Stabilization of linear systems with bounded controls*

Yudi Yang, Héctor J. Sussmann, and Eduardo D. Sontag
SYCON - Rutgers Center for Systems and Control
Department of Mathematics, Rutgers University, New Brunswick, NJ 08903

Abstract. Linear systems with saturated controls can be smoothly stabilized as long as certain natural controllability and eigenvalue conditions hold, as shown in a previous paper by the authors. Generalizing an idea of Teel, we show here how to obtain these stabilizers using very simple combinations of saturation nonlinearities —in neural network terms, sigmoidal feedforward nets.

Key words: nonlinear stabilization, saturated actuators, neural nets

1. Introduction

Linear systems with saturated controls can be stabilized continuously —and even smoothly— as long as certain natural controllability and eigenvalue conditions hold. (This was shown in our previous paper [4].) The construction in [4] relied on a complicated and far from explicit inductive procedure. On the other hand, in particular cases, such as one- and two-dimensional systems, it is easy to obtain stabilizing laws using a saturated linear feedback $u = \sigma(Fx)$, where σ is a saturation function. It is therefore natural to ask if such simple control laws can also be used in the general case. Unfortunately, Fuller established in [1] that this is already false for *triple integrators*, for certain classes of functions σ. A stronger negative result, which applies to basically arbitrary σ's, was given in [5]. One is thus motivated to search for other simple control rules, such as *linear combinations* and *compositions* of saturation nonlinearities. (In the language of neural networks, one wants control laws that are implementable by feedforward nets with several "hidden layers").

Recently, in his beautiful paper [6], Teel showed that, in the particular case of single-input multiple integrators of any dimension, such combinations of saturations are indeed sufficient to obtain stabilizing feedback controllers. (Roughly, $n/2$ saturations are needed.) In this work, we show how to extend the results of [6] to obtain an analogous solution for the general case. While the basic idea borrows heavily from [6], the technical details become far more complicated, because of the need to consider nonzero eigenvalues and to deal with multiple inputs. We remark that the approach in [6], as well as the one pursued here, is heavily based on the notion of CICS ("convergent-input implies convergent state") stability, which was studied in [2].

For reasons of space, this paper will only consider control laws with a rather special class of nonlinearities, for which the proofs are much simpler than for the general case. In a more detailed version (see [7]), we show that the same results hold for almost arbitrary "sigmoidal" nonlinearities.

1.1. Statement of the main result

We consider linear systems of the type

$$\sum : \quad \dot{x} = Ax + Bu.$$

It was stated (and a proof was sketched) in [4] that $\sum$ can be stabilized by means of a smooth bounded feedback if and only if:

(a) all eigenvalues of A have nonpositive real part, and

(b) all eigenvalues of the uncontrollable part of $\sum$ have strictly negative real parts.

Here we investigate the possibility of stabilizing linear systems satisfying the above conditions through the use of a special type of bounded feedback, which generalizes that introduced in [6] for the special case of multiple integrators. To be precise, we first assume given, for each pair (a, b) of positive real numbers such that $a \le 2b$, a nonempty class $\mathcal{S}_{a,b}$ consisting of Lipschitz functions $\sigma : \mathbb{R} \to \mathbb{R}$ which satisfy:

- $s\sigma(s) > 0$ for $s \neq 0$;
- $\sigma(s) = s$ for $|s| < \frac{a}{2}$;
- $|\sigma(s)| = b$ for $|s| \ge a$;
- $|\sigma(s)| \le b$ for all s.

Now let $\mathcal{F} = \{\mathcal{F}_n : n = 1, 2, \ldots\}$ be a collection of sets $\mathcal{F}_n$ of functions such that

- every $f \in \mathcal{F}_n$ is a globally Lipschitz function from $\mathbb{R}^n$ to $\mathbb{R}$ that vanishes at 0;
- each $\mathcal{F}_n$ contains the zero function $\mathbb{R}^n \to \mathbb{R}$;
- if $f \in \mathcal{F}_n$, $h : \mathbb{R}^n \to \mathbb{R}$ is linear, and $\sigma \in \mathcal{S}_{a,b}$ for some (a, b) such that $a \le 2b$, then it follows that the function $x \to \sigma(h(x) + f(x))$ is in $\mathcal{F}_n$;
- if $\nu \le n$ and $f \in \mathcal{F}_\nu$, then the function $(x_1, \ldots, x_n) \to f(x_1, \ldots, x_\nu)$ belongs to $\mathcal{F}_n$.

For any family $\mathcal{F} = \{\mathcal{F}_n\}$ as above and any $\varepsilon > 0$, we let $\mathcal{F}_{n,\varepsilon}$ denote the set of functions in $\mathcal{F}_n$ whose absolute value is bounded by ε. Our main result is as follows:

Theorem 1 *Let $\sum$ be a linear system $\dot{x} = Ax + Bu$ with state space $\mathbb{R}^n$ and input space $\mathbb{R}^m$. Suppose that all eigenvalues of A have nonpositive real part and all eigenvalues of the uncontrollable part of $\sum$ have strictly negative real parts. Then, for each $\varepsilon > 0$ there exist functions $k_i \in \mathcal{F}_{n,\varepsilon}$, $i = 1, 2, \cdots, m$ such that, with the feedback*

$$u_i = k_i(x), \qquad i = 1, 2, \cdots, m,$$

*This research was supported in part by US Air Force Grants 91-0343 and 91-0346 and by NSF Grant DMS-8902994.

the origin is a globally asymptotically stable equilibrium of the the resulting closed-loop system. Furthermore, the closed-loop system has the converging-input converging-state (CICS) property: if $e : [0, +\infty) \to \mathbb{R}^n$ is any bounded measurable function which converges to zero as $t \to +\infty$, then all the solutions of the system

$$\dot{x} = Ax + B(k_1(x), \cdots, k_m(x))' + e(t)$$

converge to zero as $t \to +\infty$.

