Application of Interval Predictor Model Into Model Predictive Control

WANG JIANHONG[1] .''RICARDO A. RAMIREZ-MENDOZA[2]

[1]School of Electronic Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, 343100, CHINA

[2]School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, Ave. Eugenio Garza Sada 2501, Monterrey, N.L., 64849, MEXICO

Abstract: In this paper, interval prediction model is studied for model predictive control (MPC) strategy with unknown but bounded noise. After introducing the family of models and some basic information, some computational results are presented to construct interval predictor model, using linear regression structure whose regression parameters are included in a sphere parameter set. A size measure is used to scale the average amplitude of the predictor interval, then one optimal model that minimizes this size measure is efficiently computed by solving a linear programming problem. The active set approach is applied to solve the linear programming problem, and based on these optimization variables, the predictor interval of the considered model with sphere parameter set can be directly constructed. As for choosing a fixed non-negative number in our given size measure, a better choice is proposed by using the Karush-Kuhn-Tucker (KKT) optimality conditions. In order to apply interval prediction model into model predictive control, the midpoint of that interval is substituted in a quadratic optimization problem with inequality constrained condition to obtain the optimal control input. After formulating it as a standard quadratic optimization and deriving its dual form, the Gauss-Seidel algorithm is applied to solve the dual problem and convergence of Gauss-Seidel algorithm is provided too. Finally simulation examples confirm our theoretical results.

1. Introduction

Model predictive control (MPC) has developed considerably over the past two decades, both within the control field and in industry. This success can be attributed to the fact that model predictive control is perhaps the most general way of posing constrain control problem in the time domain. Model predictive control formulation integrates optimal control, stochastic control, control of processes with dead time, multivariable control and future references when available. One important advantage of model predictive control is that because of the finite control horizon used, constraints and, in general nonlinear processes which are frequently found in industry, can be handled. The rationale behind model predictive control is the following: at each time step, an L2 or alternative variation of the cost function is locally optimized over time to obtain the open loop control as a function of time instant, only a small portion of which is actually applied to the system. The time horizon is then shifted, and the process is repeated at a later time step based on state feedback (Marko Tanaskovic, 2014). Although model predictive control has been found to be quite a robust type of control in most reported applications, some new and very promising results allow one to think that this control technique will experience greater expansion within this community in near years (X Zhang, 2017). However although lots of applications have been reported in both industries and research institutions, model predictive control has not yet reached

popular in industry.

The most important element in model predictive control is the prediction of output value. After deriving the prediction of output value by prediction error method and then substituting in into one considered cost function, we take the derivative of the cost function with respect to input value to obtain one optimal input (Marko Tanaskovic, 2017). But the problem of deriving the prediction of output value is dependent of external noise, which is always assumed to be independent and identically distributed white noise. Due to white noise is an ideal case, it does not exist in engineering and in additional, deriving statistical properties of noise is often very difficult in practice as it is usually not possible to measure noise directly (M Casini, 2014). To relax this strict probabilistic description on noise, assumption on the noise bound is less restrictive, as noise is bounded and the bound can be roughly calculated from the specification of the used sensor. Here we investigate model predictive control in presence of bounded noise, which is similar to set membership predictive control (Vito Cerone, 2014). The idea of set membership predictive control is from set membership identification in classical system identification theory. As this paper is different from set membership identification, we apply interval predictor model, coming from classical system identification theory, into model predictive control. To have a better understanding about how to apply interval predictor model into model predictive control, firstly we give a short review about interval predictor model.

In classical system identification theory, one parametric model structure corresponding to an identified system is selected firstly, and then the parameters in the parametric model structure are estimated using the measured input-output data. During the system identification process, many identification methods are proposed to identify these unknown parameters, for example the classical least squares method, instrumental variable method, maximum likelihood estimation method, prediction error method, Bayesian method, etc. (J M

Bravo, 2016). One common property in these identification methods is that the prior information about noise is known. Based on some probabilistic assumptions on noise, the unknown parameters are identified as the specific numerical values. But the probabilistic assumptions on noise are not realistic and these probabilistic assumptions are not realized easily in reality. So in order to relax the probabilistic assumptions on noise, we always assume that the noises are unknown but bounded. This unknown but bounded assumption is weaker than the original probabilistic assumption, as it needs not any prior distribution on noise. The common used method applied to solve the unknown but bounded case is called set membership identification (M.Milanese, 2004). In set membership identification, the obtained result is not a detailed numerical value, but a guaranteed interval with respect to each parameter. This guaranteed interval means that each parameter can be included in this interval with one guaranteed accuracy which is assessed by some probabilistic inequalities. From the idea of set membership identification, after the unknown parameters are identified, the identified parametric model may be applied to determine one prediction for the output value of the system, together with probabilistic confidence intervals around the prediction (M.C.Campi, 1998). As we expect the confidence interval can accurately describe the actual probability that the future predictor will fall into the obtained interval. The future predictor is important for the next controller design and state estimation, so in (M.C.Campi, 2009), a novel approach for the construction of prediction models is proposed. The advantage of this novel approach is that instead of using a standard identification way, where one constructs a parametric model by minimizing an identification cost, and the identified model is used to design the prediction interval (J M Bravo, 2017). This novel approach directly considers interval model and applies measured data to ascertain the reliability of such interval predictor model (M.C.Campi, 2001). It means that we directly obtain the interval predictor

model from measured data and avoid the identification process for the parametric model structure.

In this paper, we continue to do a deep research for the construction of interval prediction model and then apply it in model predictive control strategy. The contribution of the paper is two fold. First, after introducing the family of models under study and some basic information about the interval predictor model, we present the computational results for the construction of interval predictor model, using linear regression structure, whose regression parameters are included in a sphere. Given a size measure to scale the average amplitude of the predictor interval, one optimal model that minimizes a size measure is efficiently computed by solving a linear programming problem, our first contribution is to apply the active set approach to solve this linear programming problem and then propose a Newton iterative form for the optimization variables. Based on these optimization variables, the predictor interval of the considered model with sphere parameter set can be directly constructed. Furthermore as for a fixed non-negative number coming from the size measure, we propose a good choice by using the Karush-Kuhn-Tucker (KKT) optimality conditions. Second, on the basis of this constructed interval predictor model, the midpoint of this interval predictor can be used as a central estimate, then the prediction of the output value can be obtained by that central estimate. As the goal of model predictive control is to control the system in order to track a desired output reference and reject disturbances. Moreover, the considered controller may enforce input and output constraints. So after introducing that central estimate corresponding to the prediction of the output value, such a control objective can be formulated as a quadratic programming problem with inequality constrained condition. It is well known that the dual of the quadratic programming problem is an unconstrained optimization problem (M Casini, 2017). After simple but tedious calculations, we formulate those

input and output constraints into a standard inequality form and give a detailed process about how to derive the dual of the quadratic programming problem. As the dual problem has a simple constraint set, so it is amenable to the use of the Gauss-Seidel algorithm, whose convergence can be shown, if the step-size parameter is chosen appropriately.

