Sliding Mode Control under State and Control Constraints

Hirohumi Tanizawa and Yuzo Ohta

Abstract—In this paper, we propose a method of continuous-time sliding mode control (SMC) system design so as not to violate state and control constraints. To circumvent chattering problem, we use continuous control law. And, we employ the inner approximation of maximal admissible set for a nonlinear continuous-time system to guarantee the satisfaction of constraints. This approach has the advantages that it is robust over initial state error and that save the on-line computing time required to inclusion check of maximal admissible sets. Moreover, we propose a control strategy of switching “sliding” hyperplanes to achieve better performance.

I. INTRODUCTION

For almost all practical control systems, we need to take into account the existence of constraints on state and/or control input caused by amplitude limitation of state variables and saturation property of actuators. If we ignore these constraints, the real performance of the system degrade because of the wind-up phenomena, or in worst cases the control system become unstable. In these respect, many researchers have studied constrained systems [1]-[10]. In [1], the reference shaping (RS) method was proposed and could give good response characteristics. But, this method is not robust over initial state error because it solves an optimization problem which depends on initial state. On the other hand, the method using reference governor (RG) is robust over initial state error since it is a on-line method. Moreover, in [10] it was proposed to using outer LQ feedback to improve the performance.

We note that both the RS method and the method using outer LQ feedback force system variables to reach a hyper plane corresponding to the output as fast as possible. Then, it is natural to apply the sliding mode control (SMC) to this kind of control problem, and we face the issue of treating constraints in the context of SMC.

So far, in this paper, we consider SMC of systems with state and control constraints. This approach has the advantages that it is robust over initial state error and that save the on-line computing time required to inclusion check of maximal admissible sets (MASs). The main issue of this method is take into account constraints. To approach this issue, we employ the inner approximation of maximal output admissible set (MAS) concept for a nonlinear continuous-time system [11]. Moreover, we propose a control strategy of switching “sliding” hyperplanes to achieve better performance.

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C), 2005-2006, 15503777.

H. Tanizawa is with the Graduate School of Science and Technology, Kobe University, 1-1 Rokkodai Nada, Kobe 657-8501, JAPAN choro@cs23.cs.kobe-u.ac.jp

Y. Ohta is with the Department of Computer and Systems, Faculty of Engineering, Kobe University, 1-1 Rokkodai Nada, Kobe 657-8501, JAPAN ohta@cs.kobe-u.ac.jp

II. SYSTEM DESCRIPTION

Consider the single-input single-output linear system Σ

\[ \dot{x} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} x + \begin{bmatrix} 0_{n-1} \\ B_2 \end{bmatrix} u, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \]

\[ y = Cx, \]

\[ \sigma = Gx, \]

where \( x \in \mathbb{R}^n \) is the system state, \( u \in \mathbb{R} \) is the control input, \( y \in \mathbb{R} \) is the output, \( \sigma \) is the switching function and \( A_{11} \in \mathbb{R}^{(n-1)\times(n-1)}, A_{12} \in \mathbb{R}^{n-1}, A_{21} \in \mathbb{R}^{1\times(n-1)}, A_{22} \in \mathbb{R}, B_2 \in \mathbb{R}, C, G \in \mathbb{R}^{1\times n}, X_0 \subseteq \mathbb{R}^n \) is the set in which initial state belongs. We assume that \((A, B)\) is controllable.

The constraints are given by

\[ M_x x \leq m_x \]

\[ M_u u \leq m_u \]

where \( M_x \in \mathbb{R}^{p_x \times n}, m_x \in \mathbb{R}^{p_x}, m_x > 0, M_u \in \mathbb{R}^{p_u}, m_u \in \mathbb{R}^{p_u}, m_u > 0 \), and \( p_x \) and \( p_u \) are numbers of constraints for state and control input, respectively.

In this paper, we consider the regulation problem, and our aim is to control the system so that \( y(t) \to 0 \) as fast as possible.