2. Proof of the main theorem

Before proving Theorem 1, we need some preliminary results. First, we notice that under the conditions of the theorem there exists a linear change of coordinates of the state space that transforms $\sum$ into the block form

$$\sum : \quad \left\{ \begin{array}{rcl} \dot{x}_1 & = & A_1 x_1 + B_1 u, \quad x_1 \in \mathbb{R}^{n_1}, \\ \dot{x}_2 & = & A_2 x_2 + B_2 u, \quad x_2 \in \mathbb{R}^{n_2}, \end{array} \right.$$

where (i) $n_1 + n_2 = n$, (ii) all the eigenvalues of A_1 have zero real part, (iii) all the eigenvalues of A_2 have negative real part, and (iv) (A_1, B_1) is a controllable pair. Suppose that we find a feedback law $u = k(x_1) = (k_1(x_1), \cdots, k_m(x_1))'$, with $k_i \in \mathcal{F}_{n_1,\varepsilon}$ for $i = 1, 2, \cdots, m$, such that the required property is satisfied for the system $\dot{x}_1 = A_1 x_1 + B_1 k(x_1)$. Then this same feedback law will work for $\sum$ as well. Indeed, if $e_1(t), e_2(t)$ are vector functions of the appropriate sizes that converge to zero as $t \to +\infty$ then, if $t \to x_1(t)$ is any solution of the system $\dot{x}_1 = A_1 x_1 + B_1 k(x_1) + e_1(t)$, it follows that all the solutions $t \to x_2(t)$ of $\dot{x}_2 = A_2 x_2 + \sum_{i=1}^m b_i k_i(x_1(t)) + e_2(t)$ converge to zero (see Proposition 6.3.3 and Lemma 6.3.2 in [3]). Thus, in order to stabilize $\sum$, it is enough to stabilize the "critical subsystem" $\dot{x}_1 = A_1 x_1 + B_1 u$. Without loss of generality, in our proof of the theorem we will suppose that $\sum$ is already in this form.

Next we state two lemmas, whose proofs are deferred to the next section.

Lemma 2.1 Let $\sum : \dot{x} = Ax + ub$, $x \in \mathbb{R}^n$, be a controllable single-input linear system such that all the eigenvalues of A have zero real part.

(i) If 0 is an eigenvalue of A, then there is a linear change of coordinates $Tx = (y_1, \ldots, y_n)' = (y', y_n)'$ of $\mathbb{R}^n$ that puts $\sum$ in the form

$$\begin{array}{rcl} \dot{y} & = & A_1 y + (y_n + u) b_1, \\ \dot{y}_n & = & u, \end{array} \tag{1}$$

where the pair (A_1, b_1) is controllable.

(ii) If A has an eigenvalue of the form $i\omega$, with $\omega > 0$, then there is a linear change of coordinates $Tx = (y_1, \ldots, y_{n-1}, y_n)' = (y', y_{n-1}, y_n)'$ of $\mathbb{R}^n$ that puts $\sum$ in the form

$$\begin{array}{rcl} \dot{y} & = & A_1 y + (y_n + u) b_1, \\ \dot{y}_{n-1} & = & \omega y_n, \\ \dot{y}_n & = & -\omega y_{n-1} + u, \end{array} \tag{2}$$

where the pair (A_1, b_1) is controllable.

Lemma 2.2 Assume given real numbers $\omega > 0$ and $\varepsilon \geq \delta > 0$ such that the following conditions hold

- $\varepsilon < \frac{1}{8}\rho$,
- $\delta < \frac{\varepsilon}{\omega}$,

where $\rho = \min\{1, \omega\}$. Let $\sigma(s)$ be a continuous function satisfying:

- $|\sigma(s)| \leq \varepsilon$ for all s;
- $\sigma(s) = \varepsilon$ if $s > \delta$;
- $\sigma(s) = -\varepsilon$ if $s < -\delta$.

Let S denote the unit disk in $\mathbb{R}^2$. Pick any measurable function $u(t) : [0, \infty) \to [-\delta, \delta]$ and any two bounded measurable functions $\alpha(t), \beta(t)$ such that both $\alpha(t), \beta(t) \to 0$ as $t \to \infty$. Consider any solution $x(\cdot), y(\cdot)$ of the system

$$\begin{array}{rcll} \dot{x} & = & \omega y & + \alpha(t), \\ \dot{y} & = & -\omega x - \sigma(y + u(t)) & + \beta(t). \end{array} \tag{3}$$

Then there exists a $T > 0$ such that $(x(t), y(t)) \in S$ for all $t > T$.

2.1. The single-input case

We start with the single-input case, and prove the theorem by induction on the dimension of the system. As discussed earlier, we may assume that all eigenvalues of A have zero real part and the pair (A, B) is controllable.

For dimension zero, there is nothing to prove. Now assume given a single-input n-dimensional system, $n \geq 1$, and suppose that Theorem 1 has been established for all single-input systems of dimension $\leq n-1$. We consider separately the following two possibilities:

(i) zero is an eigenvalue of A,

(ii) zero is not an eigenvalue of A.

We want to prove the existence of a feedback $u = k(x)$, $k \in \mathcal{F}_{n,\varepsilon}$ that stabilizes the system $\sum : \dot{x} = Ax + Bu + e(t)$ for every bounded measurable vector function $e : [0, \infty) \to \mathbb{R}^n$ that converges to zero as $t \to +\infty$.

In the first case, we apply Lemma 2.1 (i), and rewrite our system in the form

$$\begin{array}{rcll} \dot{y} & = & A_1 y + (y_n + u) b_1 & + \tilde{e}(t), \\ \dot{y}_n & = & u & + e_n(t), \end{array} \tag{4}$$

where $\tilde{e}(t)$ and $e_n(t)$ still converge to zero as $t \to \infty$. (Note that if $n = 1$, only the second equation occurs.)

