10 Nonlinear Observers

Key points

- All the control methodologies covered so far require full state information. This can be impossible, or expensive, to obtain.
- Nonlinear observers:
  - Deterministic:
    - Lyapunov based: Thau, Raghavan
    - Geometric
    - Sliding
  - Stochastic: Extended Kalman Filter (EKF)

Introduction to Nonlinear Observers

Motivation

The big weakness of all the control methodologies that we have learned so far is that they require the full state.
- Sometimes impossible to measure.
- Sometimes, possible, but expensive.
Methodologies learned so far:

- Linearization
- I/O and I/S feedback linearization
- Sliding control (robust I/O linearization)
- Integrator backstepping
- Dynamic surface control

Note: make sure you understand what the similarities and differences between all those methodologies are!

In general, we only have access to p sensor outputs, that is:

\[ z = M \cdot x + v(t) \]

where:
- \( z \) is the measurement (of size p\times 1)
- \( M \) is the measurement matrix (of size p\times n)
- \( x \) is the state (of size n\times 1)
- \( v(t) \) represents measurement noise (of size p\times 1)

Even in nonlinear systems, the measurements will be linearly related to the state in general (property of any useful sensor).

The best-known methodology for dealing with a full state feedback controller is to separate the problem into a static controller (for example \( u = -k \cdot x \)) and a dynamic observer. We then:

a. Design the controller as if \( x = \dot{x} \).

b. Design the observer so that \( x \rightarrow \dot{x} \) as quickly as possible.
**Review of Linear Observers**

Process: \[ \dot{x} = Ax + Bu + w(t) \]
Measurement: \[ z = Mx + v(t) \]

Let: \[ \dot{x} = A\hat{x} + Bu + L(z - M\hat{x}) \] where the last term is basically a correction term \( z - \hat{z} \)

The error dynamics are expressed by:
\[ \tilde{x} = x - \hat{x} \]
\[ \dot{\tilde{x}} = \dot{x} - \dot{\hat{x}} = A\tilde{x} + Bu + w - A\hat{x} - Bu - L(M\tilde{x} + v - M\hat{x}) = (A - LM)\tilde{x} + w(t) - Lv(t) \]

There are two different classical approaches to dealing with the choice of \( L \):

**A. Deterministic (Luenberger observer)**

Ignore \( w(t), v(t) \).
\[ \dot{\tilde{x}} = (A - LM)\tilde{x} \] and if \( (A, M) \) is observable, then the eigenvalues of \( (A - LM) \) can be placed arbitrarily.

**B. Stochastic (Kalman-Bucy filter)**

\( L \) is chosen to minimize the variance of \( x - \hat{x} \), the state estimation error.
\[ \dot{x} = A\hat{x} + Bu + L(z - M\hat{x}) \]
\[ L = PM^TR^{-1} \]
(Intuition: if the variance is small make the gains large)
R and Q are noise statistics \( (R \) is associated with \( v \))
Note that \( R^{-1} \) is identical to \( 1/r \) for scalars.
\[ \dot{P} = AP + PA^T + Q - PM^TR^{-1}MP \]

The steady-state Kalman filter is usually implemented. In that case the last equation becomes:
and is usually referred to as the Algebraic Ricatti Equation (are).

**Nonlinear Observers**

- **Overview**

**Thau’s method:** a common sense approach. Mimic the nonlinear dynamics and add a linear correction term. Not constructive. An analysis method. The control analog is linearization.

**Raghavan’s method:** Uses the same observer structure as Thau’s method, but provides a constructive iteration to get the gains.

**Geometric observers:** The observer analog to I/S linearization. Find a state transformation so that the nonlinear system looks linear, and apply well known linear techniques. Seek a new state so that when we observe this new state, the error dynamics are linear.

**Sliding observers:** Explicitly account for imperfect models. It’s the observer analog to sliding mode control.

They are generally of two types:
- deterministic
- stochastic

The dominant stochastic nonlinear observer is called an **Extended Kalman Filter**

If $\dot{x} = f(x) + g(x)u$ and $y = h(x)$, use a standard Kalman filter, but with:

$$A = \frac{\partial f}{\partial x}$$ taken at the current operating point
\[ C = \frac{\partial h}{\partial x} \] taken at the current operating point

**Justification:** the nonlinearities are treated as process noise.

As a preliminary to discussing deterministic nonlinear observers, let’s review what “Lipschitz” means.

