

基于平衡截断的降阶模型

黄政宇

北京大学北京国际数学研究中心
北京大学国际机器学习研究中心



本堂课大纲

- 线性动力系统
 - 输入输出映射
 - 解析解
- 状态分类
 - 可达状态
 - 可控状态
 - 可观测状态
- 平衡截断法
 - 算法
 - 误差分析



线性动力系统

➤ 线性模型

$$\begin{aligned}\frac{d}{dt}\mathbf{u}(t) &= A\mathbf{u}(t) + Bw(t) \\ y(t) &= C\mathbf{u}(t) + Dw(t)\end{aligned}$$

➤ 输入函数

$$t \rightarrow w(t) \in R^q$$

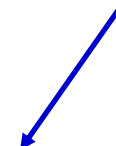
➤ 输出函数

$$t \rightarrow y(t) \in R^p$$

➤ 假设输入输出映射满足

$$w(t) \rightarrow y(t) \quad y(t) = \int_{-\infty}^{\infty} h(t, \tau)w(\tau)d\tau$$

核函数 $h(t, \tau) \in R^{p \times q}$





线性动力系统

➤ 输入输出映射

$$w(t) \rightarrow y(t) \quad y(t) = \int_{-\infty}^{\infty} h(t, \tau) w(\tau) d\tau$$

➤ 系统响应与其输入之间有因果关系 (Causality)

$$h(t, \tau) = 0 \quad \forall \tau > t$$

➤ 系统的特性不随时间的变化而变化 (Time invariant)

$$h(t, \tau) = h(t - \tau)$$

$$y(t) = \int_{-\infty}^{\infty} h(t - \tau) w(\tau) d\tau = (h * w)(t)$$

➤ 线性因果 (Causality) 时不变 (Time invariant)

$$y(t) = \int_{-\infty}^t h(t - \tau) w(\tau) d\tau = (h * w)(t)$$



线性动力系统

核函数分解

$$h(t) = h_0 \delta(t) + h_a(t)$$

脉冲响应部分： $h_0 \delta(t)$

平滑核函数部分： $h_a(t)$

$$h_a(t) = h_1 + h_2 \frac{t}{1} + h_3 \frac{t^2}{2!} \cdots h_k \frac{t^{k-1}}{(k-1)!} \cdots \in R^{p \times q}$$

输入输出函数

线性因果
(Causality)
时不变
(Time
invariant) 的
线性动力系统



Markov 参数
 $\Sigma = (h_0, h_1, h_2, \cdots, h_k \cdots)$



线性动力系统

➤ Laplace变化

$$H(\xi) = (\mathcal{L}h)(\xi) = \int_0^{\infty} h(t)e^{-\xi t} d\tau$$

$$h(t) = (\mathcal{L}^{-1}H)(t) = \frac{1}{2\pi i} \lim_{T \rightarrow \infty} \int_{\gamma - iT}^{\gamma + iT} H(\xi)e^{\xi t} d\xi$$

➤ 输入输出映射

$$y(t) = \int_{-\infty}^t h(t - \tau)w(\tau)d\tau = (h * w)(t)$$

$$(\mathcal{L}h)(\xi) = h_0 + h_1\xi^{-1} + h_2\xi^{-2} + h_3\xi^{-3} \dots h_k\xi^{-k} \dots$$

$$\begin{aligned} (\mathcal{L}y)(\xi) &= \int_0^{\infty} \int_{-\infty}^t h(t - \tau)w(\tau)d\tau e^{-\xi t} dt \\ &= (\mathcal{L}h)(\xi) (\mathcal{L}w)(\xi) \end{aligned}$$



线性动力系统

➤ 线性模型

$$\begin{aligned}\frac{d}{dt}\mathbf{u}(t) &= A\mathbf{u}(t) + Bw(t) \\ y(t) &= C\mathbf{u}(t) + Dw(t) \\ \mathbf{u}(t_0) &= \mathbf{u}_0\end{aligned}$$

