Feedback Linearization

Ravi N Banavar
March 26, 2013

These notes were based on my readings of the two books


Consider a dynamical system of the form

\[ \dot{x} = f(x) + g(x)u \quad x(t) \in U \subset \mathbb{R}^n \]  (1)

where \( f(\cdot) \), \( g(\cdot) \) are smooth functions and \( u \) takes values in a subset of \( \mathbb{R} \). The vector \( f(x) \) is called the drift vector field and the vectors \( g_i(x) \) are called the control vector fields.

Now let us examine the behaviour of the system in a region \( U \subset \mathbb{R} \) containing a point \( x_0 \) in the state-space \( (x_0 \in U) \). Associated with system (1) let us define the output as

\[ y = h(x) \]

The first derivative of the output is expressed as

\[ y^{(1)} = L_f h(x) + L_g h(x)u \]

where \( L_f \) and \( L_g \) denote the Lie derivatives along the vector fields \( f \) and \( g \) respectively and \( y^{(1)} \) denotes the first derivative with respect to time. Let us assume that \( L_g h(x) \) is zero in \( U \) and further, that

\[ L_g L^k_f h(x) = 0 \quad \text{for} \quad k = 0, 1, \ldots, r - 2 (r \leq n) \quad \forall x \in U \]

and

\[ L_g L^{r-1}_f h(x_0) \neq 0 \]

**Lemma 0.1** The following two equalities are equivalent.

\[ L_g h(x) = L_g L^1_f h(x) = \cdots = L_g L^k_f h(x) = 0 \quad \forall x \in U \]

\[ L_g h(x) = L_{[f,g]} h(x) = \cdots = L_{ad^k_g} h(x) = 0 \quad \forall x \in U \]

where \( ad^k_g \triangleq [f, \ldots, [f,g]] \) \((k\)-brackets).
Proof:
Let us assume that the first equality is true. The proof is based on an iteration. From \( L_g h(x) = 0 \) and \( L_g L_f h(x) = 0 \) we have

\[
L_{\text{ad}} f g h(x) = [L_f, [f, g]] h(x) = [L_f L_g - L_g L_f] h(x) = 0
\]

Similarly,

\[
L_{\text{ad}}^2 f g h(x) = [-[L_f L_g - L_g L_f] L_f] h(x) = 0
\]

Based on the above arguments, it follows that

\[
L_{\text{ad}}^k f g L_{\text{ad}}^l f h(x) = 0 \quad \text{if} \quad L_g L_{\text{ad}}^k f h(x) = 0
\]

Going on this way, we have

\[
L_g h(x) = L_{[f, g]} h(x) = \cdots = L_{\text{ad}}^k f g h(x) = 0 \quad \forall x \in U
\]

\( \square \).

**Lemma 0.2** For a system with relative degree \( r \) at \( x_0 \), the \( r \) row vectors \( dh(x_0), d(L_f h(x_0)), \ldots, d(L_f^{r-1} h(x_0)) \) are linearly independent.

**Proof:**

Note that

\[
<dL_f^j h(x), ad_f g(x) > = L_{ad_f g(x)} L_f^j h(x)
\]

and

\[
L_{ad_f^k g(x)} L_f^j h(x) \neq 0 \quad \forall k + l = r - 1
\]