The paper is organized as follows. In section 2, some preliminaries are formulated about interval predictor model, providing a linear programming problem with respect to three variables. In section 3, one choice of a fixed non-negative number is proposed by using the optimality KKT sufficient and necessary condition. In section 4, Newton method is applied to solve that linear programming problem so that one interval predictor model is constructed. In section 5, the midpoint of the obtained interval predictor model is used in model predictive control strategy, where one quadratic programming problem with inequality constrained condition is needed to be solved. After reformulating that quadratic programming problem with inequality constrained condition and deriving its dual problem, the Gauss-Seidel algorithm is used to obtain the optimal control input and the convergence of the algorithm is also given. In section 6, simulation examples illustrate the effectiveness of the proposed interval predictor model. Section 7 ends the paper with final conclusion and points out the next topic, which we are studying in our next paper.

2. Interval predictor model

Interval predictor model returns an interval as output. The following concepts can be seen in (M.C.Campi, 2009) and (M.Vidyasagar, 2008). Define $\Phi \subseteq R^n$ and $Y \subseteq R$ be given sets, and they are denoted as the instance set and outcome set. The interval predictor model is a rule that assigns to each instance vector $\varphi \in \Phi$ a corresponding output interval. An interval predictor model is a set valued map.

$$I : \varphi \to I(\varphi) \subseteq Y$$

where φ is a regressor vector, $I(\varphi)$ is the predictor interval, also $I(\varphi)$ is called an informative interval. Consider the parametric model family M, the output of a system is expressed as $y = M(\varphi, q)$, for some parameters $q \in Q \subseteq R^{n_q}$. Through selecting a feasible set Q, an interval predictor model is obtained as the following relation.

$$M = \left\{ y = M(\varphi, q), q \in Q \subseteq R^{n_q} \right\}$$

In a dynamic setting, at each time instant the instance vector φ may contain past values of input and output measurements, then behaving as a linear regression function. From standard auto-regressive structures, a parametric interval predictor model is derived.

$$y(k) = \varphi^T(k)\theta(k) + e(k), |e(k)| \le \gamma \quad (3)$$

where $y(k)$ and $\varphi(k)$ denote the output measurement and regressor vector at time instant k, $\theta(k) \in R^n$ is the time varying unknown parameter, $e(k)$ is the external noise. But here any prior probability information of noise $e(k)$ is unknown. We only assume that noise $e(k)$ is unknown but bounded and γ is its magnitude bound.

Assume time varying unknown parameter $\theta(k) \in R^n$ satisfies.

$$\theta(k) \in \Delta \subseteq R^n \quad (4)$$

where Δ is one assigned bounded set. Here we assume Δ is a sphere with center θ and radius r.

$$\Delta = \left\{ \theta + \delta : \theta, \delta \in R^n, \|\delta\| \le r \right\} \quad (5)$$

Combining equations (2), (3) and (5), the parameters indicating the feasible set Q are the center θ and radius r of sphere Δ, and the magnitude bound γ on noise $e(k)$.

substituting (5) into (3), we obtain the output of the system.

$$y(k) = \varphi^T(k)(\theta + \delta) + e(k) = \varphi^T(k)\theta + \varphi^T(k)\delta + e(k)$$
$$(6)$$

Using the bounded radius r and magnitude bound γ, regression vector $\varphi(k)$, the output of the parametric model is one interval.

$$I(\varphi(k)) = \left\lfloor \varphi^T(k)\theta - (r\|\varphi(k)\| + \gamma), \varphi^T(k)\theta + (r\|\varphi(k)\| + \gamma) \right\rfloor$$
$$(7)$$

Equation (7) is one interval model, it contains the output of the parametric model $y(k)$ with some

guaranteed probability. When the observations are collected in the data sequence $D_N = \left\{ \varphi(k), y(k) \right\}_{k=1}^N$, whatever open or closed loop, the following relation holds.

$$y(k) \in I(\varphi(k)), for \ k = 1, 2 \cdots N$$

where equation (8) means that the interval $I(\varphi(k))$ is consistent with a given data sequence D_N. $\quad (2)$

Observing the interval (7) again, we see that this interval is dependent on three parameters- (θ, r, γ).

So if these three parameters are identified, the interval can be constructed based on equation (7). In order to obtain these three parameters, one linear programming problem is constructed (G.Calafiore, 2005).

Introducing a size measure $\mu = \gamma + ar$, where a is a fixed non-negative number, the optimal model that minimizes μ can be derived by solving a linear programming problem.

Theorem 1 (M.C.Campi,2009) Given an observed data sequence $D_N = \left\{ \varphi(k), y(k) \right\}_{k=1}^N$, a model order n, and a size objective $\mu = \gamma + ar$, three parameters used to construct the optimal interval predictor model are computed by solving the following linear programming problem with respect to three variables.

$$\theta \in R^n, \gamma, r$$

$$\begin{cases} \min_{\theta, \gamma, r} \quad \gamma + ar \\ subject \ to \ \varphi^T(k)\theta - r\|\varphi(k)\| - \gamma \le y(k) \\ -\varphi^T(k)\theta - r\|\varphi(k)\| - \gamma \le -y(k), k = 1, 2 \cdots N \end{cases}$$

According to linear programming problem (9), there are no any references on how to solve it, so here the main contributions of the next two sections are to solve this linear programming problem and choose an appropriate fixed non-negative number.

