Stabilization of Nonlinear Systems via State Feedback

Part I: Relative Degree, Normal Forms, Zero Dynamics

© Alberto Isidori, 2019

We consider the class of single-input single-output nonlinear systems that can be modeled by equations of the form

$$\begin{aligned} \dot{x} &= f(x) + g(x)u \\ y &= h(x) \end{aligned} \tag{1}$$

in which $x \in \mathbb{R}^n$ and in which $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^n$ are smooth maps and $h : \mathbb{R}^n \to \mathbb{R}$ is a smooth function.

If the coordinates in which the state space description is provided are appropriately transformed, the equations describing the system can be brought to a form in which the design of feedback laws is facilitated.

In the case of a linear system, a change of coordinates consists in replacing the original state vector x with a new vector $\tilde{x}$ related to x by means of linear transformation $\tilde{x} = Tx$ in which T is a nonsingular matrix.

If the system is nonlinear, it is more appropriate to allow also for nonlinear changes of coordinates. A nonlinear change of coordinates is a transformation $\tilde{x} = \Phi(x)$ in which $\Phi(\cdot)$ is a map $\mathbb{R}^n \to \mathbb{R}^n$. Such a map qualifies for a change of coordinates if:

(i) $\Phi(\cdot)$ is invertible, i.e. there exists a map $\Phi^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ such that $\Phi^{-1}(\Phi(x)) = x$ for all $x \in \mathbb{R}^n$ and $\Phi(\Phi^{-1}(\tilde{x})) = \tilde{x}$ for all $\tilde{x} \in \mathbb{R}^n$

(ii) $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are both smooth mappings, i.e. have continuous partial derivatives of any order.

A transformation of this type is called a global diffeomorphism. A transformation defined only in a neighborhood of a given point is called a local diffeomorphism.

Relative degree

The nonlinear system (1) is said to have relative degree r at a point x° if: [1]

(i) $L_g L_f^k h(x) = 0$ for all x in a neighborhood of x° and all $k < r-1$

(ii) $L_g L_f^{r-1} h(x^\circ) \neq 0$.

Note that the concept thus introduced is a local concept, namely r may depend on the specific point x° where the functions $L_g L_f^k h(x)$ are evaluated. The value of r may be different at different points of $\mathbb{R}^n$ and there may be points where a relative degree cannot be defined. This occurs when the first function of the sequence

$$L_g h(x),\ L_g L_f h(x), \ldots,\ L_g L_f^k h(x), \ldots$$

which is not identically zero (in a neighborhood of x°) is zero exactly at the point $x = x^\circ$. However, since $f(x), g(x), h(x)$ are smooth, the set of points where a relative degree can be defined is an open and dense subset of $\mathbb{R}^n$.

[1] Let λ be real-valued function and f a vector field, both defined on a subset U of $\mathbb{R}^n$. The function $L_f \lambda$ is the real-valued function defined as

$$L_f \lambda(x) = \sum_{i=1}^{n} \frac{\partial \lambda}{\partial x_i} f_i(x) := \frac{\partial \lambda}{\partial x} f(x)\,.$$

This function is sometimes called the derivative of λ along f. If g is another vector field, the notation $L_g L_f \lambda(x)$ stands for the derivative of the real-valued function $L_f \lambda$ along g and the notation $L_f^k \lambda(x)$ stands for the derivative of the real-valued function $L_f^{k-1} \lambda$ along f.

