计误差协方差矩阵 Σ 的 l 个最大特征值的正交特征向量, Σ 由 Riccati 方程(9)得出.

至此,降阶模型(2)的参数可通过式(4)获得.注意到式(14)给出的集聚矩阵是由正交特征向量组成的,显然满足 $(\Lambda \Lambda^T)^{-1}$ 的每个元素都是非负的条件.在降阶模型的阶次给定的情况下,由式(2),(4)和(14)得出的降阶模型平均条件信息损失最小.

4 分析与讨论(Analysis and discussion)

4.1 稳定性(Stability)

设 $D(\alpha, \gamma)$ 为复平面上单位圆内一个以实轴上的 α 为圆心、 γ 为半径的圆域.

引理 1^[17] 考虑矩阵方程

$$-\alpha A P - \alpha P A^T + A P A^T + (\alpha^2 - \gamma^2) P = -W. \quad (15)$$

其中 W 为任意正定矩阵.则当且仅当方程(16)存在正定解 P 时,矩阵 A 的特征值在圆域 $D(\alpha, \gamma)$ 内.

设全阶模型(1)满足式(15),既渐近稳定系统(1)的特征值在圆域 $D(\alpha, \gamma)$ 内.对式(15)两边左乘 Λ 和右乘 Λ^T (式(14)),得

$$\begin{aligned} & -\alpha \Lambda A P \Lambda^T - \alpha \Lambda P A^T \Lambda^T + \Lambda A P A^T \Lambda^T + \\ & (\alpha^2 - \gamma^2) \Lambda P \Lambda^T = -\Lambda W \Lambda^T. \end{aligned}$$

由于 $\Lambda\Lambda = \Lambda_r\Lambda$, 因而有

$$-\alpha\Lambda_r\Lambda\Lambda^T - \alpha\Lambda\Lambda^T\Lambda_r^T + \Lambda_r\Lambda\Lambda^T\Lambda_r^T + (\alpha^2 - \gamma^2)\Lambda\Lambda^T = -\Lambda\mathbf{W}\Lambda^T. \quad (16)$$

设 $\Lambda' = [\Lambda^T \quad \eta_{l+1} \quad \cdots \quad \eta_n]^T$, 其中 $\eta_{l+1}, \dots, \eta_n$ 为相应于 Σ 的 $n-l$ 个最小特征值的正交特征向量. 由于 Λ' 是由正交向量组成的, 对于正定阵 $\mathbf{P}$, 可知 $\Lambda'\mathbf{P}(\Lambda')^T$ 也是正定的. 而 $\Lambda\Lambda^T$ 为 $\Lambda'\mathbf{P}(\Lambda')^T$ 主对角线上的 $l \times l$ 子阵. 所以当 $\mathbf{P}$ 和 $\mathbf{W}$ 正定时, $\Lambda\Lambda^T$ 和 $\Lambda\mathbf{W}\Lambda^T$ 也是正定的. 由引理 1 和式 (16) 可知降阶模型 (2) 的特征值也在圆域 $D(\alpha, \gamma)$ 内. 因而, 当原全阶模型是渐近稳定时, 由本文方法得出的降阶模型也是渐近稳定的.

另外, 由系统稳定性理论可知: 原系统 (1) 为渐近稳定的, 当且仅当对于任意给定的正定矩阵 $\mathbf{W}$, 方程 $\Lambda\mathbf{P}\Lambda^T + \mathbf{W} = \mathbf{P}$ 存在正定解 $\mathbf{P}$. 对该方程两边左乘 Λ 和右乘 Λ^T , 通过与上面相同的分析方法也可得出这样的结论: 由本文方法得出的降阶模型保持了原模型的稳定性.

4.2 关于降阶模型阶次的选择 (Selection of the order for approximation model)

全阶模型状态估计误差的平均信息量为

$$\begin{aligned} \bar{H}(\bar{\mathbf{x}}) &= \frac{n}{2} \ln(2\pi e) + \frac{1}{2} \ln \det \Sigma = \\ &= \frac{n}{2} \ln(2\pi e) + \frac{1}{2} \sum_{i=1}^n \ln \lambda_i. \end{aligned} \quad (17)$$

