

随机数据驱动控制：统一理论框架、系统理论知识与非线性扩展

殷明周

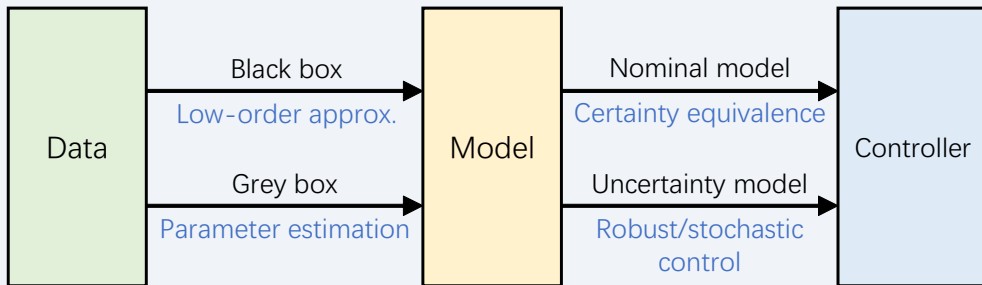
2026 年 9 月 14 日

- 1 数据驱动控制概览
- 2 数据驱动控制的随机数据扩展
- 3 Hammerstein-Wiener 系统的随机数据驱动预测与控制
- 4 建筑能源系统应用

什么是数据驱动控制?

- 广义来说, 几乎所有控制应用都是数据驱动的, 但数据利用受限于可用的**计算能力**和**控制设计工具**
- 经典控制理论不直接利用数据, 而是以**模型**作为中间载体

系统辨识范式



参数化、非参数化与过参数化模型

- 不同控制算法利用数据的能力不同
 - PID 控制：系统特征标量
 - H_∞ 鲁棒控制：标称模型 + 不确定性模型
 - 模型预测控制 (MPC)：任何能获得预测的模型
- 这些模型传统上都是**参数化**的：模型参数维度 $\ll$ 数据集大小
- **系统辨识**是利用系统理论知识，将**高维数据**回归 / 拟合 / 近似到**低维参数空间**的过程
- 另一方面，机器学习得到的输入输出关系也被称作（大）模型
- 这类模型通常是**过参数化**的：模型参数维度 $\gg$ 数据集大小
- **模型训练**结合高复杂度模型与正则化，目标是得到不依赖系统理论知识的出色**泛化能力**
- **非参数化**模型处于两者之间：模型参数维度 $\approx$ 数据集大小
- 跳过系统辨识 / 模型训练步骤，直接用数据作为模型

基于模型、数据驱动与基于学习的控制

控制方法	基于模型的控制	数据驱动控制	基于学习的控制
所用模型	参数化模型	非参数化模型	过参数化模型
模型大小	$n \ll N$	$n \approx N$	$n \gg N$
典型方法	LQR, MPC, H_∞	VRFT, DDPC, data informativity	RL, learning-based MPC
主要问题	模型结构、维度选择	非线性、随机性	过拟合、计算复杂度
领域知识依赖	高	中	低
数据需求	低	中	高
可解释性	高	中	低
安全证明	成熟	发展中	困难

Willems (基本) 引理 & 行为系统理论

LTI 系统的**可能行为**可用**有限数据**表达:

$H = [z_1^d \quad \cdots \quad z_M^d] \sim$ 信号矩阵

$$= \begin{bmatrix} u_{t_1}^d & u_{t_2}^d & \cdots & u_{t_M}^d \\ u_{t_1+1}^d & u_{t_2+1}^d & \cdots & u_{t_M+1}^d \\ \vdots & \vdots & \ddots & \vdots \\ u_{t_1+L-1}^d & u_{t_2+L-1}^d & \cdots & u_{t_M+L-1}^d \\ \hline y_{t_1}^d & y_{t_2}^d & \cdots & y_{t_M}^d \\ y_{t_1+1}^d & y_{t_2+1}^d & \cdots & y_{t_M+1}^d \\ \vdots & \vdots & \ddots & \vdots \\ y_{t_1+L-1}^d & y_{t_2+L-1}^d & \cdots & y_{t_M+L-1}^d \end{bmatrix}$$

每列构成长 L 的信号轨迹

$$= \begin{bmatrix} \mathcal{H}_L(\mathbf{u}^d) \\ \mathcal{H}_L(\mathbf{y}^d) \end{bmatrix}$$

- 轨迹的任意线性组合仍是有效轨迹

$\forall g \in \mathbb{R}^M, Hg$ 仍是有效轨迹

- 如果有足够好的数据……

一条长 L 的轨迹有 $(n_u L + n_x)$ 个自由度,
如果 $\text{rank}(H) = n_u L + n_x$ 覆盖所有自由度

- ……轨迹的线性组合可以覆盖所有 LTI 系统的可能行为

$\forall$ 有效轨迹 $\mathbf{z}, \exists g \in \mathbb{R}^M, \mathbf{z} = Hg$

基于 Willems 引理的数据驱动控制

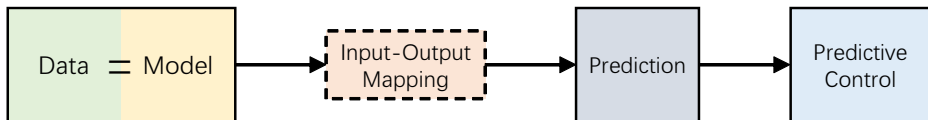
- Jan Willems 提出了充分的**持续激励**条件 [Wil+05], 并建立了**行为系统理论**的数学框架
 - 系统是可控的 & $\mathcal{H}_{L+n_x}(\mathbf{u}^d)$ 是行满秩的
- Ivan Markovsky 意识到 Willems 引理可用于**数据驱动预测** [MR08]
 - 如果我们给定输入轨迹 $\mathbf{u} \in \mathbb{R}^{n_u L'}$ 和足够长的初始条件 $\mathbf{u}_{\text{ini}} \in \mathbb{R}^{n_u L_0}, \mathbf{y}_{\text{ini}} \in \mathbb{R}^{n_y L_0}, n_y L_0 \geq n_x$
 - 输出可以根据 Willems 引理预测:

$$\mathbf{y} = f(\mathbf{u}; \mathbf{u}_{\text{ini}}, \mathbf{y}_{\text{ini}}) : \begin{bmatrix} \mathbf{u}_{\text{ini}} \\ \mathbf{y}_{\text{ini}} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} U_p \\ Y_p \\ U_f \end{bmatrix} g, \quad \mathbf{y} = Y_f g \quad \left| \quad H = \begin{bmatrix} U \\ Y \end{bmatrix} = \begin{bmatrix} U_p \\ U_f \\ Y_p \\ Y_f \end{bmatrix}$$

[MR08] I. Markovsky and P. Rapisarda. "Data-driven simulation and control". In: *International Journal of Control* 81.12 (2008), pp. 1946–1959.

