On Maximizing the Convergence Rate for Linear Systems With Input Saturation

Tingshu Hu, Zongli Lin, and Yacov Shamash

***Abstract*—In this note, we consider a few important issues related to the maximization of the convergence rate inside a given ellipsoid for linear systems with input saturation. For continuous-time systems, the control that maximizes the convergence rate is simply a bang–bang control. Through studying the system under the maximal convergence control, we reveal several fundamental results on set invariance. An important consequence of maximizing the convergence rate is that the maximal invariant ellipsoid is produced. We provide a simple method for finding the maximal invariant ellipsoid, and we also study the dependence of the maximal convergence rate on the Lyapunov function.**

***Index Terms*—Convergence rate, invariant set, saturation, stability.**

I. Introduction

Fast response is always a desired property for control systems. The time optimal control problem was formulated for this purpose (see, e.g., [10] and [11]). Although it is well known that the time optimal control is a bang–bang control, this control strategy is rarely implemented in real systems. The main reason is that it is generally impossible to characterize the switching surface. A notion directly related to fast response is the convergence rate of the state trajectories. For a linear system, the convergence rate is determined by the real part of the pole which is closest to the imaginary axis. For linear systems subject to actuator saturation, efforts have been made to increase the convergence rate in various heuristic ways. For example, the Q matrix in linear quadratic regulator (LQR) design can be increased piecewisely [12] as the state trajectory converges to the origin.

For better understanding of the convergence rate and its related problems, we need a precise definition of the convergence rate for a general nonlinear system. Consider a nonlinear system

$$\dot{x} = f(x).$$

Assume that the system is asymptotically stable at the origin. Given a Lyapunov function $V(x)$, let $L_V(\rho)$ be a level set $L_V(\rho) = \{x \in \mathbf{R}^n : V(x) \leq \rho\}$. Suppose that $\dot{V}(x) < 0$ for all $x \in L_V(\rho) \setminus \{0\}$. Then, the overall convergence rate of $V(x)$ on $L_V(\rho)$ can be defined as

$$\alpha := \frac{1}{2} \inf \left\{ -\frac{\dot{V}(x)}{V(x)} : x \in L_V(\rho) \setminus \{0\} \right\}. \tag{1}$$

In recent years, control systems with actuator saturation have been extensively studied (see the special issue on this topic [1] and the references therein). In this note, we will investigate issues related to the maximization of the convergence rate for a linear system subject to actuator saturation. We will be interested in quadratic Lyapunov functions, whose level sets are ellipsoids. A very important consequence of maximizing the convergence rate is that the maximal invariant ellipsoid of a given shape is produced. As pointed out in [2], set invariance is a very important notion and a powerful tool in studying the stability and other performances of systems (see also [3], [5], and the references therein). Recent years have witnessed a surge of interest in this topic. In [3], [4], [6], [8], and [9], invariant ellipsoids are used to estimate the domain of attraction and to study disturbance rejection capability of the closed-loop system. Various criteria have been derived for determining if an ellipsoid is invariant under a given saturated linear feedback law and efforts have been made to design controllers that result in large invariant ellipsoids (see, e.g., [4], [6], [8], [9], and [12]). To explore the full potential of saturating actuators, i.e., to design a controller that will produce the largest invariant ellipsoid, we need to answer the fundamental question: what is the largest ellipsoid that can be made invariant with the bounded control delivered by a saturating actuator? We will address this issue in this note through studying the system under the maximal convergence control. It turns out that the maximal convergence rate is limited by the shape of the ellipsoid, or, the P matrix in the Lyapunov function. We will develop a method to raise the limit by suitably choosing the P matrix.

This note is organized as follows. Section II shows that the maximal convergence control is a bang–bang type control with a simple switching scheme and that it produces the maximal invariant ellipsoid of a given shape. A method for determining the largest invariant ellipsoid is also given in this section. Section III reveals some properties and limitations about the overall convergence rate and provides methods to deal with these limitations. A brief concluding remark is made in Section IV.

Throughout this note, we will use standard notation. For a vector $u \in \mathbf{R}^m$, we use $|u|_\infty$ to denote the ∞-norm. We use $\mathrm{sat}(\cdot)$ to denote the standard saturation function, i.e., $\{\mathrm{sat}(s)\}_i = \mathrm{sign}(s_i)\min\{1, |s_i|\}$. We use $\mathrm{sign}(\cdot)$ to denote the sign function which takes value $+1$ or -1.

II. Maximal Convergence Rate Control and Maximal Invariant Ellipsoid

Consider a linear system subject to actuator saturation

$$\dot{x} = Ax + Bu \quad x \in \mathbf{R}^n \quad u \in \mathbf{R}^m \quad |u|_\infty \leq 1. \tag{2}$$

Assume that the system is stabilizable and that B has a full-column rank. Denote the ith column of B as b_i. In this note, we study the convergence rate of a quadratic Lyapunov function. Given a positive definite matrix $P > 0$, let $V(x) = x^{\mathrm{T}}Px$. For a positive number ρ, the level set associated with $V(x)$ is the ellipsoid

$$\mathcal{E}(P, \rho) = \{x \in \mathbf{R}^n : x^{\mathrm{T}}Px \leq \rho\}.$$

Along the trajectory of (2)

$$\begin{aligned}\dot{V}(x, u) &= 2x^{\mathrm{T}}P(Ax + Bu) \\ &= x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x + 2\sum_{i=1}^{m} x^{\mathrm{T}}Pb_iu_i.\end{aligned}$$