III. CONTROL LAW

The control law is given by

\[ u = F_L x - \frac{\kappa}{GB} \text{sat}(\sigma/\varepsilon), \]

\[ F_L = -GA/(GB), \]

\[ \text{sat}(\sigma/\varepsilon) = \begin{cases} 1 & \text{if } \sigma > \varepsilon, \\ \sigma/\varepsilon & \text{if } |\sigma| \leq \varepsilon, \\ -1 & \text{if } \sigma < -\varepsilon, \end{cases} \]

where \( \varepsilon \) is a positive constant.

Since the control law (4) is continuous control law, the chattering is eliminated, and, hence, in the exact sense, the sliding mode does not occur. Therefore, we apply explicitly a linear feedback \( F_L x = -GA/(GB)x \), which is usually called as the equivalent control law.

Suppose that

\[ G = \alpha \begin{bmatrix} g & 1 \end{bmatrix}, \]
and let us consider a transformation given by
\[
\bar{x} = \begin{bmatrix}
\bar{x}_1 \\
\bar{x}_2
\end{bmatrix} = Tx, \quad T = \begin{bmatrix}
I & 0 \\
g & 1
\end{bmatrix},
\]
so that \(\bar{x}_1 = x_1\) and \(\bar{x}_2 = \sigma\). Then, we have
\[
\dot{x} = TAx + TBF_Lx - \frac{\kappa}{GB}TBs\text{at}(\sigma/\varepsilon)
\]
\[
= \begin{bmatrix}
A_{11} - A_{12g} & A_{12} \\
0 & 0
\end{bmatrix} \bar{x} + \begin{bmatrix} 0 \\
-\kappa/\alpha \end{bmatrix} \text{sat}(\sigma/\varepsilon) \tag{7}
\]
\[
\dot{x}_1 = (A_{11} - A_{12g})x_1 + A_{12}\sigma, \quad x_1(0) = [x_0]_1 \tag{8}
\]
\[
\sigma = -\frac{\kappa}{\alpha} \text{sat}(\sigma/\varepsilon), \quad \sigma(0) = \sigma_0 = Gx_0. \tag{9}
\]
Obviously, we have \(\sigma(t) \to 0\) as \(t \to \infty\) for any \(\sigma_0\) if \((\kappa/\alpha) > 0\), and, hence, \(x_1(t) \to 0\) as \(t \to \infty\) if \(A_{11} - A_{12g}\) is a stable matrix.

A. Design of \(G\)

Let \(C = [C_1 \ C_2], \ C_1 \in \mathbb{R}^{n \times 1}, \) and \(C_2 \in \mathbb{R}\).

(DG-1) If \(C_2 = 0\), then \(y = Cx = C_1x_1\), and, hence, we design \(g\) so that minimize

\[
J = \int_0^\infty (x_1^T Q x_1 + R|\sigma|^2) \, dt, \quad Q = C_1^T C_1 \tag{10}
\]
subject to the system (8). We note that \((A_{11}, A_{12})\) is controllable since \((A,B)\) is controllable and \(B = [0 \ 0]^T\). We solve the algebraic Riccati matrix equation

\[
PA_{11} + A_{11}^T P - PA_{12} R^{-1} A_{12}^T P + Q = 0, \tag{11}
\]
and set

\[
g \triangleq -R^{-1} A_{12}^T P, \quad \alpha \triangleq 1. \tag{12}
\]

(DG-2) When \(C_2 \neq 0\), let us denote \(C = [C'_1 \ 1] C_2\), where \(C'_1 = C_1/C_2\). Let \(\alpha = C_2, \ g = C'_1\).

(DG-2-1) If \(A_{11} - A_{12}C'_1\) is a stable matrix, then we choose \(G = C\), because it is natural to expect that \(y(t) = \sigma(t)\) will converge to 0 very fast by the nature of SMC.

(DG-2-2) If \(A_{11} - A_{12}C'_1\) is not stable, we solve the algebraic Riccati matrix equation

\[
\hat{P}A_{11} + A_{11}^T \hat{P} - \hat{P}A_{12} R^{-1} A_{12}^T \hat{P} + Q = 0, \tag{13}
\]
where \(Q = Q^T\) is a positive definite matrix, and set

\[
g \triangleq -R^{-1} A_{12}^T \hat{P}, \quad \alpha \triangleq 1. \tag{14}
\]
In this case, we have not special relation such as \(y = C_1 x_1\) or \(y = \sigma\), and, hence, we need to design \(g\) so that \(x_1\) converges to 0 as fast as possible. This is the reason why we assume that \(Q\) is positive definite in this case.

