第 38 卷第 11 期
2021 年 11 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 38 No. 11
Nov. 2021

具有安全性保证的基于增强主动学习的模型预测控制

任 瑞, 邹媛媛, 李少远[†]

(上海交通大学 自动化系; 系统控制与信息处理教育部重点实验室, 上海 200240)

摘要: 本文提出了一种基于主动学习的增强模型预测控制方法. 该方案克服了大多数基于学习的方法的缺点, 即只能被动地利用可获得的系统数据并导致学习缓慢. 首先应用高斯过程来评估残差模型的不确定性并构建多步预测模型. 然后提出了一个两阶段主动学习策略, 通过在优化问题中引入信息增益作为对偶目标来激励系统探测. 最后, 基于鲁棒不变集定义了安全控制输入集保证了状态约束满足与系统安全性. 本文提出的方法在保证系统安全的情况下提高了学习能力和闭环控制性能, 实验说明了本文方案的优越性.

关键词: 模型预测控制; 主动学习; 高斯过程回归; 对偶控制; 信息增益

引用格式: 任瑞, 邹媛媛, 李少远. 具有安全性保证的基于增强主动学习的模型预测控制. 控制理论与应用, 2021, 38(11): 1735 – 1742

DOI: 10.7641/CTA.2021.10796

Enhanced active learning for model-based predictive control with safety guarantees

REN Rui, ZOU Yuan-yuan, LI Shao-yuan[†]

(Key Laboratory of System Control and Information Processing, Ministry of Education of China;
Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: This paper proposes an active learning-based MPC scheme that overcomes the shortcomings of most learning-based methods which passively leverage the available system data and result in slow learning. We first apply Gaussian process regression to assess the residual model uncertainty and construct multi-step predictive model. Then we propose a two-step active learning strategy and reward the system probing by introducing information gain as dual objective in the optimization problem. Finally, the safe control input set is defined based on robust admissible input set to robustly guarantee state constraint satisfaction. The proposed method improves the learning ability and closed-loop performance with safety guarantees. The advantages of our proposed active learning-based MPC scheme are illustrated in the experiments.

Key words: model predictive control; active learning; Gaussian process regression; dual control; information gain

Citation: REN Rui, ZOU Yuanyuan, LI Shaoyuan. Enhanced active learning for model-based predictive control with safety guarantees. *Control Theory & Applications*, 2021, 38(11): 1735 – 1742

1 Introduction

Model predictive control (MPC) [1–2], as the main control method to systematically deal with system constraints, has achieved remarkable success in many different fields, such as process control [3], autonomous driving [4–6] and robotics [7]. MPC relies heavily on a suitable and sufficiently accurate model that describes the dynamics of the system. However, in many practical scenarios, models based on principles or data-driven approaches are subject to certain uncertainties due to incomplete knowledge of the system and changes in the dynamics over time, which can potentially lead to constraint violation, performance deterioration, as well as instability [8–9].

In the last few years, learning-based model predictive control (LB-MPC) [10–14] has become an active research topic, one direction of which considers the automatic adjustment of the system model, whether it is during operation or between different operation instances. Most researches on learning-based MPC focus on the automatic correction or uncertainty description of predictive models based on data, which is the most obvious component that affects the performance of MPC. Aswani et al. [10] firstly proposed a framework of LB-MPC which decoupled the safety and performance using two models: a model with bounds on its uncertainty and a model updated by statistical methods. This LB-MPC scheme improved the system performance

Received 25 August 2021, accepted 21 October 2021.

[†]Corresponding author. E-mail: syli@sjtu.edu.cn; Tel.: +86 21-34204011.

Recommended by Associate Editor: HONG Yi-guang.

Supported by the National Key R&D Program of China (2018AAA0101701).

through learning model while ensuring the robustness. Terzi et al. [12] constructed a multi-step predictive model with model uncertainties using set-membership, and proposed a robust MPC law in the control design phase. The authors in [13] obtained the predictive model using a nonparametric machine learning technique and then proposed a novel stabilizing robust predictive controller without terminal constraint. These studies require a prior strict bound on uncertainty, which is conservative in practice. To reduce conservatism, Di Cairano et al. [14] proposed a learning-based stochastic MPC for automotive controls using Markov chains. Cautious MPC was proposed in [15] that applied Gaussian process regression (GPR) to learn the model error between the true dynamic and prior model. To solve the problem of constraints, the author used chance constraints on both states and inputs. Based on this, Hewing et al. [4] applied GPR in the control of autonomous race cars which showed significant improvement on the performance in varying racing tasks. In [16], the authors reviewed these LB-MPC methods in detail and divided them into three categories: model learning, controller learning and safe MPC. As far as we are aware of, most LB-MPC techniques are passive learning methods. They account for the modeling challenge only by passively relying on available history process data, which cannot provide effective information as well as directly incentivizing any form of learning. In this work, we try to solve this problem by introducing the notion of active learning.