Under the constraint that $|u|_\infty \leq 1$, the control that maximizes the convergence rate, or minimizes $\dot{V}(x, u)$, is simply

$$u_i = -\mathrm{sign}\left(b_i^{\mathrm{T}}Px\right), \qquad i = 1, 2, \ldots, m. \tag{3}$$

Under this bang–bang control, we have

$$\dot{V}(x) = x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - 2\sum_{i=1}^{m} x^{\mathrm{T}}Pb_i\,\mathrm{sign}\left(b_i^{\mathrm{T}}Px\right).$$

Now, consider the closed-loop system

$$\dot{x} = Ax - \sum_{i=1}^{m} b_i \mathrm{sign}\left(b_i^{\mathrm{T}}Px\right). \tag{4}$$

Manuscript received January 19, 2001; revised September 7, 2001 and October 28, 2002. Recommended by Associate Editor P. A. Iglesias. This work was supported in part by the Office of Naval Research Young Investigator Program under Grant N00014-99-1-0670.

T. Hu and Z. Lin are with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904-4743 USA (e-mail: th7f@virginia.edu; zl5y@virginia.edu).

Y. Shamash is with the Department of Electrical Engineering, State University of New York, Stony Brook, NY 11794 USA (e-mail: yshamash@notes.sunysb.edu).

Digital Object Identifier 10.1109/TAC.2003.814271

The discontinuity of the bang–bang control may cause technical problems like nonexistence of solution. Since this problem can be handled by using a high gain saturated feedback to replace the bang–bang control (see [7] for more detail), in what follows, we use the bang–bang control law to investigate the possibility that an ellipsoid can be made invariant with a bounded control $|u|_\infty \leq 1$. In the sequal, we simply say "a bounded control."

Recall that an ellipsoid $\mathcal{E}(P,\rho)$ is invariant for a system $\dot{x} = f(x)$ if all the trajectories starting from it will stay inside of it. It is contractively invariant if

$$\dot{V}(x) = 2x^{\mathrm{T}}Pf(x) < 0 \qquad \forall\, x \in \mathcal{E}(P,\rho) \setminus \{0\}.$$

Since the bang–bang control (3) minimizes $\dot{V}(x,u)$ at each x, we have the following obvious fact.

Fact 1: An ellipsoid $\mathcal{E}(P,\rho)$ can be made contractively invariant for (2) with a bounded control if and only if it is contractively invariant for (4), i.e., the following condition is satisfied:

$$\dot{V}(x) = x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - 2\sum_{i=1}^{m} x^{\mathrm{T}}Pb_i \mathrm{sign}\left(b_i^{\mathrm{T}}Px\right) < 0 \qquad \forall\, x \in \mathcal{E}(P,\rho) \setminus \{0\}. \tag{5}$$

It is clear from Fact 1 that the maximal convergence rate control produces the maximal invariant ellipsoid. For an arbitrary matrix $P > 0$, there may exists no ρ such that $\mathcal{E}(P,\rho)$ can be made invariant. In what follows, we give a condition on P such that $\mathcal{E}(P,\rho)$ can be made invariant for some ρ and provide a method for finding the largest ρ.

Proposition 1: For a given matrix $P > 0$, the following three statements are equivalent.

a) There exists a $\rho > 0$ such that (5) is satisfied.

b) There exists an $F \in \mathbf{R}^{m\times n}$ such that

$$(A + BF)^{\mathrm{T}}P + P(A + BF) < 0. \tag{6}$$

c) There exists a $k > 0$ such that

$$(A - kBB^{\mathrm{T}}P)^{\mathrm{T}}P + P(A - kBB^{\mathrm{T}}P) < 0. \tag{7}$$

Proof: b) $\rightarrow$ a). If (6) is satisfied, then there exists a $\rho > 0$ such that

$$\mathcal{E}(P,\rho) \subset \{x \in \mathbf{R}^n : |Fx|_\infty \leq 1\}.$$

If $x_0 \in \mathcal{E}(P,\rho)$, then under the control $u = Fx$, $x(t)$ will stay in $\mathcal{E}(P,\rho)$ and we also have $|u|_\infty \leq 1$ for all $t \geq 0$. This means that $\mathcal{E}(P,\rho)$ can be made contractively invariant with a bounded control. Hence, by Fact 1, we have (5).

c) $\rightarrow$ b). It is obvious.

a) $\rightarrow$ c). Let us assume that $PB = \left[\begin{smallmatrix}0\\R\end{smallmatrix}\right]$, where R is an $m \times m$ nonsingular matrix. If not so, we can use a state transformation, $\bar{x} = Tx$, with T nonsingular such that

$$P \rightarrow \bar{P} = (T^{-1})^{\mathrm{T}}PT^{-1}$$
$$B \rightarrow \bar{B} = TB$$

and

$$PB \rightarrow \bar{P}\bar{B} = (T^{-1})^{\mathrm{T}}PB = \begin{bmatrix}0\\R\end{bmatrix}.$$

Recall that we have assumed that B has a full-column rank. Also, let us accordingly partition x as $\left[\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right]$ and $A^{\mathrm{T}}P + PA$ and P as

$$A^{\mathrm{T}}P + PA = \begin{bmatrix}Q_1 & Q_{12}\\ Q_{12}^{\mathrm{T}} & Q_2\end{bmatrix} \quad P = \begin{bmatrix}P_1 & P_{12}\\ P_{12}^{\mathrm{T}} & P_2\end{bmatrix}.$$

