Lecture 14: Conditions for Input-State Linearizability

Recall that a given nonlinear control system in the form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})u, \tag{1}$$

is *input-state linearizable* if there exists a diffeomorphism $\boldsymbol{T}(\cdot)$ such that the change of coordinates $\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{x})$ converts the system (1) into the form

$$\dot{\boldsymbol{z}} = \boldsymbol{A}_c\boldsymbol{z} + \boldsymbol{B}_c\frac{u - \alpha(\boldsymbol{x})}{\beta(\boldsymbol{x})}$$

with

$$\boldsymbol{A}_c = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad \boldsymbol{B}_c = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

and $\beta(\boldsymbol{z}) \neq 0$. Comparing the $\boldsymbol{z}$-dynamics with the original system dynamics, we find that

$$\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{x}}\left(\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})u\right) = \boldsymbol{A}_c\boldsymbol{T}(\boldsymbol{x}) + \boldsymbol{B}_c\frac{u - \alpha(\boldsymbol{x})}{\beta(\boldsymbol{x})}. \tag{2}$$

To simplify notation, define the *Lie derivative* of a function $\phi(\boldsymbol{x})$ along the flow of a vector field $\boldsymbol{p}(\boldsymbol{x})$:

$$L_p\phi = \frac{\partial \phi}{\partial \boldsymbol{x}}\boldsymbol{p}(\boldsymbol{x}).$$

So, for example, if $\boldsymbol{x}(t)$ evolves according to the dynamics $\dot{\boldsymbol{x}} = \boldsymbol{p}(\boldsymbol{x})$, then the Lie derivative of the scalar function $\phi(\boldsymbol{x})$ is simply its total time derivative. In this new notation, equation (2) becomes

$$\begin{aligned} L_fT_1 + (L_gT_1)\,u &= T_2 \\ L_fT_2 + (L_gT_2)\,u &= T_3 \\ &\vdots \\ L_fT_{n-1} + (L_gT_{n-1})\,u &= T_n \\ L_fT_n + (L_gT_n)\,u &= \frac{u - \alpha(\boldsymbol{x})}{\beta(\boldsymbol{x})}. \end{aligned} \tag{3}$$

Setting $u = 0$ gives $n-1$ equations defining the components T_i in terms of the single function $T_1(\boldsymbol{x})$ for $i \in \{2, \ldots, n\}$ and a single equation defining $\frac{\alpha(\boldsymbol{x})}{\beta(\boldsymbol{x})}$. Subtracting these equations from (3) gives $n-1$ partial differential equations which the functions T_i must satisfy and a single equation for $\beta(\boldsymbol{x}) \neq 0$. A natural question to ask is: *What are the conditions under which these PDEs can be solved?*

To answer this question, we must introduce some new machinery. Define the Lie bracket of two smooth vector fields $\boldsymbol{f}$ and $\boldsymbol{g}$ as follows:

$$[\boldsymbol{f}, \boldsymbol{g}] = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\boldsymbol{f} - \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\boldsymbol{g}.$$

Aside: The Lie bracket of vector fields satisfies the following properties:

- bilinearity:

$$[\alpha\boldsymbol{f} + \beta\boldsymbol{g}, \boldsymbol{h}] = \alpha[\boldsymbol{f}, \boldsymbol{g}] + \beta[\boldsymbol{f}, \boldsymbol{h}],$$

- skew-symmetry:

$$[\boldsymbol{f},\boldsymbol{g}] = -[\boldsymbol{g},\boldsymbol{f}],$$

- the Jacobi identity:

$$[[\boldsymbol{f},\boldsymbol{g}],\boldsymbol{h}] + [[\boldsymbol{h},\boldsymbol{f}],\boldsymbol{g}] + [[\boldsymbol{g},\boldsymbol{h}],\boldsymbol{f}] = 0.$$

The vector space $\mathcal{X}$ of smooth vector fields, together with the Lie bracket, forms an *algebra.* □