**Definition and Intuition for “Lipschitz”**

\[ f : \mathbb{R}^n \rightarrow \mathbb{R}^n \]

f is globally Lipschitz if there exists a \( \gamma \) (called the Lipschitz constant of f) such that:

\[ \forall (x_1, x_2), \quad \| f(x_1) - f(x_2) \| \leq \gamma \| x_1 - x_2 \| \]

**Interpretation (in the scalar case)**

\[ \forall (x_1, x_2), \quad | f(x_1) - f(x_2) | \leq \gamma | x_1 - x_2 | \]

\( \Rightarrow \) bounded derivative:

if \( x_2 = x_1 + h \) \( \Rightarrow \]

\[ | f(x_1) - f(x_1 + h) | \leq \gamma | h | \]

\( \Rightarrow \) \( \forall h \)

\[ \left| \frac{f(x_1) - f(x_1 + h)}{h} \right| \leq \gamma \]

In particular,

\[ \lim_{h \to 0} \left| \frac{f(x_1) - f(x_1 + h)}{h} \right| \leq \gamma \iff \left| \frac{\partial f}{\partial x} \right| \leq \gamma \]

\( \Rightarrow \) In the scalar case, if f has a bounded derivative, f is Lipschitz.
Deterministic, Lyapunov-Based Nonlinear Observers:

Thau’s Method

Let the plant be:
\[ \dot{x} = Ax + g(t, u, z) + f(x) \]
\[ z = Mx \]

Assumptions:
No noise
The nonlinearity, f, depends only on x.

Notes:
Ax is the LTI part
The “g” part can be cancelled out (we know t, u and z).
The “f” part represents the rest of the stuff.

Let the observer be:
\[ \dot{\hat{x}} = A\hat{x} + g(t, u, z) + f(\hat{x}) + L(z - M\hat{x}) \]

Given an L and the Lipschitz constant of f, γ, Thau’s method tells us whether or not the L gives asymptotically stable error dynamics.

This method is NOT constructive: it gives a yes/no answer.

The error dynamics are given by:
\[ \ddot{x} = x - \hat{x} \]
\[ \dot{x} - \dot{\hat{x}} = (A - LM)\ddot{x} + f(x) - f(\hat{x}) \]

Assumptions:
(A,M) is an observable pair
f is globally Lipschitz
Theorem (Thau, 1973):
Under the above assumptions, the system given by:
\[ \dot{x} = (A - LM)x + f(x) - f(\dot{x}) \]
is globally asymptotically stable (\( \lim_{t \to \infty} x = 0 \) as \( t \to \infty \)) if:
\[ \gamma < \frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} \]
where \( P \) and \( Q \) are defined by:
\[ (A - LM)^T P + P(A - LM) = -Q \]

Proof:
The proof is Lyapunov based.
Select
\[ V = \tilde{x}^T P\tilde{x} \]
\[ \dot{V} = \tilde{x}^T P\tilde{x} + \tilde{x}^T P \dot{x} \]
Then
\[ \dot{V} = \tilde{x}^T [(A - LM)^T P + P(A - LM)]\tilde{x} + 2\tilde{x}^T P[f(x) - f(\dot{x})] \]

(A,M) is an observable pair \( \Rightarrow \) there exists \( L \) such that (A-LM) is stable \( \Rightarrow \) there exists (P,Q) such that:
\[ (A - LM)^T P + P(A - LM) = -Q \]
and \( P \) and \( Q \) are positive definite.
\[ \Rightarrow \dot{V} = -\tilde{x}^T Q\tilde{x} + 2\tilde{x}^T P[f(x) - f(\dot{x})] \]
Using the Lipschitz condition and the fact that:
\[ \lambda_{\min}(Q)|\tilde{x}|^2 \leq \tilde{x}^T Q\tilde{x} \leq \lambda_{\max}(Q)|\tilde{x}|^2 \]
\[ \Rightarrow \dot{V} \leq -\lambda_{\text{min}}(Q)|\bar{x}|^2 + 2\bar{x}^TP\bar{x} \gamma \]
\[ \leq -\lambda_{\text{min}}(Q)|\bar{x}|^2 + 2\gamma \lambda_{\text{max}}(P)|\bar{x}|^2 \]
\[ \leq [2\gamma \lambda_{\text{max}}(P) - \lambda_{\text{min}}(Q)]|\bar{x}|^2 \]

So \( \dot{V} \leq 0 \) if \( \gamma < \frac{\lambda_{\text{min}}(Q)}{2\lambda_{\text{max}}(P)} \)

FACT: \( \frac{\lambda_{\text{min}}(Q)}{2\lambda_{\text{max}}(P)} \) is maximized for \( Q=I \).