For all $x = \left[\begin{smallmatrix}x_1\\0\end{smallmatrix}\right] \in \partial\mathcal{E}(P,\rho)$, we have $x^{\mathrm{T}}PB = 0$. So, if a) is true, then (5) holds for some $\rho > 0$, which implies that $x_1^{\mathrm{T}}Q_1x_1 < 0$, for all x_1 such that $x_1^{\mathrm{T}}P_1x_1 = \rho$. It follows that $Q_1 < 0$. Hence, there exists a $k > 0$ such that

$$(A - kBB^{\mathrm{T}}P)^{\mathrm{T}}P + P(A - kBB^{\mathrm{T}}P) = \begin{bmatrix}Q_1 & Q_{12}\\ Q_{12}^{\mathrm{T}} & Q_2 - kRR^{\mathrm{T}}\end{bmatrix} < 0.$$

This shows that c) is true. □

Now assume that we have a $P > 0$ such that the conditions in Proposition 1 are satisfied. Given a $\rho > 0$, we would like to determine if $\mathcal{E}(P,\rho)$ is contractively invariant for the closed-loop system (4). Let us start with the single input case. In this case, (5) simplifies to

$$\dot{V}(x) = x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - 2x^{\mathrm{T}}PB\mathrm{sign}(B^{\mathrm{T}}Px) < 0 \qquad \forall x \in \mathcal{E}(P,\rho) \setminus \{0\}. \tag{8}$$

We claim that (8) is equivalent to

$$x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - 2x^{\mathrm{T}}PB\mathrm{sign}(B^{\mathrm{T}}Px) < 0 \qquad \forall\, x \in \partial\mathcal{E}(P,\rho). \tag{9}$$

To see this, we consider kx for $k \in (0,1]$ and $x \in \partial\mathcal{E}(P,\rho)$. Suppose that

$$x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - 2x^{\mathrm{T}}PB\mathrm{sign}(B^{\mathrm{T}}Px) < 0.$$

Since $-2x^{\mathrm{T}}PB\mathrm{sign}(B^{\mathrm{T}}Px) \leq 0$, we have

$$x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - \frac{2x^{\mathrm{T}}PB}{k}\mathrm{sign}(B^{\mathrm{T}}Px) < 0, \ \forall k \in (0,1].$$

Therefore

$$\begin{aligned}&(kx)^{\mathrm{T}}(A^{\mathrm{T}}P + PA)(kx) - 2(kx)^{\mathrm{T}}PB\mathrm{sign}(B^{\mathrm{T}}Pkx)\\ &\quad = k^2\left(x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - \frac{2x^{\mathrm{T}}PB}{k}\mathrm{sign}(B^{\mathrm{T}}Px)\right)\\ &\quad < 0\end{aligned}$$

for all $k \in (0,1]$. This shows that (8) is equivalent to (9). Based on this equivalence relation, we have the following necessary and sufficient condition for the contractive invariance of a given ellipsoid.

Theorem 1: Assume that $m = 1$. Suppose that $\mathcal{E}(P,\rho)$ can be made contractively invariant for some $\rho_0 > 0$. Let $\lambda_1, \lambda_2, \ldots, \lambda_{\mathcal{J}} > 0$ be real numbers such that

$$\det\begin{bmatrix}\lambda_j P - A^{\mathrm{T}}P - PA & P\\ \rho^{-1}PBB^{\mathrm{T}}P & \lambda_j P - A^{\mathrm{T}}P - PA\end{bmatrix} = 0 \tag{10}$$

and

$$B^{\mathrm{T}}P(A^{\mathrm{T}}P + PA - \lambda_j P)^{-1}PB > 0. \tag{11}$$

Then, $\mathcal{E}(P,\rho)$ is contractively invariant for (4) iff

$$\lambda_j\rho - B^{\mathrm{T}}P(A^{\mathrm{T}}P + PA - \lambda_j P)^{-1}PB < 0 \qquad \forall\, j = 1,2,\ldots,\mathcal{J}.$$

If there exists no $\lambda_j > 0$ satisfying (10) and (11), then $\mathcal{E}(P,\rho)$ is contractively invariant.

Here, we note that all the λ_j's satisfying (10) are the eigenvalues of the matrix

$$\begin{bmatrix}P^{-\frac{1}{2}}A^{\mathrm{T}}P^{\frac{1}{2}} + P^{\frac{1}{2}}AP^{-\frac{1}{2}} & -I\\ -\rho^{-1}P^{\frac{1}{2}}BB^{\mathrm{T}}P^{\frac{1}{2}} & P^{-\frac{1}{2}}A^{\mathrm{T}}P^{\frac{1}{2}} + P^{\frac{1}{2}}AP^{-\frac{1}{2}}\end{bmatrix}.$$

Hence, the condition of Theorem 1 can be easily checked.