On the other hand, active learning or dual effect in control was first proposed by Feldbaum [17] that control inputs must have a probing effect that generates informative closed-loop data. [18–22] considered simultaneous identification and control of uncertain systems through dual MPC. Mesbah and Ali [19] summarized MPC with active learning and dual control. This article divided dual control into implicit and explicit methods: in implicit dual control, the optimal control problem is solved approximately; in the explicit one, the probing effect of the controller is directly taken into account in the control scheme in the form of additive cost function or persistent excitation. Heirung [20] presented two approaches to dual MPC, in which the controller is calculated based on minimization of parameter estimate variance and maximization of information. In [21], the authors not only considered MPC with active learning for systems with parametric uncertainty, but also dealt with the problem with model-structure uncertainty. In terms of robust research, A robust dual MPC with constraint satisfaction was proposed for linear systems subject to parametric and additive uncertainty in [22]. In [23], the controller's robustness was achieved through the multi-stage approach which uses a scenario-tree representation of the propagation of the uncertainties over the prediction horizon. A new development in dual MPC is the emergence of control-oriented methods to obtain model uncertainty descriptions related to pre-specified control performance [24–25], and we will not go into details in this work. However, these state-of-the-art approaches are limited to simple linear system dynamics which cannot be applied in complex systems, and most of them also fail to provide theoretical guarantees on safety and closed-loop performance.

In this paper, with the aim to improve the information quality of system operating data and enhance learning ability, we propose an active learning-based MPC (ALB-MPC) scheme based on information measures with safety guarantees. Contributions of our work are as follows: Firstly, the GPR mean function is used to learn the model error and construct the conventional LB-MPC scheme. Secondly, the state constraint satisfaction is robustly guaranteed by selecting control inputs from safety input set. Thirdly, we introduce information gain as dual objective in the optimal control problem and propose a two-stage procedure for ALB-MPC. The next section presents the problem formulation and GPR method. Section 3 presents the common learning-based approach based on GPR and gives the notions of safety guarantee. Section 4 provides the definitions of relevant information measures and the two-step active learning control scheme. Finally, we illustrate the results with some numerical examples in Section 5 and end with the concluding remarks in Section 6.

2 Preliminaries

In this section, we define the notation, problem formulation and the basic content of Gaussian process regression.

2.1 Notation

A normally distributed vector y with mean μ and variance Σ is given by $y \sim N(\mu, \Sigma)$ and so a GP of y is represented by $y \sim gp()$. k_{**} is short for $k(z_*, z_*)$ and $[K]_{ij}$ means the i-th row and j-th column element of matrix K. The superscript d of m_b means that m have d elements. $x_{i|k}$ represents the i-step-ahead prediction of the state at the time step k. $\mathrm{Proj}_X(S)$ means the orthogonal projection of the set S onto X. $A\backslash B$ is the set difference between A and B. $I(z; D)$ denotes the information content of data set D after adding new data z. $H(D)$ is the entropy of data set D and $H(z, D)$ means the differential entropy at any data point z.

2.2 Problem formulation

In this paper, we consider a discrete-time, nonlinear dynamical system

$$x_{k+1} = f(x_k, u_k) = h(x_k, u_k) + g(x_k, u_k), \quad (1)$$

with observable system state $x_k \in \mathbb{R}^{n_x}$ and control input $u_k \in \mathbb{R}^{n_u}$ at time step $k \in \mathbb{N}$, where n_x, n_u is the dimension of the state and input. We assume that the true system f is not exactly known and use the sum of a prior nominal model and a learned model to represent it. Here, $h(x_k, u_k)$ is a simple and fixed nominal linear model that could be achieved by first principles or people's prior knowledge. $g(x_k, u_k)$ is a learned part that represents the model error between the true behavior of the system and the prior model. We can use machine learning methods to model the learned part by collecting observations from the system during operation.Note that both the state and input are required to satisfy the following mixed constraints:

$$(x, u) \in Z \subset X \times U. \tag{2}$$

2.3 Gaussian process regression

Gaussian process regression (GPR) provides an explicit estimate of the model uncertainty that is used to derive probabilistic bounds in control settings. In this section, we will briefly introduce the concept of GPR and use it to learn the model g. Gaussian process can be viewed as a collection of random variables with a joint Gaussian distribution for any finite subset. Given noisy observations y of function $g : \mathbb{R}^{n_z} \mapsto \mathbb{R}^{n_d}$

$$y = g(z) + \varepsilon,$$

where $n_z = n_x + n_u$ and n_d is the dimension of output of g. $\varepsilon \sim N(0, \sigma_{n_g}^2)$ is Gaussian noise with zero mean and diagonal variance $\sigma_{n_g}^2$. A GP of y is fully described by its mean function $m(z)$ and covariance function $k(z, z')$, denoted by $y \sim gp(m(z), k(z, z'))$. Given a set of m input vectors $Z = [z_0^{\mathrm{T}} \ \cdots \ z_{m-1}^{\mathrm{T}}] \in \mathbb{R}^{m \times n_z}$ and the corresponding output $Y = [y_0^{\mathrm{T}} \ \cdots \ y_{m-1}^{\mathrm{T}}] \in \mathbb{R}^{m \times n_d}$, then we define the training data by $D = (Z, Y)$. We assume that each output dimension $a \in \{1, \cdots, n_d\}$ is independent and the posterior distribution in a at a test point z_* is Gaussian with mean and variance given by