[Wil+05] J. C. Willems, P. Rapisarda, I. Markovsky, and B. L. De Moor. "A note on persistency of excitation". In: *Systems & Control Letters* 54.4 (2005), pp. 325–329.

基于 Willems 引理的数据驱动控制



- Jeremy Coulson 提出直接运用轨迹刻画方程 $\mathbf{z} = Hg$ 进行**数据驱动预测控制** (DDPC / DeePC) [CLD19], 并提出不准确数据条件下的经验正则化项
- 在不准确数据条件下 H 几乎必然行满秩, 导致数据驱动预测器定义不良

$$\begin{aligned}
 & \min_{\mathbf{u}} \quad J_{\text{ctr}}(\mathbf{u}, \mathbf{y}) + \lambda_g \|\mathbf{g}\|_1 + \lambda_y \|\sigma_y\|_1 \\
 & \text{s.t.} \quad \begin{bmatrix} \mathbf{u}_{\text{ini}} \\ \mathbf{u} \\ \mathbf{y}_{\text{ini}} + \sigma_y \end{bmatrix} = \begin{bmatrix} U_p \\ U_f \\ Y_p \end{bmatrix} g, \quad \mathbf{y} = Y_f g, \quad \mathbf{u} \in \mathcal{U}, \quad \mathbf{y} \in \mathcal{Y}.
 \end{aligned}$$

[CLD19]

J. Coulson, J. Lygeros, and F. Dörfler. "Data-enabled predictive control: In the shallows of the DeePC". In: *European Control Conference (ECC)*. 2019, pp. 307–312.

[HTML] A note on persistency of excitation

[JC Willems](#), [P Rapisarda](#), I Markovsky... - Systems & Control ..., 2005 - Elsevier

... Crucial in our result is the **persistency** of **excitation** of a component (typically, the input component) of w . The signal $f = [1, T] \cap \mathbb{N} \rightarrow \mathbb{R}^f$ is said to be persistently exciting of order $L : \Leftrightarrow \dots$

☆ Save  Cite Cited by 1775 Related articles All 19 versions

Data-enabled **predictive control**: In the shallows of the DeePC

[J Coulson](#), [J Lygeros](#), [F Dörfler](#) - 2019 18th European **control** ..., 2019 - ieeexplore.ieee.org

... MPC based on Dynamic Matrix **Control** has been historically used as a **data-driven control** ... non-parametric **predictive control** model. Other non-parametric **predictive** models have been ...

☆ Save  Cite Cited by 1157 Related articles All 7 versions

Formulas for data-driven control: Stabilization, optimality, and robustness

[C De Persis](#), P Tesi - IEEE Transactions on Automatic Control, 2019 - ieeexplore.ieee.org

In a paper by Willems et al., it was shown that persistently exciting data can be used to represent the input-output behavior of a linear system. Based on this fundamental result, we ...

☆ Save  Cite Cited by 1462 Related articles All 10 versions

[PDF] Data-driven model predictive control with stability and robustness guarantees

[J Berberich](#), [J Köhler](#), [MA Müller](#)... - IEEE transactions on ..., 2020 - ieeexplore.ieee.org

We propose a robust data-driven model predictive control (MPC) scheme to control linear time-invariant systems. The scheme uses an implicit model description based on behavioral ...

☆ Save  Cite Cited by 1138 Related articles All 8 versions

研究现状

数据驱动预测控制

- 把各类 MPC 算法迁移到数据驱动框架里

$$\begin{aligned}
 \min_{\mathbf{u}} \quad & \sum_{k=0}^{L-2n-1} \|u_k - u^s\|_R^2 + \|y_k - y^s\|_Q^2 + \lambda_g \|g\|_2^2 + \lambda_y \|\sigma_y\|_2^2 \quad [\text{Ber+20}] \\
 \text{s.t.} \quad & \begin{bmatrix} \mathbf{u}_{\text{ini}} \\ \mathbf{u} \\ \mathbf{y}_{\text{ini}} \\ \mathbf{y} \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma_y \end{bmatrix} = \begin{bmatrix} U_p \\ U_f \\ Y_p \\ Y_f \end{bmatrix} g, \\
 & \|y_k\|_\infty + a_k^1 \|\mathbf{u}\|_1 + a_k^2 \|g\|_1 + a_k^3 \|\sigma_y\|_\infty + a_k^4 \leq y_{\max}, \quad k = 0, \dots, L - 2n - 1, \\
 & \begin{bmatrix} u_k \\ y_k \end{bmatrix} = \begin{bmatrix} u^s \\ y^s \end{bmatrix}, \quad k = L - 2n, \dots, L - n - 1
 \end{aligned}$$

- 主要难点：固有多步预测、预测误差模型

[Ber+20]

J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer. "Data-driven model predictive control with stability and robustness guarantees". In: *IEEE Transactions on Automatic Control* 66.4 (2020), pp. 1702–1717.

研究现状

数据信息性理论

- 基于**状态空间矩阵**控制设计的**数据矩阵**等价替代
- 持续激励条件不满足时数据驱动控制的能力与局限性

$$U = [u_0 \ u_1 \ \cdots \ u_{N-1}], \quad X_- = [x_0 \ x_1 \ \cdots \ x_{N-1}], \quad X_+ = [x_1 \ x_2 \ \cdots \ x_N]$$

定理：基于 LMI 的数据驱动镇定 [DT19]

数据集 (U, X) 对状态反馈镇定具有信息性，当且仅当存在矩阵 $\Theta \in \mathbb{R}^{N \times n}$ ，使得 $X_- \Theta = (X_- \Theta)^\top$ ，

$$\begin{bmatrix} X_- \Theta & X_+ \Theta \\ \Theta^\top X_+^\top & X_- \Theta \end{bmatrix} \succ 0.$$

当且仅当 $K = U_- \Theta (X_- \Theta)^{-1}$ ，状态反馈控制律 $u = Kx$ 能够镇定该系统。

- 类似结果包括系统性质分析、LQR、鲁棒控制、输出反馈、跟踪与调节等。

[DT19]

C. De Persis and P. Tesi. "Formulas for data-driven control: stabilization, optimality, and robustness". In: *IEEE Transactions on Automatic Control* 65.3 (2019), pp. 909–924.

非线性拓展

- 用**非线性提升**或**核化表示**代替原始输入输出信号

定理： Hammerstein 系统的 Willems 引理 [BA20]

考虑 Hammerstein 系统 $y = G(z)f(u)$ 。如果输入非线性可以分解为有限维基函数： $f(u) = a^\top \psi(u)$ ，定义 $\bar{u} = \psi(u)$ 。当系统可控，且 $\bar{u}$ 满足 $(L + n_x)$ 阶持续激励条件时，Willems 引理对提升系统 $\bar{u} \rightarrow y$ 仍成立。

- 类似方法可被用于**双线性系统**、**反馈线性化系统**、**有限维 Koopman 算子**等情况
- **主要难点**：持续激励条件满足、近似误差分析

[BA20]

J. Berberich and F. Allgöwer. "A trajectory-based framework for data-driven system analysis and control". In: *European Control Conference (ECC)*. 2020, pp. 1365–1370.