Consider the product

\[
\begin{bmatrix}
  dh(x_0) \\
  d(L_f h(x_0)) \\
  \vdots \\
  d(L_f^{r-1} h(x_0))
\end{bmatrix}
\begin{bmatrix}
  g(x_0) & ad_f g(x_0) & \cdots & ad_f^{r-1} g(x_0)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
  0 & \cdots & < dh(x_0), ad_f^{r-1} g(x_0) > \\
  0 & \cdots & < d(L_f h(x_0)), ad_f^{r-1} g(x_0) > \\
  \vdots & \vdots & \vdots \\
  < d(L_f^{r-1} h(x_0)), g(x_0) > & \cdots & < d(L_f^{r-1} h(x_0)), ad_f^{r-1} g(x_0) >
\end{bmatrix}
\]
In the first row, \( < dh(x_0), ad_f^{r-1}g(x_0) > \neq 0 \). Consider the second row. The elements are 
\[
0,0,\ldots, < d(L_fh(x_0)), ad_f^{r-2}g(x_0) > , < d(L_fh(x_0)), ad_f^{r-1}g(x_0) > 
\]
\[
= 0,0,\ldots, L_ad_f^{r-2}g L_fh(x_0), L_ad_f^{r-1}g L_fh(x_0)
\]
The last but one \((r-1)th\) element of the row is non-zero. Similarly the third row has its last but second \((r-2)th\) element to be non-zero. Thus the \((r-i+1)th\) element of row \(i\) is non-zero and the matrix is lower triangular. Hence the matrix has rank \(r\). From the result on the product of two linear transformations (or matrices)
\[
\text{Rank}(AB) = \text{minimum}\{\text{Rank}(A), \text{Rank}(B)\}
\]
it follows that the two matrices on the LHS have at least rank \(r\). □.

The purpose of the lemma is that it helps us to define a new set of coordinates in a neighbourhood around \(x_0\) as

\[
z = \Phi(x) = \begin{bmatrix} h(x) \\ L_fh(x) \\ \vdots \\ L_f^{r-1}h(x) \\ \phi_1(x) \\ \vdots \\ \phi_{n-r}(x) \end{bmatrix}
\]
where the \(\phi_i\)s are \(n-r\) smooth functions such that \(\Phi(\cdot)\) is a diffeomorphism and the Jacobian
\[
\frac{\partial \Phi(x)}{\partial x} \bigg|_{x=x_0}
\]
has full rank. The system evolution in the new set of coordinates is
\[
\dot{z} = \frac{d}{dt} \Phi(x) = \frac{\partial \Phi(x)}{\partial x}(f(x) + g(x)u) \\
\dot{z} = \frac{d}{dt} \Phi(x) = \frac{\partial \Phi(x)}{\partial x} |_{x=\Phi^{-1}(z)}(f(\Phi^{-1}(z)) + g(\Phi^{-1}(z))u)
\]

**Question:** Why do we have the requirements - 1. that \(\Phi\) be a diffeomorphism? 2. that the Jacobian have full rank at \(x_0\)? What would happen if these were not satisfied?

Let us now define the dynamics of the system in this new set of coordinates
\[
\dot{z}_1 = \frac{d}{dt} h(x) = L_fh(x) = z_2 \quad \dot{z}_2 = \frac{d}{dt} L_fh(x) = L_f^2h(x) = z_3
\]
and
\[
\dot{z}_r = \frac{d}{dt}L_f^{\tau - 1}h(x) = L_f^r h(x) + L_g L_f^{\tau - 1}h(x)u = b(z) + a(z)u
\]

where \( a(z) \triangleq L_g L_f^{\tau - 1}h(\Phi^{-1}(z)) \) and \( b(z) \triangleq L_f^r h(\Phi^{-1}(z)) \). If the \( \phi_i \)'s are chosen such that \( L_g \phi_i(x) = 0 \) (this is called the normal form), and
\[
\frac{d}{dt}\phi_i(x) = L_f \phi_i(x) = q_i(z)
\]
then the system in the new coordinates looks like
\[
\begin{bmatrix}
\dot{z}_1 \\
\vdots \\
\dot{z}_r \\
\dot{z}_{r+1} \\
\vdots \\
\dot{z}_n
\end{bmatrix}
= \begin{bmatrix}
z_2 \\
\vdots \\
b(z) + a(z)u \\
q_{r+1}(z) \\
\vdots \\
q_n(z)
\end{bmatrix}
\]