$$m^a(z_*) = k_*^a(K^a + I\sigma_a^2)^{-1}Y^a, \tag{3}$$

$$\Sigma^a(z_*) = k_{**}^a - k_*^a(K^a + I\sigma_a^2)^{-1}k_*^{a\mathrm{T}}, \tag{4}$$

where $k_{**}^a = k(z_*, z_*) \in \mathbb{R}$, $k_*^a = k(z_*, Z) \in \mathbb{R}^m$, K is the covariance matrix with elements $[K]_{ij} = k^a(z_i, z_j)$ and I is the identity matrix, Y^a is the output at each dimension a. In this model, we consider the squared exponential kernel

$$k^a(z, \overline{z}) = \sigma_{f,a}^2 \exp(-(z - \overline{z})^{\mathrm{T}} L^a (z - \overline{z})),$$

here $\sigma_{f,a}^2$ denotes the squared signal variance and $L^a \in \mathbb{R}^{n_z \times n_z}$ is a positive diagonal length scale matrix, which can be achieved by maximizing the Marginal Log-likelihood. The resulting GP approximation of the function g is given by

$$g(z) \sim N(m^d(z), \Sigma^d(z)), \tag{5}$$

with $m^d = [m_1^d \ \cdots \ m_{n_d}^d]$ and $\Sigma^d = \mathrm{diag}\{[\Sigma_1^d \ \cdots \ \Sigma_{n_d}^d]\}$.

Note that when GPR is applied in control, many research consider the propagation of uncertainty. In this paper, however, we employ the mean function to perform multi-step ahead predictions without considering the propagation of uncertainty for simplicity. Also, sparse Gaussian processes can reduce computation, but here we do not consider this method. For more specific knowledge about GPR, readers can refer to literature [26–27].

3 Learning-based MPC and safety guarantees

In this section, we apply GPR mentioned above to learn the model error and combine it with the prior nominal model to design learning-based MPC scheme. Also, we introduce the concept of robust control invariant set and safe control input set which guarantee the safety of the system.

3.1 Model learning and LB-MPC

The training data y_k is generated from the mismatch between measurements of x_{k+1} and the prior nominal model during operation

$$y_k = x_{k+1} - h(x_k, u_k) = g(z_k), \tag{6}$$

where $z_k = [x_k \ u_k]^{\mathrm{T}}$. We then denote the recorded data set including all past control inputs and states available at time step k

$$D_k = \{(x_k, z_{k-1}), (x_{k-1}, z_{k-2}), \cdots, (x_1, z_0)\}.$$

We use the data set D_k to update the GPR model and make multi-step prediction combined with the nominal model at every time step. The model constructed is assumed to be equivalent to the true dynamics of the system. Then, the control inputs are determined knowing that the best prediction is available given the current system information. This method is a kind of certainty-equivalence approaches and the learning here is passive.

GPR mean function is used in the passively learning-based MPC approach. From equations (3)–(6) and the data set D_k, the one-step-ahead predictive model can be constructed as follows:

$$x_{k+1} = h(x_k, u_k) + m^d(x_k, u_k). \tag{7}$$

Then we can formulate the closed-loop LB-MPC problem. Considering the following finite-horizon object function:

$$\min_u J_{\mathrm{task}} = \sum_{i=0}^{N-1} l(x_{i|k}, u_{i|k}) + L_N(x_N), \tag{8}$$

here, J_{task} means the cost function of control task, $i = 0, 1, \cdots, N-1$, $x_{i|k}$ is the i-step-ahead prediction of the state initialized at $x_{0|k} = x(k)$, and $u_{i|k}$ is of the same. This cost function can be selected as $l(x_{i|k}, u_{i|k}) = x_{i|k}^{\mathrm{T}} Q x_{i|k} + u_{i|k}^{\mathrm{T}} R u_{i|k}$ or modified to include set-points for both states and actions. The optimization problem is minimized at time step k and the model above can be used to predict the effect of the control inputs on the system state as follows:

$$x_{i+1|k} = h(x_{i|k}, u_{i|k}) + m^d(x_{i|k}, u_{i|k}),$$

$x_{0|k} = x(k)$.

Constraints on the inputs and states from (2) can be formulated as follows and these constraints are usually chosen based on physical hardware limitations, desired performance or safety considerations:

$$(x_{i|k}, u_{i|k}) \in Z \subset X \times U.$$

At every time step k, the current system states is measured, then we can get the error between the real output of the system and the one of the learned model in the previous time step. Afterwards, this information is combined with past information to learn the new predictive model, and the open-loop optimization problem formed by (8) is solved. The solution to the optimal control problem is the open-loop input sequence $\{u_{i|k}\}_{i=0}^{N-1}$ and the first element of this sequence is used as an input to the system $u^*_{0|k}$. Then the process is repeated at each sampling time.

3.2 Safety consideration

In the LB-MPC problem, satisfaction of the state constraints cannot be guaranteed and the chosen input may not be safe. As illustrated above, we apply GPR mean function to learn the model error. Then, the confidence bounds of the GPR can be used to characterize model uncertainty.

$$G(x,u) = \{g|(g - m^d(x,u))^{\mathrm{T}}(\Sigma^d(x,u))^{-1}(g - m^d(x,u)) \leqslant \chi_n^2(p)\}, \tag{9}$$

where $\chi_n^2(p)$ is the quantile function for the chi-squared distribution with n degrees of freedom and $p \in (0,1)$ is a tuning parameter. $G(x,u)$ can be found through offline learning. Then we make the following assumption:

Assumption 1 At every time step k, the learned model satisfies constraint $g(x_k, u_k) \in G(x_k, u_k)$ in (9) which is determined through offline learning.