Let σ be any function in $\mathcal{S}_{\delta,\varepsilon}$, where $\delta \leq \varepsilon$ and ε is as in the statement of the theorem. We will use a feedback law of the type $u = -\sigma(y_n - v)$, where v will be chosen below so as to satisfy the constraint that $|v(t)| \leq \bar{\delta} < \frac{1}{4}\delta$ for all t. Since $e_n \to 0$ as $t \to \infty$, there exists $T_0 > 0$ so that $|e_n(t)| < \min\{|\sigma(s)| : s \geq \frac{1}{4}\delta - \bar{\delta}\}$ for all $t > T_0$. Therefore when $t > T_0$, along every trajectory of (4), with this choice of u, $\frac{d}{dt}(y_n^2) = -2y_n\big(\sigma(y_n - v) - e_n(t)\big)$, so it follows that $\frac{d}{dt}(y_n^2) < 0$

for all $t > T_0$ as long as $|y_n| \geq \frac{1}{4}\delta$. We conclude that for each trajectory $y(\cdot)$ there is some finite $T \geq T_0$ so that $y_n(t) \in (-\frac{\delta}{4}, \frac{\delta}{4})$ for all $t \geq T$. For such t the feedback law becomes linear, $u = -y_n + v$. Thus, for each trajectory $(y(\cdot), y_n(\cdot))$ of the closed-loop system there exists a time T such that the equations

$$\begin{array}{rcll} \dot{y} & = & A_1 y + v b_1 & + \tilde{e}(t) , \\ \dot{y}_n & = & -y_n - v & + e_n(t) \end{array} \tag{5}$$

are satisfied for $t \geq T$.

We must still pick $v = k(y)$. Applying the inductive hypothesis to the system $\dot{y} = A_1 y + v b_1$, we know that there exists a $k \in \mathcal{F}_{n-1,\delta}$ such that the system $\dot{y} = A_1 y + k(y) b_1 + \check{e}(t)$ is globally asymptotically stable for all possible $\check{e}(t) \to 0$. (In the special case $n = 1$, we let $v = 0$.) So, if we substitute $v = k(y)$ into (5), we see that $y(t) \to 0$ as $t \to \infty$ whenever $\tilde{e}(t) \to \infty$. Therefore $y_n(t)$ in (5) with $v = k(y(t))$ goes to zero as well, because it is a trajectory of a stable linear system with an input that converges to zero. Thus all solutions of (5) converge to zero, whenever $e(t) \to 0$. We conclude that $u = -\sigma(y_n - k(y))$ is as required. (This case is basically that treated in [6], with essentially the same proof.)

We now turn to case (ii). Since $n > 0$, A has an eigenvalue of the form $i\omega$, with $\omega > 0$, so Lemma 2.1 (ii) applies. We can then make a linear transformation of the state space that puts $\sum$ in the form

$$\begin{array}{rcll} \dot{y} & = & A_1 y + (y_n + u) b_1 & + \tilde{e}(t) , \\ \dot{y}_{n-1} & = & \omega y_n & + e_{n-1}(t) , \\ \dot{y}_n & = & -\omega y_{n-1} + u & + e_n(t) , \end{array} \tag{6}$$

where $y = (y_1, y_2, \cdots, y_{n-2})'$ and $\tilde{e}(t)$, $e_{n-1}(t)$, $e_n(t)$ converge to zero as $t \to \infty$. (Naturally, in the special case when $n = 2$, the first equation will be missing.)

Consider the last two equations in (6). Let $\varepsilon > 0$ be as in the statement of the theorem. From Lemma 2.2, we know that there is a $\sigma \in \mathcal{S}_{\delta,\varepsilon}$, for some $\delta \leq \varepsilon$ such that, as long as $|v(t)| < \delta$, every trajectory $t \to (y_{n-1}(t), y_n(t))$ of the system

$$\begin{array}{rcll} \dot{y}_{n-1} & = & \omega y_n & + e_{n-1}(t) , \\ \dot{y}_n & = & -\omega y_{n-1} - \sigma(y_n - v) & + e_n(t) \end{array} \tag{7}$$

enters the unit disk S and stays inside after a finite time.

More generally, consider the system

$$\dot{\xi} = f(\xi, v, v_1, v_2) , \tag{8}$$

where $\xi = (\xi_1, \xi_2)'$, and

$$f(\xi, v, v_1, v_2) = \begin{pmatrix} \omega\xi_2 + v_1 \\ -\omega\xi_1 - \sigma(\xi_2 - v) + v_2 \end{pmatrix} .$$

Since (8) is globally asymptotically stable when $v = v_1 = v_2 = 0$, there exists a strict smooth Lyapunov function $V(\xi)$ for the system $\dot{\xi} = f(\xi, 0, 0, 0)$. So, if we let $\alpha(\xi, v, v_1, v_2) = \bigtriangledown V(\xi) \cdot f(\xi, v, v_1, v_2)$, then $\alpha(\xi, 0, 0, 0) < 0$ when $\xi \neq 0$. Consider two level surfaces $V(\xi) = L_1$ —with $L_1 > 0$— and $V(\xi) = L_2$ in $\mathbb{R}^2$, such that $V(\xi) \leq L_1$ implies $||\xi|| < \frac{1}{4}\delta$, and $||\xi|| < 1$ implies $V(\xi) \leq L_2$. Then, $C = \{\xi \in \mathbb{R}^2 : L_1 \leq V(\xi) \leq L_2\}$ is compact and $0 \notin C$. Since $\alpha(\xi, 0, 0, 0) < 0$ for $\xi \in C$, there exists a $\bar{\delta} < \frac{\delta}{4}$ such that $\alpha(\xi, v, v_1, v_2) < 0$ whenever $\xi \in C$ and $|v|, |v_1|, |v_2| < \bar{\delta}$. Since $e_{n-1}(t), e_n(t) \to 0$ as $t \to \infty$, there exists a $T_0 > 0$ such that $|e_i(t)| \leq \bar{\delta}$, $i = n-1, n$, for $t \geq T_0$. Therefore, if we choose $v(t)$ so that $|v(t)| \leq \bar{\delta}$, then any trajectory of (7) enters the set where $V(\xi) \leq L_1$ in finite time $T > T_0$, and stays there for $t \geq T$. For such t, the solution stays inside the circle $y_{n-1}^2 + y_n^2 = (\frac{\delta}{4})^2$, so in particular $|y_n(t)| < \frac{1}{4}\delta$. Since $|v(t)| \leq \bar{\delta}$ for all t, we conclude that $\sigma(y_n - v) = y_n - v$ for $t \geq T$. So, if $u = -\sigma(y_n - v)$ and $|v| < \bar{\delta}$, then (6) becomes

$$\begin{array}{rclll} \dot{y} & = & A_1 y & + v b_1 & + \tilde{e}(t) , \\ \dot{y}_{n-1} & = & \omega y_n & & + e_{n-1}(t) , \\ \dot{y}_n & = & -\omega y_{n-1} - y_n & + v & + e_n(t) . \end{array} \tag{9}$$

