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

$$\frac{d}{dt}\boldsymbol{u}(t) = A\boldsymbol{u}(t) + Bw(t)$$
$$y(t) = C\boldsymbol{u}(t) + Dw(t)$$

➢ 输入函数

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

➢ 输出函数

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

核函数 $h(t,\tau) \in R^{p\times q}$

➢ 假设输入输出映射满足

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

线性动力系统

- 输入输出映射

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

- 系统响应与其输入之间有因果关系（Causality）

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

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

- Laplace变化

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

- 输入输出映射的计算

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

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

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

线性动力系统

➢ 线性模型

$$\frac{d}{dt}\boldsymbol{u}(t) = A\boldsymbol{u}(t) + Bw(t)$$
$$y(t) = C\boldsymbol{u}(t) + Dw(t)$$
$$\boldsymbol{u}(t_0) = \boldsymbol{u}_0$$

➢ 解析解

$$\boldsymbol{u}(t, w; t_0, \boldsymbol{u}_0) = e^{A(t-t_0)}\boldsymbol{u}_0 + \int_{t_0}^{t} e^{A(t-\tau)}B\, w(\tau)d\tau$$

$$y(t) = Ce^{A(t-t_0)}\boldsymbol{u}_0 + C\int_{t_0}^{t} e^{A(t-\tau)}B\, w(\tau)d\tau + Dw(t)$$
$$:= \phi(t, w; t_0, \boldsymbol{u}_0)$$

线性动力系统

- 线性模型

$$\frac{d}{dt}\boldsymbol{u}(t) = A\boldsymbol{u}(t) + Bw(t)$$
$$y(t) = C\boldsymbol{u}(t) + Dw(t)$$
$$\boldsymbol{u}(-\infty) = \boldsymbol{0}$$

- 解析解

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

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

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

线性动力系统

- 线性模型

$$\frac{d}{dt}\boldsymbol{u}(t) = A\boldsymbol{u}(t) + Bw(t)$$
$$y(t) = C\boldsymbol{u}(t) + \mathrm{D}w(t)$$
$$\boldsymbol{u}(-\infty) = \boldsymbol{0}$$

- 输入输出映射

$$H(\xi) = D + C(\xi I - A)^{-1}B$$
$$= h_0 + h_1\xi^{-1} + h_2\xi^{-2} + h_3\xi^{-3} \cdots h_k\xi^{-k} \cdots$$
$$\Sigma = \left(D, CB, CAB, CA^2B \cdots, CA^{k-1}B \cdots\right)$$

坐标变化 $\widetilde{\boldsymbol{u}} = T\boldsymbol{u} \quad (\det(T) \neq 0)$

$$\tilde{\Sigma} = \left(\widetilde{D}, \tilde{C}\tilde{B}, \tilde{C}\tilde{A}\tilde{B}, \tilde{C}\tilde{A}^2\tilde{B} \cdots, \tilde{C}\tilde{A}^{k-1}\tilde{B} \cdots\right) = \Sigma$$

线性动力系统

线性模型

$$\frac{d}{dt}\boldsymbol{u}(t) = A\boldsymbol{u}(t) + Bw(t)$$
$$y(t) = C\boldsymbol{u}(t) + Dw(t)$$

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

Markov 参数：

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

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$\bar{\boldsymbol{u}} \in R^N$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$\bar{\boldsymbol{u}}$

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

定义所有可达状态的集合为X^{reach}，如果$X^{\text{reach}} = R^N$，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

$$R(A,B) = [B\ AB\ A^2B\ \cdots A^kB \cdots],$$

和有限可达Gramians矩阵

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

那么

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

Cayley-Hamilton 定理：

$$R(A,B) = [B, AB, A^2B, \cdots A^{n-1}B]$$

可达状态

可达状态的最小能量

给定四元组$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$，对于可达状态$\bar{\boldsymbol{u}} \in R^N$，给定$\bar{T}$，如果存在一个有限能量的输入函数$\overline{w}$ ，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 $\bar{\boldsymbol{u}}$，那么

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

等号取到的$\overline{w}$满足

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

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$\overline{\boldsymbol{u}} \in R^N$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$\overline{\boldsymbol{u}}$的情况下，系统状态能够变为 0

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

定义所有可控状态的集合为X^{contr}，那么

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

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$\bar{\boldsymbol{u}} \in R^N$是不可观测的，如果以$\bar{\boldsymbol{u}}$为初始状态，采用0输入函数，系统观测永远为0