Note that this assumption will not formally hold since the normal distribution has infinite support, but it is useful in practice and the confidence bounds are commonly used to model uncertainty. Combining it with the previous constraint on the state and input, we define set in (10):

$$\Omega := \{(x,u,g)|(x,u) \in Z \cap g \in G(x,u)\}. \tag{10}$$

It can be viewed as the subset of the graph G where the state and input constraints are both satisfied. Hence, we have $G = \mathrm{Proj}_{X\times U}(\Omega)$ which is the orthogonal projection of set Ω. Then the state-dependent set of admissible inputs can be defined as

$$U(x) := \{u|(x,u) \in Z\},$$

such that the set of admissible states is then

$$X := \{x|\exists u, (x,u) \in Z\} = \mathrm{Proj}_X(Z).$$

Based on these sets, we define the notions of robust control invariant set and safe control input set.

Definition 1 (Robust control invariant set) A set $C \subseteq X$ is a robust control invariant set (RCI) for (1)(2), if for any, there exists a $u \in U(x)$ such that

$$x \in C \Rightarrow \exists u \in U(x) : h(x,u) + g \subset C,$$
$$\forall g \in G(x,u).$$

The set $C^\infty \subseteq X$ is the maximal RCI set if all other RCI sets are contained in it. Based on definition (1), we define the safe input set.

Definition 2 (Safe control input set) Given an maximal RCI, the safe control input set (SCIS) for state $x \in X$ is

$$\Pi_{\mathrm{safe}}(x) := \{u \in U(x)|h(x,u) + g \in C^\infty, \forall g \in G(x,u)\}\}. \tag{11}$$

As a result, any control inputs can be chosen in the SCIS and keep the system safe.

Theorem 1 Based on Assumption 1, Definitions 1 and 2, at every time step k, system $x_{k+1} = h(x_k, us_k) + m^d(x_k, us_k)$ is safe that it always satisfies constraint (2) for any safe control input $us_k \in \Pi_{\mathrm{safe}}$.

Proof Since C is an RCI set and SCIS is not empty, from Definition 1 and 2, any control input selected from the SCIS guarantees that C is an RCI set for system system $x_{k+1} = h(x_k, us_k) + m^d(x_k, us_k)$ and constraint (2). □

So the challenge is how to compute the safe input set. First, we adopt the notion of predecessor set (or one-step set). Given a set $\Omega \subset X$, the predecessor set is

$$\mathrm{Pre}(\Upsilon) := \{x|\exists u \in U(x), f(x,u,G(x,u)) \in \Upsilon\}.$$

RCI sets can be computed recursively from the target set (also called terminal constraint set) $X_f \subseteq X$, then we can get

$$\begin{cases} X_0 = X_f, \\ X_{i+1} = \mathrm{Pre}(X_i). \end{cases} \tag{12}$$

For more details, some important and famous results that related to the recursion and the computation of invariant sets are in the surveys [28–29]. The next main problem in this section is how to calculate predecessor set $\mathrm{Pre}(\Upsilon)$.

Theorem 2 Given the set of admissible state-input pairs $\Sigma(\Upsilon)$ and the set of triplets $\Phi(\Upsilon)$, the predecessor set of Υ is given by

$$\mathrm{Pre}(\Upsilon) = \mathrm{Proj}_X(\Sigma(\Upsilon)),$$

where

$$\Sigma(\Upsilon) = Z \backslash \mathrm{Proj}_{X\times U}(\Omega \backslash \Phi(\Upsilon)),$$
$$\Phi(\Upsilon) = f^{-1}(\Upsilon) = \{(x,u,g)|h(x,u) + g \in \Upsilon\}.$$

Then the sets $\mathrm{Pre}(\Upsilon)$ can be calculated and the proof can be referred to paper [28,32]. Due to the confidence bounds in (9) are non-convex union of ellipsoids, so we need to construct polyhedral cover outer approximation of set $\varOmega$. Firstly, we make polyhedral partition for $Z = \bigcup\limits_{i=1}^{N_\mathrm{c}} Z_i$, N_c is the number of regions used to construct the polyhedral cover. Then, we calculate the minimum and maximum value $g_i^{\min}, g_i^{\max}$ according to $(x,u) \in Z_i$ and $g \in G(x,u)$. Finally, we can get $\varOmega = \bigcup\limits_{i=1}^{N_\mathrm{c}} \varOmega_i$.

Given this polyhedral cover representations of the sets and the assumption that the nominal model is linear or piecewise affine, then the calculation of $\mathrm{Pre}(\varUpsilon)$ is outlined by the following procedures:

A) Compute the projection: $Z = \mathrm{Proj}_{X\times U}(\varOmega)$.

B) Compute the inverse map: $\varPhi = f^{-1}(\varUpsilon)$.