In the proof of Theorem 1, we will use the following algebraic fact. Suppose that X_1 and X_4 are square matrices and are nonsingular, then

$$\begin{aligned}\det\begin{bmatrix}X_1 & X_2\\ X_3 & X_4\end{bmatrix} &= \det(X_1)\det(X_4 - X_3X_1^{-1}X_2)\\ &= \det(X_4)\det(X_1 - X_2X_4^{-1}X_3).\end{aligned} \tag{12}$$

Proof of Theorem 1: Denote $g(x) = x^{\mathrm{T}}(A^{\mathrm{T}}P + PA)x - 2x^{\mathrm{T}}PB$. By the equivalence of (8) and (9), the contractive invariance of $\mathcal{E}(P,\rho)$ is equivalent to

$$\max\left\{g(x) : B^{\mathrm{T}}Px \geq 0, x^{\mathrm{T}}Px = \rho\right\} < 0. \tag{13}$$

Since $\mathcal{E}(P,\rho)$ can be made contractively invariant for some $\rho>0$, we must have $g(x)<0$ for all $B^{\mathrm{T}}Px=0$. In this case, the contractive invariance of $\mathcal{E}(P,\rho)$ is equivalent to that all the extrema of $g(x)$ in the surface $x^{\mathrm{T}}Px=\rho$, $B^{\mathrm{T}}Px>0$, if any, are less than zero.

By the Lagrange multiplier method, an extremum of $g(x)$ in the surface $x^{\mathrm{T}}Px=\rho$, $B^{\mathrm{T}}Px>0$, must satisfy

$$(A^{\mathrm{T}}P+PA-\lambda P)x=PB,\ \ x^{\mathrm{T}}Px=\rho,\ \ x^{\mathrm{T}}PB>0 \qquad (14)$$

for some real number λ. And at the extremum, we have $g(x)=\lambda\rho-x^{\mathrm{T}}PB$. If $\lambda\le 0$, then $g(x)<0$ since $x^{\mathrm{T}}PB>0$. So, we only need to consider $\lambda>0$.

Now, suppose that $\lambda>0$. From $(A^{\mathrm{T}}P+PA-\lambda P)x=PB$, we conclude that $\det(A^{\mathrm{T}}P+PA-\lambda P)\neq 0$. To show this, we assume, without loss of generality, that

$$A^{\mathrm{T}}P+PA=\begin{bmatrix}Q_1 & Q_{12}\\ Q_{12}^{\mathrm{T}} & q_2\end{bmatrix},\ P=\begin{bmatrix}P_1 & P_{12}\\ P_{12}^{\mathrm{T}} & p_2\end{bmatrix},\ PB=\begin{bmatrix}0\\ r\end{bmatrix}$$

as in the proof of Proposition 1, it follows that $Q_1<0$. Since $\lambda>0$, $Q_1<0$ and $P_1>0$, $Q_1-\lambda P_1$ is nonsingular. Let $x=\begin{bmatrix}x_1\\ x_2\end{bmatrix}$, $x_2\in\mathbf{R}$, and suppose that $x\neq 0$ satisfies

$$(A^{\mathrm{T}}P+PA-\lambda P)x=PB$$

then

$$x_1=-(Q_1-\lambda P_1)^{-1}(Q_{12}-\lambda P_{12})x_2$$

and

$$\Big(-(Q_{12}^{\mathrm{T}}-\lambda P_{12}^{\mathrm{T}})(Q_1-\lambda P_1)^{-1}\times(Q_{12}-\lambda P_{12})+q_2-\lambda p_2\Big)x_2=r.$$

Multiplying both sides with $\det(Q_1-\lambda P_1)$ and applying (12), we obtain

$$\det(A^{\mathrm{T}}P+PA-\lambda P)x_2=\det(Q_1-\lambda P_1)r.$$

Since $r\neq 0$ and $\det(Q_1-\lambda P_1)\neq 0$, we must have $\det(A^{\mathrm{T}}P+PA-\lambda P)\neq 0$.

Thus, for all $\lambda>0$ and x satisfying (14), we have $x=(A^{\mathrm{T}}P+PA-\lambda P)^{-1}PB$, and from $x^{\mathrm{T}}Px=\rho$, we obtain

$$B^{\mathrm{T}}P(A^{\mathrm{T}}P+PA-\lambda P)^{-1}P(A^{\mathrm{T}}P+PA\lambda P)^{-1}PB=\rho. \qquad (15)$$

Denote $\Phi=\lambda P-A^{\mathrm{T}}P-PA$, then the (15) can be written as

$$B^{\mathrm{T}}P\Phi^{-1}P\Phi^{-1}PB=\rho.$$

By invoking (12), we obtain

$$\det\begin{bmatrix}\rho & -B^{\mathrm{T}}P\Phi^{-1}\\ -\Phi^{-1}PB & P^{-1}\end{bmatrix}=0$$

$$\Updownarrow$$

$$\det\left(\begin{bmatrix}\rho & 0\\ 0 & P^{-1}\end{bmatrix}-\begin{bmatrix}B^{\mathrm{T}}P & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}\Phi^{-1} & 0\\ 0 & \Phi^{-1}\end{bmatrix}\begin{bmatrix}0 & I\\ PB & 0\end{bmatrix}\right)=0$$

$$\Updownarrow$$

$$\det\left(\begin{bmatrix}\Phi & 0\\ 0 & \Phi\end{bmatrix}-\begin{bmatrix}0 & I\\ PB & 0\end{bmatrix}\begin{bmatrix}\rho^{-1} & 0\\ 0 & P\end{bmatrix}\begin{bmatrix}B^{\mathrm{T}}P & 0\\ 0 & I\end{bmatrix}\right)=0$$

$$\Updownarrow$$

$$\det\begin{bmatrix}\lambda P-A^{\mathrm{T}}P-PA & P\\ \rho^{-1}PBB^{\mathrm{T}}P & \lambda P-A^{\mathrm{T}}P-PA\end{bmatrix}=0.$$

This last equation is (10).