Because the Lie bracket of two smooth vector fields generates a new vector field, we may compute the Lie derivative of a function, say T_1, along the flow of this new vector field:

$$\begin{aligned} L_{[f,g]}T_1 &= \frac{\partial T_1}{\partial \boldsymbol{x}}[\boldsymbol{f},\boldsymbol{g}] \\ &= \frac{\partial T_1}{\partial \boldsymbol{x}}\left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\boldsymbol{f} - \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\boldsymbol{g}\right) \end{aligned}$$

Also, notice that

$$\begin{aligned} L_f\left(L_gT_1\right) - L_g\left(L_fT_1\right) &= \frac{\partial}{\partial \boldsymbol{x}}\left(\frac{\partial T_1}{\partial \boldsymbol{x}}\boldsymbol{g}\right)\boldsymbol{f} - \frac{\partial}{\partial \boldsymbol{x}}\left(\frac{\partial T_1}{\partial \boldsymbol{x}}\boldsymbol{f}\right)\boldsymbol{g} \\ &= \left(\frac{\partial^2 T_1}{\partial \boldsymbol{x}^2}\boldsymbol{g} + \frac{\partial T_1}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\right)\boldsymbol{f} - \left(\frac{\partial^2 T_1}{\partial \boldsymbol{x}^2}\boldsymbol{f} + \frac{\partial T_1}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)\boldsymbol{g} \\ &= \boldsymbol{f}^T\frac{\partial^2 T_1}{\partial \boldsymbol{x}^2}\boldsymbol{g} - \boldsymbol{g}^T\frac{\partial^2 T_1}{\partial \boldsymbol{x}^2}\boldsymbol{f} + \left(\frac{\partial T_1}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\right)\boldsymbol{f} - \left(\frac{\partial T_1}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)\boldsymbol{g} \\ &= \left(\frac{\partial T_1}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}}\right)\boldsymbol{f} - \left(\frac{\partial T_1}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)\boldsymbol{g} \end{aligned}$$

so we have

$$L_{[f,g]}T_1 = L_f\left(L_gT_1\right) - L_g\left(L_fT_1\right).$$

Now, according to (3), the system is input-state linearizable only if $L_gT_1 = 0$ for some smooth function $T_1(\boldsymbol{x})$. Thus, we require

$$L_{[f,g]}T_1 = -L_g\left(L_fT_1\right).$$

But $T_2 = L_fT_1$ and input-state linearizability requires $L_gT_2 = 0$, as well. Thus, we see that

$$\begin{aligned} L_gT_1 &= 0 \\ L_gT_2 &= 0 \end{aligned} \qquad \Leftrightarrow \qquad \begin{aligned} L_gT_1 &= 0 \\ L_{[f,g]}T_1 &= 0. \end{aligned}$$

Next, one may check that

$$L_gT_3 = 0 \qquad \Leftrightarrow \qquad L_{[f,[f,g]]}T_1 = 0.$$

To simplify expressions, define the repeated Lie bracket operation

$$\mathrm{ad}_f^j\boldsymbol{g} = [\boldsymbol{f}, \ldots\ j \text{ times } \ldots [\boldsymbol{f},\boldsymbol{g}]]$$

with $\mathrm{ad}_f^0\boldsymbol{g} = \boldsymbol{g}$. Continuing as before, we find that the PDEs of the form $L_gT_i = 0$ for $i \in \{1, \ldots, n-1\}$ can be rewritten as

$$\begin{aligned} L_{\mathrm{ad}_f^0 g}T_1 &= 0 \\ L_{\mathrm{ad}_f^1 g}T_1 &= 0 \\ &\vdots \\ L_{\mathrm{ad}_f^{n-2} g}T_1 &= 0. \end{aligned} \tag{4}$$