So for the observer structure:
\[ \hat{x} = A\hat{x} + g(t,u,z) + f(\hat{x}) + L(z - \hat{M}x) \]
\[ z = Mx \]

Choose an \( L \) (of size \( nxp \)) such that \( (A - LM)^TP + P(A - LM) = -I \) and \( \gamma < \frac{1}{2\lambda_{\text{max}}(P)} \).

**Alternative Approach: Ragahvan's Method**

Reference: Raghavan, IJC

\[ \dot{x} = Ax + g(t,u,z) + f(x) \]
\[ z = Mx \]

\[ \dot{x} = A\hat{x} + g(t,u,z) + f(\hat{x}) + L(z - \hat{M}x) \]

\[ \tilde{x} = x - \hat{x} \]
\[ \hat{x} = (A - LM)\tilde{x} + f(x) - f(\hat{x}) \]

Define \( A_0 = A - LM \)

\[ V = \tilde{x}^TP\tilde{x} \]
\[ \Rightarrow \dot{V} = -\tilde{x}^T[A_0^TP + PA_0]\tilde{x} + 2\tilde{x}^TP[f(x) - f(\hat{x})] \]

Fact 1: \[ 2\tilde{x}^TP[f(x) - f(\hat{x})] \leq 2|P\tilde{x}||y|\tilde{x} |
(from I Lipschitz)

Fact 2: (name?)

\[ ab \leq \frac{a^2}{4} + b^2 \]

Let:

\[
\begin{align*}
    a & \equiv 2\gamma |P\bar{x}| \\
    b & \equiv |\bar{x}| 
\end{align*}
\]

Then, \(2\gamma |P\bar{x}| \leq \gamma^2 \bar{x}^T PP\bar{x}^T + \bar{x}^T \bar{x}\)

\[ \Rightarrow \dot{V} \leq \bar{x}^T \left [ A_0^T P + PA_0 + \gamma^2 PP + I \right ] \bar{x} \]

Fact 3: if \( \left [ A_0^T P_1 + P_1 A_0 + \gamma^2 P_1 P_1 + I \right ] < 0 \) (negative definite) then there exists a \( P_2 \) such that:

\[ \left [ A_0 P_2 + P_2 A_0^T + \gamma^2 P_2 P_2 + I \right ] < 0 \]

(This is proven in the IJC paper).

Let \( A_0 P + P A_0^T + \gamma^2 PP + I \equiv -\epsilon I \) with \( \epsilon \in \mathbb{R}^+ \)

Then, we have \( \dot{V} \leq -\epsilon \bar{x}^T \bar{x} < 0 \) so \( \bar{x} \to 0 \) exponentially.

\[(A - LM)P + P(A - LM)^T + \gamma^2 PP + I(\epsilon + 1) = 0 \]

Let \( L \equiv \frac{PM^T}{2\epsilon} \)

Then \( AP + PA^T + P \left ( \gamma^2 I - \frac{M^T M}{\epsilon} \right ) P + I(\epsilon + 1) = 0 \) Ricatti equation

If we set: \( R_1 \equiv \left ( \gamma^2 I - \frac{M^T M}{\epsilon} \right ) \) and \( Q_1 \equiv (\epsilon + 1)I \)

Then we have:

\[ AP + PA^T + PR_1 P + Q_1 = 0 \]

Standard algebraic Ricatti equation ("are" in Matlab)

**Design Steps**

a. Set \( \epsilon \) to a small positive value.
b. Solve the ARE 

c. Check if $P$ is symmetric and positive definite 

(i) If yes, set $L \equiv \frac{PM^T}{2\varepsilon}$ 

(ii) If no, set $\varepsilon = \frac{\varepsilon}{2}$ and repeat 

It is shown that if the algorithm fails, then there does not exist a $P$, $L$ such that $\dot{V} < 0$ for all $f$ with Lipschitz constant $\gamma$.