➤ 解析解

$$\begin{aligned}\mathbf{u}(t, w; t_0, \mathbf{u}_0) &= e^{A(t-t_0)}\mathbf{u}_0 + \int_{t_0}^t e^{A(t-\tau)}Bw(\tau)d\tau \\ y(t) &= Ce^{A(t-t_0)}\mathbf{u}_0 + C \int_{t_0}^t e^{A(t-\tau)}Bw(\tau)d\tau + Dw(t) \\ &:= \phi(t, w; t_0, \mathbf{u}_0)\end{aligned}$$



线性动力系统

➤ 线性模型

$$\begin{aligned}\frac{d}{dt}\mathbf{u}(t) &= A\mathbf{u}(t) + Bw(t) \\ y(t) &= C\mathbf{u}(t) + Dw(t) \\ \mathbf{u}(-\infty) &= \mathbf{0}\end{aligned}$$

➤ 解析解

$$\mathbf{u}(t, w; -\infty, \mathbf{0}) = \int_{-\infty}^t e^{A(t-\tau)} B w(\tau) d\tau$$

$$y(t) = C \int_{-\infty}^t e^{A(t-\tau)} B w(\tau) d\tau + Dw(t) = \int_{-\infty}^t h(t-\tau) w(\tau) d\tau$$

$$h(t) = \begin{cases} Ce^{At}B + \delta(t)D, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad H(\xi) = D + C(\xi I - A)^{-1}B$$



线性动力系统

➤ 线性模型

$$\begin{aligned}\frac{d}{dt}\mathbf{u}(t) &= A\mathbf{u}(t) + Bw(t) \\ y(t) &= C\mathbf{u}(t) + Dw(t) \\ \mathbf{u}(-\infty) &= \mathbf{0}\end{aligned}$$

➤ 输入输出映射

$$\begin{aligned}H(\xi) &= D + C(\xi I - A)^{-1}B \\ &= h_0 + h_1\xi^{-1} + h_2\xi^{-2} + h_3\xi^{-3} \dots h_k\xi^{-k} \dots \\ \Sigma &= (D, CB, CAB, CA^2B \dots, CA^{k-1}B \dots)\end{aligned}$$

坐标变化 $\tilde{\mathbf{u}} = T\mathbf{u}$ ($\det(T) \neq 0$)

$$\tilde{\Sigma} = (\tilde{D}, \tilde{C}\tilde{B}, \tilde{C}\tilde{A}\tilde{B}, \tilde{C}\tilde{A}^2\tilde{B} \dots, \tilde{C}\tilde{A}^{k-1}\tilde{B} \dots) = \Sigma$$



线性动力系统

➤ 线性模型

$$\begin{aligned}\frac{d}{dt}\mathbf{u}(t) &= A\mathbf{u}(t) + Bw(t) \\ y(t) &= C\mathbf{u}(t) + Dw(t)\end{aligned}$$

四元组： $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$

Markov 参数：

$$\Sigma = (D, CB, CAB, CA^2B \dots, CA^{k-1}B \dots)$$

核函数 $h(t)$ ，输入输出映射：

$$(\mathcal{L}y)(\xi) = (\mathcal{L}h)(\xi) (\mathcal{L}w)(\xi)$$



本堂课大纲

- 线性动力系统
 - 输入输出映射
 - 解析解

- 状态分类
 - 可达状态
 - 可控状态
 - 可观测状态

- 平衡截断法
 - 算法
 - 误差分析



可达状态

可达状态(reachability)

给定四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，状态 $\bar{\mathbf{u}} \in \mathbf{R}^N$ 是可达的，如果存在一个有限能量的输入函数 w 和一个有限时间 \bar{T} ，使得在该输入作用下且初始状态为零的情况下，系统状态能够变为 $\bar{\mathbf{u}}$

$$\bar{\mathbf{u}} = \mathbf{u}(\bar{T}, w; t_0, 0)$$

定义所有可达状态的集合为 X^{reach} ，如果 $X^{\text{reach}} = \mathbf{R}^N$ ，那么该系统是完全可达的。



可达状态

可达状态(reachability)的判定

给定四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，我们定义可达矩阵

$$R(A, B) = [B, AB, A^2B, \dots, A^k B \dots],$$

和有限可达Gramians矩阵

$$\mathcal{P}(t) = \int_0^t e^{A\tau} B B^* e^{A^* \tau} d\tau$$

那么

$$X^{\text{reach}} = \text{im } R(A, B) = \text{im } \mathcal{P}(t) \quad \forall t > 0$$