其中 $\lambda_i (i = 1, 2, \dots, n)$ 为 Σ 的特征值 (为正实数), 设它们排列为 $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$. 由上式可知每个特征值对平均信息量的贡献不同. 由于 Λ 由正交向量组成, 所以

$$\begin{aligned} \bar{H}(\bar{\mathbf{x}}_r) &= \frac{l}{2} \ln(2\pi e) + \frac{1}{2} \sum_{i=1}^l \ln \rho_i = \\ &= \frac{l}{2} \ln(2\pi e) + \frac{1}{2} \sum_{i=1}^l \ln \lambda_i. \end{aligned} \quad (18)$$

即 $\rho_1, \rho_2, \dots, \rho_l$ 为 Σ 的 l 个最大特征值 $\lambda_1, \lambda_2, \dots, \lambda_l$. 因而模型降阶带来的平均信息损失为

$$\begin{aligned} \bar{H}(\mathbf{x}, \mathbf{x}_r) &= \bar{H}(\bar{\mathbf{x}}) - \bar{H}(\bar{\mathbf{x}}_r) = \\ &= \frac{n-l}{2} \ln(2\pi e) + \frac{1}{2} \sum_{i=l+1}^n \ln \lambda_i. \end{aligned} \quad (19)$$

由式 (19) 可知, 除了阶次减小的直接因素 (式中第一项) 外, 平均信息损失体现在模型降阶中“丢弃”的特征值 $\lambda_{l+1}, \lambda_{l+2}, \dots, \lambda_n$ 上 (式中第二项), $\lambda_j, j = l+1, \dots, n$ 越大, 信息损失越多. 从这个因素出发, 若相对其他两个相邻特征值的差距来说 $\lambda_{l1} \gg \lambda_{l1+1}$, 则可考虑将 $l1$ 作为降阶模型的阶次 (即取 $l = l1$),

这样因“丢弃”特征值 $\lambda_{l+1}, \lambda_{l+2}, \dots, \lambda_n$ 带来的信息损失相对较小. 这一选择方法与文献 [18] 建议的方法类似. 文献 [18] 的方法考虑的是标准相关矩阵 (canonical correlation matrix) 的特征值, 而这里则考察估计误差协方差阵的特征值. 当然, 降阶模型阶次的确定还须考虑设计任务 (对降阶模型阶次的要求)、误差允许范围及阶次减小带来的直接信息损失等其他因素.

5 例子 (Example)

针对文献 [3] 所给的离散 SISO 模型, 将本文的结果与文献 [3] 所得的结果进行比较. 系统的全阶状态方程模型为

$$\begin{cases} \mathbf{x}(k+1) = \begin{bmatrix} -1.1 & 1 & 0 & 0 \\ 0.01 & 0 & 1 & 0 \\ 0.275 & 0 & 0 & 1 \\ 0.06 & 0 & 0 & 0 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} w(k), \\ E[w(k) \quad w(k)] = 1, \end{cases} \quad (20)$$

$$\begin{cases} y(k) = [1 \quad 0 \quad 0 \quad 0] \mathbf{x}(k) + v(k), \\ E[v(k) \quad v(k)] = 1. \end{cases} \quad (21)$$

$w(k)$ 和 $v(k)$ 为相互独立的零均值高斯白噪声. 系统极点为 $\{0.5, -0.8, -0.5, -0.3\}$, 系统是渐近稳定的. 设一个复平面上单位圆内的以点 -0.15 (在实轴上) 为原点、以 0.6501 为半径的圆域, 记为 $D(-0.15, 0.6501)$. 全阶模型的极点都在圆域 $D(-0.15, 0.6501)$ 内.

对于式 (20) 和 (21) 描述的 4 阶模型, 文献 [3] 由协方差等价实现方法得出的两个 2 阶近似模型为

$$\begin{cases} M_{m1}, M_{m2}: \mathbf{x}_{mi}(k+1) = \mathbf{A}_{mi} \mathbf{x}_{mi}(k) + \mathbf{B}_{mi} w(k), \\ y_{mi}(k) = \mathbf{C}_{mi} \mathbf{x}_{mi}(k) + v(k), \\ i = 1, 2. \end{cases} \quad (22)$$

其中

$$\begin{aligned} \mathbf{A}_{m1} &= \begin{bmatrix} 0 & 1 \\ 0.7640 & -0.0855 \end{bmatrix}, \\ \mathbf{A}_{m2} &= \begin{bmatrix} 0 & 1 \\ -0.7622 & -1.7093 \end{bmatrix}, \\ \mathbf{B}_{m1} &= \mathbf{B}_{m2} = \begin{bmatrix} 1 \\ -1.1 \end{bmatrix}, \quad \mathbf{C}_{m1} = \mathbf{C}_{m2} = [1 \quad 0]. \end{aligned}$$

M_{m1} 和 M_{m2} 的极点分别为 $\{0.8324, -0.9179\}$ 和 $\{-0.8547 + 0.1783i, -0.8547 - 0.1783i\}$ (都不在圆域 $D(-0.15, 0.6501)$ 内).