C) Compute the projection: $\varPsi = \mathrm{Proj}_{X\times U}(\varOmega\backslash \varPhi(\varUpsilon))$.

D) Compute the set difference: $\varSigma(\varUpsilon) = Z\backslash\varPsi$.

E) Compute the projection: $\mathrm{Pre}(\varUpsilon) = \mathrm{Proj}_X (\varSigma(\varUpsilon))$.

In order to achieve the results, we can use some important tools such as CPLEX, MPT3 for computing inverse images, set differences and projections.

4 Safety guaranteed active learning-based MPC

In this section, we propose a two-step procedure for active learning-based MPC. We first compute the most informative input sequence from the safety control set aiming to maximize the information of new data. Then, deviations from the desired input sequence are penalized in the constrained optimization problem. Specific methods are as follows.

4.1 Information gain and active dynamics learning

As previously presented, GPR is a non-parametric model which means we cannot use parameter estimate variance or Fisher information (FI) as the measure of model uncertainty reduction. So, we introduce an explicit information content objective to measure the reduction in estimated model uncertainty $I(x_\mathrm{new}, u_\mathrm{new}; D)$, which denotes the information content of new data x_new, u_new adding to the history data set D. Here, we employ the concept of information gain which is commonly used to qualify reduction in estimated uncertainty.

Definition 3 Given the observation set D, when new data $z_\mathrm{new} = [x_\mathrm{new}\ u_\mathrm{new}]^\mathrm{T}$ is available, the information gain of the data is defined as

$$I(z_\mathrm{new}; D) = H(D) - H(D \cup z_\mathrm{new}), \tag{13}$$

where $H(D)$ denotes the entropy before observation and $H(D \cup z_\mathrm{new})$ is the entropy when adding new data. $I(z_\mathrm{new}; D)$, which is also known as mutual information, is often greater than zero. The greater the value is, the more information we have gained and the more uncertainty reduction is achieved.

As a result, we want to find the new data which maximize the information gain. Due to the fact that equation (13) is hard to be optimized [30–31] and needs to be approximated as follows:

$$I(z_\mathrm{new}; D) \geqslant H(z_\mathrm{new}, D) = \frac{d}{2}\log(2\pi e \varSigma^d(z_\mathrm{new})), \tag{14}$$

this equation illustrates that the information gain approximates to the log value of the output variance at new data point. Furthermore, when a new data point is collected, both the GPR model and information gain change. We then define the active dynamics learning problem.

$$\begin{aligned}
\max_{U_a} J_a = & \sum_{i=0}^{N_a-1} H((x_i, u_{a,i}), D),\\
\text{s.t.}\ \ & x_{i+1} = h(x_i, u_{a,i}) + m^d(x_i, u_{a,i}),\\
& x_0 = x_k,\\
& u_{a,i} \in \varPi_\mathrm{safe}(x_i),\\
& \text{(3)(4)(14)},
\end{aligned} \tag{15}$$

here, N_a is the active learning horizon. Solving this problem at each time step k, we get the most informative input sequence $U_a := \{u_{a,1}, \cdots, u_{a,N_a-1}\}$ and guarantee the safety of active exploration because the control inputs are selected from the safety control input set.

4.2 Safe active learning-based MPC

In dual control paradigm, the control inputs not only need to satisfy control task performance but also have probing effect on system dynamics. So, we consider two objectives: one is the control task objective J_task of equation (8) and the other is dual objective J_dual which is achieved by penalizing the deviations from the desired input sequence. These two objectives are conflicting and we need to achieve a balance between them: $J = J_\mathrm{task} + J_\mathrm{dual}$.

The safe active learning-based MPC optimization problem can be stated in (16)

$$\begin{aligned}
\min_{U} J = & \sum_{i=0}^{N-1} l(x_{i|k}, u_{i|k}) + L_N(x_N) + \\
& \alpha\|u_{i|k} - u_{a,i}\|_2^2,\\
\text{s.t.}\ \ & x_{i+1|k} = h(x_{i|k}, u_{i|k}) + m^d(x_{i|k}, u_{i|k}),\\
& x_{0|k} = x_k,\\
& u_{i|k} \in \varPi_\mathrm{safe}(x_i),\\
& \text{(3)(4)(14)},
\end{aligned} \tag{16}$$

where α is a tuning parameter which determines the amount of dual effect. When $\alpha = 0$, it means no dual effect or no active learning and it is a common control problem. Also, the control inputs are limited in the safe set which guarantees the safety of our ALB-MPC approach.

Remark 1 The advantages of this two-step strategy lie that, (16) can still generates safe control inputs using U_a at time $k-1$ if (15) fails at time step k. This optimization problem is simplified because we just penalize the deviations from the desired input sequence.

Remark 2 The dual control problem can be divided into two phases: a control phase and an identification phase. Switch between the two phases is based on the model uncertainty that there is no need for active exploration when the uncertainty becomes small enough.

Finally, the safe active learning-based MPC (ALB-MPC) strategy is outlined in Algorithm 1.

Algorithm 1 Safety guaranteed active learning-based MPC scheme.

Offline: Calculate (9) and determine the safe control input set in Section 4.2.

Online: Update data set, adjust the GPR model and design controller.