By defining a new control input \( v \) as
\[
v \triangleq b(z) + a(z)u
\]
we notice that the sub-system "from \( v \) to \( z \)" is a linear system. By means of the feedback
\[
u = \frac{1}{a(z)}(v - b(z)) \quad \text{this is state-feedback}
\]
the nonlinear system has been "partially linearized" as
\[
\begin{bmatrix}
\dot{z}_1 \\
\vdots \\
\dot{z}_r
\end{bmatrix}
= \begin{bmatrix}
0 & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 1
\end{bmatrix}
\begin{bmatrix}
z_1 \\
\vdots \\
z_r
\end{bmatrix}
+ \begin{bmatrix}
0 \\
\vdots \\
1
\end{bmatrix}v
\begin{bmatrix}
\dot{z}_{r+1} \\
\vdots \\
\dot{z}_n
\end{bmatrix}
= \begin{bmatrix}
q_{r+1}(z) \\
\vdots \\
q_n(z)
\end{bmatrix}
\]

If the relative degree \( r = n \) (order of the system), then the original nonlinear system can be fully feedback linearized.

**Zero dynamics**

Defining
\[
\xi = \begin{bmatrix}
z_1 \\
z_2 \\
\vdots \\
z_r
\end{bmatrix} \quad \eta = \begin{bmatrix}
z_{r+1} \\
\vdots \\
z_n
\end{bmatrix}
\]
the system can be written as
\[
\begin{bmatrix}
\dot{z}_1 \\
\vdots \\
\dot{z}_r
\end{bmatrix} =
\begin{bmatrix}
z_2 \\
\vdots \\
b(\xi, \eta) + a(\xi, \eta)u
\end{bmatrix}
\quad \dot{\eta} = q(\xi, \eta)
\]
Suppose the output \( y = z_1 \) is identically zero. Then it follows that
\[
z_1(t) \equiv z_2(t) \equiv \ldots = z_r(t) \equiv 0 (\Rightarrow \xi \equiv 0)
\]
and
\[
\dot{z}_r \equiv 0 = b(0, \eta) + a(0, \eta)u
\]
The input history \( u(t) \) is then given by
\[
u(t) = -\frac{b(0, \eta(t))}{a(0, \eta(t))}
\]
where \( \eta(t) \) is the solution of the differential equation
\[
\dot{\eta} = q(0, \eta) \quad \eta(0) = \eta_0
\]
The dynamics of the above equation are called the zero dynamics of the system.

**Zero dynamics of a linear system**

Consider the transfer function
\[
G(s) = \frac{b_0 + b_1 s + \cdots + b_{n-r} s^{n-r}}{a_0 + a_1 s + \cdots + s^n}
\]
The state-space realization in a canonical form is
\[
A = \begin{bmatrix}
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-a_0 & -a_1 & \cdots & -a_{n-1}
\end{bmatrix} \quad B = \begin{bmatrix}
0 \\
0 \\
\vdots \\
1
\end{bmatrix} \quad C = \begin{bmatrix}
b_0 & b_1 & \cdots & b_{n-r-1} & 1 & 0 & \cdots & 0
\end{bmatrix}
\]
Following a similar procedure as before
\[
z_1 = y = b_0 x_1 + b_1 x_2 + \cdots + b_{n-r-1} x_{n-r} + x_{n-r+1}
\]
\[
z_2 = \dot{z}_1 = b_0 x_2 + b_1 x_3 + \cdots + b_{n-r-1} x_{n-r+1} + x_{n-r+2}
\]
\[
\vdots
\]
\[
z_r = \dot{z}_{r-1} = b_0 x_r + b_1 x_{r+1} + \cdots + b_{n-r-1} x_{n-1} + x_n
\]
To ensure that the rank of the jacobian is \( n \), choose the other \( n - r \) coordinates as

\[
z_{r+1} = x_1 \quad z_{r+2} = x_2 \quad \cdots \quad z_n = x_{n-r}
\]

(Check on the rank of the jacobian).