下面运用本文的方法考察模型 (20) (21).

对于模型(20)(21),通过求解 Riccati 方程(式(9))获得其状态估计误差协方差阵,进一步得该协方差阵的特征值为

$$\lambda_1 = 1.77332720, \lambda_2 = 0.03009852,$$

$$\lambda_3 = 0.02325786, \lambda_4 = 0.00000333.$$

可以看出,相对其他两个相邻特征值来说, λ_3 和 λ_4 的差距最大,即 $\lambda_3 \gg \lambda_4$. 据 4.2 部分中分析,当选择降阶模型的阶次为 $l = 3$ 时,因“丢弃”特征值所带来的信息损失会很小(当然此时因阶次降低而直接带来的信息损失也是最小的). 基于这点考虑,同时为了与文献[3]得出的 2 阶模型进行比较,这里用本文算法(式(2),(4)和(14))分别得出 3 阶和 2 阶近似模型 M_{r1} 和 M_{r2} :

$$\begin{cases} M_{r1}, M_{r2}: \mathbf{x}_{ri}(k+1) = \mathbf{A}_{ri} \mathbf{x}_{ri}(k) + \mathbf{B}_{ri} w(k), \\ y_{ri}(k) = \mathbf{C}_{ri} \mathbf{x}_{ri}(k) + v(k), \\ i = 1, 2. \end{cases} \quad (23)$$

其中

$$\mathbf{A}_{r1} = \begin{bmatrix} -1.0835 & 0.8998 & 0.5215 \\ -0.1296 & -0.4246 & 0.5075 \\ 0.0875 & -0.4190 & 0.6292 \end{bmatrix},$$

$$\mathbf{B}_{r1} = \begin{bmatrix} 0.9942 \\ -0.0868 \\ 0.0637 \end{bmatrix},$$

$$\mathbf{C}_{r1} = [0.9942 \quad -0.0868 \quad 0.0637],$$

$$\mathbf{A}_{r2} = \begin{bmatrix} -1.0835 & 0.8998 \\ -0.1296 & -0.4246 \end{bmatrix},$$

$$\mathbf{B}_{r2} = \begin{bmatrix} 0.9942 \\ -0.0868 \end{bmatrix}, \mathbf{C}_{r2} = [0.9942 \quad -0.0868].$$

降阶模型 M_{r1} 和 M_{r2} 的系统极点分别为 $\{-0.7802, -0.5829, 0.4841\}$, $\{-0.7541 + 0.0902i, -0.7541 - 0.0902i\}$ 显然都在圆域内 $D(-0.15, 0.6501)$, 因而 M_{r1} 和 M_{r2} 是渐近稳定的. 这与 4.1 部分中的分析是一致的.

运用 MATLAB 进行仿真. 首先将本文得出的 2 阶模型 M_{r2} 与文献[3]中近似性能较好的 2 阶模型 M_{m2} 进行比较. 图 1 和图 2 分别为降阶模型 M_{m2} , M_{r2} 与原模型的 Bode 图(图中所示为 Nyquist 频率以下频段的幅、相频特性). 虽然 M_{r2} 与 M_{m2} 的幅频和相频特性在中、低频率段与原模型的吻合程度大致相同,但在高频段 M_{r2} 与原模型几乎完全一致,而 M_{m2} 与原模型还是有差距的. 所以对于高频随机的响应, M_{r2} 将比 M_{m2} 更接近原模型(这也将在下文的响应误差指标 D_{m2} 和 D_{r2} 上得到体现). 图中: F 为频率,

M 为增益, P 为相角.