可达状态

可达状态的最小能量

给点四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，对于可达状态 $\bar{\mathbf{u}} \in R^N$ ，给定 \bar{T} ，如果存在一个有限能量的输入函数 \bar{w} ，使得在该输入作用下且初始状态为零的情况下，系统状态能够在 \bar{T} 时间之内变为 $\bar{\mathbf{u}}$ ，那么

$$\|\bar{w}\|^2 = \int_0^{\bar{T}} \bar{w}^*(t)\bar{w}(t)dt \geq \bar{\mathbf{u}}^* \mathcal{P}(\bar{T})^{-1} \bar{\mathbf{u}}$$

等号取到的 \bar{w} 满足

$$\bar{w}(t) = B^* e^{A^*(\bar{T}-t)} \mathcal{P}(\bar{T})^{-1} \bar{\mathbf{u}}$$



可控状态

可控状态 (controllability)

给定四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，非零状态 $\bar{\mathbf{u}} \in R^N$ 是可控的，如果存在一个有限能量的输入函数 w 和一个有限时间 \bar{T} ，使得在该输入作用下且初始状态为 $\bar{\mathbf{u}}$ 的情况下，系统状态能够变为 0

$$0 = \mathbf{u}(\bar{T}, w; t_0, \bar{\mathbf{u}})。$$

定义所有可控状态的集合为 X^{contr} ，那么

$$X^{\text{reach}} = X^{\text{contr}}。$$

如果 $X^{\text{contr}} = R^N$ ，那么该系统是完全可控的。



可观测状态

可观测状态 (observability)

给定四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，状态 $\bar{\mathbf{u}} \in R^N$ 是不可观测的，如果以 $\bar{\mathbf{u}}$ 为初始状态，采用 0 输入函数，系统观测永远为 0

$$y(t) = Cu(t, 0; 0, \bar{\mathbf{u}})$$

定义所有不可观测状态的集合为 X^{unobs} ，如果 $X^{\text{unobs}} = \{0\}$ ，那么该系统是完全可观测的。



可观测状态

观测状态的判定

给定四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，我们定义可观测矩阵

$$\mathcal{O}(A, C) = [C, CA, CA^2, \dots, CA^k \dots]$$

有限可观测Gramians矩阵

$$Q(T) = \int_0^T e^{A^* \tau} C^* C e^{A \tau} d\tau$$

$$X^{\text{unobs}} = \text{Ker } \mathcal{O}(A, C) = Q(t) \quad \forall t$$



可观测状态

输出函数的能量

给点四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，状态 $\bar{\mathbf{u}} \in R^N$ 是可观测的，如果以 $\bar{\mathbf{u}}$ 为初始状态，采用 0 输入函数，系统观测为 $y(t)$ ，那么观测能量为

$$\|y\|^2 = \int_0^{\bar{T}} y(t)^* y(t) dt = \bar{\mathbf{u}}^* Q(\bar{T}) \bar{\mathbf{u}}$$



可达性可观测性规范分解

可达性可观测性规范分解(Canonical decomposition)

给定四元组 $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ ，存在坐标变换

$$\begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix} = \begin{pmatrix} A_{r\bar{o}} & A_{12} & A_{13} & A_{14} & B_{r\bar{o}} \\ 0 & A_{r\bar{o}} & 0 & A_{24} & B_{r\bar{o}} \\ 0 & 0 & A_{\bar{r}\bar{o}} & A_{34} & 0 \\ 0 & 0 & 0 & A_{\bar{r}\bar{o}} & 0 \\ 0 & C_{r\bar{o}} & 0 & C_{\bar{r}\bar{o}} & \tilde{D} \end{pmatrix}$$