- 1 数据驱动控制概览
- 2 数据驱动控制的随机数据扩展
- 3 Hammerstein-Wiener 系统的随机数据驱动预测与控制
- 4 建筑能源系统应用

数据驱动控制的随机数据扩展

核心问题：当数据具有不确定性时， H 几乎必然行满秩，任何轨迹都可以表示为已有轨迹的线性叠加，确定性行为系统理论**定义不良**

已有多种方法将**确定性**数据驱动表述推广至**随机数据**

- **恢复低秩结构：**子空间辨识 (Van Overschee & De Moor, 2012)，投影 (Breschi et al., 2023)，低秩矩阵近似 (Markovsky & Dörfler, 2022)
- **针对特定任务设计的方法：**分布鲁棒优化 (Coulson et al., 2022)，多项式混沌展开 (Pan et al., 2023)，ARX 核方法 (Chiuso et al., 2025)，最大似然估计 (Yin et al., 2023)

新贡献 [Yin+26]:

- 构建**统一贝叶斯框架**，通过**最大后验 (MAP) 估计**将行为系统理论推广至随机数据
- 适用于各种数据驱动任务：平滑、预测 & 控制，支持输入误差 & 相关误差
- 建立通用的估计误差分析方法

[Yin+26] M. Yin, A. Iannelli, S. A. Nazari, and M. A. Müller. "A unified Bayesian framework for data-driven smoothing, prediction, and control". In: *arXiv preprint arXiv:2512.01475* (2026).

轨迹估计问题

- 具有输入扰动 & 输出噪声的 LTI 系统: $y_t^0 = Gu_t^0$, $\underbrace{\begin{bmatrix} u_t \\ y_t \end{bmatrix}}_{z_t \in \mathbb{R}^n} = \underbrace{\begin{bmatrix} u_t^0 \\ y_t^0 \end{bmatrix}}_{z_t^0} + \underbrace{\begin{bmatrix} w_t \\ v_t \end{bmatrix}}_{\nu_t}$
- 利用离线数据代替模型 G 刻画系统:

长度为 L 的信号: $\mathbf{z}_i^d := \begin{bmatrix} z_{t_i+1}^d \\ \vdots \\ z_{t_i+L}^d \end{bmatrix}$, 信号矩阵: $H := [\mathbf{z}_1^d \quad \mathbf{z}_2^d \quad \cdots \quad \mathbf{z}_M^d] = H^0 + \Delta_H$

- 考虑估计长度为 L 的轨迹 $\mathbf{z}^0 := \text{col}(z_1^0, z_2^0, \dots, z_L^0)$ 的问题, 所用信息为 (部分) 线性观测 ζ & 信号矩阵 H :

$$\hat{\mathbf{z}} = \hat{\mathbf{z}}(\zeta, H), \quad \zeta = \Phi \mathbf{z}^0 + \underbrace{\epsilon}_{\text{观测误差}}, \quad \mathbf{z}^0 = H^0 \mathbf{g} = H \mathbf{g} - \underbrace{\Delta_H \mathbf{g}}_{\text{离线数据不确定性}}$$

- 正确性 (可辨识性) 由以下条件保证: $\text{rank}(H^0) = \text{rank}(\Phi H^0) = n_u L + n_x$ (Willems 引理)

数据驱动任务示例

- 平滑：整条轨迹已知

$$\zeta = \text{col}(z_1, z_2, \dots, z_L), \quad \epsilon = \text{col}(\nu_1, \nu_2, \dots, \nu_L)$$

- 预测：所有输入 & 前 L_0 个输出已知 ($L = L_0 + L'$)

$$\zeta = \text{col}(z_1, \dots, z_{L_0}, u_{L_0+1}, \dots, u_L), \quad \epsilon = \text{col}(\nu_1, \dots, \nu_{L_0}, w_{L_0+1}, \dots, w_L)$$

- 控制：前 L_0 个输入 & 输出已知，后 L' 个输入 & 输出以以下二次型控制代价为目标：

$$\sum_{t=L_0+1}^L \left(\|u_t^0 - u_t^{\text{ref}}\|_R^2 + \|y_t^0 - y_t^{\text{ref}}\|_Q^2 \right)$$

$$\zeta = \text{col}(z_1, \dots, z_{L_0}, \zeta_{L_0+1}^{\text{ctr}}, \dots, \zeta_L^{\text{ctr}}), \quad \epsilon = \text{col}(\nu_1, \dots, \nu_{L_0}, \epsilon^{\text{ctr}}),$$

其中 $\zeta_t^{\text{ctr}} = \text{col}(u_t^{\text{ref}}, y_t^{\text{ref}})$, ϵ^{ctr} : 跟踪误差。

不确定性描述

- 考虑零均值、平稳、高斯输入扰动 ϵ_t 输出噪声：

$$\begin{bmatrix} \nu_t^d \\ \nu_{t+\tau}^d \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \Sigma_\nu^d(0) & \Sigma_\nu^d(\tau) \\ \Sigma_\nu^d(\tau) & \Sigma_\nu^d(0) \end{bmatrix} \right), \quad \begin{bmatrix} \nu_t \\ \nu_{t+\tau} \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \Sigma_\nu(0) & \Sigma_\nu(\tau) \\ \Sigma_\nu(\tau) & \Sigma_\nu(0) \end{bmatrix} \right)$$

- 参考跟踪引入的不确定性由下式表示： $\Sigma^{\text{ctr}} = \text{blkdiag} (R^{-1}, Q^{-1})$
- 刻画在线观测 ϵ 离线数据中的不确定性：

$$\zeta = \Phi \mathbf{z}^0 + \epsilon, \quad \mathbf{z}^0 = H^0 \mathbf{g} = H \mathbf{g} - \Delta_H \mathbf{g}$$

- 在线观测：** $\epsilon \sim \mathcal{N}(\mathbf{0}_p, \Sigma_\epsilon)$. 令 $(\Sigma_z)_{i,j} = \Sigma_\nu(|i-j|)$, $\Gamma = \begin{bmatrix} \mathbf{0} & \mathbb{I} \end{bmatrix}$.

$$\text{平滑: } \Sigma_\epsilon = \Sigma_z, \quad \text{预测: } \Sigma_\epsilon = \Phi \Sigma_z \Phi^\top, \quad \text{控制: } \Sigma_\epsilon = \Sigma_z + \Gamma^\top (\mathbb{I}_{L'} \otimes \Sigma^{\text{ctr}}) \Gamma$$