Assume the system at time $t=0$ is in the state $x(0)=x^\circ$ and let's calculate the value of the output $y(t)$ and of its derivatives with respect to time $y^{(k)}(t)$, for $k=1,2,\ldots$, at $t=0$. We obtain

$$y(0)=h(x(0))=h(x^\circ)$$

and

$$y^{(1)}(t)=\frac{\partial h}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}=\frac{\partial h}{\partial x}[f(x(t))+g(x(t))u(t)]=L_f h(x(t))+L_g h(x(t))u(t)\,.$$

At time $t=0$,

$$y^{(1)}(0)=L_f h(x^\circ)+L_g h(x^\circ)u(0)\,,$$

from which it is seen that, if $r=1$, the value $y^{(1)}(0)$ is an affine function of $u(0)$. Otherwise, suppose r is larger than 1. If $|t|$ is small, $x(t)$ remains in a neighborhood of x° and hence $L_g h(x(t))=0$ for all such t. As a consequence

$$y^{(1)}(t)=L_f h(x(t))\,.$$

This yields

$$y^{(2)}(t)=L_f h\frac{\mathrm{d}x}{\mathrm{d}t}=L_f h[f(x(t))+g(x(t))u(t)]=L_f^2 h(x(t))+L_g L_f h(x(t))u(t)\,.$$

At time $t=0$,

$$y^{(2)}(0)=L_f^2 h(x^\circ)+L_g L_f h(x^\circ)u(0)\,,$$

from which it is seen that, if $r=2$, the value $y^{(2)}(0)$ is an affine function of $u(0)$.

Otherwise, if r larger than 2, for all t near $t = 0$ we have $L_g L_f h(x(t)) = 0$ and

$$y^{(2)}(t) = L_f^2 h(x(t)) .$$

Continuing in this way, we get

$$\begin{aligned} y^{(k)}(t) &= L_f^k h(x(t)) \qquad \text{for all } k < r \text{ and all } t \text{ near } t = 0 \\ y^{(r)}(0) &= L_f^r h(x^\circ) + L_g L_f^{r-1} h(x^\circ) u(0) . \end{aligned}$$

Thus, the integer r is exactly equal to the number of times one has to differentiate the output $y(t)$ at time $t = 0$ in order to have the value $u(0)$ of the input explicitly appearing.

The calculations above suggest that the functions $h(x)$, $L_f h(x), \ldots, L_f^{r-1} h(x)$ must have a special importance. As a matter of fact, such functions can be used in order to define, at least partially, a local coordinates transformation around x°.

Proposition. Suppose the system has relative degree r at x°. Then $r \le n$. If r is strictly less than n, it is always possible to find $n - r$ more functions $\psi_1(x), \dots, \psi_{n-r}(x)$ such that the mapping

$$\Phi(x) = \begin{pmatrix} \psi_1(x) \\ \cdots \\ \psi_{n-r}(x) \\ h(x) \\ L_f h(x) \\ \cdots \\ L_f^{r-1} h(x) \end{pmatrix}$$

has a jacobian matrix which is nonsingular at x° and therefore qualifies as a local coordinates transformation in a neighborhood of x°. The value at x° of these additional functions can be fixed arbitrarily. Moreover, it is always possible to choose $\psi_1(x), \dots, \psi_{n-r}(x)$ in such a way that

$$L_g \psi_i(x) = 0 \qquad \textit{for all } 1 \le i \le n - r \textit{ and all } x \textit{ around } x^\circ.$$

The description of the system in the new coordinates is found very easily. Set

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \cdots \\ z_{n-r} \end{pmatrix} = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \\ \cdots \\ \psi_{n-r}(x) \end{pmatrix}, \quad \xi = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \cdots \\ \xi_r \end{pmatrix} = \begin{pmatrix} h(x) \\ L_f h(x) \\ \cdots \\ L_f^{r-1} h(x) \end{pmatrix}$$

and

$$\tilde{x} = \mathrm{col}(z_1, \ldots, z_{n-r}, \xi_1, \ldots, \xi_r) := \Phi(x)\,.$$

Bearing in mind the previous calculations, it is seen that

$$\begin{aligned} \frac{\mathrm{d}\xi_1}{\mathrm{d}\,t} &= \frac{\partial h}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\,t} = L_f h(x(t)) = \xi_2(t) \\ &\cdots \\ \frac{\mathrm{d}\xi_{r-1}}{\mathrm{d}\,t} &= \frac{\partial(L_f^{r-2}h)}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\,t} = L_f^{r-1}h(x(t)) = \xi_r(t)\,. \end{aligned}$$

while for ξ_r we obtain

$$\frac{\mathrm{d}\xi_r}{\mathrm{d}\,t} = L_f^r h(x(t)) + L_g L_f^{r-1} h(x(t)) u(t).$$