3. One choice of a fixed non-negative number

In linear programming problem (9), as the optimization variables r and γ denote the radius of sphere Δ and magnitude bound on noise $e(k)$, so these two optimization variables must satisfy that.

$$r \geq 0 \; and \; \gamma \geq 0$$

Combining linear programming problem (9) and inequality constraints (10), we rewrite the new linear programming problem as that.

$$\begin{cases} \min\limits_{\theta,\gamma,r} \quad \gamma + ar \\ subject \; to \; \varphi^T(k)\theta - r\|\varphi(k)\| - \gamma \leq y(k) \\ -\varphi^T(k)\theta - r\|\varphi(k)\| - \gamma \leq -y(k), k = 1,2\cdots N \\ r \geq 0 \; and \; \gamma \geq 0 \end{cases} \quad (11)$$

Define the Lagrangian function L corresponding to the above linear programming problem by

$$L\left(\theta,\gamma,r,\lambda_1,\lambda_2,\mu_k^+,\mu_k^-\right) = \gamma + ar - \lambda_1\gamma - \lambda_2 r - \sum_{k=1}^{N}\mu_k^+\left(y(k) - \varphi^T(k)\theta + r\|\varphi(k)\| + \gamma\right)$$
$$-\sum_{k=1}^{N}\mu_k^-\left(-y(k) + \varphi^T(k)\theta + r\|\varphi(k)\| + \gamma\right)$$
$$(12)$$

We refer to $\lambda_1, \lambda_2, \left\{\mu_k^+, \mu_k^-\right\}_{k=1}^{N}$ as the Lagrangian multipliers. By applying the optimality KKT sufficient and necessary condition on Lagrangian function, then some equality relations for the optimal solution hold.

$$\frac{\partial L}{\partial \theta} = \sum_{k=1}^{N}\left(\mu_k^+ - \mu_k^-\right)\varphi(k) = 0 \quad (13)$$

$$\frac{\partial L}{\partial \gamma} = 1 - \lambda_1 - \sum_{k=1}^{N}\left(\mu_k^+ + \mu_k^-\right) = 0$$

$$\frac{\partial L}{\partial r} = a - \lambda_2 - \sum_{k=1}^{N}\left(\mu_k^+ + \mu_k^-\right)\|\varphi(k)\| = 0 \quad (14)$$

$$(15)$$

$$\lambda_1\gamma = 0, \quad \lambda_2 r = 0$$

$$\mu_k^+\left(y(k) - \varphi^T(k)\theta + r\|\varphi(k)\| + \gamma\right) = 0 \quad (16)$$

$$\mu_k^-\left(-y(k) + \varphi^T(k)\theta + r\|\varphi(k)\| + \gamma\right) = 0 \quad (17)$$

$$(18)$$

Also as optimization variables r and γ denote the radius of sphere Δ and magnitude bound on noise $e(k)$ respectively, if $\gamma = 0$, then that $|e(k)| = 0$ means no noise exists in the standard auto-regressive structure (3). If $r = 0$, then sphere Δ reduces to its center θ, so here for interval predictor model, we want to satisfy that.

$$r > 0 \; and \; \gamma > 0 \quad (19)$$

When r and γ are all equal to zero, then $I(\varphi(k))$ is not an interval, but a fixed output value. The midpoint of the interval predictor model is a central estimate here.

$$I\left(\varphi(k)\right) = \frac{\varphi^T(k)\theta - \left(r\|\varphi(k)\| + \gamma\right) + \varphi^T(k)\theta + \left(r\|\varphi(k)\| + \gamma\right)}{2} = \varphi^T(k)\theta$$

Comparing (16) and (19), we see that equation (16) holds unless Lagrangian multipliers $\{\lambda_1, \lambda_2\}$ must

satisfy by that.   (10)

$$\lambda_1 = \lambda_2 = 0$$

Further in equation (13), assume regression vectors $\varphi(1), \varphi(2)\cdots\varphi(N)$ at different instants are linearly independent. So in order to let equation (13) hold, the $\left\{\mu_k^+, \mu_k^-\right\}_{k=1}^{N}$ need to satisfy by.

$$\mu_k^+ = \mu_k^- = \mu_k, k = 1,2\cdots N$$

substituting (20) and (21) into (14) and (15), we obtain the following simplified forms.

$$\begin{cases} \sum\limits_{k=1}^{N}\mu_k = \frac{1}{2} \\ \sum\limits_{k=1}^{N}\mu_k\|\varphi(k)\| = \frac{a}{2} \end{cases}$$

From the idea of equation (22), one choice of this fixed non-negative number a is given here. As $\sum\limits_{k=1}^{N}\mu_k = \frac{1}{2}$ holds, we set $\mu_k = \frac{1}{2N}$. Then after substituting $\mu_k = \frac{1}{2N}$ into equality $\sum\limits_{k=1}^{N}\mu_k\|\varphi(k)\| = \frac{a}{2}$, we have.

$$\frac{1}{2N}\left[\|\varphi(1)\| + \|\varphi(2)\| + \cdots + \|\varphi(N)\|\right] = \frac{a}{2}$$
$$(23)$$

It means that

$$a = \frac{1}{N}\left[\|\varphi(1)\| + \|\varphi(2)\| + \cdots + \|\varphi(N)\|\right]$$

Equation (24) is one of the choices of that fixed non-negative number a. As regressor vectors $\left\{\varphi(k)\right\}_{t=1}^{N}$ and the number of data are given, so the form (24) of a can be computed easily.

4. Newton method for interval predictor model

Because interval predictor model (7) is dependent of the linear programming problem (11) with respect to three kinds of optimization variables $\{\theta, \gamma, r\}$, so the important step in constructing interval predictor model (7) is to solve that linear programming problem (11). First we rewrite the linear programming problem (11) as its standard form. Define a new vector $x \in R^{n+2}$ as $x = \{\theta, \gamma, r\}$. Based on the new optimization vector $x \in R^{n+2}$, the cost function in (11) can be rewritten as.

$$\gamma + ar = \begin{pmatrix} 0 & 1 & a \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \\ r \end{pmatrix} = C^T x$$

$$C^T = \begin{pmatrix} 0 & 1 & a \end{pmatrix} \quad (25)$$

Also each inequality can be rewritten as.

$$\begin{cases} \varphi^T(k)\theta - r\|\varphi(k)\| - \gamma \le y(k) \\ \Leftrightarrow \begin{pmatrix} \varphi^T(k) & -1 & -\|\varphi(k)\| \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \\ r \end{pmatrix} \le y(k), \\ -\varphi^T(k)\theta - r\|\varphi(k)\| - \gamma \le -y(k) \\ \Leftrightarrow \begin{pmatrix} -\varphi^T(k) & -1 & -\|\varphi(k)\| \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \\ r \end{pmatrix} \le -y(k), \quad (26) \\ r \ge 0 \Leftrightarrow \begin{pmatrix} 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \\ r \end{pmatrix} \le 0, \quad \gamma \ge 0 \\ \Leftrightarrow \begin{pmatrix} 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} \theta \\ \gamma \\ r \end{pmatrix} \le 0 \end{cases}$$

Define some matrices to merge all inequities in equation (26).

$$A = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ \varphi^T(1) & -1 & -\|\varphi(1)\| \\ \vdots & \vdots & \vdots \\ \varphi^T(N) & -1 & -\|\varphi(N)\| \\ -\varphi^T(1) & -1 & -\|\varphi(1)\| \\ \vdots & \vdots & \vdots \\ -\varphi^T(N) & -1 & -\|\varphi(N)\| \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ y(1) \\ \vdots \\ y(N) \\ -y(1) \\ \vdots \\ -y(N) \end{bmatrix}$$

(27)

where $A \in R^{(2N+2)\times 3}, B \in R^{(2N+2)\times 1}$. Given regression vector $\{\varphi(k)\}_{k=1}^{N}$ and output measured data $\{y(k)\}_{k=1}^{N}$, the above two matrices A, B are known.