- 离线数据：** $\Delta_H \mathbf{g} | \mathbf{g} \sim \mathcal{N}(\mathbf{0}_{nL}, \Sigma_g(\mathbf{g}))$

$$(\Sigma_g(\mathbf{g}))_{i,j} = \sum_{k=1}^M \sum_{m=1}^M g_k g_m \Sigma_\nu^d(|t_k + i - (t_m + j)|)$$

经验贝叶斯方法

- 由此得到如下贝叶斯模型：

$$\zeta | \mathbf{z}^0 \sim \mathcal{N}(\Phi \mathbf{z}^0, \Sigma_\epsilon), \quad \mathbf{z}^0 | \mathbf{g} \sim \mathcal{N}(H \mathbf{g}, \Sigma_g(\mathbf{g}))$$

- 通常采用经验贝叶斯方法求解，即：1) 将 $\mathbf{g}$ 作为超参数，通过最大边缘似然（MML）进行估计；2) 以估计值 $\hat{\mathbf{g}}$ 为条件，求最大后验（MAP）估计 / 后验均值。
- MML 步骤：** $\zeta = \Phi H \mathbf{g} - \Phi \Delta_H \mathbf{g} + \epsilon \rightarrow \zeta | \mathbf{g} \sim \mathcal{N}(\Phi H \mathbf{g}, \Phi \Sigma_g(\mathbf{g}) \Phi^\top + \Sigma_\epsilon)$

$$\hat{\mathbf{g}} = \operatorname{argmax}_{\mathbf{g}} p(\zeta | \mathbf{g}) = \operatorname{argmin}_{\mathbf{g}} \log \det(\Psi(\mathbf{g})) + \|\zeta - \Phi H \mathbf{g}\|_{\Psi^{-1}(\mathbf{g})}^2, \quad \Psi(\mathbf{g}) := \Phi \Sigma_g(\mathbf{g}) \Phi^\top + \Sigma_\epsilon$$

- MAP 步骤：** $p(\mathbf{z}^0 | \zeta, \hat{\mathbf{g}}) \propto p(\zeta | \mathbf{z}^0) p(\mathbf{z}^0 | \hat{\mathbf{g}})$

$$\hat{\mathbf{z}} = \operatorname{argmax}_{\mathbf{z}^0} p(\zeta | \mathbf{z}^0) p(\mathbf{z}^0 | \hat{\mathbf{g}}) = \operatorname{argmax}_{\mathbf{z}^0} \|\zeta - \Phi \mathbf{z}^0\|_{\Sigma_\epsilon^{-1}}^2 + \|\mathbf{z}^0 - H \hat{\mathbf{g}}\|_{\Sigma_g^{-1}(\hat{\mathbf{g}})}^2$$

估计值与后验均值一致：

$$\begin{bmatrix} \mathbf{z}^0 \\ \zeta \end{bmatrix} \Big| \hat{\mathbf{g}} \sim \mathcal{N} \left(\begin{bmatrix} H \hat{\mathbf{g}} \\ \Phi H \hat{\mathbf{g}} \end{bmatrix}, \begin{bmatrix} \Sigma_g(\hat{\mathbf{g}}) & \Sigma_g(\hat{\mathbf{g}}) \Phi^\top \\ \Phi \Sigma_g(\hat{\mathbf{g}}) & \Psi(\hat{\mathbf{g}}) \end{bmatrix} \right), \quad \mathbf{z}^0 | \zeta, \hat{\mathbf{g}} \sim \mathcal{N}(\hat{\mathbf{z}}, \Omega)$$

非凸优化问题求解

- 非凸 MML 步骤求解：半正定松弛 + 上界最小化 (MM) 算法

$$\hat{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} \underbrace{\log \det (\Psi(\mathbf{g}))}_{\text{concave} \circ \text{quadratic}} + \underbrace{\|\zeta - \Phi H \mathbf{g}\|_{\Psi^{-1}(\mathbf{g})}^2}_{\text{convex} \circ \text{quadratic}}$$

$$G = \mathbf{g} \mathbf{g}^\top, \mathcal{A}(G) = \Psi(\mathbf{g}) = \Phi \left(\sum_{k=1}^M \sum_{m=1}^M (G)_{k,m} B_{km} \right) \Phi^\top + \Sigma_\epsilon, (B_{km})_{i,j} = \Sigma_\nu^d(|t_k + i - (t_m + j)|)$$

- 半正定松弛：

$$(\hat{\mathbf{g}}, \hat{G}) = \underset{\mathbf{g}, G}{\operatorname{argmin}} \log \det (\mathcal{A}(G)) + \|\zeta - \Phi H \mathbf{g}\|_{\mathcal{A}^{-1}(G)}^2 + \eta(\operatorname{tr}(G) - \|\mathbf{g}\|_2^2) \quad \text{s.t.} \quad \begin{bmatrix} G & \mathbf{g} \\ \mathbf{g}^\top & 1 \end{bmatrix} \succeq 0$$

- 上界最小化算法：

$$(\hat{\mathbf{g}}^{i+1}, \hat{G}^{i+1}) = \underset{\mathbf{g}, G}{\operatorname{argmin}} \operatorname{tr}(\mathcal{A}^{-1}(G^i) \mathcal{A}(G)) + t + \eta \operatorname{tr}(G) - 2\eta(\mathbf{g}^i)^\top \mathbf{g}$$

$$\text{s.t.} \quad \begin{bmatrix} \mathcal{A}(G) & \zeta - \Phi H \mathbf{g} \\ (\zeta - \Phi H \mathbf{g})^\top & t \end{bmatrix} \succeq 0, \quad \begin{bmatrix} G & \mathbf{g} \\ \mathbf{g}^\top & 1 \end{bmatrix} \succeq 0$$

误差量化

- 后验分布 $\mathbf{z}^0 | \zeta, \hat{\mathbf{g}} \sim \mathcal{N}(\hat{\mathbf{z}}, \Omega)$ 通过 Ω 量化轨迹估计误差，但该后验分布以 $\hat{\mathbf{g}}$ 为条件，忽略了 $\hat{\mathbf{g}}$ 的误差
- 解决方案：通过层次贝叶斯方法联合估计 $\mathbf{z}^0$ 和 $\mathbf{g}$

$$\zeta | \mathbf{z}^0 \sim \mathcal{N}(\Phi \mathbf{z}^0, \Sigma_\epsilon), \quad \mathbf{z}^0 | \mathbf{g} \sim \mathcal{N}(H\mathbf{g}, \Sigma_g(\mathbf{g})), \quad \mathbf{g} \sim \mathcal{N}(\mathbf{0}, \lambda \mathbb{I})$$