On the right-hand side of this equation, x must be replaced by its expression as a function of $\tilde{x}$, which will be written as $x = \Phi^{-1}(z, \xi)$. Thus, setting

$$\begin{aligned} q(z,\xi) &= L_f^r h(\Phi^{-1}(z,\xi)) \\ b(z,\xi) &= L_g L_f^{r-1} h(\Phi^{-1}(z,\xi)) \end{aligned}$$

the equation in question can be rewritten as

$$\frac{\mathrm{d}\xi_r}{\mathrm{d}\,t} = q(z(t),\xi(t)) + b(z(t),\xi(t))u(t)\ .$$

Note that, by definition of relative degree, at the point $\tilde{x}^\circ = \mathrm{col}(z^\circ, \xi^\circ) = \Phi(x^\circ)$, we have $b(z^\circ, \xi^\circ) = L_g L_f^{r-1} h(x^\circ) \neq 0$. Thus, the coefficient $b(z,\xi)$ is nonzero for all (z,ξ) in a neighborhood of (z°, ξ°).

As far as the other new coordinates are concerned, we cannot expect any special structure for the corresponding equations. However, if $\psi_1(x), \ldots, \psi_{n-r}(x)$ have been chosen in such a way that $L_g\psi_i(x) = 0$, then

$$\frac{\mathrm{d}z_i}{\mathrm{d}\,t} = \frac{\partial \psi_i}{\partial x}[f(x(t)) + g(x(t))u(t)] = L_f\psi_i(x(t)) + L_g\psi_i(x(t))u(t) = L_f\psi_i(x(t)).$$

Setting

$$f_0(z,\xi) = \begin{pmatrix} L_f\psi_1(\Phi^{-1}(z,\xi)) \\ \cdots \\ L_f\psi_{n-r}(\Phi^{-1}(z,\xi)) \end{pmatrix}$$

the latter can be rewritten as

$$\frac{\mathrm{d}z}{\mathrm{d}\,t} = f_0(z(t),\xi(t))\ .$$

Thus, in summary, in the new (local) coordinates the system is described by equations of the form

$$\begin{aligned}
\dot{z} &= f_0(z,\xi) \\
\dot{\xi}_1 &= \xi_2 \\
\dot{\xi}_2 &= \xi_3 \\
&\cdots \\
\dot{\xi}_{r-1} &= \xi_r \\
\dot{\xi}_r &= q(z,\xi) + b(z,\xi)u\,.
\end{aligned} \tag{2}$$

In addition to these equations one has to specify how the output of the system is related to the new state variables. Being $y = h(x)$, it is immediately seen that

$$y = \xi_1\,. \tag{3}$$

The equations thus introduced are said to be in strict normal form. They are useful in understanding how certain control problems can be solved.

The equations in question can be given a compact expression if we use the three matrices $\hat{A} \in \mathbb{R}^r \times \mathbb{R}^r$, $\hat{B} \in \mathbb{R}^r \times \mathbb{R}$ and $\hat{C} \in \mathbb{R} \times \mathbb{R}^r$ defined as

$$\hat{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad \hat{B} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix}, \qquad \hat{C} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

With the aid of such matrices, the equations (2) and (3) can be re-written as

$$\begin{aligned} \dot{z} &= f_0(z, \xi) \\ \dot{\xi} &= \hat{A}\xi + \hat{B}[q(z, \xi) + b(z, \xi)u] \\ y &= \hat{C}\xi \,. \end{aligned} \tag{4}$$

Assumptions are known under which a normal form exists globally.