Applying the inductive hypothesis to the system $\dot{y} = A_1 y + v b_1 + \check{e}(t)$, we know that there is a feedback $v = k(y) \in \mathcal{F}_{n-2,\bar{\delta}}$ such that all the trajectories of $\dot{y} = A_1 y + k(y) b_1 + \check{e}(t)$ go to zero as $t \to \infty$, for every vector function $\check{e}$ that converges to zero. (If $n = 2$, we just let $v = 0$.) It then follows that the closed-loop system (6) with $u = -\sigma(y_n - k(\bar{y}))$ is globally asymptotically stable, just as in the case of zero eigenvalues. The proof for the single-input case is complete. □

2.2. The multi-input case

Next, we deal with the general case of $m > 1$ inputs, and prove Theorem 1 by induction on m.

First, we know from Section 2.1 that the theorem is true if $m = 1$. Assume that Theorem 1 has been established for all k-input systems, for all $k \leq m - 1$, and take an m-input system $\sum : \dot{x} = Ax + Bu + e(t)$ where $e(t)$ is any decaying vector function.

Assume without loss of generality that the first column b_1 of B is nonzero, and consider the Kalman controllability decomposition of the system $\sum_1 : \dot{x} = Ax + b_1 u$ (see [3], Lemma 3.3.3). We conclude that, under a change of coordinates $y = T^{-1}x$, $\sum_1$ has the form

$$\begin{array}{rcl} \dot{y}_1 & = & A_1 y_1 + A_2 y_2 + \bar{b}_1 u , \\ \dot{y}_2 & = & A_3 y_2 , \end{array}$$

where $(A_1, \bar{b}_1)$ is a controllable pair. In these coordinates, $\sum$ has the form

$$\begin{array}{rll} \dot{y}_1 = & A_1 y_1 + A_2 y_2 & + \bar{b}_1 u_1 + \bar{B}_1 \bar{u} + \tilde{e}_1(t) , \\ \dot{y}_2 = & A_3 y_2 & + \bar{B}_2 \bar{u} + \tilde{b}e_2(t) , \end{array} \tag{10}$$

where $\bar{u} = (u_2, \cdots, u_m)'$ and $\bar{B}_1, \bar{B}_2$ are appropriate matrices. Let n_1, n_2 denote the dimensions of y_1, y_2 respectively. Since (10) is controllable, we conclude that the (m-1)-input subsystem $\dot{y}_2 = A_3 y_2 + \bar{B}_2 \bar{u}$ is controllable as well. By the inductive hypothesis, this subsystem can be stabilized with a feedback

$$\bar{u} = (k_2(y_2), \cdots, k_m(y_2))', k_i \in \mathcal{F}_{n_2,\varepsilon}, 2 \leq i \leq m, \tag{11}$$

and the CICS property holds. Therefore all solutions of the system

$$\dot{y}_2 = A_3 y_2 + \bar{B}_2(k_2(y_2), \cdots, k_m(y_2))' + \tilde{e}_2(t) \tag{12}$$

converge to zero. Similarly, for the single-input controllable system $\dot{y}_1 = A_1 y_1 + \bar{b}_1 u_1$ there is a feedback

$$u_1 = k_1(y_1) \in \mathcal{F}_{n_1,\varepsilon} \tag{13}$$

such that the resulting closed-loop system is globally asymptotically stable and the CICS property holds. Since $A_2 y_2(t) + \bar{B}_1(k_2(y_2(t)), \cdots, k_m(y_2(t)))' + \tilde{e}_1(t)$ still converges to zero for any solution $y_2(t)$ of (12), we conclude that the coordinate $y_1(t)$ in equation (10), with $u_1, \bar{u}$ given by (11) and (13), converges to zero. Therefore, the feedback given in (11) and (13) globally stabilizes the system $\dot{x} = Ax + Bu$, and the CICS property holds. □

Remark 2.3 The feedback constructed here has the CICS property. This was necessary for the inductive step of the proof for the multi-input case, and will also be needed below to deal with the partially observed case. If we are only interested in stabilizing a single-input system, then we could use more general feedbacks. It is not hard to see that not all stabilizing feedbacks have the CICS property, even if they are linear near the origin. To illustrate this, consider the double integrator: $\dot{x} = y$, $\dot{y} = u$.

Let $\sigma(s)$ be an odd continuous function such that $s\sigma(s) > 0$ for $s \neq 0$, $\sigma(s) = s$ for $|s| < \frac{1}{2}$ and $\sigma(s) = \frac{1}{s}$ for $s > 1$. Then the feedback $u = -\sigma(x+y)$ stabilizes the double integrator. (This can be established by verifying that $V(x,y) = \int_0^{x+y} \sigma(s)\,ds + \frac{1}{2}y^2$ is a Lyapunov function for the closed-loop system with $u = -\sigma(x+y)$, and applying the LaSalle Invariance Principle.) Let $e(t) = \sigma\left(\log(t+1) + \frac{1}{t+1}\right) - \frac{1}{(t+1)^2}$. Then clearly $e(t) \to 0$ as $t \to \infty$. But not every trajectory of the system

$$\begin{aligned} \dot{x} &= y\ , \\ \dot{y} &= -\sigma(x+y) + e(t) \end{aligned} \tag{14}$$

converges to zero. For instance, $x(t) = \log(t+1)$, $y(t) = \frac{1}{t+1}$ is a solution of (14), but $x(t) \to \infty$ as $t \to \infty$. □