Applying two matrices A, B, all inequities in equation (26) are obtained in a simplified form.

$$Ax \le B \quad (28)$$

Then based on (25) and (28), the formal linear programming problem (11) can be formulated into a standard linear programming form.

$$\begin{cases} \min_{x} & C^T x \\ \text{subject to } Ax \le B & (29) \end{cases}$$

Now we propose a Newton method to solve the above standard linear programming problem. Our

Newton method introduces the active set approach into classical Newton approach. The active set approach introduced here is based on a transformation by means of which the optimality KKT conditions are converted into a system of nonlinear equations.

Firstly for a fixed scalar $c > 0$, consider the open set $S_c^* \subset R^{2N+2} \times R^{2N+2}$ defined by.

$$S_c^* = \{(x,m)/m_j + cA_j x, j = 1, 2 \cdots 2N+2\}$$

(30)

where A_j is the j column of matrix A, and the system of equations on S_c^*.

$$\begin{cases} C + \nabla_x g^+(x,m,c)m = 0 \\ g^+(x,m,c) = 0 \end{cases}$$

where the function $g^+(x,m,c)$ is defined by.

$$g^+(x,m,c) = \begin{bmatrix} g_1^+(x,m_1,c) \\ \vdots \\ g_{2N+2}^+(x,m_{2N+2},c) \end{bmatrix}$$

$$g_j^+(x,m_j,c) = \max\left\{ A_j x, -\frac{m_j}{c} \right\}, j = 1, 2 \cdots 2N+2$$

(32)

Note that $g^+(x,m,c)$ is differentiable on S_c^* as $Ax - B$, so equations (31), (32) are well defined. $g^+(x,m,c)$ appears in the definition of the augmented Lagrangian function, which takes the form as that.

$$L_c(x,m,c) = C^T x + m'g^+(x,m,c) + \frac{1}{2}c\left|g^+(x,m,c)\right|^2$$

Secondly consider the implementation of Newton approach. Define for $(x.m) \in S_c^*$.

$$L^+(x,m,c) = C^T x + m'g^+(x,m,c)$$

$$A_c(x,m) = \left\{ j / A_j x - B_j > -\frac{m_j}{c}, j = 1, 2 \cdots 2N+2 \right\}$$

Assume without loss of generality that $A_c(x,m) = \{1 \cdots p\}$ for one integer p. This integer p depends on x and m. We view $A_c(x,m)$ as the active index set, in the sense that indices in $A_c(x,m)$ are predicted by the algorithm to be active at the solution. By differentiation in equation (31), we propose that Newton method consists of the iteration.

$$\bar{x} = x + \Delta x, \quad \bar{m} = m + \Delta m$$

where $(\Delta x, \Delta m)$ is the solution of the following system.

$$\begin{bmatrix} \nabla^2_{xx}\{m'g^+(x,m,c)\} & N(x,m,c) & 0 \\ N(x,m,c) & 0 & 0 \\ 0 & 0 & -\dfrac{1}{c}I \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta m_1 \\ \mathrm{M} \\ \Delta m_p \\ \Delta m_{p+1} \\ \mathrm{M} \\ \Delta m_{2N+2} \end{bmatrix} =$$

$$-\begin{bmatrix} C+\nabla_x\{m'g^+(x,m,c)\} \\ g_1^+(x,m,c) \\ \mathrm{M} \\ g_p^+(x,m,c) \\ g_{p+1}^+(x,m,c) \\ \mathrm{M} \\ g_{2N+2}^+(x,m,c) \end{bmatrix}$$

(35)

where $N(x,m,c)$ is the $(2N+2)\times p$ matrix having as columns the gradients $A_j, j \in A_c(x,m)$, I is the $(2N+2-p)\times(2N+2-p)$ identity matrix, and the zero matrices have appropriate dimension. Since we see that.

$$g_j^+(x,m,c) = -\frac{m_j}{c}, \forall j \notin A_c(x,m)$$

It follows that

$$\bar{m}_j = 0, \forall j \notin A_c(x,m)$$

It follows from equation (35) that the remaining variables Δx and $\Delta m_1 \cdots \Delta m_p$ are obtained by solving the reduced system.

$$\begin{bmatrix} \nabla^2_{xx}\{m'g^+(x,m,c)\} & N(x,m,c) \\ N(x,m,c) & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta m_1 \\ \mathrm{M} \\ \Delta m_p \end{bmatrix} =$$

$$-\begin{bmatrix} C+\nabla_x\{m'g^+(x,m,c)\} \\ A_1 x - B_1 \\ \mathrm{M} \\ A_p x - B_p \end{bmatrix}$$

(36)

where we make use of the fact that.

$$g_j^+(x,m,c) = g_j(x), \forall j \in A_c(x,m)$$

$$\nabla_x L^+(x,m,c) = C + \sum_{j \in A_c(x,m)} (A_j - B_j)m_j \quad (37)$$

From above equations, we see that the proposed Newton iteration can be described in a simpler manner.

5. Application interval predictor model into model predictive control

Now in this section, we start to apply the above obtained interval predictor model into model predictive control strategy. Set the optimal variables corresponding to the linear programming problem (9) as follows.

$$\hat{\theta}, \hat{\gamma}, \hat{r}$$

Then based on these three optimal variables $(\hat{\theta}, \hat{\gamma}, \hat{r})$, the interval predictor model $I(\varphi(k))$ is defined as that.

$$I(\varphi(k)) = \left[\varphi^T(k)\hat{\theta} - (\hat{r}\|\varphi(k)\| + \hat{\gamma}), \ \varphi^T(k)\hat{\theta} + (\hat{r}\|\varphi(k)\| + \hat{\gamma})\right]$$

(38)

Then from probability theory (Carlo Novara, 2016), then predictor of output value will be included in this confidence interval with one guaranteed accuracy, i.e.

$$y(k) \in I(\varphi(k)), \forall k = 0,1,2\cdots N$$

(39)

5.1 Model predictive control problem

As the goal of model predictive control is to control the system in order to track a desired output reference and reject disturbances from $k = 0$ up to some finite time step N, where this time step N can be very large.