IV. Trajectory of \(\sigma\) and Constraints

The typical trajectory of \(\sigma(t)\) is shown in Fig. 1. More precisely, we have the following:

**Lemma 1** Consider the system (9). Let \(\beta = \kappa/\alpha > 0\) and \(\gamma = \beta/\varepsilon > 0\). We have:

(A) If \(\sigma(0) > \varepsilon\), we have

\[
\sigma(t) = \begin{cases}
\sigma(0) - \beta t \geq \varepsilon, & t \in [0, t^+_\varepsilon], \\
\varepsilon e^{-\gamma(t-t^+_\varepsilon)} > 0, & t \geq t^+_\varepsilon
\end{cases}
\]
where \(t^+_\varepsilon = (\sigma(0) - \varepsilon)/\beta\).

(B) If \(\sigma(0) \in (0, \varepsilon)\), we have

\[
\sigma(t) = \sigma(0)e^{-\gamma t} > 0, \quad t \geq 0
\]

(C) If \(\sigma(0) = 0\), we have

\[
\sigma(t) = 0, \quad t \geq 0
\]

(D) If \(\sigma(0) \in (-\varepsilon, 0)\), we have

\[
\sigma(t) = \sigma(0)e^{-\gamma t} < 0, \quad t \geq 0
\]

(E) If \(\sigma(0) < -\varepsilon\), we have

\[
\sigma(t) = \begin{cases}
\sigma(0) + \beta t \leq -\varepsilon, & t \in [0, t^-_\varepsilon], \\
-\varepsilon e^{-\gamma(t-t^-_\varepsilon)} > 0, & t \geq t^-_\varepsilon
\end{cases}
\]
where \(t^-_\varepsilon = (\sigma(0) + \varepsilon)/\beta\).

![Fig. 1. Examples of trajectories of \(\sigma(t)\).](image-url)
Remark 1 \( \kappa \) must be selected so that
\[
    m_u + \frac{\kappa}{GB} M_u > 0, \quad m_u - \frac{\kappa}{GB} M_u > 0
\]
holds, which is satisfied if and only if
\[
    \kappa < \min_{i=1, \ldots, p_u} \frac{|GB|[m_u]_i}{[M_u]_i}
\]

In view of Lemmas 1 and 2, we consider 3 cases of constraints according to the sign of \( \sigma(0) = Gx(0) \).

i) \( \sigma(0) > 0 \)
\[
    \begin{bmatrix} M_x \\ M_u F_L \end{bmatrix} x \leq \begin{bmatrix} m_x \\ m_u, p \end{bmatrix} m
\]

ii) \( \sigma(0) = 0 \)
\[
    \begin{bmatrix} M_x \\ M_u F_L \end{bmatrix} x \leq \begin{bmatrix} m_x \\ m_u \end{bmatrix} m
\]

iii) \( \sigma(0) < 0 \)
\[
    \begin{bmatrix} M_x \\ M_u F_L \end{bmatrix} x \leq \begin{bmatrix} m_x \\ m_u, n \end{bmatrix} m
\]

We estimate the region of state, in which the system state is satisfied constraints (24), (25), or (26), using the inner approximation of MAS for nonlinear continuous-time system [11].

V. INNER APPROXIMATION OF MAS
Consider a nonlinear continuous-time system
\[
    \dot{x} = f(x), \quad x(0) = x_0
\]
where \( x \in \mathbb{R}^n \), we assume that the origin is the equilibrium and consider constraints
\[
    Mx \leq m,
\]
where \( M \in \mathbb{R}^{p \times n}, m \in \mathbb{R}^p \), and \( m > 0 \).

Definition 1 The MAS for the system (27) is defined as
\[
    \Omega_{\infty} \triangleq \left\{ x_0 \in \mathbb{R}^n \mid Mx(t; x_0) \leq m, \forall t \in \mathbb{R}_+ \right\},
\]
where \( x(t; x_0) \) is the state of the system (27) at time \( t \) for the initial state \( x(0) = x_0 \).