3. The Proofs of Lemma 2.1 and Lemma 2.2

Proof of Lemma 2.1

We first prove (i). If 0 is an eigenvalue of A, then there exists a nonzero n-dimensional row vector v such that $vA = 0$. Let $\xi : \mathbb{R}^n \to \mathbb{R}$ be the linear function $x \to vx$. Then, along trajectories of $\sum$, $\dot{\xi} = (vb)u$. Controllability of (A,b) implies that $vb \neq 0$. So we may assume that $vb = 1$. Make a linear change of coordinates $Tx = (z', z_n)'$ so that $z_n \equiv \xi$. Then the system equations are of the form $\dot{z} = A_1 z + z_n \tilde{b}_1 + u\tilde{b}_2$, $\dot{z}_n = u$. It is clear that every eigenvalue of A_1 also has zero real part. Now change coordinates again by letting $y = z + z_n \tilde{b}_3$, $y_n = z_n$, where the vector $\tilde{b}_3$ will be chosen below. Then the system equations become $\dot{y} = A_1 y + y_n(\tilde{b}_1 - A_1\tilde{b}_3) + u(\tilde{b}_2 + \tilde{b}_3)$, $\dot{y}_n = u$. Choose $\tilde{b}_3$ to be a solution of $A_1\tilde{b}_3 + \tilde{b}_3 = \tilde{b}_2 - \tilde{b}_1$. (This is possible because $A_1 + I$ is nonsingular.) Let $b_1 = \tilde{b}_1 - A_1\tilde{b}_3 = \tilde{b}_2 + \tilde{b}_3$. The equations now become $\dot{y} = A_1 y + (y_n + u)b_1$, $\dot{y}_n = u$, as desired.

We now prove (ii). Let $i\omega$, $\omega > 0$, be an eigenvalue of A. Then $-\omega^2$ is an eigenvalue of A^2. So there is a nonzero n-dimensional row vector $\tilde{v}$ such that $\tilde{v}A^2 = -\omega^2\tilde{v}$. Let $\tilde{w} = \omega^{-1}\tilde{v}A$. Then $\tilde{w}A = -\omega\tilde{v}$ and $\tilde{w}A^2 = -\omega^2\tilde{w}$. Moreover, $\tilde{w}$ cannot be a multiple of $\tilde{v}$ because, if $\tilde{w} = \lambda\tilde{v}$, then $\lambda\omega$ would be a nonzero real eigenvalue of A. So the linear span S of $\tilde{v}$ and $\tilde{w}$ is a two-dimensional subspace all whose members v satisfy $vA^2 = -\omega^2 v$. In particular, we can choose $v \in S$ such that $vb = 0$ but $v \neq 0$. If we then define w by $w = \omega^{-1}vA$, we have $wA = -\omega v$. Moreover, wb cannot vanish for, if it did, the subspace $\{x : vx = wx = 0\}$ would contain b and be invariant under A, contradicting controllability. So, after multiplying both v and w by a constant, if necessary, we may assume that $wb = 1$. Let ξ, η be the linear functionals $x \to vx$, $x \to wx$. Then, along trajectories of $\sum$, $\dot{\xi} = \omega\eta$ and $\dot{\eta} = -\omega\xi + u$. Make a linear change of coordinates $Tx = (z', z_{n-1}, z_n)'$ so that $z_{n-1} \equiv \xi$, $z_n \equiv \eta$. Then the system equations are of the form $\dot{z} = A_1 z + z_{n-1}\tilde{b}_1 + z_n\tilde{b}_2 + u\tilde{b}_3$, $\dot{z}_{n-1} = \omega z_n$, $\dot{z}_n = -\omega z_{n-1} + u$, and every eigenvalue of A_1 has zero real part. Now change coordinates again by letting $y = z + z_{n-1}\tilde{b}_4 + z_n\tilde{b}_5$, $y_{n-1} = z_{n-1}$, $y_n = z_n$, where the vectors $\tilde{b}_4$, $\tilde{b}_5$ will be chosen below. Then the system equations become

$$\begin{aligned} \dot{y} &= A_1 y + y_{n-1}(\tilde{b}_1 - A_1\tilde{b}_4 - \omega\tilde{b}_5) \\ &\quad + y_n(\tilde{b}_2 - A_1\tilde{b}_5 + \omega\tilde{b}_4) + u(\tilde{b}_3 + \tilde{b}_5), \\ \dot{y}_{n-1} &= \omega y_n\ , \\ \dot{y}_n &= -\omega y_{n-1} + u. \end{aligned} \tag{15}$$

If we could choose $\tilde{b}_4$, $\tilde{b}_5$ such that

$$\tilde{b}_1 - A_1\tilde{b}_4 - \omega\tilde{b}_5 = 0 \tag{16}$$

and

$$\tilde{b}_2 - A_1\tilde{b}_5 + \omega\tilde{b}_4 = \tilde{b}_3 + \tilde{b}_5, \tag{17}$$

then we could let

$$b_1 = \tilde{b}_2 - A_1\tilde{b}_5 + \omega\tilde{b}_4 = \tilde{b}_3 + \tilde{b}_5 \tag{18}$$

and the system equations would become $\dot{y} = A_1 y + (y_n + u)b_1$, $\dot{y}_{n-1} = \omega y_n$, $\dot{y}_n = -\omega y_{n-1} + u$, as desired. To prove the existence of $\tilde{b}_4$ and $\tilde{b}_5$, we rewrite (17) as $(A_1 + I)\tilde{b}_5 = \tilde{b}_2 - \tilde{b}_3 + \omega\tilde{b}_4$, multiply both sides by ω, and plug in the value of $\omega\tilde{b}_5$ given by (16), namely, $\omega\tilde{b}_5 = \tilde{b}_1 - A_1\tilde{b}_4$. We end up with the equation $(A_1^2 + A_1 + \omega^2)\tilde{b}_4 = A_1\tilde{b}_1 + \tilde{b}_1 - \omega\tilde{b}_2 + \omega\tilde{b}_3$. Since the eigenvalues of A_1 have zero real part, the matrix $A_1^2 + A_1 + \omega^2$ is nonsingular, so $\tilde{b}_4$ exists. □

Proof of Lemma 2.2

We begin the proof with the simpler case in which $\alpha(t) = \beta(t) = 0$ for all t. Then (3) takes the form

$$\begin{aligned} \dot{x} &= \omega y\ , \\ \dot{y} &= -\omega x - \sigma(y + u(t))\ . \end{aligned} \tag{19}$$

First, we establish the following facts for (19):

(I) Consider any trajectory that starts from a point (x_0, y_0) with $x_0 > \frac{1}{2}$, $y_0 = 2\delta$. Then, either there is some time $t > 0$ such that $x(t) = \frac{1}{2}$ and $|y(t)| \leq 2\delta$ or there is a $t > 0$ such that $y(t) = -2\delta$ and $\frac{1}{2} < x(t) < x_0 + 2\delta$.