Assuming that the control input $u(l), l = -1, -2, \cdots$ are known, then a control objective can be formalized by one optimization problem.

$$\min_{u(0),u(1)\cdots u(N)} \sum_{k=0}^{N} \left[y(k) - y_{des}(k)\right]^T Q_1\left[y(k) - y_{des}(k)\right] + u^T(k)S_1u(k)$$

(40)

where $y_{des}(k)$ is the desired output reference, Q_1, S_1 are positive semi-definite weighting matrices selected by the designer. Observing (39) and (40), the predictor of output value $y(k), k = 0,1,2\cdots N$ exists in cost function (40), but the only knowledge about the predictor of output value is that $y(k)$ is included in confidence interval $I(\varphi(k))$, so there are two cases needed to be

considered here. One case is to expand the cost function (40), then a numerical value about $y(k)$ is needed to substitute into cost function, i.e. the midpoint of that confidence interval can be used as the prediction of output value $y(k)$.

$$I(\varphi(k)) = \frac{\varphi^T(k)\hat{\theta} - (\hat{r}\|\varphi(k)\| + \hat{\gamma}) + \varphi^T(k)\hat{\theta} + (\hat{r}\|\varphi(k)\| + \hat{\gamma})}{2} = \varphi^T(k)\hat{\theta}$$

(41)

Replacing $y(k)$ as $\varphi^T(k)\hat{\theta}$ into cost function (40), we obtain an explicit form corresponding to optimization problem.

$$\min_{u(0),u(1)\cdots u(N)} \sum_{k=0}^{N} \left[\varphi^T(k)\hat{\theta} - y_{des}(k)\right]^T Q_1 \left[\varphi^T(k)\hat{\theta} - y_{des}(k)\right] + u^T(k)S_1 u(k)$$

(42)

The second case is to use confidence interval $I(\varphi(k))$ into optimization problem directly, but another max operation is added, i.e. a robust model predictive control problem is obtained.

$$\min_{u(0),u(1)\cdots u(N)} \max_{y(k)\in I(\varphi(k))} \sum_{k=0}^{N} \left[y(k) - y_{des}(k)\right]^T Q_1 \left[y(k) - y_{des}(k)\right] + u^T(k)S_1 u(k)$$

(43)

Here in this paper, we only consider the first case (41) and the second case (42) corresponding to a robust model predictive control problem will be solved later. Further the controller will satisfy input and output constraints as that.

$$\begin{cases} u_{\min} \le u(i) \le u_{\max} \\ y_{\min} \le y(i) \le y_{\max} \end{cases} \quad i = 0,1\cdots N \qquad (44)$$

Combining (42) and (44), the model prediction control problem is formalized by the following optimization problem with inequality constrained condition.

$$\min_{u(0),u(1)\cdots u(N)} \sum_{k=0}^{N} \left[\varphi^T(k)\hat{\theta} - y_{des}(k)\right]^T Q_1 \left[\varphi^T(k)\hat{\theta} - y_{des}(k)\right] + u^T(k)S_1 u(k)$$

$$subject\ to \begin{cases} u_{\min} \le u(i) \le u_{\max} \\ y_{\min} \le y(i) \le y_{\max} \end{cases} \quad i = 0,1\cdots N$$

(45)

where $u_{\min}, u_{\max}$ and $y_{\min}, y_{\max}$ denote the lower and upper bounds on input and output respectively. Before solving optimization problem (45), we analyze the cost function and inequality constrained condition in (45). The second term in cost function can be rewritten as.

$$\sum_{k=0}^{N} u^T(k)S_1 u(k) = \begin{bmatrix} u(0) & u(1) & \cdots & u(N) \end{bmatrix} \begin{bmatrix} S_1 & 0 & \cdots & 0 \\ 0 & S_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & S_1 \end{bmatrix} \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(N) \end{bmatrix} = u^T S u$$

(46)

where vector u and matrix S are defined as that.

$$u = \begin{bmatrix} u(0) & u(1) & \cdots & u(N) \end{bmatrix}^T, S = diag\begin{pmatrix} S_1 & S_1 & \cdots & S_1 \end{pmatrix}$$

It means that vector u includes all optimal input at all time step $i, i = 0,1\cdots N$. The first term in cost function can be also reformulated as that.

$$\min_{u(0),u(1)\cdots u(N)} \sum_{k=0}^{N} \left[\varphi^T(k)\hat{\theta} - y_{des}(k)\right]^T Q_1 \left[\varphi^T(k)\hat{\theta} - y_{des}(k)\right]$$

$$= \sum_{k=0}^{N} \varphi^T(k)\hat{\theta} Q_1 \varphi^T(k)\hat{\theta} - 2\varphi^T(k)\hat{\theta} Q_1 y_{des}(k) + \sum_{k=0}^{N} y_{des}(k)^T Q_1 y_{des}(k)$$

As no input $u(i)$ exists in term $\sum_{k=0}^{N} y_{des}(k)^T Q_1 y_{des}(k)$, then this term can be omitted. For clarity of presentation, set the control inputs $u(l), l = -1,-2,\cdots$ be zero, it not, we use coordinate transformation to satisfy it. Due to $\hat{\theta} \in R^n$, then regressor vector $\varphi(k)$ and $\hat{\theta}$ can be defined as follows.

$$\begin{cases} \varphi(k) = \begin{bmatrix} u(k-1) & u(k-2) & \cdots & u(k-N) \end{bmatrix}^T \\ \hat{\theta} = \begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_N \end{bmatrix}^T \end{cases}$$

From above descriptions, the following relations hold.