使得 $\begin{pmatrix} A_{r\bar{o}} & B_{r\bar{o}} \\ 0 & * \end{pmatrix}$ 是完全可达但不可观测的， $\begin{pmatrix} A_{r\bar{o}} & B_{r\bar{o}} \\ C_{r\bar{o}} & * \end{pmatrix}$

是完全可达可观测的， $\begin{pmatrix} A_{\bar{r}\bar{o}} & B_{\bar{r}\bar{o}} \\ C_{\bar{r}\bar{o}} & * \end{pmatrix}$ 是不可达不可观测的，

$\begin{pmatrix} A_{\bar{r}\bar{o}} & 0 \\ C_{\bar{r}\bar{o}} & * \end{pmatrix}$ 是不可达但完全可观测的。



本堂课大纲

- 线性动力系统
 - 输入输出映射
 - 解析解

- 状态分类
 - 可达状态
 - 可控状态
 - 可观测状态

- 平衡截断法
 - 算法
 - 误差分析



平衡截断法(Balanced truncation)

➤ 可达、可观测Gramians矩阵

$$\mathcal{P}(t) = \int_0^t e^{A\tau} B B^* e^{A^*\tau} d\tau \quad \mathcal{Q}(t) = \int_0^t e^{A^*\tau} C^* C e^{A\tau} d\tau$$

$$\mathcal{P}(t_1) \geq \mathcal{P}(t_2) \quad \mathcal{Q}(t_1) \geq \mathcal{Q}(t_2) \quad (t_1 > t_2)$$

➤ 无限时间Gramians矩阵

$$\mathcal{P} = \int_0^\infty e^{A\tau} B B^* e^{A^*\tau} d\tau \quad \mathcal{Q} = \int_0^\infty e^{A^*\tau} C^* C e^{A\tau} d\tau$$

满足

$$A\mathcal{P} + \mathcal{P}A^* + B B^* = 0$$

$$A^*\mathcal{Q} + \mathcal{Q}A + C^* C = 0$$



平衡截断法(Balanced truncation)

- 对于一个可达的系统，基于 \mathcal{P}^{-1} 的内积表征了将状态从0导向 \mathbf{u} 所需的最小能量，当时间 t 趋向于无穷大时。

$$\|\bar{\mathbf{w}}\|^2 = \int_0^{\bar{T}} \bar{\mathbf{w}}^*(t)\bar{\mathbf{w}}(t)dt \geq \mathbf{u}^T \mathcal{P}(\bar{T})^{-1} \mathbf{u} \geq \|\mathbf{u}\|_{\mathcal{P}^{-1}}^2 = \mathbf{u}^T \mathcal{P}^{-1} \mathbf{u}$$

- 基于 Q 的内积表示了无输入条件下，观察与初始状态 \mathbf{u}_0 相对应的系统输出所能产生的最大能量。

$$\|\mathbf{y}\|^2 = \int_0^{\bar{T}} \mathbf{y}(t)^* \mathbf{y}(t) dt = \mathbf{u}_0^* Q(\bar{T}) \mathbf{u}_0 \leq \|\mathbf{u}_0\|_Q^2 = \mathbf{u}_0^T Q \mathbf{u}_0$$



平衡截断法(Balanced truncation)

➤ 降阶模型(近似状态 \mathbf{u})

- 难以到达的状态，即需要很大的能量 $\mathbf{u}^T \mathcal{P}^{-1} \mathbf{u}$ 才能到到达
- 难以观测的状态，即观测到的能量 $\mathbf{u}^T \mathcal{Q} \mathbf{u}$ 很小

忽略这些状态

这些状态的定义和基底的选取相关

我们希望找到一组基，在这组基下，可达性和可观测性这两个概念是等价的，即系统处于平衡状态。



平衡截断法(Balanced truncation)