The additional requirement that $L_g T_n \neq 0$ is equivalent to requiring

$$L_{\mathrm{ad}_f^{n-1} g} T_1 \neq 0. \tag{5}$$

The system is input-state linearizable if and only if the $n-1$ PDEs (4) are solvable with condition (5) satisfied. So when is this possible? First, it is necessary that the n vector fields $\boldsymbol{g}$, $[\boldsymbol{f}, \boldsymbol{g}]$, $\cdots$, $\mathrm{ad}_f^{n-1}\boldsymbol{g}$ be point-wise linearly independent, that is

$$\dim\left(\mathrm{span}\left\{\boldsymbol{g},\ [\boldsymbol{f}, \boldsymbol{g}],\ \cdots\ \mathrm{ad}_f^{n-1}\boldsymbol{g}\right\}\right) = n \tag{6}$$

at every point $\boldsymbol{x}$ for which the transformation $\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{x})$ is well-defined. To better understand this requirement, consider an analogous problem in linear algebra: Find the unique vector $\boldsymbol{t} \neq \boldsymbol{0}$ satisfying

$$\boldsymbol{a}_i \cdot \boldsymbol{t} = 0 \qquad \forall\ \ i \in \{0, \ldots, n-2\}$$

and $\boldsymbol{a}_{n-1} \cdot \boldsymbol{t} \neq 0$. Putting the system of linear equations in matrix form,

$$\begin{bmatrix} \boldsymbol{a}_0^T \\ \vdots \\ \boldsymbol{a}_{n-2}^T \\ \boldsymbol{a}_{n-1}^T \end{bmatrix} \boldsymbol{t} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ * \end{pmatrix}.$$

There is a unique solution $\boldsymbol{t}$ if and only if each $\boldsymbol{a}_i$ is linearly independent.

Condition (6) is a controllability condition, similar in spirit to requiring that the controllability matrix have full rank for linear time-invariant systems. (In fact, that condition can be shown to be equivalent to (6) in the case that $\boldsymbol{f} = \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{g} = \boldsymbol{B}$.) Having found a transformation $\boldsymbol{T}(\boldsymbol{x})$, this condition allows us to choose linearizing feedback.

So when can we find a transformation $\boldsymbol{T}(\boldsymbol{x})$? The partial differential equations (4) are solvable if and only if the Lie bracket of any two vector fields contained in the *distribution*

$$\Delta(\boldsymbol{x}) = \mathrm{span}\left\{\boldsymbol{g},\ \mathrm{ad}_f\boldsymbol{g},\ \cdots\ \mathrm{ad}_f^{n-2}\boldsymbol{g}\right\} \tag{7}$$

is also contained in the distribution $\Delta(\boldsymbol{x})$. In this case, one calls the distribution $\Delta(\boldsymbol{x})$ *involutive.* If (6) holds and $\Delta(\boldsymbol{x})$ is involutive, then one may solve the $n-1$ PDEs (4) for a single function $T_1(\boldsymbol{x})$ which defines the transformation of variables $\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{x})$.

Theorem. The affine nonlinear control system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})u$$

is input-state feedback linearizable in a region D_0 if and only if

$$\dim\left(\mathrm{span}\left\{\boldsymbol{g},\ \mathrm{ad}_f\boldsymbol{g},\ \cdots\ \mathrm{ad}_f^{n-1}\boldsymbol{g}\right\}\right) = n$$

and

$$\Delta(\boldsymbol{x}) = \mathrm{span}\left\{\boldsymbol{g},\ \mathrm{ad}_f\boldsymbol{g},\ \cdots\ \mathrm{ad}_f^{n-2}\boldsymbol{g}\right\}$$

is involutive within D_0.

Example. Recall the pendulum with actuator dynamics. The equations of motion are

$$\dot{\boldsymbol{x}} = \begin{pmatrix} x_2 \\ -bx_2 + \frac{g}{l}\sin(x_1) + x_3 \\ -\frac{1}{\tau}x_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \frac{1}{\tau} \end{pmatrix} u.$$