$$\varphi^T(0)\hat{\theta} = \begin{bmatrix} u(-1) & u(-2) & \cdots & u(-N) \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_N \end{bmatrix} = 0$$

$$\varphi^T(1)\hat{\theta} = \begin{bmatrix} u(0) & u(-1) & \cdots & u(1-N) \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_N \end{bmatrix} = \begin{bmatrix} u(0) & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_N \end{bmatrix}$$

$$= \begin{bmatrix} u(0) & u(1) & \cdots & u(N) \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_N \end{bmatrix} = u^T I_1 \hat{\theta}$$

$$\varphi^T(2)\hat{\theta}=\begin{bmatrix}u(1)&u(0)&\cdots&0\end{bmatrix}\begin{bmatrix}\hat{\theta}_1\\\hat{\theta}_2\\\vdots\\\hat{\theta}_N\end{bmatrix}=\begin{bmatrix}u(0)&u(1)&\cdots&u(N)\end{bmatrix}\begin{bmatrix}0&1&\cdots&0\\1&0&\cdots&0\\\vdots&&\vdots\\0&0&\cdots&0\end{bmatrix}\begin{bmatrix}\hat{\theta}_1\\\hat{\theta}_2\\\vdots\\\hat{\theta}_N\end{bmatrix}=u^T I_2 \hat{\theta}$$

$$\vdots$$

$$\varphi^T(N)\hat{\theta}=\begin{bmatrix}u(N-1)&u(N-2)&\cdots&u(N-n)\end{bmatrix}\begin{bmatrix}\hat{\theta}_1\\\hat{\theta}_2\\\vdots\\\hat{\theta}_N\end{bmatrix}=\begin{bmatrix}u(0)&u(1)&\cdots&u(N)\end{bmatrix}\begin{bmatrix}0&0&\cdots&1&0\\0&0&\cdots&0&0\\\vdots&&\vdots\\0&1&\cdots&0&0\\1&0&\cdots&0&0\\0&0&\cdots&0&0\end{bmatrix}\begin{bmatrix}\hat{\theta}_1\\\hat{\theta}_2\\\vdots\\\hat{\theta}_N\end{bmatrix}=u^T I_N \hat{\theta}$$

substituting above relations into the first term of the cost function, we obtain.

$$\sum_{k=0}^{N}\varphi^T(k)\hat{\theta}Q_1\varphi^T(k)\hat{\theta}-2\varphi^T(k)\hat{\theta}Q_1 y_{des}(k)=\sum_{k=0}^{N}u^T\left[I_k\hat{\theta}Q_1 I_k^T\right]u+u^T S u-2\left[I_k Q_1 y_{des}(k)\right]u$$

(47)

Set

$$\sum_{k=0}^{N}\left[I_k\hat{\theta}Q_1 I_k^T+S\right]=\frac{1}{2}Q,\quad 2\left[I_k Q_1 y_{des}(k)\right]=b$$

Then the cost function in optimization problem (45) can be simplified as that.

$$\min_u \frac{1}{2}u^T Q u - bu \qquad (48)$$

Those inequality constraints on input and output can be also reformulated as that.

$$u(i)\le u_{max},i=0,1L\ N$$

$$\Rightarrow\begin{bmatrix}1&0&L&0\\0&1&L&0\\M&&M\\0&0&L&1\end{bmatrix}\begin{bmatrix}u(0)&u(-1)&L&u(1-N)\end{bmatrix}\le\begin{bmatrix}u_{max}\\u_{max}\\M\\u_{max}\end{bmatrix}$$

$$\Rightarrow Iu\le U_{max}$$

$$-u(i)\le u_{min},i=0,1L\ N$$

$$\Rightarrow-\begin{bmatrix}1&0&L&0\\0&1&L&0\\M&&M\\0&0&L&1\end{bmatrix}\begin{bmatrix}u(0)&u(-1)&L&u(1-N)\end{bmatrix}\le\begin{bmatrix}-u_{min}\\-u_{min}\\M\\-u_{min}\end{bmatrix}$$

$$\Rightarrow-Iu\le-U_{min}$$

$$y(i)\le y_{max},i=0,1L\ N\Rightarrow\varphi^T(i)\theta\le y_{max}$$

$$\Rightarrow\theta^T\underbrace{\begin{bmatrix}I_1&0&L&0\\0&I_2&L&0\\M&&M\\0&0&L&I_N\end{bmatrix}}_{I_\Delta}u\le\begin{bmatrix}y_{max}\\y_{max}\\M\\y_{max}\end{bmatrix}\Rightarrow I_\Delta u\le Y_{max}$$

Similarly we have that.

$$-y(i)\le-y_{min},i=0,1\cdots N\Rightarrow-I_\Delta u\le-Y_{min}$$

where I denotes identify matrix, $I_i,i=1,2\cdots N$ is from cost function , I_Δ is defined as.

$$I_\Delta=diag\begin{bmatrix}I_1&I_2&\cdots&I_N\end{bmatrix}\theta$$

For notational clarity, combining above four inequalities to obtain.

$$\begin{bmatrix}I\\-I\\I_\Delta\\-I_\Delta\end{bmatrix}u\le\begin{bmatrix}U_{max}\\-U_{min}\\Y_{max}\\-Y_{min}\end{bmatrix}\Rightarrow Eu\le F$$

with E and F labeled underneath.

Merging (48) and (49), the following quadratic programming problem is considered.

$$\min_u \frac{1}{2}u^T Q u - bu$$
$$subject\ to\ Eu\le F$$

(50)

In order to obtain the optimal control input u by solving above quadratic programming problem (50), the following Gauss-Seidel algorithm is used.

5.2 Gauss-Seidel algorithm

To solve that quadratic programming problem (50) with inequality constraint, as the dual of the quadratic programming problem is an unconstrained optimization problem (J M Bravo, 2006), so it is easy to solve its dual problem.

The dual function is defined as that.

$$q(z)=\inf_u\left\{\frac{1}{2}u^T Q u - bu + z(Eu-F)\right\}$$

The infimum is attained for $u=Q^{-1}(b-Ez)$ and after substitution of this expression in the preceding relation for $q(z)$, a straightforward calculation yields.

$$q(z)=\frac{1}{2}Q^{-1}(b-Ez)^T QQ^{-1}(b-Ez)$$
$$=\frac{1}{2}(b-Ez)^T Q^{-1}(b-Ez)-bQ^{-1}(b-Ez)+z\left[EQ^{-1}(b-Ez)-F\right]$$
$$=-\frac{1}{2}z^T EQ^{-1}E^T z - zF + zEQ^{-1}b$$

(51)

Then the dual of the quadratic programming problem is given by

$$\min_z \frac{1}{2}z^T G z + dz$$
$$subject\ to\quad z\ge 0$$

where

$$G=EQ^{-1}E^T,\quad d=f-EQ^{-1}b$$

(53)