1) Initialize training dataset D_k, control inputs and states, controller parameters

2) for $i = 1 : N$, do

3) Measure the current state x_k at every time step k;

4) Calculate the model error $x_k - h(x_{k-1}, u_{k-1})$ and update data set D_k;

5) Learn the model and construct multi-step prediction model using (3)(4)(7);

6) Solve problem (15) according to the information content using predictive new data (13)(14);

7) Solve problem (16) and apply the first element of control input;

8) end for

5 Numerical examples

Two numerical examples are considered in this section. An Van der Pol oscillator and a cart-pole balancing task. Both examples are constructed such that we are able to illustrate advantages of our proposed active LB-MPC.

5.1 Van der Pol oscillator

The equation of the system dynamics is as follows:

$$\dot{x}_1 = (1 - x_2^2)x_1 - x_2 + u,$$
$$\dot{x}_2 = x_1.$$

In the model learning part, the initial states of the system are $x_1 = 1, x_2 = 0$ and the lower and upper bounds on the inputs are $u_{\min} = -0.75, u_{\max} = 1$. We first discretize this ODE equation to get a nonlinear discrete model. Then we also get a linear model of the system using successive linearization methods and treat it as the prior nominal model: $x_{k+1} = Ax_k + Bu_k$, where $A = [1.4766\ -0.6221; 0.6221\ 0.8544]$ and $B = [0.6221; 0.1456]$. We use this nominal model to start the control process and collect related information to learn the model error. The initial training data is zero and updated online, then we learn a GPR model based on these data and the hyper-parameters are optimized by maximizing the marginal log-likelihood. Finally, we get the whole predictive model. In the MPC design, the modeling horizon N_s is 20, the prediction horizon is chosen equal to control horizon: $N_{\rm p} = N_t = 10$, the weight matrices are $Q = \mathrm{diag}\{[1\ 1]\}$ and $R = 1$. The active learning horizon is selected $N_a = 5$ and the tuning parameter of dual effect α is chosen 0 and 10 (when $\alpha = 0$, it means no probing), then we will make comparison when α is chosen these two different values.

Figure 1 shows the changes of the system state, they eventually converge to the origin from the initial state. when $\alpha = 10$, the convergence speed of the states is faster than that when $\alpha = 0$, suggesting that active learning scheme not only introduces additional excitation to the system but also makes the system learn actively. The results show that the tracking effect of the controller is improved effectively and the learning of active scheme is faster. In order to illustrate the dual effect more clearly, we introduce an index to represent the excitation level defined as $I_u = J_u^{\rm ALBMPC}/J_u^{\rm LBMPC} = (\sum\limits_{k=1}^{N_A} u_{Ak}^2/N_A)/(\sum\limits_{k=1}^{N_L} u_{Lk}^2/N_L)$. This index describes the comparison of the value of control inputs obtained by different methods before reaching steady state, and we can get the excitation level here $I_u = 1.14$. Note that we do not consider the sate constraints here and we will verify safety in the second example.

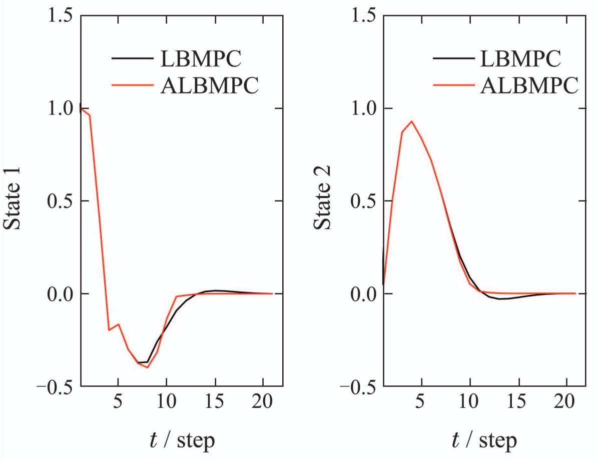

Fig. 1 Comparison of state trajectory in Vdp

5.2 Cart-pole balancing task

The schematic diagram of the inverted pendulum example is shown in Figure 2 and we aim to achieve an upright pendulum position of the pole by applying force to the cart. The continuous-time dynamics of the pendulum are given as follows:

$$(M_{\rm c} + M_{\rm p})\ddot{x} + b\dot{x} + \frac{1}{2}M_{\rm p}l\ddot{\theta}\cos\theta - \frac{1}{2}M_{\rm p}l\dot{\theta}^2\sin\theta = F,$$

$$(I + M_{\mathrm{p}}(\frac{l}{2})^2)\ddot{\theta} - \frac{1}{2}M_{\mathrm{p}}\mathrm{g}l\sin\theta + M_{\mathrm{p}}l\ddot{x}\cos\theta = 0.$$