(II) For any trajectory γ starting at a point (x_1, y_1) with $x_1 > \frac{1}{2}$, $y_1 = -2\delta$, there is a $t > 0$ such that $y(t) = -2\delta$ and $x(t) = -x_1 + \frac{2\varepsilon}{\omega}$.

We introduce the following regions:

$$\begin{aligned}
\mathcal{R} &= \{(x,y) : |y| \le 2\delta\} , \\
\mathcal{R}^+ &= \{(x,y) : y > 2\delta\} , \\
\mathcal{R}^- &= \{(x,y) : y < -2\delta\} , \\
\mathcal{R}_0 &= \{(x,y) : |x| \le \tfrac{1}{2}, |y| \le 2\delta\} .
\end{aligned}$$

To prove (I), we consider a trajectory starting at an initial point (x_0, y_0) with $x_0 > \frac{1}{2}$, $y_0 = 2\delta$. Notice that $\dot{x} = \omega y$, so when the trajectory is above the x-axis, the x-coordinate is increasing. On the other hand, $\dot{y} = -\omega x - \sigma(y+u) < -\frac{1}{2}\omega + \varepsilon < -\frac{1}{4}\omega$, so y is decreasing and therefore also $y(t) < 2\delta$ for $t > 0$. It follows that $\frac{\dot{y}}{\dot{x}} < \frac{-\frac{1}{4}\omega}{2\delta\omega} < -1$ (as long as $0 < y(t) \le 2\delta$). Let l_1 denote the line through (x_0, y_0) with slope -1. Then the trajectory must eventually enter the region $y < 0$ at a point below l_1. So if $(x_1, 0)$ is the first intersection between the trajectory and the x-axis, then it must hold that $x_1 < x_0 + 2\delta$. Let l_2 denote the vertical line $x = x_0 + 2\delta$. It is clear that the trajectory cannot cross l_2 in $\mathcal{R}$ since $\dot{x} = \omega y < 0$ there. Furthermore, $y(t)$ is decreasing in the region $(\mathcal{R} - \mathcal{R}_0) \cap \{x > 0\}$ with rate $\dot{y}(t) = -\omega x(t) - \sigma(y(t) + u) \le -\frac{1}{2}\omega + \varepsilon < -\frac{1}{4}\omega$. Therefore $\frac{\dot{y}}{\dot{x}} = \frac{\dot{y}(t)}{\omega y(t)} > \frac{\frac{1}{4}\omega}{2\delta\omega} > 1$ as long as $-2\delta \le y(t) < 0, x(t) > \frac{1}{2}$. Thus, if the trajectory does not enter $\mathcal{R}_0$ it must hit the line $y = -2\delta$ at $(x_2, -2\delta)$ with $x_2 < x_1$. This establishes (I).

To prove (II), we notice that γ must satisfy $\dot{y}(0) < 0$. So γ is in the region $\mathcal{R}^-$ for small $t > 0$. Let $T > 0$ be as large as possible so that $(x(t), y(t)) \in \mathcal{R}^-$ for all $0 < t < T$. Then for $t < T$ we have $y(t) \le -2\delta$ and $y(t) + u(t) \le -\delta$. Therefore $\sigma(y(t) + u(t)) = -\varepsilon$ for $t < T$ and (19) becomes a linear system with a constant perturbation:

$$\begin{aligned}
\dot{x} &= \omega y , \\
\dot{y} &= -\omega x + \varepsilon .
\end{aligned} \tag{20}$$

With $A = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}$, the solution of (20) with initial condition (x_1, y_1) is

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = e^{At} \begin{pmatrix} x_1 - \frac{\varepsilon}{\omega} \\ y_1 \end{pmatrix} + \begin{pmatrix} \frac{\varepsilon}{\omega} \\ 0 \end{pmatrix} . \tag{21}$$

So for $t < T$, the trajectory of (19) is also given by (21). Notice that the first term on the right hand side of (21) is just the solution of the system $\dot{\xi} = A\xi$ starting at $(x_1 - \frac{\varepsilon}{\omega}, y_1)$, which is a part of a circle centered at the origin and passing through $(x_1 - \frac{\varepsilon}{\omega}, y_1)$. Therefore T is finite, and when $t = T$ we have $y(t) = -2\delta$, $x(t) = -(x_1 - \frac{\varepsilon}{\omega}) + \frac{\varepsilon}{\omega} = -x_1 + \frac{2\varepsilon}{\omega}$. So (II) is proved.

Notice that (I) and (II) imply that a trajectory starting from (x_0, y_0) with $x_0 > \frac{1}{2}$, $y_0 = 2\delta$ either enters $\mathcal{R}_0$ in finite time or hits $y = -2\delta$ at $(\bar{x}_0, \bar{y}_0)$ with $-x_0 + 2(\frac{\varepsilon}{\omega} - \delta) < \bar{x}_0 < 0$. A similar analysis starting at $(\bar{x}_0, \bar{y}_0)$ shows that the trajectory either enters $\mathcal{R}_0$ or hits $y = 2\delta$ with an x-coordinate which is positive but closer to zero by at least $2(\frac{\varepsilon}{\omega} - \delta)$. Repeating this argument we conclude that after a finite time the trajectory enters $\mathcal{R}_0$. Since any part of each trajectory of system (19) which is outside $\mathcal{R}$ is an arc of a circle centered at $(\pm\frac{\varepsilon}{\omega}, 0)$, it follows that trajectories starting in $\mathcal{R}^+$ and $\mathcal{R}^-$ must cross the boundary of $\mathcal{R}$. Similarly, from the proof of (I), we conclude that any trajectory starting inside $\mathcal{R}$ either enters $\mathcal{R}_0$ or hits the boundary of $\mathcal{R}$. Therefore the proof of the simple case will be complete if we show:

(III) Any trajectory starting from $\mathcal{R}_0$ is completely contained in S.