- 联合 MAP 估计： $(\hat{\mathbf{z}}, \hat{\mathbf{g}}) = \underset{\mathbf{z}^0, \mathbf{g}}{\operatorname{argmax}} p(\zeta | \mathbf{z}^0) p(\mathbf{z}^0 | \mathbf{g}) p(\mathbf{g})$
- 半正定松弛 + 上界最小化 (MM) 算法： $\mathcal{B}(G) = \Sigma_g(\mathbf{g}) = \sum_{k=1}^M \sum_{m=1}^M (G)_{k,m} B_{km}$

$$(\hat{\mathbf{z}}^{i+1}, \hat{\mathbf{g}}^{i+1}, \hat{G}^{i+1}) = \underset{\mathbf{z}^0, \mathbf{g}, G}{\operatorname{argmin}} \operatorname{tr}(\mathcal{B}^{-1}(G^i) \mathcal{B}(G)) + \|\zeta - \Phi \mathbf{z}^0\|_{\Sigma_\epsilon^{-1}}^2 + t + \lambda^{-1} \|g\|_2^2 + \eta \operatorname{tr}(G) - 2\eta(\mathbf{g}^i)^\top \mathbf{g}$$

$$\text{s.t.} \quad \begin{bmatrix} \mathcal{B}(G) & \mathbf{z}^0 - H\mathbf{g} \\ (\mathbf{z}^0 - H\mathbf{g})^\top & t \end{bmatrix} \succeq 0, \quad \begin{bmatrix} G & \mathbf{g} \\ \mathbf{g}^\top & 1 \end{bmatrix} \succeq 0$$

- 得到 $(\hat{\mathbf{z}}, \hat{\mathbf{g}})$ 后，通过 Laplace 近似得到同时考虑 $\hat{\mathbf{z}}$ 和 $\hat{\mathbf{g}}$ 不确定性的后验方差估计。

后续工作

- **滚动时域实现**: 将时刻 t 的 $\hat{\mathbf{z}}, \hat{\Sigma}_z$ 移位后, 用于初始化时刻 $(t+1)$ 的问题。
 - 在预测 & 控制中, 需要增加一次滤波, 以最优地融合最新的输出测量值与先前的估计结果。
- **控制中的探索**: 在控制任务中, 借鉴贝叶斯优化的思想, 可以在 $\beta \hat{\Sigma}_z$ 置信区域内搜索最优的可能轨迹并进行探索, 而非直接施加 $\hat{\mathbf{z}}$ 中的后验均值输入。

$$\begin{aligned} \hat{\mathbf{z}}_{\text{explore}} = \underset{\mathbf{z}}{\operatorname{argmin}} \quad & \sum_{t=L_0+1}^L \left(\|u_t - u_t^{\text{ref}}\|_R^2 + \|y_t - y_t^{\text{ref}}\|_Q^2 \right) \\ \text{s.t.} \quad & \|\mathbf{z} - \hat{\mathbf{z}}\|_{\hat{\Sigma}_z^{-1}}^2 \leq \beta \end{aligned}$$

- 1 数据驱动控制概览
- 2 数据驱动控制的随机数据扩展
- 3 Hammerstein-Wiener 系统的随机数据驱动预测与控制
- 4 建筑能源系统应用

多步数据驱动预测器：两种极端情形

- **问题：**用数据 $(\mathbf{u}^d, \mathbf{y}^d)$ ，构造多步数据驱动预测器 $\mathbf{y} = f(\tilde{\mathbf{u}}, \mathbf{y}_{\text{ini}})$, $\tilde{\mathbf{u}} := \begin{bmatrix} \mathbf{u}_{\text{ini}} \\ \mathbf{u} \end{bmatrix}$
- **线性系统 + 无噪声数据：Willems 引理**

$$\text{存在 } g, \text{ 使得 } \mathbf{y} \text{ 满足 } \begin{bmatrix} U \\ Y_p \\ Y_f \end{bmatrix} g = \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{y}_{\text{ini}} \\ \mathbf{y} \end{bmatrix}, \quad \text{i.e., } \mathbf{y} = [\Gamma_1 \ \Gamma_2] \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{y}_{\text{ini}} \end{bmatrix}, \quad [\Gamma_1 \ \Gamma_2] := Y_f \begin{bmatrix} U \\ Y_p \end{bmatrix}^\dagger$$

- **纯黑箱模型：高斯过程**

$$\eta := \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{y}_{\text{ini}} \end{bmatrix}, \quad f(\eta) \sim \mathcal{GP}(\mathbf{m}(\cdot), \mathbf{k}(\cdot, \cdot))$$

训练输入: $\eta_k^d = \text{col}(u_k^d, \dots, u_{k+L-1}^d, y_k^d, \dots, y_{k+L_0-1}^d)$

输出: $\chi_k^d = \text{col}(y_{k+L_0}^d, \dots, y_{k+L-1}^d)$

介于两者之间……

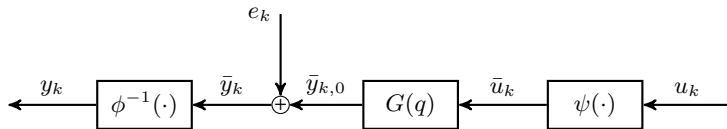


Figure: Hammerstein-Wiener 系统

- 能否直接推广 Willems 引理？
- 对于 Hammerstein 系统，采用基函数分解： $\psi(u) = \sum_{i=1}^r a_i \psi_i(u)$
- 对增广输入 $\psi(u)$ ，通过 Willems 引理进行线性预测
- 该方法不适用于 Wiener 部分，因为 $\phi(y) = \sum_{i=1}^q b_i \phi_i(y)$

$$\begin{cases} x_{k+1} &= Ax_k + \bar{B}\psi(u_k), \\ \mathbf{b}^\top \phi(y_k) &= Cx_k + \bar{D}\psi(u_k), \end{cases}$$

$\mathbf{b}^\top$ 和 $\phi(\cdot)$ 均不可逆

另一种思路 [YM26]

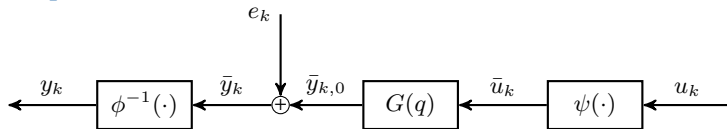


Figure: Hammerstein-Wiener 系统

- 将显式预测器结构推广至 H-W 系统:

$$\mathbf{y} = [\Gamma_1 \ \Gamma_2] \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{y}_{\text{ini}} \end{bmatrix} + [-\Gamma_2 \ \mathbb{I}] \mathbf{e} \quad \Rightarrow \quad \mathbf{0} = [\Gamma_1 \ \Gamma_2 \ -\mathbb{I}] \begin{bmatrix} \Psi(\tilde{\mathbf{u}}) \\ \Phi(\mathbf{y}_{\text{ini}}) \\ \Phi(\mathbf{y}) \end{bmatrix} + [-\Gamma_2 \ \mathbb{I}] \mathbf{e}$$

$$\eta := \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{y}_{\text{ini}} \\ \mathbf{y} \end{bmatrix}, \quad f(\eta) = [\Gamma_1 \ \Gamma_2 \ -\mathbb{I}] \begin{bmatrix} \Psi(\tilde{\mathbf{u}}) \\ \Phi(\mathbf{y}_{\text{ini}}) \\ \Phi(\mathbf{y}) \end{bmatrix}, \quad \begin{cases} \psi(\cdot) \sim \mathcal{GP}(m_u(\cdot), k_u(\cdot, \cdot; \theta_u)) \\ \phi(\cdot) \sim \mathcal{GP}(m_y(\cdot), k_y(\cdot, \cdot; \theta_y)) \end{cases}$$