➤ 平衡变换

$$\tilde{\mathbf{u}} = T\mathbf{u}$$
$$\begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix} = \begin{pmatrix} TAT^{-1} & TB \\ CT^{-1} & D \end{pmatrix} \quad \longrightarrow \quad \tilde{\mathcal{P}} = \tilde{\mathcal{Q}}$$

$$\tilde{\mathcal{P}} = \int_0^{\infty} e^{\tilde{A}\tau} \tilde{B} \tilde{B}^* e^{\tilde{A}^*\tau} d\tau = T\mathcal{P}T^*$$

$$\tilde{\mathcal{Q}} = \int_0^{\infty} e^{\tilde{A}^*\tau} \tilde{C}^* \tilde{C} e^{\tilde{A}\tau} d\tau = T^{-*}\mathcal{Q}T^{-1}$$



平衡截断法(Balanced truncation)

➤ 平衡变换

$$\tilde{\mathcal{P}} = T_{bal} \mathcal{P} T_{bal}^* = T_{bal}^{-*} \mathcal{Q} T_{bal}^{-1} = \tilde{\mathcal{Q}}$$

Cholesky分解：

$$\mathcal{P} = LL^*$$

特征值分解：

$$L^* \mathcal{Q} L = K \Sigma^2 K^*$$

其中对角矩阵 Σ 中的元素按降序排列。

$$T_{bal} = \Sigma^{1/2} K^* L^{-1}$$

$$T_{bal}^{-1} = L K \Sigma^{-1/2}$$

$$\tilde{\mathcal{P}} = \tilde{\mathcal{Q}} = \Sigma$$

$\Sigma = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_N\}$ 是这个系统的Hankel奇异值。



平衡截断法(Balanced truncation)

➤ 降阶模型

在平衡变换 $\tilde{\mathbf{u}} = T_{bal}\mathbf{u}$ 下

$$\begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix} = \begin{pmatrix} T_{bal}AT_{bal}^{-1} & T_{bal}B \\ CT_{bal}^{-1} & D \end{pmatrix} = \begin{pmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \tilde{B}_1 \\ \tilde{A}_{21} & \tilde{A}_{22} & \tilde{B}_2 \\ \tilde{C}_1 & \tilde{C}_2 & \tilde{D} \end{pmatrix}$$

下标为1的块对应于最可达和可观测的状态。

下标为2的块对应于最不可达和最不可观测的状态。



平衡截断法(Balanced truncation)

➤ 降阶模型

$$\begin{pmatrix} A_r & B_r \\ C_r & D_r \end{pmatrix} = \begin{pmatrix} \tilde{A}_{11} & \tilde{B}_1 \\ \tilde{C}_1 & \tilde{D} \end{pmatrix}$$

➤ Petrov-Galerkin 基于投影的降阶模型

$$\frac{d\mathbf{q}(t)}{dt} = (W^T V)^{-1} W^T (A V \mathbf{q}(t) + B w(t))$$

$$y(t) \approx C V \mathbf{q}(t) + D w(t)$$

$$W = T_{bal}^*[:, 1:k] \quad V = T_{bal}^{-1}[:, 1:k]$$



平衡截断法(Balanced truncation)

输出函数的能量

给定完全可达可观测的稳定线性系统 $\Sigma = \begin{pmatrix} A & B \\ C & 0 \end{pmatrix}$ ，以及基于平衡截断法的降阶模型 $\Sigma_r = \begin{pmatrix} A_r & B_r \\ C_r & 0 \end{pmatrix}$ ，那么模型 Σ_r 稳定，并且

$$\|\Sigma_r - \Sigma\|_{H_\infty} \leq 2(\sigma_{k+1} + \sigma_{k+2} + \cdots \sigma_m)$$

让系统 Σ 的不同Hankel奇异值记为 σ_i ，每个奇异值的重数为 m_i ，其中 $i = 1, 2, \dots, m$ 。令降阶系统 Σ_r 的奇异值为 σ_i ， $i = 1, 2, \dots, k$ ，奇异值的重数同样是 m_i 。

H_∞ 范数为 $\|\Sigma_r - \Sigma\|_{H_\infty} = \sup_w \frac{\sqrt{\int_0^\infty \|\Delta y(t)\|_2^2 dt}}{\sqrt{\int_0^\infty \|w(t)\|_2^2 dt}}$ ， Δy 为两个模型观测的差异，初始值均为 $\mathbf{u}_0 = 0$ 。