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

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

可观测状态的判定

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

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

有限可观测Gramians矩阵

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

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

Cayley-Hamilton 定理：

$$\mathcal{O}(A,C) = [C; CA; CA^2; \cdots CA^{n-1}]$$

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$\overline{\boldsymbol{u}} \in R^N$是可观测的，如果以$\overline{\boldsymbol{u}}$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

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

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

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

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，$\begin{pmatrix} A_{ro} & B_{ro} \\ C_{r0} & * \end{pmatrix}$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,$\begin{pmatrix} A_{\bar{r}o} & 0 \\ C_{\bar{r}0} & * \end{pmatrix}$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

$$\mathcal{P}(t) = \int_0^t e^{A\tau}BB^*e^{A^*\tau}d\tau \qquad \mathcal{Q}(t) = \int_0^t e^{A^*\tau}C^*Ce^{A\tau}d\tau$$

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

- 无限时间Gramians矩阵

$$\mathcal{P} = \int_0^\infty e^{A\tau}BB^*e^{A^*\tau}d\tau \qquad \mathcal{Q} = \int_0^\infty e^{A^*\tau}C^*Ce^{A\tau}d\tau$$

满足Lyapunov 方程：

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

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 $\boldsymbol{u}$ 所需的最小能量，其中时间 t 趋向于无穷大时。

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

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

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

平衡截断法(Balanced truncation)

- 降阶模型(近似状态$\boldsymbol{u}$)

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

$$\tilde{\boldsymbol{u}} = T\boldsymbol{u}$$

$$\begin{pmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{pmatrix} = \begin{pmatrix} TAT^{-1} & TB \\ CT^{-1} & D \end{pmatrix} \Longrightarrow \tilde{\mathcal{P}} = \tilde{\mathcal{Q}}$$

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

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

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：

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

其中对角矩阵Σ中的元素按降序排列。

$$T_{bal} = \Sigma^{1/2}K^{*}L^{-1}$$
$$T_{bal}^{-1} = LK\Sigma^{-1/2}$$
$$\tilde{\mathcal{P}} = \tilde{\mathcal{Q}} = \Sigma$$

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

平衡截断法(Balanced truncation)

- 降阶模型

下标为1的块对应于最可达和可观测的状态。

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

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

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

最小能量可达状态：$\operatorname{argmin}\dfrac{\overline{\boldsymbol{u}}^*\tilde{\mathcal{P}}^{-1}\overline{\boldsymbol{u}}}{\overline{\boldsymbol{u}}^*\overline{\boldsymbol{u}}} = \operatorname{argmax}\dfrac{\overline{\boldsymbol{u}}^*\tilde{\mathcal{P}}\overline{\boldsymbol{u}}}{\overline{\boldsymbol{u}}^*\overline{\boldsymbol{u}}}$

最大观测能量状态：$\operatorname{argmax}\dfrac{\overline{\boldsymbol{u}}^*\tilde{\mathcal{Q}}\overline{\boldsymbol{u}}}{\overline{\boldsymbol{u}}^*\overline{\boldsymbol{u}}}$

降阶子空间为：$\tilde{V} = [e_1\ e_2 \cdots e_k]$ $\tilde{\boldsymbol{u}} = \tilde{V}q$

$$\frac{dq(t)}{dt} = \left(\tilde{A}_{11}q(t) + \tilde{B}_1 w(t)\right)$$

$$y(t) = \tilde{C}_1 q(t) + \tilde{D}w(t)$$

平衡截断法(Balanced truncation)

- 降阶模型

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

- Petrov-Galerkin 基于投影的降阶模型

$$\frac{dq(t)}{dt} = (W^T V)^{-1} W^T \big(AVq(t) + Bw(t)\big)$$

$$y(t) \approx CVq(t) + Dw(t)$$

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

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$\Sigma_r$稳定，并且

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

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 $i = 1,2,\cdots m$ 。令降阶系统 Σ_r 的奇异值为 σ_i ，$i = 1,2,\cdots k$ ，奇异值的重数同样是 m_i 。 H_∞范数为$\|\Sigma_r - \Sigma\|_{H_\infty} = \sup\limits_{w} \frac{\sqrt{\int_0^\infty \|\Delta y(t)\|_2^2 dt}}{\sqrt{\int_0^\infty \|w(t)\|_2^2 dt}}$ ，Δy为两个模型观测的差异，初始值均为 $\boldsymbol{u}_0 = 0$ 。

扩展阅读

- 文献

Antoulas, Athanasios C. " Approximation of large-scale dynamical systems, (Chapter 5). " Society for Industrial and Applied Mathematics, 2005.