If z^* solve the dual problem (52), then $u^*=Q^{-1}(b-Ez^*)$ solves the primal problem (T

Alamo, 2005). Let g_j denote the j th column of E, assume that g_j is nonzero for all j. Since Q is symmetric and positive define, the first partial derivative of the dual cost function with respect to z_j is given by.

$$d_j + \sum_{k=1}^{N} g_{jk} z_k \qquad (54)$$

where g_{jk} and d_j are the corresponding elements of the matrix G and the vector d, respectively. Set the derivative to be zero, the unconstrained minimum of the dual cost function along the j th coordinate starting from z is attained at $\tilde{z}_j$ given by.

$$\tilde{z}_j = -\frac{1}{g_{jj}}\left(d_j + \sum_{k \neq j} g_{jk} z_k\right) = z_j - \frac{1}{g_{jj}}\left(d_j + \sum_{k=1} g_{jk} z_k\right)$$

Taking into account the non-negativity constraint $z_j \geq 0$, when the j th coordinate is updated, the Gauss-Seidel iteration has the form.

$$\begin{cases} z_j = \max\{0, \tilde{z}_j\} = \max\left\{0, z_j - \frac{1}{g_{jj}}\left(d_j + \sum_{k=1} g_{jk} z_k\right)\right\} \\ z_i = z_i, \quad \forall i \neq j \end{cases}$$
$$(55)$$

Consider a linear projection Jacobi method, it is a special case of the Gauss-Seidel algorithm. Taking into account the first partial derivation of the dual cost function with respect to z_j, another iteration formula is given by.

$$z_j(t+1) = \max\left\{0, z_j(t) - \frac{\eta}{g_{jj}}\left(d_j + \sum_{k=1} g_{jk} z_k(t)\right)\right\}$$
$$(56)$$

where $\eta > 0$ is the stepsize parameter. The above iteration is more suitable for parallelization than the Gauss-Seidel iteration. On the other hand, for convergence, the stepsize η should be chosen sufficiently small, and some experimentations may be needed to obtain the appropriate range for η. In the dual problem (52), let k_G be the largest eigenvalue of G, and assume that $k_G > 0$, then the Lipschitz condition corresponding to the dual function $q(z) = \frac{1}{2} z^T G z + dz$ is written as.

$$\|\nabla q(z_1) - \nabla q(z_2)\| \leq k_G \|z_1 - z_2\|$$

or equivalently

$$\|G(z_1 - z_2)\| \leq k_G \|z_1 - z_2\|$$

Then from the parallel distribution algorithm (T

Alamo, 2005), we obtain the following theorem 2 easily.

Theorem 2 (Convergence of the Gauss-Seidel algorithm) Suppose that dual function $q(z)$ satisfies the Lipchitz condition, if $0 < \eta < \frac{1}{k_G} g_{jj}$ and if z^* is a limit point of the sequence $\{z(t)\}$ generated by the Gauss-Seidel algorithm, then $(z - z^*)\nabla q(z^*) \geq 0$ for all $z \geq 0$. Moreover if $q(z)$ is convex on the set $\{z / z \geq 0\}$, then z^* minimizes $q(z)$ over the set $\{z / z \geq 0\}$

6. Simulation examples

Now we propose a simulation example to illustrate the nature of the above results. In this simulation example, the unknown regressive structure is assumed as follows.

$$y(k) = 0.1u(k-1) + 0.2u(k-2) + e(k) =$$
$$\left(u(k-1) \quad u(k-2)\right)\begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix} + e(k)$$

Setting regression vector as
$$\varphi^T(k) = \left(u(k-1) \quad u(k-2)\right)$$
Seeking an explanatory interval predictor model of the form.

$$y(k) = \varphi^T(k)\theta(k) + e(k), |e(k)| \leq \gamma$$

In order to fit this above interval predictor model to the measured data, we choose $u(k) = \cos(k)$ and collect $N = 500$ observations as the data sequence $D_N = \{\varphi(k), y(k)\}_{k=1}^{N}$. Choosing that size measurement as $\mu = \gamma + 0.8r$, and solving the linear programming problem by Newton method on the basis of our measured data $D_N = \{\varphi(k), y(k)\}_{k=1}^{N}$, we obtain one optimal center $\theta = \begin{pmatrix} 0.1 \\ 0.18 \end{pmatrix}$ with bounded radius $r = 0.2$, and level of magnitude bound $\gamma = 0.1$. The resulting iterative estimations of the unknown parameter is shown in Figure 1, with the iterative estimations are also clustered around at the point $\theta = \begin{pmatrix} 0.1 \\ 0.18 \end{pmatrix}$.

In Figure 1, the center of the circle is the optimal or true value, denoted by red triangle. The black

triangles are denoted as iterative estimations, obtained by Newton method. From Figure.1, we see that the iterative estimations will converge to its optimal or true value with increasing iteration steps. After substituting the optimal center into equation (7), our interval predictor model for output value is obtained. This interval predictor model contains the output of the parametric model with some guaranteed probability. The whole output frequency response curves are showed in Figure 2, based on estimated model parameters. The red curve is the actual true amplitude curve from Bode plot tool. When the estimated parameters are contained in the uncertainty bound with probability level 0.99, the amplitude curves lie above or low the red curve. From Figure 2, we see these three curves are very close and the red amplitude curve lies between two confidence amplitude curves with probability level 0.99.