We assume that there exists a set of matrices \( \{A_i\}^{q_1} \) such that
\[
    f(x) \in \text{co} \left\{ A_i x \right\}_{i=1}^{q_1}
\]
Consider the corresponding linear continuous-time-varying system
\[
    \dot{x} = A(t)x \in \text{co} \left\{ A_i x \right\}_{i=1}^{q_1},
\]
and its Euler approximated system
\[
    x[k+1] = \tilde{A}[k] x[k], \quad \tilde{A}[k] \in \text{co} \left\{ \tilde{A}_i \right\}_{i=1}^{q_1},
\]
where \( \tilde{A}[k] = I + \tau A[k], \tilde{A}_i = I + \tau A_i \) and \( \tau \) is step width of the forward Euler method.

Definition 2 MAS for the system (29) is defined as
\[
    \tilde{\Omega}_{\infty} \triangleq \left\{ x_0 \in \mathbb{R}^n \mid Mx[k; x_0] \leq m, \forall k \in \mathbb{Z}_+ \right\},
\]
where \( x[k; x_0] \) is the \( k \) step later state of the system (29) for an initial state \( x(0) = x_0 \).

Lemma 3 [11] We have
\[
    \tilde{\Omega}_{\infty} \subseteq \Omega_{\infty}
\]

In [13], the method to compute MAS \( \tilde{\Omega}_{\infty} \) for the system (29) was proposed.

Lemma 4 [13] \( \tilde{\Omega}_{\infty} \) is given by
\[
    \tilde{\Omega}_{k} = \left\{ x_0 \in \mathbb{R}^n \mid M \prod_{j=1}^{k} \tilde{A}_j x_0 \leq m, \ 0 \leq k \leq k \right\}
\]
\[
    \tilde{\Omega}_{\infty} = \lim_{k \to \infty} \tilde{\Omega}_{k}
\]
where \( \tilde{A}_j \in \{ \tilde{A}_1, \ldots, \tilde{A}_q \} \).
Moreover, if the system (29) is robustly stable, then there exists \( k > 0 \) such that \( \tilde{\Omega}_{\infty} = \Omega_k \).

Now we consider the case when
\[
    f(x) = (A + BF_L)x - B \frac{\kappa}{GB} \text{sat}(\sigma/\varepsilon), \quad \sigma = Gx.
\]
We treat \( f(x) \) in the region given by
\[
    \mathcal{X}_0 = \left\{ x \in \mathbb{R}^n \mid |\sigma| = |Gx| \leq \delta \right\},
\]
where \( \delta \) is chosen so that
\[
    \mathcal{X}_0 \subseteq \mathcal{X}_\delta.
\]
Then, as shown in Fig. 2, we have
\[
    \text{sat}(\sigma/\varepsilon) \in \text{co}\{\sigma/\varepsilon, \sigma/\delta\},
\]
and the corresponding extreme matrices \( A_i, i = 1, 2 \) are
\[
    A_1 = A + BF_L - \frac{\kappa}{GB} BG, \quad A_2 = A + BF_L - \frac{\kappa}{GB} BG.
\]
Therefore, \( \tilde{A}_i \) of Lemma 4 is given by
\[
    \tilde{A}_i = I + \tau A_i, \quad i = 1, 2
\]
where \( \tau \) is step width of the forward Euler method.

Let \( \Omega_{\infty}, \Omega_{\infty, 0} \) and \( \Omega_{\infty, n} \) stand for \( \Omega_{\infty} \) corresponding to \( M \) defined by (24), (25) and (26), respectively. We compute these 3 MASs off-line and use one of them at online computation, which we will explain in the next section, according to \( \sigma(0) = Gx(0) \).
VI. CONTROL STRATEGY

Our controllers have parameters $g$, $\kappa$, and $\varepsilon$, where $g = C_1$ for Case (DG-2-1) and $g$ is determined when $Q$ and $R$ are given for Cases (DG-1) and (DG-2-2). We set $Q = C_1 C_1$ for Case (DG-1) and $Q = I$ for Case (DG-2-2). Therefore, design parameters are $R$, $\kappa$, and $\varepsilon$.