- 训练 $f(\eta)$ 的隐式高斯过程模型, 输入: $\eta_k^d = \text{col}(u_k^d, \dots, u_{k+L-1}^d, y_k^d, \dots, y_{k+L-1}^d)$, 输出: $\chi_k^d = \mathbf{0}$, 噪声协方差: $\Sigma = \text{cov}([- \Gamma_2 \ \mathbb{I}] \mathbf{e}) = \sigma^2 (\Gamma_2 \Gamma_2^\top + \mathbb{I})$, 假设 $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbb{I})$

[YM26]

M. Yin and M. A. Müller. "Data-driven prediction and control of Hammerstein-Wiener systems with implicit Gaussian processes". In: *arXiv preprint arXiv:2501.15849* (2026).

高斯过程回归

- 学习如下未知函数

$$f : \mathbb{R}^{n_\eta} \rightarrow \mathbb{R}^{n_\chi} \sim \mathcal{GP}(m(\cdot), k(\cdot, \cdot; \Theta))$$

所用数据为

$$(\boldsymbol{\eta}^d, \boldsymbol{\chi}^d) = (\boldsymbol{\eta}_k^d, \boldsymbol{\chi}_k^d)_{k=1}^M, \quad \boldsymbol{\chi}_k^d = f(\boldsymbol{\eta}_k^d) + \boldsymbol{\epsilon}_k, \quad \boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

- 联合分布:

$$\begin{bmatrix} f(\boldsymbol{\eta}) \\ \boldsymbol{\chi}^d \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(\boldsymbol{\eta}) \\ \mathbf{m}(\boldsymbol{\eta}^d) \end{bmatrix}, \begin{bmatrix} k(\boldsymbol{\eta}, \boldsymbol{\eta}) & \mathbf{k}^\top(\boldsymbol{\eta}^d, \boldsymbol{\eta}) \\ \mathbf{k}(\boldsymbol{\eta}^d, \boldsymbol{\eta}) & K(\boldsymbol{\eta}^d, \boldsymbol{\eta}^d) + \bar{\Sigma} \end{bmatrix} \right), \quad \bar{\Sigma} := \mathbb{I} \otimes \Sigma$$

- 后验分布:

$$f(\boldsymbol{\eta}) | \boldsymbol{\chi}^d \sim \mathcal{N} \left(\underbrace{m_p(\boldsymbol{\eta})}_{\text{名义预测}}, \underbrace{k_p(\boldsymbol{\eta})}_{\text{预测误差}} \right)$$

$$m_p(\boldsymbol{\eta}) := m(\boldsymbol{\eta}) + \mathbf{k}^\top(\boldsymbol{\eta}^d, \boldsymbol{\eta}) (K + \bar{\Sigma})^{-1} (\boldsymbol{\chi}^d - \mathbf{m}(\boldsymbol{\eta}^d))$$

$$k_p(\boldsymbol{\eta}) := k(\boldsymbol{\eta}, \boldsymbol{\eta}) - \mathbf{k}^\top(\boldsymbol{\eta}^d, \boldsymbol{\eta}) (K + \bar{\Sigma})^{-1} \mathbf{k}(\boldsymbol{\eta}^d, \boldsymbol{\eta})$$

与纯黑箱模型比较

纯黑箱模型

$$\eta = \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{y}_{\text{ini}} \end{bmatrix}, \quad \chi = \mathbf{y}$$

$$f(\eta) \sim \mathcal{GP}(\mathbf{m}(\cdot), \mathbf{k}(\cdot, \cdot))$$

- 纯黑箱模型学习整个映射

$$f: \mathbb{R}^{n_u L + n_y L_0} \rightarrow \mathbb{R}^{n_y L'}$$

将其视为由高斯过程建模的一般非线性函数

- 结构化灰箱模型仅学习静态非线性映射

$$\psi: \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_u}, \quad \phi: \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_y}$$

将其视为由高斯过程建模的一般非线性函数

- 线性结构由以下线性算子保持: $[\Gamma_1 \ \Gamma_2 \ -\mathbb{I}]$
- 结构化灰箱模型学习的是隐函数
 $\Rightarrow$ 无法直接得到显式预测

结构化灰箱模型

$$\eta = \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{y}_{\text{ini}} \\ \mathbf{y} \end{bmatrix}, \quad \chi = \mathbf{0}$$

$$f(\eta) = [\Gamma_1 \ \Gamma_2 \ -\mathbb{I}] \begin{bmatrix} \Psi(\tilde{\mathbf{u}}) \\ \Phi(\mathbf{y}_{\text{ini}}) \\ \Phi(\mathbf{y}) \end{bmatrix}$$

$$\begin{cases} \psi(\cdot) \sim \mathcal{GP}(m_u(\cdot), k_u(\cdot, \cdot; \theta_u)) \\ \phi(\cdot) \sim \mathcal{GP}(m_y(\cdot), k_y(\cdot, \cdot; \theta_y)) \end{cases}$$

隐式高斯过程的解释

- 由后验分布可得:

$$[\Gamma_1 \ \Gamma_2] \begin{bmatrix} \Psi(\tilde{\mathbf{u}}) \\ \Phi(\mathbf{y}_{\text{ini}}) \end{bmatrix} - \Phi(\mathbf{y}) \sim \mathcal{N}(m_p(\eta), k_p(\eta)) \quad (\text{认知误差})$$

- 由模型方程可得:

$$[\Gamma_1 \ \Gamma_2] \begin{bmatrix} \Psi(\tilde{\mathbf{u}}) \\ \Phi(\mathbf{y}_{\text{ini}}) \end{bmatrix} - \Phi(\mathbf{y}_0) = [\Gamma_2 \ \mathbf{0}] \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Gamma_2 \Gamma_2^\top), \quad \mathbf{y}_0 : \text{无噪声输出} \quad (\text{偶然误差})$$

- 任意预测均具有如下 **投影预测误差**=(认知误差)+(偶然误差)

$$\Phi(\mathbf{y}_0) - \Phi(\mathbf{y}) \sim \mathcal{N}(m_p(\eta), \hat{k}_p(\eta)), \quad \hat{k}_p(\eta) = k_p(\eta) + \sigma^2 \Gamma_2 \Gamma_2^\top$$

- 最优预测:**

$$\mathbf{y} = \underset{\mathbf{y}}{\operatorname{argmin}} \operatorname{tr}(\hat{k}_p(\eta)) + m_p^\top(\eta) m_p(\eta) \quad (\text{Min-MSE})$$