As before, partition the new set of variables into \( \xi = (z_1, \cdots, z_r) \) and \( \eta = (z_{r+1}, \cdots, z_n) \). The dynamics are

\[
\begin{align*}
\dot{z}_1 &= z_2 \\
\vdots & \quad \vdots \\
\dot{z}_{r-1} &= z_r \\
\dot{z}_r &= R\xi + S\eta + u \\
\dot{\eta} &= P\xi + Q\eta
\end{align*}
\]

If we identically zero the output (\( \xi \equiv 0 \)), then

\[
\dot{\eta} = Q\eta
\]

where

\[
Q = \begin{bmatrix}
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-b_0 & -b_1 & \cdots & -b_{n-r-1}
\end{bmatrix}
\]

The eigen values of \( Q \) correspond to the zeros of the transfer function!

**Remark 1** Please note that “zero dynamics” is a concept related to identically zeroing the output with a particular control. The reason we are concerned about the “hidden” part (the “\( \eta \)” part) is that we do not wish that these variables get unbounded when the output is identically zero. Hence we insist on “asymptotic stability” of this subsystem. So whatever the initial condition of “\( \eta \),” this variable then decays to zero.

As seen from above, the same concepts apply to a linear system. Though one can draw comparisons to observability/detectability, please remember that here we are using an “input.” We are not studying an output due to initial conditions alone.

In the situation of the output identically being zero, as we have seen, the following identities hold

\[
y(t) = h(x(t)) \equiv 0 \quad y^{(1)}(t) = L_f h(x(t)) \equiv 0 \quad \cdots \quad y^{(r-1)}(t) = L_f^{r-1} h(x(t)) \equiv 0
\]

and

\[
y^{(r)}(t) = L_f^r h(x(t)) + L_{g_f} L_f^{r-1} h(x(t)) u(t) \equiv 0
\]

(Note: Each independent algebraic constraint on the system dynamics reduces the dimension of the manifold on which the system trajectories evolve by 1.) We recall that the \( r \) row vectors \( dh(x_0), d(L_f h(x_0)), \cdots, d(L_f^{r-1} h(x_0)) \) are linearly independent and this holds in a
region around $x_0$ (since the functions are smooth) and hence the zero-dynamics trajectory lies in an $n - r$ dimensional submanifold ($Z$) of the original $n$-dimensional manifold.

The control input required to realize the output zeroing is

$$u^*(x(t)) = -\frac{L_f^r h(x(t))}{L_g L_f^{r-1} h(x(t))}$$

and we have

$$\begin{bmatrix}
  dh(x) \\
  d(L_f h(x)) \\
  \vdots \\
  d(L_f^{r-1} h(x))
\end{bmatrix} [f(x) + g(x)u^*(x)] = \begin{bmatrix}
  h(x) \\
  (L_f h(x)) \\
  \vdots \\
  L_f^{r-1} h(x) \\
  0
\end{bmatrix}$$

For all $x \in Z$, the output is zero and hence

$$\begin{bmatrix}
  dh(x) \\
  d(L_f h(x)) \\
  \vdots \\
  d(L_f^{r-1} h(x))
\end{bmatrix} [f(x) + g(x)u^*(x)] = 0$$

which implies that the vector field $[f(x) + g(x)u^*(x)]$ is tangential to $Z$. The system trajectories for a zero output thus evolve on $Z$ - the $n - r$ dimensional manifold.

**Local asymptotic stability**

Zero dynamics are helpful in determining the local asymptotic stability of a system around a critical point (when the linearization contains eigen values on the imaginary axis which are not controllable)

**Proposition 0.1** Consider the system (1) with an equilibrium $x_0 = 0$ and where the relative degree is $r$. Without loss of generality assume that $(\xi, \eta) = (0, 0)$ is the equilibrium point in the transformed coordinates.

Suppose the equilibrium $\eta = 0$ of the zero dynamics of the system is locally asymptotically stable and the feedback linearized subsystem is stable, then the system with the feedback control law is locally asymptotically stable at $(\xi, \eta) = (0, 0)$.