As using Matlab simulation tool to simulate the output response of Bode plot in closed loop, the phase plot is get with amplitude plot simultaneously. The confidence interval phase plot is given in Figure 3, and the red phase curve lies also between two confidence phase curves with the probability level 0.99.This is similar to the derivation of Figure 3.The output error between true output value and actual output value will converge to zero with time increases in Figure.4, then this model error can be neglected in our next process in designing controllers.
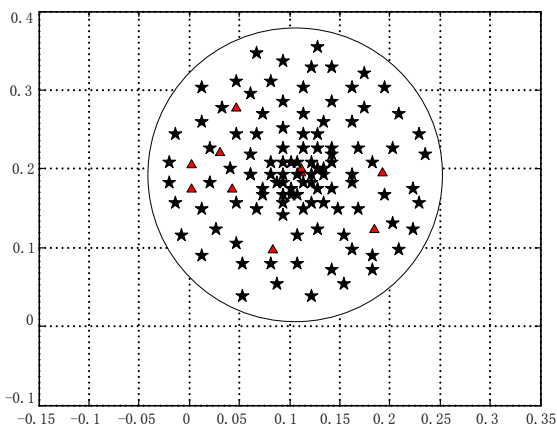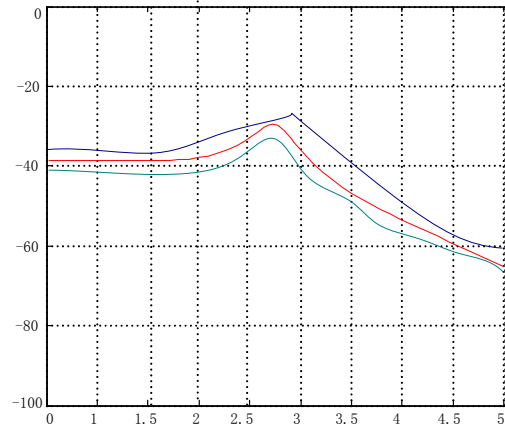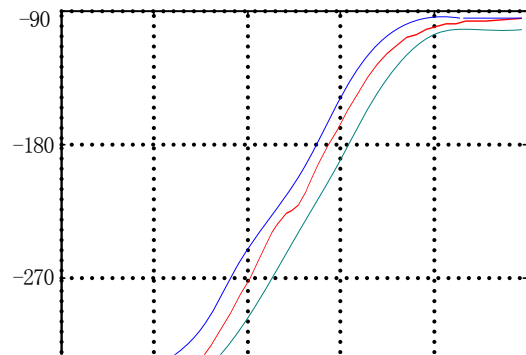

Figure. 1. Iterative parameter

Figure. 2 Confidence interval of amplitude in Bode plot
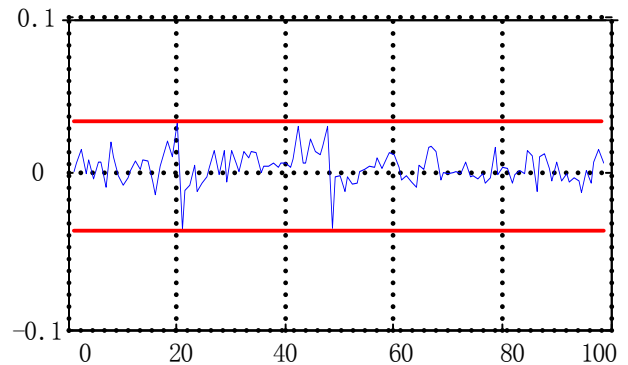

Figu

Figure.4. One bound of the model error

7. Conclusion

In this paper, we apply interval predictor model into model predictive control strategy. As the interval predictor model is dependent of three unknown parameters, one linear programming problem is constructed to obtain this interval predictor model. We propose to choose one fixed non-negative number in size measure and give a Newton method

to solve that linear programming problem. Further in model predictive control, the midpoint of the obtained interval is used in one quadratic programming problem. After some tedious calculations, the dual problem of that quadratic programming problem is solved by Gauss-Seidel algorithm. But the midpoint of the obtained interval is a special case of confidence interval, as here the prediction of output value is dependent only on one identified parameter. Generally the confidence interval needs to construct the robust model predictive control, i.e. the problem of applying dynamic programming to solve a robust model predictive control is our next topic.

Acknowledgements

Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

References

[1] M C Campi, G Calafiore. "Interval predictor models: identification and reliability," Automatica, vol 45, no.2, pp. 382-393, 2009.

[2] M C Campi, M Vidyasagar. "Learning with prior information," IEEE Transaction on Automatic Control, vol 46, no.11, pp. 1682-1694, 2001.

[3] M Vidyasagar, Rajeeva L Karandikar. "A learning theory approach to system identification and stochastic adaptive control," Journal of Process Control, vol 18, no.3, pp. 421-430, 2008.

[4] M C Campi, P R Kumar. "Learning dynamical systems in a stationary environment," Systems & Control Letters, vol 34, no.3, pp. 125-132, 1998.

[5] G Calafiore, M C Campi. "Uncertain convex programs: randomized solutions and confidence levels," Mathematical Programming, vol 102,
no.11, pp. 25-46, 2005.

[6] M Milanese, C Novara. "Set membership identification of nonlinear systems," Automatica, vol 40, no.6, pp. 957-975, 2004.

[7] T Alamo, J M Bravo, E F Camacho. "Guaranteed state estimation by zonotopes," Automatica, vol 41, no.6, pp. 1035-1043, 2005.

[8] J M Bravo, T Alamo, E F Camacho. "Bounded error identification of systems with time varying parameters," IEEE Transaction on Automatic Control, vol 51, no.7, pp. 1144-1150, 2006.

[9] J M Bravo, A Suarez, M Vasallo. "Slide window bounded error tome varying systems identification," IEEE Transaction on Automatic Control, vol 61, no.8, pp. 2282-2287, 2016.

[10] J M Bravo, T Alamo, M Vasallo. "A general framework for predictions based on bounding techniques and local approximation," IEEE Transaction on Automatic Control, vol 62, no.7, pp. 3430-3435, 2017.

[11] Marko Tanaskovic, Lorenzo Fagiano, Roy Smith. "Adaptive receding horizon control for constrained MIMO systems," Automatica, vol 50, no.12, pp. 3019-3029, 2014.

[12] Marko Tanaskovic, Lorenzo Fagiano, Carlo Novara. "Data driven control of nonlinear systems: an on line direct approach," Automatica, vol 75, no.1, pp. 1-10, 2017.

[13] Carlo Novara, Simone Formentin, Sergio M Savaresi. "Data driven design of two degrees of freedom nonlinear controllers," Automatica, vol 72, no.10, pp. 19-27, 2016.

[14] M Casini, A Garulli, A Vicino. "Feasible parameter set approximation for linear models with bounded uncertain regressions," IEEE Transaction on Automatic Control, vol 59, no.11, pp. 2910-2920, 2014.

[15] Vito Cerone, Jean Bernard Lasserre, Dario Piga. "A unified framework for solving a general class of conditional and robust set membership estimation problems," IEEE Transaction on Automatic Control, vol 59, no.11, pp. 2897-2909, 2014.

[16] M Casini, A Garulli, A Vicino. "A linear

programming approach to online set membership parameter estimation for linear regression models," International Journal of Adaptive Control and Signal Processing, vol 31, no.3, pp. 360-378, 2017.

[17] X Zhang, M Kamgarpour, A Georghiou. "Robust optimal control with adjustable uncertainty sets," Automatica, vol 75, no.1, pp. 249-259, 2017.