- Intuition for Thau’s method

Setup and theorem:

Plant: 

$$
\begin{align*}
\dot{x} &= Ax + f(x) + Bu \\
y &= Cx
\end{align*}
$$

$f$ is Lipschitz with constant $\gamma$

$$\forall (x, y), \|f(x) - f(y)\| \leq \gamma \|x - y\|$$

Observer:

$$\dot{x} = A\hat{x} + f(\hat{x}) + Bu + L(y - C\hat{x})$$

Theorem:

The observer has exponentially stable error dynamics ($\dot{e} = x - \hat{x}$) if the solution pair $(P, Q)$ to:

$$(A - LC)^T P + P(A - LC) = -Q$$

satisfies: $\gamma < \frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)}$

Intuition building example (hopefully)

Plant:

$$
\begin{align*}
\dot{x} &= -2x + \sin(x) + Bu \\
y &= x
\end{align*}
$$

Note: this is pointless except for intuition since we know $x$

A=[-2], f(x)=[\sin(x)], B=[1], C=[1]
The observer is given by:
\[ \dot{x} = -2\dot{x} + \sin(\dot{x}) + u + L(x - \dot{x}) \]

The error dynamics are given by:
\[ \frac{d}{dt}(x - \dot{x}) = \dot{e} = (-2 - L)e + \sin(x) - \sin(\dot{x}) \]

We would like: \( \dot{e} = -\lambda e \)

**Common sense analysis:**

a. Bounded nonlinear term: \( \gamma \) for \( \sin(x) \) is 1

b. \[ \dot{e} = (-2 - L)e + (\gamma = 1)e \Rightarrow \dot{e} = [-2 - L + \gamma]e \]

Condition for exponentially stable error dynamics:
\[ \Rightarrow -2 - L + \gamma < 0 \Rightarrow 2 + L > \gamma \]

**We can get the same result using Thau’s theorem:**

a. \( \gamma = 1 \)

b. Lyapunov equation
\[ (-2 - L)^T P + P(-2 - L) = -Q \quad (P, \text{ scalar} = P^T) \]

So: \( 2P(-L) = -Q \)

Choose \( Q=I=1 \) \( \Rightarrow P = \frac{1}{2(2 + L)} \)

Plug into Thau’s equation:
\[ \gamma < \frac{\lambda_{\min}(Q)}{2\lambda_{\max}(P)} \]
So: \(1 < \frac{1}{2\left(2 + L\right)} \Rightarrow \gamma < (2 + L)\)

Exact same result:
\(1 < 2 + L \Rightarrow L > -1\)

In the scalar case, it is easy to see how to choose \(L\) to give exponentially stable error dynamics. In higher dimensions, it is not as easy to see how to choose \(L\).

For most problems (for example, homework 7), you need to construct a gain matrix \(L\). To do this, use the algorithm on page 519 of the Raghavan paper distributed in class.

- **Raghavan’s method**

**Setup**

Plant:
\[
\begin{cases}
  x = Ax + f(x) + Bu \\
  y = Cx
\end{cases}
\]

Observer:
\[
\dot{x} = Ax + f(x) + Bu + L(y - Cx)
\]

**Procedure** (to get \(L\))
- Initialize \(\gamma\), \(A\), \(B\), \(C\)
- Initialize a small number \(\varepsilon\) (for example on the order of 0.1)
- Initialize a flag “\(P\_not\_positive\_definite = 1\)”
- While (\(P\_not\_positive\_definite\))
  - \(\varepsilon = \varepsilon/2\)
  - \(R = \left[\gamma^2 I - \frac{1}{\varepsilon} C^T C\right]\)
  - \(Q = (I + \varepsilon I)\)
  - Solve for \(P\): \(AP + PA^T + PRP + Q = 0\)
  - Check if \(P\) is positive definite and update flag
- 
  \(L = \frac{1}{2\varepsilon} PC^T\)

**Implementation Notes**

The command “are” in MATLAB solves Ricatti equations of this form:

\[
A^TP + PA^T - PRP + Q'= 0
\]

Set:
\[
\begin{cases}
  A' = A^T \\
  R' = -R \\
  Q' = Q
\end{cases}
\]
To check positive definiteness, use the command “chol” (sets a flag).

Exact Linearization – Geometric Observers

⇒ Observers with linearized error dynamics

Suppose we are given:
\[
\begin{align*}
\dot{x} &= f(x) + g(x)u \\
y &= h(x)
\end{align*}
\]

Can we find a nonlinear change of coordinates \( z = \Phi(x) \) such that:
\[
\begin{align*}
\dot{z} &= Az + K(y,u) \\
y &= Cz
\end{align*}
\]

where (A,C) is an observable pair.

Then a design such as:
\[
\dot{\hat{z}} = A\hat{z} + K(y,u) + L(y - C\hat{z})
\]

yields the following linear error dynamics:
\[
\begin{align*}
\dot{\tilde{z}} &= (A - LC)\tilde{z} \\
\hat{x} &= \Phi^{-1}(\tilde{z})
\end{align*}
\]

(Nice linear design).