We check the first condition, with $n = 3$. That is, we check if $\boldsymbol{g}$, $[\boldsymbol{f}, \boldsymbol{g}]$, and $[\boldsymbol{f}, [\boldsymbol{f}, \boldsymbol{g}]]$ are linearly independent. We first compute

$$[\boldsymbol{f}, \boldsymbol{g}] = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} \boldsymbol{f} - \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \boldsymbol{g} = \mathbf{0} - \begin{pmatrix} 0 & 1 & 0 \\ \frac{g}{l} \cos x_1 & -b & 1 \\ 0 & 0 & -\frac{1}{\tau} \end{pmatrix} \boldsymbol{g} = \begin{pmatrix} 0 \\ -\frac{1}{\tau} \\ \frac{1}{\tau^2} \end{pmatrix},$$

which is linearly independent of $\boldsymbol{g}$. Next, we compute

$$[\boldsymbol{f}, [\boldsymbol{f}, \boldsymbol{g}]] = \frac{\partial [\boldsymbol{f}, \boldsymbol{g}]}{\partial \boldsymbol{x}} \boldsymbol{f} - \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} [\boldsymbol{f}, \boldsymbol{g}] = \mathbf{0} - \begin{pmatrix} 0 & 1 & 0 \\ \frac{g}{l} \cos x_1 & -b & 1 \\ 0 & 0 & -\frac{1}{\tau} \end{pmatrix} [\boldsymbol{f}, \boldsymbol{g}] = \begin{pmatrix} \frac{1}{\tau} \\ -\frac{b}{\tau} - \frac{1}{\tau^2} \\ \frac{1}{\tau^3} \end{pmatrix}.$$

These three vector fields are clearly linearly independent, so the first condition is satisfied.

Next we check involutivity of Δ. In this example, we have

$$\Delta(\boldsymbol{x}) = \operatorname{span} \{\boldsymbol{g}, [\boldsymbol{f}, \boldsymbol{g}]\}.$$

Now, both of these vector fields are *constant* and the Lie bracket of two constant vector fields is zero (because the Jacobian of a constant vector field is zero). Thus, Δ is involutive. It follows, as we already knew from our previous discussion of this example, that this system is input-state feedback linearizable.

Introduction to I/O Linearization. Condition (6) is generally easy to satisfy. A generic nonlinear system will satisfy this controllability requirement. Condition (7), on the other hand is typically not satisfied. Most nonlinear control systems are *not* input-state linearizable. We therefore consider a more general technique, known as input-output feedback linearization, which generalizes the idea of input-state linearization.

We will continue to consider single-input systems and we will assume that there is a single output of interest. That is, we consider systems in the form

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

where $\boldsymbol{f}$, $\boldsymbol{g}$, and h are "sufficiently smooth" in some domain D_0. Suppose we define a new coordinate $\psi_1(\boldsymbol{x}) = h(\boldsymbol{x})$ which will serve as a partial change of coordinates. Differentiating the output gives

$$\begin{aligned} \dot{y} = \dot{\psi}_1 &= \frac{\partial \psi_1}{\partial \boldsymbol{x}} (\boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x}) u) \\ &= L_f \psi_1 + (L_g \psi_1) u. \end{aligned}$$

Suppose that

$$L_g \psi_1 \neq 0$$

for all $\boldsymbol{x} \in D_0$ so that the input u appears in the first derivative of y. If we choose

$$u = \frac{1}{L_g \psi_1} (-L_f \psi_1 + v),$$

then we obtain the first order linear control system

$$\dot{\psi}_1 = v. \tag{8}$$

In principle, we may choose v to drive the output $\psi_1 = y$ to whatever value we like.

Alternatively, suppose that

$$L_g\psi_1 = 0$$

for all $x \in D_0$. Define $\psi_2(\boldsymbol{x}) = \dot{\psi}_1 = L_f\psi_1$. Differentiating $\dot{\psi}_1$, we find

$$\ddot{y} = \dot{\psi}_2 = L_f\psi_2 + (L_g\psi_2)u.$$

If

$$L_g\psi_2 \neq 0$$

for all $\boldsymbol{x} \in D_0$ then we may choose

$$u = \frac{1}{L_g\psi_2}\left(-L_f\psi_2 + v\right)$$

to obtain the second order linear control system

$$\begin{aligned} \dot{\psi}_1 &= \psi_2 \\ \dot{\psi}_2 &= v \end{aligned} \tag{9}$$

Once again, in principle, we may choose v to drive the output y to whatever value we like.