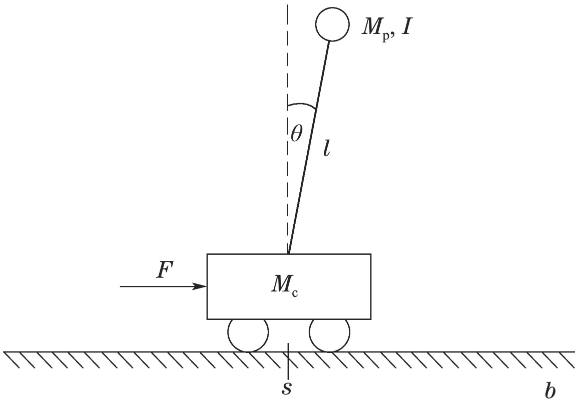

Fig. 2 Schematic diagram of Cart-pole

The mass of the carriage and the pole are given by $M_{\mathrm{c}} = 5, M_{\mathrm{p}} = 2$, the pole is defined by its length and moment of inertia $l = 3, I = 0.6$. b is the friction coefficient and $\mathrm{g} = 9.81$ is gravitational constant. The states of the system are chosen as $[\ddot{x}, \dot{x}, \ddot{\theta}, \theta]$ and $u = F$ is the control input. In the model learning part, we first linearize the equation above and get a linear nominal model. Then the model error is learning by GPR online. In the controller designing part, the control objective is to stabilize the pole at $\dot{x} = 0, \dot{\theta} = 0, \theta = 0$. The origin of the system corresponds to the pendulum standing upright and so the reference is selected as $r = [0, 0, 0]$. The objective function is modified to include set-points for the states as $l(x, u) = (x - r)^{\mathrm{T}}Q(x - r) + u^{\mathrm{T}}Ru$, where Q and R are penalizing weight matrices/values. In this part, we choose the prediction horizon: $N_{\mathrm{p}} = 10$ which equals to the control horizon. The active learning horizon is selected $N_a = 10$ and the tuning parameter is chosen $\alpha = 0, 1, 5$. In this example, the system is under control constraint $U = \{u \in \mathbb{R}| - 10 \leqslant u \leqslant 5]\}$ and state constraint $X = \{\theta \in \mathbb{R}||\theta| \leqslant \frac{\pi}{3}]\}$. In the simulation, pendulum will be unstable if the pole angle is beyond $\pi\backslash 3$. The target constraint set (12) of this problem acts as a stability constraint and keep the pendulum stable.

The evolution of state θ is depicted in Fig. 3. We can see that these methods all enable adapting the model and stabilizing the system. The passive learning-based method in black color shows slower convergence than other active learning ones. When the parameter is chosen $\alpha = 5$, the state converges faster than the one with $\alpha = 1$. This result illustrates that our algorithm really introduces active learning and the tracking performance of the controller is improved effectively. The safety can also be reflected in this experiment that the inverted pendulum has not failed during the whole process.

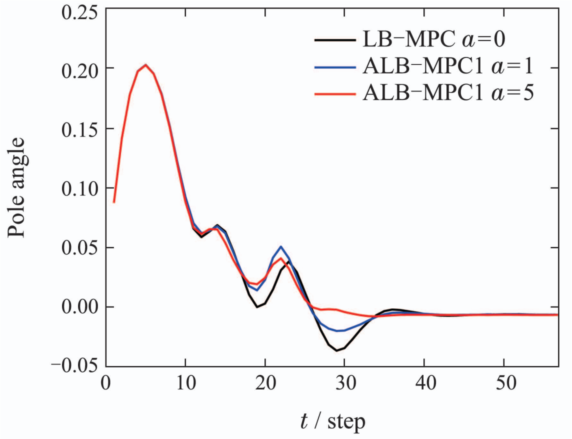

Fig. 3 Comparison of state trajectory in Cart-pole

6 Conclusion

In this work, we focus on the drawbacks of learning-based MPC methods that they lack effective data information and cannot excite any form of learning. we solve this problem by introducing active learning and dual effect, which enhances rapid learning ability and improves closed-loop control performance. The input and state constraints are also guaranteed in our method, and the advantages of the proposed method are illustrated in the simulations. Further research will be devoted to the control-oriented uncertainty description of the complex systems and the new form of dual objective. We will also study how to reduce the computational cost and focus on the application of our proposed method in process control.

References:

[1] MAYNE D Q. Model predictive control: recent developments and future promise. *Automatica*, 2014, 50(12): 2967 – 2986.

[2] RAWLINGS J B, MAYNE D Q, DIEHL M. *Model Predictive Control: Theory, Computation, and Design*. Madison, WI: Nob Hill Publishing, 2017.

[3] FORBES M G, PATWARDHAN R S, HAMADAH H, et al. Gopaluni, model predictive control in industry: challenges and opportunities. *IFAC-PapersOnLine*, 2015, 48(8): 531 – 538.

[4] HEWING L, LINIGER A, ZEILINGER M N. Cautious nmpc with Gaussian process dynamics for autonomous miniature race cars. *European Control Conference (ECC)*. Limassol, Cyprus: IEEE, 2018, 1341 – 1348.

[5] HEWING L, KABZAN J, ZEILINGER M N. Cautious model predictive control using Gaussian process regression. *IEEE Transactions on Control Systems Technology*, 2019, arXiv:1705.10702.

[6] KABZAN J, HEWING L, LINIGER A, et al. Learning-based model predictive control for autonomous racing. *IEEE Robotics and Automation Letters*, 2019, 4(4): 3363 – 3370.

[7] CARRON A, ARCARI E, WERMELINGER M, et al. Data-driven model predictive control for trajectory tracking with a robotic arm. *IEEE Robotics and Automation Letters*, 2019, 4(4): 3758 – 3765.