Also, at the extremum, we have $x^{\mathrm{T}}PB>0$. This is equivalent to (11)

$$B^{\mathrm{T}}P(A^{\mathrm{T}}P+PA-\lambda P)^{-1}PB>0.$$

Finally, at the extremum

$$\begin{aligned}g(x)&=x^{\mathrm{T}}(A^{\mathrm{T}}P+PA)x-2x^{\mathrm{T}}PB\\&=\lambda\rho-B^{\mathrm{T}}P(A^{\mathrm{T}}P+PA-\lambda P)^{-1}PB.\end{aligned}$$

Hence, the result of the theorem follows. □

Recall that (8) is equivalent to (9). This implies that there is a $\rho^*>0$ such that $\mathcal{E}(P,\rho)$ is contractively invariant if and only if $\rho<\rho^*$. Therefore, the maximum value ρ^* can be obtained by checking the condition of Theorem 1 using bisection method.

For systems with multiple inputs, we may divide the surface $\partial\mathcal{E}(P,\rho)$ into subsets. For example, consider $m=2$, the surface of $\mathcal{E}(P,\rho)$ can be divided into the following subsets:

$$S_1=\left\{x\in\mathbf{R}^n: b_1^{\mathrm{T}}Px=0,\ b_2^{\mathrm{T}}Px\ge 0,\ x^{\mathrm{T}}Px=\rho\right\},\ -S_1$$

$$S_2=\left\{x\in\mathbf{R}^n: b_1^{\mathrm{T}}Px\ge 0,\ b_2^{\mathrm{T}}Px=0,\ x^{\mathrm{T}}Px=\rho\right\},\ -S_2$$

$$S_3=\left\{x\in\mathbf{R}^n: b_1^{\mathrm{T}}Px>0,\ b_2^{\mathrm{T}}Px>0,\ x^{\mathrm{T}}Px=\rho\right\},\ -S_3$$

$$S_4=\left\{x\in\mathbf{R}^n: b_1^{\mathrm{T}}Px>0,\ b_2^{\mathrm{T}}Px<0,\ x^{\mathrm{T}}Px=\rho\right\},\ -S_4.$$

With this partition, $\mathcal{E}(P,\rho)$ is contractively invariant iff

$$\max_{x\in S_1}\dot V(x)<0\quad\max_{x\in S_2}\dot V(x)<0 \qquad (16)$$

and all the local extrema of $\dot V(x)$ in S_3 and S_4 are negative.

In S_1, $\dot V(x)=x^{\mathrm{T}}(A^{\mathrm{T}}P+PA)x-2x^{\mathrm{T}}Pb_2$. Let $N\in\mathbf{R}^{n\times(n-1)}$ be a matrix of rank $n-1$ such that $b_1^{\mathrm{T}}PN=0$, i.e., $\{Ny: y\in\mathbf{R}^{n-1}\}$ is the kernel of $b_1^{\mathrm{T}}P$. The constraint $b_1^{\mathrm{T}}Px=0$ can be replaced by $x=Ny$, $y\in\mathbf{R}^{n-1}$. Thus

$$\max_{x\in S_1}\dot V(x)=\max\Big\{y^{\mathrm{T}}N^{\mathrm{T}}(A^{\mathrm{T}}P+PA)Ny - 2y^{\mathrm{T}}N^{\mathrm{T}}Pb_2: b_2^{\mathrm{T}}PNy\ge 0, y^{\mathrm{T}}N^{\mathrm{T}}PNy=\rho\Big\}.$$

This is similar to the optimization problem (13) in the proof of Theorem 1 except with a reduced order. The second optimization problem in (16) can be handled in the same way.

In S_3, $\dot V(x)=x^{\mathrm{T}}(A^{\mathrm{T}}P+PA)x-2x^{\mathrm{T}}P(b_1+b_2)$. All the local extrema of $\dot V(x)$ in S_3 (and in S_4) can be obtained like those of $g(x)$ in the proof of Theorem 1. For systems with more than two inputs, we need to divide the surface into $m\times 2^{m-1}$ closed sets like S_1 and S_2 and 2^m sets like S_3 and S_4 (with all strict inequalities except for $x^{\mathrm{T}}Px=\rho$). This indicates that the computational burden increases exponentially as m increases.

III. Overall Convergence Rate

We now consider (4) under the maximal convergence control

$$\dot x=Ax-\sum_{i=1}^{m}b_i\operatorname{sign}(b_i^{\mathrm{T}}Px). \qquad (17)$$

Assume that $\mathcal{E}(P,\rho)$ is contractively invariant for (17). We would like to know the overall convergence rate in $\mathcal{E}(P,\rho)$. We will see later that as ρ decreases (note that a trajectory goes into smaller $\mathcal{E}(P,\rho)$ as time goes by), the overall convergence rate increases but is limited by the shape of $\mathcal{E}(P,\rho)$. This limit can be raised by choosing P properly.