系统理论知识的运用：稳定线性模型参数

- 除常规超参数 θ_u 、 θ_y 、 σ^2 外，线性模型参数 Γ_1 、 Γ_2 也成为超参数
- MML 方法：

$$\begin{aligned}\Theta &= \operatorname{argmax}_{\Theta} p(\chi^d | \eta^d, \Theta) & \Theta &:= \{\theta_u, \theta_y, \sigma^2, \Gamma_1, \Gamma_2\} \\ &= \operatorname{argmin}_{\Theta} \log \det(\Lambda(\Theta)) + \mathbf{m}^\top(\eta^d) \Lambda^{-1}(\Theta) \mathbf{m}(\eta^d), & \Lambda(\Theta) &:= K(\eta^d, \eta^d; \Theta) + \bar{\Sigma}\end{aligned}$$

容易发生拟合，原因是超参数 Θ 的维数过高： $[\Gamma_1 \ \Gamma_2] \in \mathbb{R}^{n_y L' \times (n_u L + n_y L_0)}$

- 解决方案：** 将 Γ_1 、 Γ_2 视为具有超先验分布的随机变量 $\Rightarrow$ 联合最大后验 / 最大似然问题：

$$\Theta = \operatorname{argmin}_{\Theta} \log \det(\Lambda(\Theta)) + \mathbf{m}^\top(\eta^d) \Lambda^{-1}(\Theta) \mathbf{m}(\eta^d) + \log p(\Gamma | \zeta), \quad \zeta : \text{超先验的参数 (‘超超参数’)}$$

- 如何设计稳定线性模型参数的先验分布 $p(\Gamma | \zeta)$ 是系统辨识领域已解决的问题 [Pil+22]

[Pil+22] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, L. Ljung, et al. *Regularized system identification: learning dynamic models from data*. Springer, 2022.

从预测到预测控制

- 采用如下期望控制代价的滚动时域控制：

$$J_{\text{ctr}} := \|\mathbf{u}^k\|_R^2 + \mathbb{E} \left[\|\Phi(\mathbf{y}_0^k) - \Phi(\mathbf{r}^k)\|_Q^2 \right], \mathbf{r}^k := \begin{bmatrix} r_k \\ \vdots \\ r_{k+L'-1} \end{bmatrix}, \mathbf{u}^k := \begin{bmatrix} u_k \\ \vdots \\ u_{k+L'-1} \end{bmatrix}, \mathbf{y}_0^k := \begin{bmatrix} y_{k,0} \\ \vdots \\ y_{k+L'-1,0} \end{bmatrix}$$

- $\Phi(\mathbf{y}_0^k)$ 的分布由预测误差量化结果给出

$$\Phi(\mathbf{y}_0^k) \sim \mathcal{N} \left(\Phi(\mathbf{y}^k) + m_p(\eta^k), \hat{k}_p(\eta^k) \right),$$

- 期望控制代价变为：

$$J_{\text{ctr}} = \|\mathbf{u}^k\|_R^2 + \|\Phi(\mathbf{y}^k) + m_p(\eta^k) - \Phi(\mathbf{r}^k)\|_Q^2 + \text{tr} \left(Q \hat{k}_p(\eta^k) \right)$$

- 未知非线性函数值 $\Phi(\mathbf{y}^k)$ 、 $\Phi(\mathbf{r}^k)$ 分别用其后验估计 $\mathbf{m}_{yp}(\mathbf{y}^k)$ 、 $\mathbf{m}_{yp}(\mathbf{r}^k)$ 替代

机会约束的满足

引理：机会约束的满足

假设输出非线性映射 $\phi^{-1}(\cdot)$ 满足 Lipschitz 连续性，即

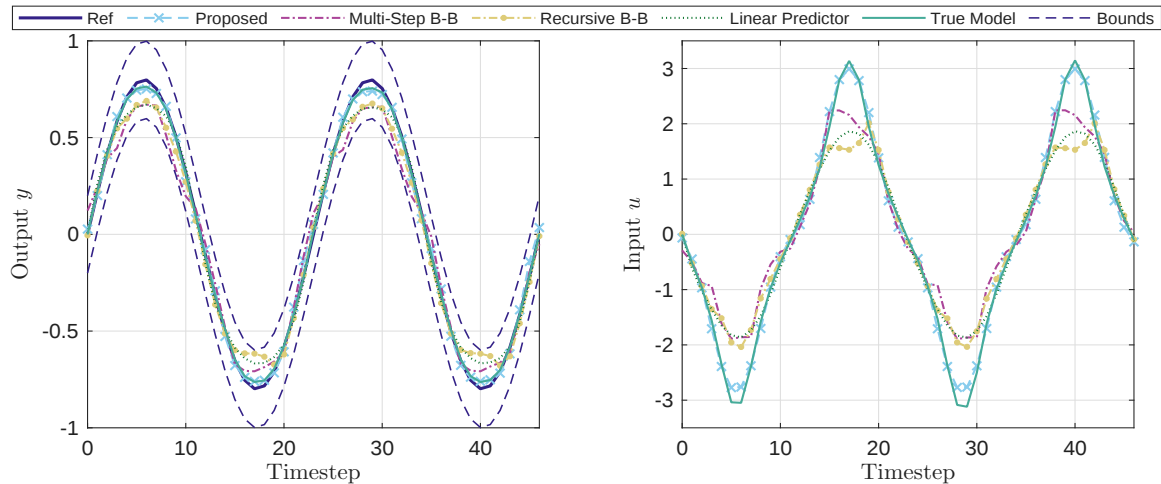
$$\|\phi^{-1}(\bar{y}_1) - \phi^{-1}(\bar{y}_2)\|_2 \leq M \|\bar{y}_1 - \bar{y}_2\|_2.$$

以下条件可保证机会约束 $\Pr(H\mathbf{y}_0^k \leq q) \geq p$ 成立：

$$H\mathbf{y}^k \leq q - c_p \sqrt{\text{diag}(HH^\top)},$$

其中 $c_p := M \left(\sqrt{\mu(p)\sigma_p(\eta^k)} + \|m_p(\eta^k)\|_2 \right)$, $\sigma_p(\eta^k)$ 为 $\hat{k}_p(\eta^k)$ 的最大特征值, $\mu(\cdot)$ 为自由度为 $n_y L'$ 的 χ^2 分布的分位数函数。

模拟算例结果：pH 值过程控制任务



- 1 数据驱动控制概览
- 2 数据驱动控制的随机数据扩展
- 3 Hammerstein-Wiener 系统的随机数据驱动预测与控制
- 4 建筑能源系统应用

建筑能源系统应用 [Yin+24]

			
控制系统	建筑空间 供暖	生活热水 供应	蓄电池 储能
平台	物理实验 高保真模拟	物理实验	物理实验
任务	资源规划	资源规划	资源规划 轨迹跟踪

- 随机扰动与测量噪声，将非线性视为扰动
- 可迁移性：几乎无需调参

[Yin+24]

M. Yin, H. Cai, A. Gattiglio, F. Khayatian, R. S. Smith, and P. Heer. "Data-driven predictive control for demand side management: theoretical and experimental results". In: *Applied Energy* 353 (2024), p. 122101.