If $R > 0$ is small, then we can expect a fast response, because eigenvalues of $A_{11} - A_{12} g$ converges to $0$ for $R$ tends to $0$. On the other hand, roughly speaking, inner approximation $\bar{\Omega}_0$ will be small for small $R$. This is a trade-off. The parameter $\kappa$ must satisfy the condition (23). It also appears in $\beta = \kappa/\alpha$ and $\gamma = \kappa/(\alpha\varepsilon)$, which are parameters to determine the amplitude of reaching control law (see Lemma 1). Parameters $\kappa$ and $\varepsilon$ is chosen so that $\kappa/\varepsilon$ is as large as possible under the condition that

\[ A_{0,0} = \{ x_0 : Gx_0 = 0, \quad x_0 \in X_0 \} \subseteq \bar{\Omega}_{\infty,0}, \quad (37) \]

\[ A_{0,0} = \{ x_0 : Gx_0 = 0, \quad x_0 \in X_0 \} \subseteq \bar{\Omega}_{\infty,0}, \quad (38) \]

\[ A_{0,0} = \{ x_0 : Gx_0 < 0, \quad x_0 \in X_0 \} \subseteq \bar{\Omega}_{\infty,n0}, \quad (39) \]

Suppose that we have designed $R_1$, $\kappa$, and $\varepsilon$ so that above inclusion relations hold. Then, we fix $\kappa$ and $\varepsilon$, and consider to change $R$. Suppose that we chose $\{R_2, R_3, \ldots, R_N\}$ such that $R_1 > R_2 > \cdots > R_N$. For each $R_i$, we compute $g$ and MASs $\bar{\Omega}_{\infty,p}, \bar{\Omega}_{\infty,0}, \bar{\Omega}_{\infty,n0}$ off-line, and to denote that these are corresponds to $R_i$ we designate them as $g_i$, $\bar{\Omega}_{\infty,p,i}$, $\bar{\Omega}_{\infty,0,i}$, and $\bar{\Omega}_{\infty,n0,i}$.

Roughly speaking, $\bar{\Omega}_{\infty,p,i}$ is larger than $\bar{\Omega}_{\infty,p,i+1}$, $\bar{\Omega}_{\infty,0,i}$ is larger than $\bar{\Omega}_{\infty,0,i+1}$, and $\bar{\Omega}_{\infty,n0,i}$ is larger than $\bar{\Omega}_{\infty,n0,i+1}$.

At the initial time $t = 0$, we apply $g = g_1$. For example, we suppose that $\sigma(0) > 0$. For each $k\tau$, we check if $x(k\tau) \in \bar{\Omega}_{\infty,0,2}$ or not. If $x(k\tau) \in \bar{\Omega}_{\infty,0,2}$, then we switch $g$ to $g_2$, and for each $k\tau$, we check if $x(k\tau) \in \bar{\Omega}_{\infty,0,3}$ or not. In this way, we switch $g$.

We note that the method in [10] requires to seek the largest number $j$ in many $j$’s such that $x(k\tau) \in \bar{\Omega}_{\infty,j}$, where $r_j$ is the $j$-th nominal reference command. Therefore, the proposing method has the advantage that the cost of on-line computing is lower than the method in [10].

VII. NUMERICAL EXAMPLE

Let us consider a position servo system considered in [9]. The state equation of the plant is given by

\[
\begin{align*}
\dot{x}_p &= \begin{bmatrix} 0 & 1 \\ 0 & -2\zeta \omega_n \end{bmatrix} x_p + \begin{bmatrix} 0 \\ \omega_n^2 \end{bmatrix} u_p \\
y_p &= \begin{bmatrix} 1 & 0 \end{bmatrix} x_p, \quad x_{p0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} 
\end{align*}
\]

where $x_p = [\theta_L \dot{\theta}_L]^\top$, $\zeta = 0.7$, $\omega_n = 7$, $\theta_L$[rad] denotes the angles of the load, and $u_p[V]$ is the input voltage. The constraint is

\[ |u_p| \leq V_{\text{max}}, \quad V_{\text{max}} = 2.4 \]

Suppose that a reference command $r^* = \pi/2$ is given. Let

\[ X_{p,0} = \{ x_p : -\pi \leq [x_p]_1 \leq \pi, \quad [x_p]_2 = 0 \} \]

where $[x_p]_1$ and $[x_p]_2$ denote the first and the second elements of $x_p$, respectively. Note that $X_{p,0}$ is the region of all stationary position since $x_p = [\theta_L \dot{\theta}_L]^\top$.