To prove this, all we need is to construct a region $\mathcal{D}$ such that $\mathcal{R}_0 \subset \mathcal{D} \subset S$, with the property that every trajectory starting on the boundary of $\mathcal{D}$ enters $\mathcal{D}$.

Let c_1, c_2 be the circles centered at $(-\frac{\varepsilon}{\omega}, 0)$ and $(\frac{\varepsilon}{\omega}, 0)$ respectively, with radius $r = \left((\frac{1}{2} + 2\delta - \frac{\varepsilon}{\omega})^2 + (2\delta)^2\right)^{\frac{1}{2}}$. Then c_1 intersects the line $y = 2\delta$ at $P_1(-\frac{1}{2} - 2\delta, 2\delta)$ and $P_2(\frac{1}{2} + 2\delta - \frac{2\varepsilon}{\omega}, 2\delta)$, and c_2 intersects the line $y = -2\delta$ at $P_3(\frac{1}{2} + 2\delta, -2\delta)$ and $P_4(-\frac{1}{2} - 2\delta + \frac{2\varepsilon}{\omega}, -2\delta)$. Let d_1, d_5 denote the small arcs $\breve{P_1P_2}$ of c_1, and $\breve{P_3P_4}$ of c_2 respectively. Let $d_2, d_3, d_4, d_6, d_7, d_8$ be the line segments from P_2 to $(\frac{1}{2}, 2\delta)$, from $(\frac{1}{2}, 2\delta)$ to $(\frac{1}{2} + 2\delta, 0)$, from $(\frac{1}{2} + 2\delta, 0)$ to P_3, from P_4 to $(-\frac{1}{2}, -2\delta)$, from $(-\frac{1}{2}, -2\delta)$ to $(-\frac{1}{2} - 2\delta, 0)$, and from $(-\frac{1}{2} - 2\delta, 0)$ to P_1 respectively. Then d_1 through d_8 enclose a compact region $\mathcal{D}$ which satisfies $\mathcal{R}_0 \subset \mathcal{D} \subset S$.

From the proof of (II), we know that trajectories starting on d_1 or d_5 continue along d_1 or d_5 and finally enter $\mathcal{D}$. From the proof of (I), we see that trajectories starting on d_3, d_4, d_7, d_8 enter $\mathcal{D}$. On d_2, notice that $x > \frac{1}{2} - 2(\frac{\varepsilon}{\omega} - \delta) > \frac{1}{4}$ because $\frac{\varepsilon}{\omega} - \delta < \frac{1}{8}$; therefore $\dot{y} < -\omega x + \varepsilon < -\frac{1}{4}\omega + \varepsilon < 0$. Similarly, $\dot{y} > 0$ on d_6. So trajectories starting on d_2, d_6 enter $\mathcal{D}$. Therefore $\mathcal{D}$ has the desired properties.

Now, we turn to the general case when $\alpha(t) \to 0$, $\beta(t) \to 0$ as $t \to \infty$. It is enough to show that there exists a $\tau > 0$ such that (I), (III) and (II)$'$ hold for (3) if $|\alpha(t)| < \tau$, $|\beta(t)| < \tau$ for all $t > 0$. Here (I) and (III) are as before and (II)$'$ is the following property:

(II)$'$ every trajectory starting at a point of the form (x_1, y_1) with $x_1 > \frac{1}{2}$, $y_1 = -2\delta$, will eventually intersect the line $y = -2\delta$ at a point of the form (x_2, y_2) with $x_2 > -x_1 + \frac{\varepsilon}{\omega} + \delta$.

To prove (I), choose $\tau > 0$ small enough so that:

- $\tau < \frac{1}{4}\omega$,
- $k = \frac{\frac{1}{4}\omega - \tau}{2\delta\omega + \tau} > \frac{1}{2}(1 + \frac{1}{8\delta}) \quad (> 1)$,
- $\frac{\frac{1}{4}\omega - \tau}{\tau} > \frac{1 + 8\delta}{1 - 8\delta} \quad (> \frac{k}{k-1})$,

where the last two inequalities can be satisfied since $\delta < \frac{1}{8}$. Let l_3 denote the line through (x_0, y_0) with slope $-k$. Then on l_3 in $\mathcal{R} - \mathcal{R}_0$, and as long as $y > 0$ we have $\dot{x} < 2\delta\omega + \tau$, and $\dot{y} < -\omega x + \varepsilon + \tau < -\frac{1}{4}\omega + \tau < 0$. It then follows that $\frac{\dot{y}}{\dot{x}} < -k$ if $\dot{x} > 0$. So either the trajectory enters $\mathcal{R}_0$ or hits the x-axis at $(x_1, 0)$ with $x_1 < x_0 + \frac{2\delta}{k}$. Similarly, let l_4 denote the line through $(x_0 + \frac{2\delta}{k}, 0)$ and $(x_0 + 2\delta, -2\delta)$. Then the slope of l_4 is $\frac{-k}{k-1}$. On l_4 and while $y < 0$, we have $\dot{x} \le \tau$, $\dot{y} < -\frac{1}{4}\omega + \tau$. So if $\dot{x} > 0$ then $\frac{\dot{y}}{\dot{x}} < \frac{-k}{k-1}$. Therefore either the trajectory enters $\mathcal{R}_0$ or hits $y = -2\delta$ with $x < x_0 + 2\delta$.