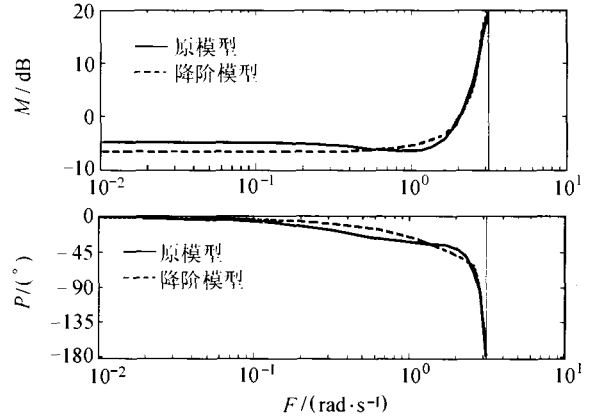


图 1 降阶模型 M_{m2} 与原模型的 Bode 图

Fig. 1 Bode plots of the reduced-order model M_{m2} and the original model

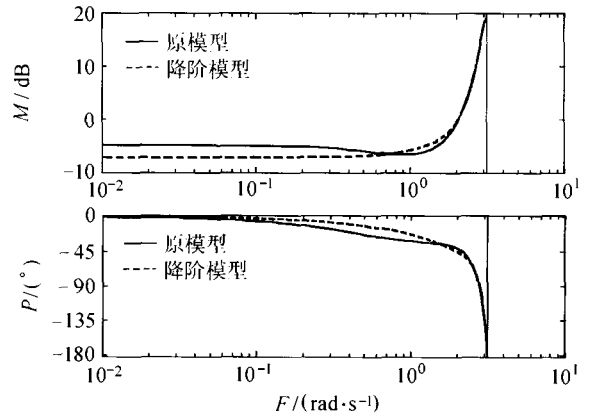


图 2 降阶模型 M_{r2} 与原模型的 Bode 图

Fig. 2 Bode plots of the reduced-order model M_{r2} and the original model

再看各模型对高斯白噪声输入 $w(k)$ ($k = 1, 2, \dots, K$) 的响应. 设检验降阶模型输出的误差指标为

$$D_* = \frac{1}{K} \sum_{k=1}^K (y(k) - y_*(k))^2.$$

其中 $*$ 分别为 m_1, m_2, r_1 和 r_2 . 仿真结果分别为 ($K = 200$)

$$D_{m1} = 0.6869, D_{m2} = 0.3622,$$

$$D_{r1} = 0.0025, D_{r2} = 0.0188.$$

图 3 所示分别为各降阶模型的响应误差: $y(k) - y_{m1}(k)$, $y(k) - y_{m2}(k)$, $y(k) - y_{r1}(k)$ 和 $y(k) - y_{r2}(k)$. 从仿真结果可以看出,同文献[3]得出的降阶模型相比,由本文方法得出的近似模型对于高斯白噪声输入的响应误差更小.

需说明的是,从随机系统一般描述考虑并与本文方法的推导过程相适应,这里所给的全阶模型(式(21))中增加了观测噪声 $v(k)$ 项. 虽然文献[3]所给

的例子中不包含该项(其他都一致),但由于文献[3]方法的推导与 $v(k)$ 无关,且这里检验模型动态的 Bode 图、输出响应误差(图3)和模型误差指标中都不包含 $v(k)$ 项,因而文献[3]的协方差等价实现方法与本文的最小信息损失方法在模型(20)和(21)上仍具有可比性。

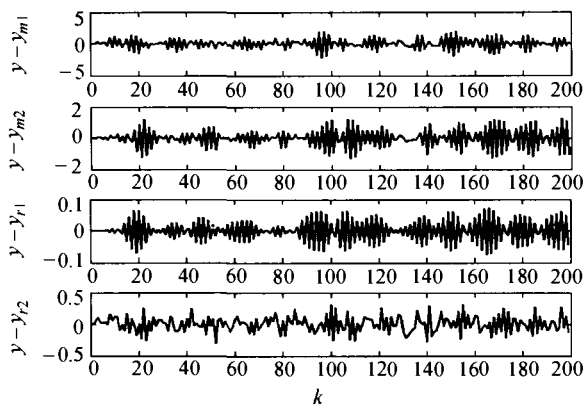


图3 降阶模型输出误差

Fig. 3 Output errors of reduced-order models

6 结束语(Conclusion)

本文针对离散线性随机系统,分析了状态方程模型的信息描述机制,讨论了系统状态集聚过程中所造成的(平均)信息损失。受模型降阶的 Kullback-Leibler 信息准则的启发,本文得出一种新的模型降阶信息论方法——基于状态集聚的最小平均信息损失方法,并讨论了降阶模型阶次的选择。分析表明,当原系统是渐近稳定时,由该方法得出的降阶模型也是渐近稳定的。与其他模型降阶方法相比(如文献[3,4,8,10]),本文提出的方法运用简单。仿真结果表明由该方法得出的降阶模型具有较小的近似误差。

参考文献(References):