**Proof:**
Integrable distributions

Consider a non-singular smooth distribution \( W = \text{span}\{f_1, \ldots, f_r\} \) of rank \( r < n \) in an open set \( U \subset \mathbb{R}^n \). This implies that at every point \( p \in U \), the \( r \) vectors 
\[
\{f_1(p), \ldots, f_r(p)\}
\]
span an \( r \)-dimensional subspace \( W(p) \) of \( \mathbb{R}^n \). From basic vector space theory, the annihilator of \( W(p) \), denoted by \( W^0(p) \), forms an \( (n-r) \)-dimensional subspace of \( \mathbb{R}^n \). So if \( y \in W^0(p) \), then 
\[
[y, f_i(p)] = 0 \quad i = 1, \ldots, r
\]
Now, as a consequence, we can find \( (n-r) \) smooth functions \( \{\alpha_1, \ldots, \alpha_{n-r}\} \) such that 
\[
[\alpha_j, f_i](x) = 0 \quad \forall i = 1, \ldots, r, j = 1, \ldots, n-r \quad \text{and} \quad \forall x \in U
\]

So, when can we say that the distribution \( W \) is integrable? If the annihilator co-vector functions \( \alpha_i \)s can be expressed as gradients of smooth functions \( \lambda_i \)s, then the distribution is said *completely integrable* in \( U \). So if there exist \( (n-r) \) functions \( \lambda_i(\cdot) : U \to \mathbb{R} \) such that 
\[
\begin{bmatrix}
- & \cdots & -d\lambda_1(x) & - & \cdots & - \\
- & \cdots & -d\lambda_1(x) & - & \cdots & - \\
\end{bmatrix}
\begin{bmatrix}
f_1(x) & \cdots & f_r(x)
\end{bmatrix} = 0
\]
holds, then the distribution \( W \) is said to be integrable.

The distribution \( W \) is said to be *involutive* if all Lie brackets lie in the distribution. Mathematically expressed 
\[
\left[ \cdots \left[ f_i, f_j \right] \right] \in W
\]
for all order Lie brackets and all \( i, j \).

**When can we completely feedback linearize?**

*Question*: At a point \( x_0 \), when can we find a scalar function \( \lambda(x) \) such that the system has relative degree \( n \) at \( x_0 \)?

*Answer*: The following two conditions are both necessary and sufficient to be able to feedback linearize the system around \( x_0 \).

1. The matrix 
\[
\begin{bmatrix}
g(x_0) & \text{ad}_f g(x_0) & \cdots & \text{ad}^{(n-2)}_f g(x_0) & \text{ad}^{(n-1)}_f g(x_0)
\end{bmatrix}
\]
has rank \( n \).
2. The distribution

\[ D = \{ g(x), \text{ad}_f^3(x), \cdots, \text{ad}_f^{(n-2)}(x) \} \]

is involutive around \( x_0 \).

Based on our previous study, the necessary conditions are

\[ L_g \lambda(x) = L_g L_f \lambda(x) = \cdots = L_g L_f^{(n-2)} \lambda(x) = 0 \quad \forall x \in U(x_0 \in U) \]

\[ L_g L_f^{(n-1)} \lambda(x_0) \neq 0 \]

The first set of conditions are (as we have seen) equivalent to

\[ L_g \lambda(x) = L_{\text{ad}_f^3} \lambda(x) = \cdots = L_{\text{ad}_f^{(n-2)}} g \lambda(x) = 0 \quad \forall x \in U(x_0 \in U) \tag{2} \]

and the second non-triviality condition is equivalent to

\[ L_{\text{ad}_f^{(n-1)}} g \lambda(x_0) \neq 0 \]

Recall lemma 0.2. These conditions imply that the vectors

\[ g(x_0), \text{ad}_f^3(x_0), \cdots, \text{ad}_f^{(n-1)} g(x_0) \tag{3} \]

are linear independent. Now the condition (3) also implies that the distribution

\[ D = \text{span}\{g(x), \text{ad}_f^3(x), \cdots, \text{ad}_f^{(n-2)} g(x)\} \tag{4} \]

is nonsingular and \((n - 1)\) dimensional around \( x_0 \). The conditions (2) can be written as

\[ d\lambda(x)[g(x), \text{ad}_f^3(x), \cdots, \text{ad}_f^{(n-2)} g(x)] = 0 \]

This implies that \( d\lambda(\cdot) \) is a basis for the one-dimensional codistribution that annihilates \( D \). By Frobenius’ theorem theorem, the distribution \( D \) is involutive.