To guarantee that (II)$'$ holds, we require in addition that $r = \frac{2\sqrt{2}\pi}{\omega}\tau < \min\{\delta, \frac{1}{2}(\frac{\varepsilon}{\omega} - \delta)\}$. Since $\dot{y}(0) < 0$ as before, the trajectory $(x(t), y(t))$ must enter $\mathcal{R}^-$. Let T_1 be as large as possible so that $(x(t), y(t)) \in \mathcal{R}^-$ for all $t < T_1$. Then, for $t < T_1$, $(x(t), y(t))$ is a solution of the system

$$\begin{array}{rcll} \dot{x} & = & \omega y & + \alpha(t) , \\ \dot{y} & = & -\omega x + \varepsilon & + \beta(t) . \end{array} \tag{22}$$

Let $(\tilde{x}(t), \tilde{y}(t))$ be the solution of the system (20) with the same initial point $(x_1, -2\delta)$. Then $(\tilde{x}(t), \tilde{y}(t))$ must hit $y = -2\delta + r$ at a time $T_2 < \frac{2\pi}{\omega}$. Therefore for $t \le \min\{T_1, T_2\}$, it follows that

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} - \begin{pmatrix} \tilde{x}(t) \\ \tilde{y}(t) \end{pmatrix} = \int_0^t e^{A(t-s)} \begin{pmatrix} \alpha(t) \\ \beta(t) \end{pmatrix} ds , \tag{23}$$

whose norm is $\le \sqrt{2}\tau\frac{2\pi}{\omega} = r$ (since $||e^{At}|| = 1$). If $T_1 \ge T_2$, then $|y(T_2) - \tilde{y}(T_2)| < r$. Therefore $y(T_2) > \tilde{y}(T_2) - r = -2\delta + r - r = -2\delta$, which is a contradiction. So $T_1 < T_2$ and $|x(T_1) - \tilde{x}(T_1)| < r$, from which $x(T_1) > \tilde{x}(T_1) - r > \tilde{x}(T_2) - r > -x_1 + \frac{2\varepsilon}{\omega} - 2r \ge -x_1 + \frac{2\varepsilon}{\omega} - (\frac{\varepsilon}{\omega} - \delta)$.

Finally, (III) is a trivial consequence of (I), (II)$'$ for (3), (III) for (19), and the property that the coordinates of the right-hand-side of (23) is bounded by r for $t < \frac{2\pi}{\omega}$. Thus the lemma is proved. □

4. Output feedback

We now turn to the problem of dynamic output feedback stabilization. In this section, we use feedbacks of the kind described in Theorem 1.

Assume we have a system $\dot{x} = Ax + Bu$ with an $\mathbb{R}^p$-valued output $y = Cx$. We assume that (A, B) is stabilizable, (A, C) is detectable, and A has no eigenvalues with strictly positive real part.

We proceed as in the classical case of stabilization by a linear feedback. That is, we consider the composite system

$$\begin{array}{rcl} \dot{x} & = & Ax + Bu , \\ y & = & Cx , \\ \dot{z} & = & Az + Bu + L(y - Cz) , \\ u & = & k(z) , \end{array} \tag{24}$$

where k is any bounded stabilizing feedback for the system: $\dot{x} = Ax + Bu$ and having the CICS property. The error $e = z - x$ then satisfies $\dot{e} = (A - LC)e$. If (A, C) is detectable, then we can choose L so that all the eigenvalues of $A - LC$ are negative. Therefore, $e(t) \to 0$ as $t \to \infty$, for any initial condition $x(0)$ and $z(0)$ of the original system and its observer. The equation for x then becomes

$$\dot{x} = Ax + Bk(x + e(t)). \tag{25}$$

We know that the trajectories of $\dot{x} = Ax + Bk(x)$ go to zero; the problem is whether the same is true for (25).

To do this for the feedback defined in Theorem 1, we recall that this feedback is globally Lipschitz. Therefore, we can rewrite (25) as

$$\dot{x} = Ax + Bk(x) + \tilde{e}, \tag{26}$$

where $\tilde{e} = B(k(x + e(t)) - k(x))$ is bounded by a constant multiple of $e(t)$. Thus $\tilde{e} \to 0$. Then Theorem 1 implies that all solutions of (26) converge to zero, i.e. the full system (24) is globally asymptotically stable. (Note that local stability is no problem.) Summarizing, we have the following result:

Theorem 2 *Consider a linear system of the form $\dot{x} = Ax + Bu$, $y = Cx$, such that A has no eigenvalues with positive real part, (A, B) is stabilizable and (A, C) is detectable. Then, for any observer system $\dot{z} = Az + Bu + L(y - Cz)$, and for any bounded stabilizing feedback $u \to k(z)$ given by Theorem 1, the composite system $\dot{x} = Ax + Bk(z)$, $\dot{z} = Az + Bk(z) + L(Cx - Cz)$ has the origin as a globally asymptotically stable equilibrium .* □

We remark that, since the stabilizing feedback given in this paper makes the resulting closed-loop system have the CICS property, Theorem 2 holds for *any* linear observer. In [4], the same problem was also considered, but the result in [4] required that all observer poles have real part less than a certain margin. So Theorem 2 answers a question posed in [4], namely whether an arbitrary linear observer can be used for stabilization.

References

[1] Fuller, A.T., "In the large stability of relay and saturated control systems with linear controllers," *Int. J. Control 1969, Vol. 10:* 457-480.

[2] Sontag, E.D., "Remarks on stabilization and input-to-state stability," *Proc. IEEE Conf. Decision and Control, Tampa, Dec. 1989*, IEEE Publications, 1989, pp. 1376-1378.

[3] Sontag, E.D., *Mathematical Control Theory*, Springer-Verlag, 1990, New York.

[4] Sontag, E.D. and H.J. Sussmann, "Nonlinear output feedback design for linear systems with saturating controls," *Proc. IEEE CDC, Honolulu, Dec. 1990,* IEEE Publications, 1990: 3414-3416.

[5] Sussmann, H. J. and Y. Yang, "On the stabilizability of multiple integrators by means of bounded feedback controls" *Proc. IEEE CDC, Brighton, UK, Dec. 1991.* IEEE Publications, 1991: 70-73.

[6] Teel, A.R., "Global stabilization and restricted tracking for multiple integrators with bounded controls," *Systems and Control Letters* **18**(1992): to appear.

[7] Yang, Y., E.D. Sontag, and H.J. Sussmann, "On the stabilization of linear systems using combinations of saturation nonlinearities," to appear.