Suppose now that

$$L_g\psi_2 = 0$$

for all $\boldsymbol{x} \in D_0$. Defining ψ_i recursively as above, suppose we find that

$$L_g\psi_i = 0 \qquad \text{for} \qquad i \in \{0, 1, \ldots, r-1\} \qquad\qquad \text{and} \qquad\qquad L_g\psi_r \neq 0$$

for all $x \in D_0$ and for some $r \leq n$. Then, choosing

$$u = \frac{1}{L_g\psi_r}\left(-L_f\psi_r + v\right)$$

we obtain a chain of r integrators

$$\begin{aligned} \dot{\psi}_1 &= \psi_2 \\ \dot{\psi}_2 &= \psi_3 \\ &\vdots \\ \dot{\psi}_{r-1} &= \psi_r \\ \dot{\psi}_r &= v. \end{aligned} \qquad\qquad \Longrightarrow \qquad\qquad \dot{\boldsymbol{\psi}} = \boldsymbol{A}_c\boldsymbol{\psi} + \boldsymbol{B}_c v,$$

where $\boldsymbol{\psi} = [\psi_1, \ \ldots \psi_r]^T$ and $\boldsymbol{A}_c$ and $\boldsymbol{B}_c$ have conforming dimensions. Notice that, if $r = n$, then the system is input-state linearizable and we have

$$\boldsymbol{z} = \boldsymbol{T}(\boldsymbol{x}) = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_n \end{pmatrix}, \qquad \alpha(\boldsymbol{x}) = -\frac{L_f\psi_n}{L_g\psi_n} \qquad \text{and} \qquad \beta(\boldsymbol{x}) = \frac{1}{L_g\psi_n}.$$

In this case, we are very fortunate to have been given a system whose output function $h(\boldsymbol{x})$ satisfies all of the PDE conditions for input-state linearizability. (Let $T_1 = h$ and verify this for yourself.) More generally, for $r \leq n$, we cannot feedback linearize the entire system, but we can feedback linearize an r-dimensional portion of it.

Definition. A nonlinear system

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

has *relative degree* r in a region D_0 if

$$L_g\psi_i = 0 \qquad \text{for} \qquad i \in \{0, 1, \ldots, r-1\} \qquad\qquad \text{and} \qquad\qquad L_g\psi_r \neq 0$$

for all $\boldsymbol{x} \in D_0$ where

$$\psi_1(\boldsymbol{x}) = h(\boldsymbol{x}) \qquad \text{and} \qquad \psi_{i+1}(\boldsymbol{x}) = L_f\psi_i, \qquad i \in \{1, 2, \ldots, r-1\}.$$

In short, the relative degree of a given system is the number of times one must differentiate the output before the input appears explicitly. Alternatively, it is the number of integrators between the input and the output. Relative degree is not always well-defined for a given system in a given domain. If it is, however, we may input-output linearize the system.

A single-input/single-output system of the form

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

is *input-output linearizable* in a domain D_0 if it has relative degree $r \leq n$ there.

In this case, the functions

$$\psi_1 = h(\boldsymbol{x}), \qquad \psi_{i+1} = \dot{\psi}_i,$$

for $i \in \{1, \ldots, r-1\}$, define a partial change of coordinates whose dynamics can easily be converted, through feedback, into a system with a single input passing through a chain of r integrators. If the system has relative degree $r = n$, then it is input-state linearizable.

This seems like a rather simple result, but there is more to the story. While the process of input-output linearization makes the problem of output regulation or tracking appear simple, it generally makes some portion of the dynamics unobservable in the output. In general, the feedback linearizing control law depends on these unobservable dynamics, so they must be well-behaved. We will discuss the $n-r$-dimensional dynamics that remain after input-output linearization in the next lecture.