The overall convergence rate, denoted by α, is defined by (1) in Section I. Here, we would like to examine its dependence on ρ, so we write

$$\alpha(\rho):=\frac{1}{2}\min\left\{-\frac{\dot V(x)}{V(x)}: x\in\mathcal{E}(P,\rho)\setminus\{0\}\right\}.$$

The main results of this section are contained in the following theorem.

Theorem 2:

a)

$$\alpha(\rho)=\frac{1}{2}\min\left\{-\frac{\dot V(x)}{\rho}: x^{\mathrm{T}}Px=\rho\right\}. \qquad (18)$$

b) $\alpha(\rho)$ increases as ρ decreases.

c) Let

$$\beta_0=\min\left\{-x^{\mathrm{T}}(A^{\mathrm{T}}P+PA)x: x^{\mathrm{T}}Px=1,\ x^{\mathrm{T}}PB=0\right\}$$

then

$$\lim_{\rho\to 0}\alpha(\rho)=\frac{\beta_0}{2}. \qquad (19)$$

Proof:

a) Consider $x \in \partial\mathcal{E}(P,\rho)$ and $k \in (0,1]$,

$$\begin{aligned} & -\frac{\dot V(kx)}{V(kx)} \\ &= -\frac{k^2 x^{\mathrm T}(A^{\mathrm T}P+PA)x - 2k\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Pkx)}{k^2 x^{\mathrm T}Px} \\ &= \frac{-x^{\mathrm T}(A^{\mathrm T}P+PA)x + \frac{2}{k}\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px)}{x^{\mathrm T}Px}. \end{aligned} \tag{20}$$

Since

$$\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px) \ge 0$$

$-(\dot V(kx))/(V(kx))$ increases as k decreases. It follows that the minimal value of $-(\dot V(x))/(V(x))$ is obtained on the boundary of $\mathcal{E}(P,\rho)$, which implies (18).

b) This follows from the proof of a).

c) From a), we see that

$$\begin{aligned} & 2\alpha(\rho) \\ &= \min\left\{\frac{-x^{\mathrm T}(A^{\mathrm T}P+PA)x + 2\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px)}{\rho} \right. \\ &\quad \left. : x^{\mathrm T}Px = \rho \right\} \\ &= \min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x \right. \\ &\quad \left. + \frac{2}{\sqrt\rho}\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px) : x^{\mathrm T}Px = 1\right\} \\ &\le \min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x \right. \\ &\quad \left. + \frac{2}{\sqrt\rho}\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px) : x^{\mathrm T}Px = 1,\ x^{\mathrm T}PB = 0\right\} \\ &= \min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x : x^{\mathrm T}Px = 1,\ x^{\mathrm T}PB = 0\right\} \\ &= \beta_0. \end{aligned}$$

It then follows that $2\alpha(\rho) \le \beta_0$ for all $\rho > 0$. To prove (19), it suffices to show that given any $\varepsilon > 0$, there exists a $\rho > 0$ such that $2\alpha(\rho) \ge \beta_0 - \varepsilon$.

Denote

$$\mathcal{X}_0 = \left\{x \in \mathrm{R}^n : x^{\mathrm T}Px = 1,\ x^{\mathrm T}PB = 0\right\}$$

and

$$\mathcal{X}(\delta) = \left\{x \in \mathbf{R}^n : x^{\mathrm T}Px = 1,\ |x^{\mathrm T}PB|_\infty \le \delta\right\}.$$

It is clear that $\lim_{\delta\to 0}\mathrm{dist}(\mathcal{X}(\delta),\mathcal{X}_0) = 0$, where $\mathrm{dist}(\cdot,\cdot)$ is the Hausdorff distance.[1] By the uniform continuity of $x^{\mathrm T}(A^{\mathrm T}P+PA)x$ on the surface $\{x \in \mathbf{R}^n : x^{\mathrm T}Px = 1\}$, we have that, given any ε, there exists a $\delta > 0$ such that

$$\min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x : x^{\mathrm T}Px=1,\ |x^{\mathrm T}PB|_\infty \le \delta\right\} \ge \beta_0 - \varepsilon. \tag{21}$$

[1]Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two bounded subsets of $\mathbf{R}^n$. Their Hausdorff distance is defined as

$$\mathrm{dist}(\mathcal{X}_1,\mathcal{X}_2) := \max\left\{\vec d(\mathcal{X}_1,\mathcal{X}_2), \vec d(\mathcal{X}_2,\mathcal{X}_1)\right\}$$

where

$$\vec d(\mathcal{X}_1,\mathcal{X}_2) = \sup_{x_1\in\mathcal{X}_1}\inf_{x_2\in\mathcal{X}_2}|x_1 - x_2|.$$

Here, the vector norm used is arbitrary.

Since $\sum_{i=1}^m x^{\mathrm T}Pb_i\mathrm{sign}(b_i^{\mathrm T}Px) \ge 0$, we have

$$\min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x + \frac{2}{\sqrt\rho}\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px) : x^{\mathrm T}Px = 1,\ |x^{\mathrm T}PB|_\infty \le \delta\right\} \ge \beta_0 - \varepsilon \tag{22}$$

for all $\rho > 0$.