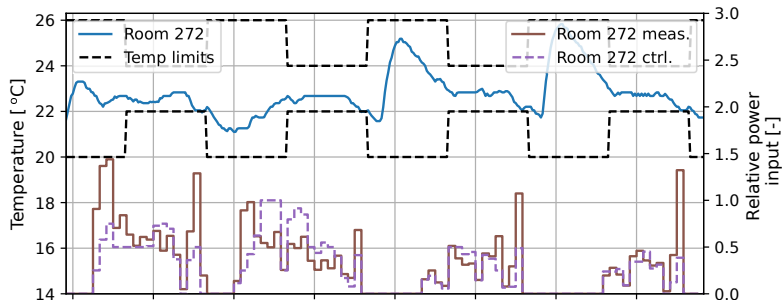


Figure: NEST 建筑能源智能控制验证平台



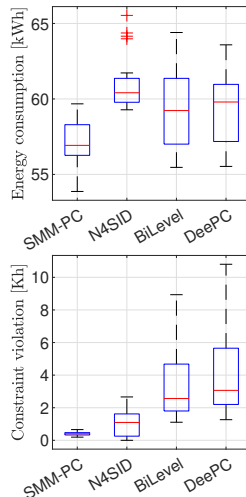
Figure: 实验使用的真实公寓结构图

空间供暖

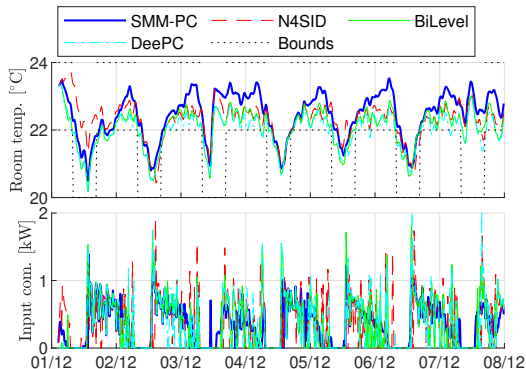


- **物理实验：**4 天内约束违反量为 $0.025^{\circ}\text{C}\cdot\text{h}$
- **高保真仿真：**约束违反减少 59% – 90%，节能 4% – 8%

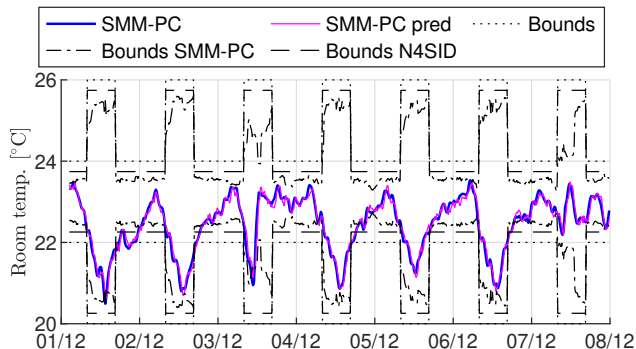
SMM-PC: 提出的方法, N4SID: 系统辨识范式方法,
 BiLevel, DeePC: 已有数据驱动预测控制方法



空间供暖

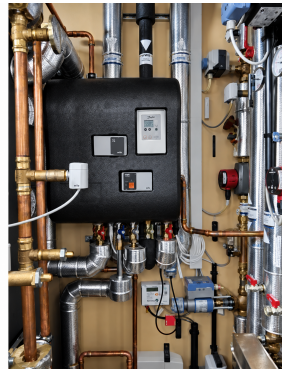
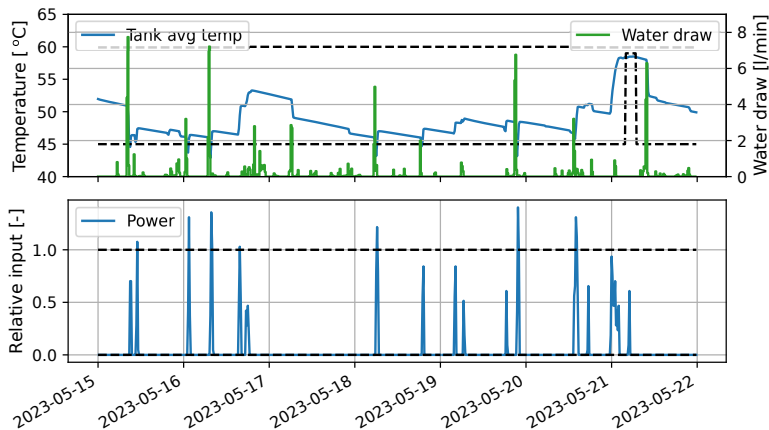


- 更平滑的控制作用
- 动态约束收紧



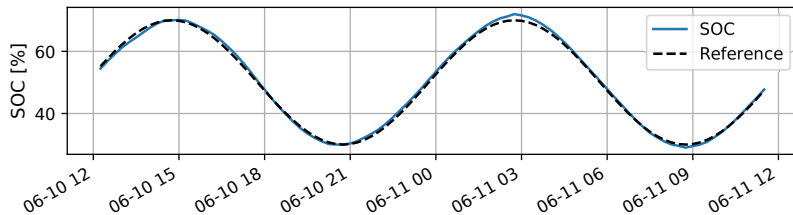
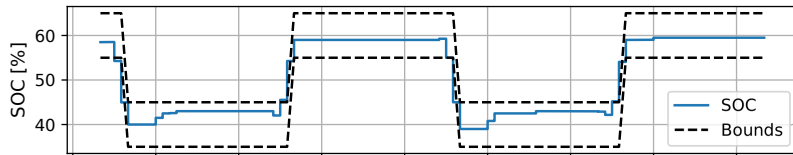
SMM-PC: 提出的方法, N4SID: 系统辨识范式方法, BiLevel, DeePC: 已有数据驱动预测控制方法

生活热水加热



- 由于缺乏用水量预测模型，系统具有很高的不确定性，但在绝大多数时间内可以正常工作

固定式储能电池



- 基于模型的控制同样可行，但数据驱动方法可以自动适应电池的损耗，无需持续参数估计

随机数据驱动控制：统一理论框架、系统理论知识与非线性扩展

- 随机数据驱动平滑、预测和控制问题可以被统一为一个贝叶斯轨迹估计问题
- 随机数据驱动问题的非线性拓展依赖统计学习理论和系统理论知识的结合



Institute of Automatic Control
Leibniz University Hannover

Mingzhou Yin (殷明周)
yin@irt.uni-hannover.de