[8] RAKOVIC S V. Model predictive control: classical, robust, and stochastic. *IEEE Control Systems Magazine*, 2016, 36(6): 102 – 105.

[9] MESBAH A. Stochastic model predictive control: An overview and perspectives for future research. *IEEE Control Systems Magazine*, 2016, 36(6): 30 – 44.

[10] ASWANI A, GONZALEZ H, SASTRY S S, et al. Provably safe and robust learning-based model predictive control. *Automatica*, 2013, 49(5): 1216 – 1226.

[11] GROS S, ZANON M. Data-driven economic nmpc using reinforcement learning. *IEEE Transactions on Automatic Control*, 2019, 65(2): 636 – 648.

[12] TERZI E, FAGIANO L, FARINA M, et al. Learning-based predictive control for linear systems: a unitary approach. *Automatica*, 2019, 108: 108473.

[13] MANZANO J M, LIMON D, DE LA PEÑA D M, et al. Robust learning-based mpc for nonlinear constrained systems. *Automatica*, 2020, 117: 108948.

[14] DI CAIRANO S, BERNARDINI D, BEMPORAD A, et al. Stochastic mpc with learning for driver-predictive vehicle control and its application to hev energy management. *IEEE Transactions on Control Systems Technology*, 2013, 22(3): 1018 – 1031.

[15] HEWING L, KABZAN J, ZEILINGER M N. *Cautious model predictive control using Gaussian process regression*. arXiv e-prints, 2017.

[16] HEWING L, WABERSICH K P, MENNER M, et al. Learning-based model predictive control: toward safe learning in control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2020, 3: 269 – 296.

[17] FELDBAUM A. Dual control theory, Part I. *Avtomatika i Telemekhanika*, 1960, 21(9): 1240 – 1249.

[18] HEIRUNG T A N, YDSTIE B E, FOSS B. Dual adaptive model predictive control. *Automatica*, 2017, 80: 340 – 348.

[19] MESBAH A. Stochastic model predictive control with active uncertainty learning: a survey on dual control. *Annual Reviews in Control*, 2018, 45: 107 – 117.

[20] HEIRUNG T A N, FOSS B, YDSTIE B E. Mpc-based dual control with online experiment design. *Journal of Process Control*, 2015, 32: 64 – 76.

[21] HEIRUNG T A N, PAULSON J A, LEE S, et al. Model predictive control with active learning under model uncertainty: why, when, and how. *AIChE Journal*, 2018, 64(8): 3071 – 3081.

[22] WEISS A, DI CAIRANO S. Robust dual control mpc with guaranteed constraint satisfaction. *The 53rd IEEE Conference on Decision and Control*. Jeju, Korea: IEEE, 2014: 6713 – 6718.

[23] THANGAVEL S, LUCIA S, PAULEN R, et al. Dual robust nonlinear model predictive control: a multi-stage approach. *Journal of Process Control*, 2018, 72: 39 – 51.

[24] TELEN D, HOUSKA B, VALLERIO M, et al. A study of integrated experiment design for NMPC applied to the droop model. *Chemical Engineering Science*, 2017, 160: 370 – 383.

[25] FENG X, HOUSKA B. Real-time algorithm for self-reflective model predictive control. *Journal of Process Control*, 2018, 65: 68 – 77.

[26] WILLIAMS C K, RASMUSSEN C E. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT press, 2006, 2(3).

[27] RASMUSSEN C E. Gaussian processes in machine learning. *Summer School on Machine Learning*. Berlin, Heidelberg: Springer, 2003, 63 – 71.

[28] RAKOVIC S V, KERRIGAN E C, MAYNE D Q, et al. Reachability analysis of discrete-time systems with disturbances. *IEEE Transactions on Automatic Control*, 2006, 51(4): 546 – 561.

[29] BONZANINI A D, GRAVES D B, MESBAH A. Learning-based smpc for reference tracking under state-dependent uncertainty: an application to atmospheric pressure plasma jets for plasma medicine. *IEEE Transactions on Control Systems Technology*, 2021, DOI: 10.1109/TCST.2021.3069825.

[30] BUISSON-FENET M, SOLOWJOW F, TRIMPE S. *Actively learning Gaussian process dynamics*. ArXiv preprint arXiv: 1911.09946, 2019.

[31] UMLAUFT J M. *Safe learning control for Gaussian process models*. Technische Universität München, 2020.

[32] BONZANINI A D, PAULSON J A, MESBAH A. Safe learning-based model predictive control under state-and input-dependent uncertainty using scenario trees. *The 59th IEEE Conference on Decision and Control (CDC)*. Jeju, Korea: IEEE, 2020: 2448 – 2454.

作者简介:

任　瑞　博士研究生, 目前研究方向为基于学习的模型预测控制, E-mail: sjturr@sjtu.edu.cn;

邹媛媛　教授, 目前研究方向为网络化系统分布式优化与控制、预测控制, E-mail: yuanyzou@sjtu.edu.cn;

李少远　教授, 目前研究方向为预测控制、自适应智能控制、模糊智能控制, E-mail: syli@sjtu.edu.cn.