Let

$$\beta_1 = \min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x : x^{\mathrm T}Px = 1,\ |x^{\mathrm T}PB|_\infty \ge \delta\right\}.$$

If $\beta_1 \ge \beta_0 - \varepsilon$, then for all $\rho > 0$

$$\min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x + \frac{2}{\sqrt\rho}\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px) : x^{\mathrm T}Px = 1,\ |x^{\mathrm T}PB|_\infty \ge \delta\right\} \ge \beta_1 \ge \beta_0 - \varepsilon.$$

Combining this with (22), we have

$$2\alpha(\rho) = \min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x + \frac{2}{\sqrt\rho} \times \sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px) : x^{\mathrm T}Px = 1\right\} \ge \beta_0 - \varepsilon. \tag{23}$$

for all $\rho > 0$. This shows that $2\alpha(\rho) \ge \beta_0 - \varepsilon, \forall\rho > 0$.

If $\beta_1 < \beta_0 - \varepsilon$, then for $\rho < ((2\delta)/(-\beta_1 + \beta_0 - \varepsilon))^2$, we have

$$\min\left\{-x^{\mathrm T}(A^{\mathrm T}P+PA)x + \frac{2}{\sqrt\rho}\sum_{i=1}^m x^{\mathrm T}Pb_i\,\mathrm{sign}(b_i^{\mathrm T}Px) : x^{\mathrm T}Px = 1,\ |x^{\mathrm T}PB|_\infty \ge \delta\right\} \ge \beta_1 + \frac{2\delta}{\sqrt\rho} > \beta_0 - \varepsilon.$$

Combining this with (22), we also obtain (23) and $2\alpha(\rho) \ge \beta_0 - \varepsilon$ for $\rho < (2\delta/(-\beta_1 + \beta_0 - \varepsilon))^2$. This completes the proof. □

Theorem 2 says that the convergence rate increases as the ellipsoid becomes smaller and it approaches an upper bound $\beta_0/2$ as ρ goes to 0. Hence, the convergence rate is bounded by $\beta_0/2$ for all ρ. For a given ρ, $\alpha(\rho)$ can be obtained by computing the maximum of $\dot V(x)$ over $\partial\mathcal{E}(P,\rho)$. For the single input case, Theorem 1 provides a method for determining if this maximum is negative. The exact value of $\alpha(\rho)$ can be computed with a procedure similar to the proof of Theorem 1.

Since the overall convergence rate is limited by $\beta_0/2$, we would like β_0 not to be too small. The following proposition will lead to an LMI approach to choosing P for maximizing β_0.

Proposition 2:

$$\begin{aligned} \beta_0 &= \sup_F \lambda \\ &\text{s.t. } (A+BF)^{\mathrm T}P + P(A+BF) \le -\lambda P. \end{aligned} \tag{24}$$

Proof: Notice that, for any F, we have

$$x^{\mathrm T}\left((A+BF)^{\mathrm T}P + P(A+BF)\right)x = x^{\mathrm T}(A^{\mathrm T}P+PA)x \quad \forall\, x^{\mathrm T}PB = 0.$$

It follows that

$$\begin{aligned} \beta_0 &= \min\left\{-x^{\mathrm T}\left((A+BF)^{\mathrm T}P \right.\right. \\ &\qquad \left.\left. + P(A+BF)\right)x : x^{\mathrm T}Px = 1,\ x^{\mathrm T}PB = 0\right\} \\ &\ge \min\left\{-x^{\mathrm T}\left((A+BF)^{\mathrm T}P \right.\right. \\ &\qquad \left.\left. + P(A+BF)\right)x : x^{\mathrm T}Px = 1\right\} \end{aligned}$$

and, hence

$$\beta_0 \ge \sup_F \min\left\{-x^{\mathrm T}\left((A+BF)^{\mathrm T}P + P(A+BF)\right)x : x^{\mathrm T}Px = 1\right\}. \tag{25}$$

We claim that

$$\beta_0 = \sup_F \min \left\{ -x^{\mathrm{T}} \left((A+BF)^{\mathrm{T}} P + P(A+BF) \right) x : x^{\mathrm{T}} P x = 1 \right\}. \quad (26)$$

In view of (25), it suffices to show that for any $\varepsilon > 0$, there exists an $F = -kB^{\mathrm{T}}P$, with $k > 0$, such that

$$\min \left\{ -x^{\mathrm{T}} \left((A+BF)^{\mathrm{T}} P + P(A+BF) \right) x : x^{\mathrm{T}} P x = 1 \right\} \geq \beta_0 - \varepsilon. \quad (27)$$

This can be proven by exploiting the same idea used in the proof of Theorem 2 c).

Denote

$$\beta(F) = \min \left\{ -x^{\mathrm{T}} \left((A+BF)^{\mathrm{T}} P + P(A+BF) \right) x : x^{\mathrm{T}} P x = 1 \right\}.$$

From (26), we have $\beta_0 = \sup_F \beta(F)$. It can be shown that

$$\beta(F) = \max \left\{ \lambda : (A+BF)^{\mathrm{T}} P + P(A+BF) \leq -\lambda P \right\}.$$