**An example: A DC motor**

**Complete feedback linearization**

The governing differential equations of a DC motor in which the rotor voltage \((V_r)\) is kept constant and the stator voltage \((V_s)\) is used as a control input are given by

\[ L_s \frac{dI_s}{dt} + RI_s = V_s \quad \text{The electrical dynamics on the stator side.} \]
\[ L_r \frac{dI_s}{dt} + RI_r = V_r - E \] The electrical dynamics on the rotor side.

\[ J \frac{d\Omega}{dt} + F\Omega = T \] The mechanical dynamics of the rotor.

where

\[ L_s, L_r \] are stator and rotor inductances,
\[ R_s, R_r \] are stator and rotor resistances,
\[ E \] is the back EMF in the rotor,
\[ J, \Omega, F, T \] are the rotor inertia, angular speed of the rotor, viscous friction and torque developed at the rotor shaft.

The back EMF, the torque and the flux are further given by the relations

\[ E = K\Phi \Omega \quad T = K\Phi I_r \quad \Phi = L_s I_s \]

Choosing the state variables as

\[ x_1 = I_s \quad x_2 = I_r \quad x_3 = \Omega \]

and the input as \( V_s \), the drift and control vector fields are

\[
\begin{align*}
  f(x) &= \begin{bmatrix}
    -\frac{R_s}{L_s} x_1 \\
    -\frac{R_r}{L_r} x_2 + \frac{V_r}{L_r} - \frac{K L_s}{L_r} x_1 x_3 \\
    F x_3 + \frac{K L_s}{J} x_1 x_2
  \end{bmatrix}, \\
  g(x) &= \begin{bmatrix}
    \frac{1}{L_s} \\
    0 \\
    0 \\
    0
  \end{bmatrix}
\end{align*}
\]

**Question 1:** Can we fully feedback linearize this system around an arbitrary \( x_0 \)?

Let us check whether the two conditions are satisfied.

**Condition 1:** Is

\[
\text{Rank}\{g(x_0) \ [f, g](x_0) \ [f, [f, g]](x_0)\} = 3?
\]

**Answer:** Yes, provided \( x_2 \neq 0 \) or \( x_3 \neq 0 \). Sounds alright since these variables correspond to the rotor current and rotor speed.

**Condition 2:**

Is the distribution\(\{g(x) \ [f, g](x)\}\) integrable around \( x_0 \)?

**Answer:** Let us try to find a \( \beta : U \to \mathbb{R} \) such that

\[
\text{grad}(\beta)[g(x) \ [f, g](x)] = 0
\]

i. e.

\[
\begin{bmatrix}
    \frac{\partial \beta}{\partial x_1} & \frac{\partial \beta}{\partial x_2} & \frac{\partial \beta}{\partial x_3}
  \end{bmatrix} [g(x) \ [f, g](x)] = (0 \ 0)
\]

Yes, one possible solution is

\[ \beta(x) = L_r x_2^2 + J x_3^2 \]
**Zero dynamics**

A natural output in a DC motor is the speed of the rotor. Suppose we wish to regulate this speed around a desired value $\Omega_d$. How would we choose an appropriate $y$ in our problem? If we wish to regulate the output to $\Omega_d$, a natural choice would be

$$y = h(x) = \Omega - \Omega_d = x_3 - x_{3d}$$

But with this output are we able to fully feedback linearize the system? If not, what happens?