Consider again system (1), assume that $f(0) = 0$ and $h(0) = 0$, suppose that the system has relative degree r at $x^\circ = 0$ and consider its normal form (4), in which $f_0(0,0) = 0$ and $q(0,0) = 0$.

Consider the problem of finding all pairs consisting of an initial state x° and of an input function $u(\cdot)$, for which the corresponding output $y(t)$ of the system is identically zero for all t in a neighborhood of $t = 0$.

Recalling that in the normal form $y(t) = \xi_1(t)$, it is seen that if $y(t) = 0$ for all t, then

$$\xi_1(t) = \xi_2(t) = \ldots = \xi_r(t) = 0 \;,$$

that is $\xi(t) = 0$ for all t.

Thus, when the output of the system is identically zero, its state is constrained to evolve in such a way that also $\xi(t)$ is identically zero.

In addition, the input $u(t)$ must necessarily be the unique solution of the equation

$$0 = q(z(t), 0) + b(z(t), 0)u(t)$$

(recall that $b(z(t), 0) \neq 0$ if $z(t)$ is close to 0).

As far as the variable $z(t)$ is concerned, it is seen that

$$\dot{z}(t) = f_0(z(t), 0) \;.$$

From this analysis we deduce the following facts. If the output $y(t)$ is identically zero, then necessarily the initial state of the system must be such that $\xi(0) = 0$, whereas $z(0) = z^\circ$ can be arbitrary. According to the value of z°, the input must be

$$u(t) = -\frac{q(z(t), 0)}{b(z(t), 0)}$$

where $z(t)$ denotes the solution of the differential equation

$$\dot{z} = f_0(z, 0) \qquad z(0) = z^\circ. \tag{5}$$

The dynamics of (5) characterize the forced state behavior of the system when input and initial conditions are chosen in such a way as to constrain the output to remain identically zero. These dynamics, which are rather important in many of the subsequent developments, are called the zero dynamics of the system.

If the system is linear, the dynamics of (5) are linear dynamics

$$\dot{z} = F_0 z$$

and it can be shown that the eigenvalues of F_0 are exactly the zeros of the transfer function of the system.

Minimum-phase systems

Definition Consider a system of the form (1), with $f(0) = 0$ and $h(0) = 0$. Suppose the system has relative degree r and possesses a globally defined normal. The system is globally minimum-phase if the equilibrium $z = 0$ of the zero dynamics

$$\dot{z} = f_0(z, 0) \tag{6}$$

is globally asymptotically stable. The system is strongly minimum-phase if system

$$\dot{z} = f_0(z, \xi)\,, \tag{7}$$

viewed as a system with input ξ and state z, is input-to-state stable.

By definition – then – a system is strongly minimum-phase if there exist a class $\mathcal{KL}$ function $\beta(\cdot,\cdot)$ and a class $\mathcal{K}$ function $\gamma(\cdot)$ such that, for any $z(0) \in \mathbb{R}^{n-r}$ and any piecewise-continuous bounded function $\xi(\cdot) : [0, \infty) \to \mathbb{R}^r$, the response $z(t)$ of (7) from the initial state $z(0)$ at time $t = 0$ satisfies,

$$\|z(t)\| \leq \beta(\|z(0)\|, t) + \gamma(\|\xi(\cdot)\|_{[0,t]}) \qquad \text{for all } t \geq 0. \tag{8}$$

Definition A system is strongly – and also locally exponentially – minimum-phase if for any $z_0 \in \mathbb{R}^{n-r}$ and any piecewise-continuous bounded function $\xi_0(\cdot) : [0, \infty) \to \mathbb{R}^r$, an estimate of the form (8) holds, where $\beta(\cdot,\cdot)$ and $\gamma(\cdot)$ are a class $\mathcal{KL}$ function and, respectively, a class $\mathcal{K}$ function bounded as

$$\gamma(r) \leq \ell\, r \qquad \beta(r, t) \leq A r e^{-\alpha t} \quad \text{for } |r| \leq d\,.$$