This brings us to (24). □

The matrix inequality constraint in Proposition 2 indicates that the ellipsoid $\mathcal{E}(P,\rho)$ has a convergence rate $\lambda/2$ under the linear state feedback $u = Fx$ (for any $\rho > 0$). From Proposition 2 and Theorem 2, we see that the convergence rate is limited by the maximal value which can be achieved with a linear state feedback and can actually approach this maximal value as ρ goes to 0. For a fixed P, β_0 is a finite value. Assume that (A, B) is controllable, then the eigenvalues of $(A+BF)$ can be arbitrarily assigned. If we also take P as a variable, then $-\beta_0/2$ can be made equal to the largest real part of the eigenvalues of $A+BF$ (see the definition, as given in Section I, of the overall convergence rate for a linear system). This means that β_0 can be made arbitrarily large. But generally, as β_0 becomes very large, the matrix P will be badly conditioned, i.e., the ellipsoid $\mathcal{E}(P,\rho)$ will become very thin in certain direction, and hence very "small," with respect to a fixed shape reference set. On the other hand, as mentioned in [8] and [9], if our only objective is to enlarge the domain of attraction with respect to a reference set, some eigenvalues of $A+BF$ will be very close to the imaginary axis, resulting in very small β_0. These two conflicting objectives can be balanced, for example, by prespecifying a lower bound on β_0 and then maximizing the invariant ellipsoid with respect to some shape reference set. This mixed problem can be described as follows:

$$\begin{aligned} &\sup_{P>0,\rho,F,H} \quad \alpha \\ &\text{s.t. a) } \alpha\mathcal{X}_{\mathrm{R}} \subset \mathcal{E}(P,\rho) \\ &\quad \text{b) } (A+BF)^{\mathrm{T}} P + P(A+BF) < 0 \\ &\quad \text{c) } \mathcal{E}(P,\rho) \subset \{x \in \mathrm{R}^n : |Fx|_\infty \leq 1\} \\ &\quad \text{d) } (A+BH)^{\mathrm{T}} P + P(A+BH) \leq -\underline{\beta} P \end{aligned} \quad (28)$$

where $\mathcal{X}_{\mathrm{R}}$ is a shape reference set (see [7]–[9]). The constraint a) means that $\mathcal{E}(P,\rho)$ contains $\alpha\mathcal{X}_{\mathrm{R}}$. By maximizing α, $\mathcal{E}(P,\rho)$ will be maximized with respect to $\mathcal{X}_{\mathrm{R}}$. The constraints b) and c) guarantee that $\mathcal{E}(P,\rho)$ can be made contractively invariant and the constraint d) guarantees a lower bound $\underline{\beta}$ on the convergence rate. This optimization problem can be transformed into one with LMI constraints as those in [7]–[9]. By solving (28), we obtain the optimal ellipsoid $\mathcal{E}(P,\rho)$ along with two feedback matrices F and H. We may actually discard both F and H but instead use the bang–bang control $u_i = -\mathrm{sign}(b_i^{\mathrm{T}} P x)$, $i = 1, 2, \ldots, m$, or the saturated high-gain control $u = -\mathrm{sat}(kB^{\mathrm{T}}Px)$. The final outcome is that under these control laws, the closed-loop system will have a contractively invariant set $\mathcal{E}(P,\rho)$ and a guaranteed limit of the convergence rate $\beta_0/2 \geq \underline{\beta}/2$.

IV. Conclusion

We have studied several issues related to the maximization of the convergence rate for continuous-time linear systems with input saturation. These issues include the maximal convergence control, the maximal ellipsoid that can be made invariant with a bounded control, the control laws that can produce the maximal invariant ellipsoid and the dependence of the maximal convergence rate on the Lyapunov function. The counterpart of these issues for discrete-time system are studied in [7], where some of the results are quite different from those in this note. For example, the maximal convergence control is continuous in the state and is determined in a quite different way.

References

[1] D. S. Bernstein and A. N. Michel, Eds., *Int. J. Robust and Nonlinear Control: Special Issue on Saturating Actuators*, 1995, vol. 5.

[2] F. Blanchini, "Set invariance in control—a survey," *Automatica*, vol. 35, no. 11, pp. 1747–1767, 1999.

[3] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory*. Philadelphia, PA: SIAM, 1994.

[4] J. M. Gomes da Silva Jr. and S. Tarbouriech, "Local stabilization of discrete-time linear systems with saturating controls: an LMI approach," *IEEE Trans. Automat. Contr.*, vol. 46, pp. 119–125, Jan. 2001.

[5] P. O. Gutman and M. Cwikel, "Admissible sets and feedback control for discrete-time linear dynamical systems with bounded control and dynamics," *IEEE Trans. Automat. Contr.*, vol. AC-31, pp. 373–376, Apr. 1986.

[6] H. Hindi and S. Boyd, "Analysis of linear systems with saturating using convex optimization," in *Proc. 37th IEEE Conf. Decision Control*, Orlando, FL, 1998, pp. 903–908.

[7] T. Hu and Z. Lin, *Control Systems With Actuator Saturation: Analysis and Design*. Boston, MA: Birkhäuser, 2001.

[8] ——, "On enlarging the basin of attraction for linear systems under saturated linear feedback," *Syst. Control Lett.*, vol. 40, no. 1, pp. 59–69, May 2000.

[9] T. Hu, Z. Lin, and B. M. Chen, "An analysis and design method for linear systems subject to actuator saturation and disturbance," *Automatica*, vol. 38, no. 2, pp. 351–359, 2002.

[10] E. B. Lee and L. Markus, *Foundations of Optimal Control*. New York: Wiley, 1967.

[11] J. Macki and M. Strauss, *Introduction to Optimal Control*. New York: Springer-Verlag, 1982.

[12] G. F. Wredenhagen and P. R. Belanger, "Piecewise-linear LQ control for systems with input constraints," *Automatica*, vol. 30, pp. 403–416, 1994.