- [1] AOKI M. Control of large-scale dynamic systems by aggregation [J]. *IEEE Trans on Automatic Control*, 1968, 13(2): 246 – 253.
- [2] MICHAILESCO G, SIRET J M, BERTRAND P. Aggregated models for high-order systems [J]. *Electronics Letters*, 1975, 11(17): 398 – 399.
- [3] WAGIE D A, SKELTON R S. A projection approach to covariance equivalent realizations of discrete systems [J]. *IEEE Trans on Automatic Control*, 1986, 31(12): 1114 – 1120.
- [4] WANG Z, UNBEHAUEN H. Model reduction based on regional pole and covariance equivalent realizations [J]. *IEEE Trans on Automatic Control*, 1999, 44(10): 1889 – 1893.
- [5] GUGERCIN S, ANTOUNAS A C. A survey of model reduction by balanced truncation and some new results [J]. *Int J Control*, 2004, 77(8): 748 – 766.
- [6] VARGA A. On stochastic balancing related model reduction [C]// *Proc of the 39th IEEE Conf on Deciaion & Control*. Sydney: IEEE Press, 2000: 2385 – 2390.
- [7] 陈雪波, STANKOVIC S S. 离散随机系统的包含原理[J]. *自动化学报*, 1997, 23(1): 94 – 97.
(CHEN Xuebo, STANKOVIC S S. Inclusion principle for discrete-time stochastic systems [J]. *Acta Automatica Sinica*, 1997, 23(1): 94 – 97.)
- [8] TJÄRNSTRÖM F, LJUNG L. L_2 model reduction and variance reduction [J]. *Automatica*, 2002, 38(9): 1517 – 1530.
- [9] BARAM Y, BE'ERI Y. Stochastic model simplification [J]. *IEEE Trans on Automatic Control*, 1981, 26(2): 379 – 390.
- [10] LELAND R. Reduced-order models and controllers for continuous-time stochastic systems: an information theory approach [J]. *IEEE Trans on Automatic Control*, 1999, 44(9): 1714 – 1719.
- [11] KULLBACK S. *Information Theory and Statistics* [M]. New York: John & Wiley, 1959.
- [12] KAPUR J N, KESAVAN H K. *Entropy Optimization Principles with Applications* [M]. San Diego: Academic Press, 1992.
- [13] 韩崇昭, 王月娟, 万百五. 随机系统理论[M]. 西安: 西安交通大学出版社, 1989.
(HAN Chongzhao, WANG Yuejuan, WANG Baiwu. *Stochastic Systems Theory* [M]. Xi'an: Xi'an Jiaotong University Press, 1989.)
- [14] COVER T M, THOMAS J A. *Elements of Information Theory* [M]. Chichester: John Wiley & Sons, 1991.
- [15] 章辉, 孙优贤. 随机系统的可估计性——信息论方法[J]. *控制理论与应用*, 1996, 13(5): 567 – 572.
(ZHANG Hui, SUN Youxian. Estimability of stochastic systems: an information theory approach [J]. *Control Theory & Applications*, 1996, 13(5): 567 – 572.)
- [16] CAINES P E. *Linear stochastic systems* [M]// *Wiley Series in Probability and Mathematical Statistics*. New York: John Wiley & Sons, 1988.
- [17] FURUTA K, KIM S B. Pole assignment in a special disk [J]. *IEEE Trans on Automatic Control*, 1987, 32(5): 423 – 427.
- [18] DESAI U B, PAL D. A realization approach to stochastic model reduction and balanced stochastic realizations [C]// *Proc of the 21st IEEE Conf on Decision & Control*. Orlando: IEEE Press, 1982: 1105 – 1112.
- [19] HICKIN J, SINKA N K. Model reduction for linear multivariable systems [J]. *IEEE Trans on Automatic Control*, 1980, 25(6): 1121 – 1127.
- [20] LAMBA S S, RAO S V. Aggregation matrix for the reduced-order continued fraction expansion model of Chen and Shieh [J]. *IEEE Trans on Automatic Control*, 1978, 23(1): 81 – 83.

作者简介:

章辉 (1967—), 男, 博士, 浙江大学控制系教师, 研究方向为随机系统、鲁棒控制、控制系统的信息论方法等, E-mail: zhanghui_iipc@zju.edu.cn;

孙优贤 (1940—), 男, 教授, 博士生导师, 中国工程院院士, 浙江大学工业自动化国家工程中心主任, 研究领域有鲁棒控制、容错控制、非线性系统、软测量技术、工业系统模型化与控制等, E-mail: yx-sun@iipc.zju.edu.cn.