The servo problem is to control position servo system so that

\[ y_p \to r^* = \pi/2 \quad \text{as} \quad t \to \infty, \quad \forall x_p(0) \in X_{p,0}. \]

The pair of equilibrium and input for step signal $r^*$ denote $(x^*, u^*)$, and can be calculated as follows.

\[ x^* = A \begin{bmatrix} 1 \\ 0 \end{bmatrix} - B \begin{bmatrix} 0 \\ \tau \end{bmatrix}, \quad x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad u^* = 0 \]

To get a system given by (1), let

\[ x = x_p - x^*, \quad u = u_p - u^*, \quad y = y_p - r^* \]

Then, we have

\[ \begin{cases} \dot{x} = \begin{bmatrix} 0 & 1 \\ 0 & -2\zeta \omega_n \end{bmatrix} x + \begin{bmatrix} 0 \\ \omega_n^2 \end{bmatrix} u \\
y = \begin{bmatrix} 1 & 0 \end{bmatrix} x 
\end{cases} \]

Our problem is to control the systems (44) so that $y(t) \to 0$ as fast as possible for any $x_0 \in X_0$ under the constraint

\[ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} V_{\text{max}} - u^* \end{bmatrix} \leq \begin{bmatrix} V_{\text{max}} \\ V_{\text{max}} \end{bmatrix}, \]

where

\[ A_0 = X_{p,0} - x^* = \{ x : -\pi \leq [x]_1 \leq \pi, \quad [x]_2 = 0 \}. \]

We selected $\kappa = 110$ so that (23) is satisfied. And, we selected $Q = 1, R_1 = 1/100$. Moreover, we selected $\delta = 50$ so that (33) is satisfied, and set $\varepsilon = 2$ so that (37) - (39) are satisfied.

We also design $g_2$, $g_3$ and $g_4$ for $R_2 = 1/200$, $R_3 = 1/400$, and $R_4 = 1/1400$, respectively.

A simulation result is shown in Fig. 3 when $x_0 = 0 - x^*$.
\[ x_0 = 0 - x^* = [-\pi/2 \ 0]^T \]. Solid line denotes simulation result of proposing method, broken line denotes that of the method in [1] with the condition that the collect initial state is known, and dotted line denotes that of the method in [10]. The settling time of the proposing method is not longer than those of methods in [1] and [10], and the overshoot does not appear. Therefore, we can conclude that the proposing method achieve better response characteristics.

In Fig. 4, we show another simulation result, where \( x_0 \) is changed to \( x_0 = [-3\pi/20 \ 0]^T \neq [-\pi/2 \ 0]^T \). In Fig. 4, broken line denotes the simulation result of the method in [1], but it uses \( u(t) \) which was computed for initial state \( [-\pi/2 \ 0]^T \). From Fig. 4, we can see that the proposing method and the method in [10] are robust over initial state error.

VIII. CONCLUSIONS

We have proposed a new method of continuous-time SMC system design so as not to violate both state and control constraints. In order to guarantee the satisfaction constraints, we use the inner approximation of MAS for nonlinear continuous-time system. The advantages of this method are that it is robust over initial state error and that the amount of on-line computing is very small. Furthermore, we have proposed a control strategy via switching “sliding” hyperplanes. The advantages of this strategy are that we have very good response characteristics, that it is robust over initial state error and that the amount of on-line computing is comparatively small.
Fig. 4. Trajectory of $x$, output $y$ and input voltage $u$ for proposing control strategy and methods in [1] and [10] when $x_0 = [-3\pi/20\ 0]^T \neq [-\pi/2\ 0]^T$

REFERENCES