Question: When can \( \begin{cases} \dot{x} = f(x) + g(x)u \\ y = h(x) \end{cases} \) be transformed to \( \begin{cases} \dot{\hat{z}} = Az + K(y,u) \\ y = Cz \end{cases} \)?

There are two primary results:
- Systems with no inputs
- Systems with inputs

Systems with no control input
The results presented here are for SISO case, but can be extended.

**Theorem:**

A transformation \( z = \Phi(x) \) exists iff:

a. The row vectors \( dh, dL_f h, \ldots, dL_f^{n-1} h \) are linearly independent (observability condition).

b. The unique vector solution, \( \bar{q}_i \), to the following equation:

\[
\begin{bmatrix}
    dh \\
    \vdots \\
    dL_f \\
    \vdots \\
    dL_f^{n-1} h
\end{bmatrix} q(x) =
\begin{bmatrix}
    0 \\
    \vdots \\
    0 \\
    1
\end{bmatrix}
\]

satisfies \( [q, ad_f^k q] = 0 \) for \( k = 1, 3, \ldots, (2n-1) \)

c. \( \bar{q}_i \) also has to satisfy:

\( [q, ad_f^k q] = 0 \) for \( k = 0, 1, \ldots, n-2 \)

**Note:** This is an important result theoretically, but this is very difficult to satisfy.

**Sliding Observers**

**Motivation**

\[
\begin{cases}
    \dot{x}_1 = x_2 \\
    \dot{x}_2 = -k_2 \text{sgn}(x_1)
\end{cases}
\]
Add a term to the first channel:

\[
\begin{align*}
\dot{x}_1 &= x_2 - k_1 \text{sgn}(x_1) \\
\dot{x}_2 &= -k_2 \text{sgn}(x_1)
\end{align*}
\]

⇒ “sliding patch”

\[ x_1 \dot{x}_1 = x_1 \left[ x_2 - k_1 \text{sgn}(x_1) \right] \leq 0 \]

Find a patch, test slightly negative and slightly positive.

Let’s try \( x_1 > 0 \) ⇒ \( (x_2 - k_1) \leq 0 \)

The sliding “patch” is \( |x_2| \leq k_1 \).
On the sliding patch, $x_{1\text{avg}} \approx 0$ and $\dot{x}_{1\text{avg}} \approx 0$

If $\dot{x}_1 = 0$, then $0 = x_{2\text{avg}} - k_1 \text{sgn}(x_1) \Rightarrow \text{sgn}(x_1) = \frac{x_{2\text{avg}}}{k_1}$

We then have: $\dot{x}_{2\text{avg}} = -\frac{k_2}{k_1} x_{2\text{avg}}$

The $k_2/k_1$ term determines the size of the sliding patch, and how long it takes to converge.

One can also add damping. We can get better system performance by adding linear “Luenberger-type” terms.

\[
\begin{align*}
\dot{x}_1 &= x_2 - h_1 x_1 - k_1 \text{sgn}(x_1) \\
\dot{x}_2 &= -h_2 x_1 - k_2 \text{sgn}(x_1)
\end{align*}
\]

(We measure $x_1$).

Second-Order Example
\[
\begin{aligned}
\begin{cases}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= f(x_1, x_2) \\
z &= x_1 + v(t)
\end{cases}
\end{aligned}
\]

Observer structure:
\[
\begin{aligned}
\tilde{x}_1 &= x_1 - \hat{x}_1 \\
\begin{cases}
\dot{\tilde{x}}_1 &= \hat{x}_2 - h_1 \tilde{x}_1 - k_1 \text{sgn}(\tilde{x}_1) \\
\dot{\tilde{x}}_2 &= f(\hat{x}_1, \hat{x}_2) - h_2 \tilde{x}_1 - k_2 \text{sgn}(\tilde{x}_1)
\end{cases}
\end{aligned}
\]

The error dynamics are given by:
\[
\begin{aligned}
\begin{cases}
\dot{\tilde{x}}_1 &= \tilde{x}_2 - h_1 \tilde{x}_1 - k_1 \text{sgn}(\tilde{x}_1) \\
\dot{\tilde{x}}_2 &= f(x_1, x_2) - f(\hat{x}_1, \hat{x}_2) - h_2 \tilde{x}_1 - k_2 \text{sgn}(\tilde{x}_1)
\end{cases}
\end{aligned}
\]

The sliding patch is given by: \[|\tilde{x}_2| \leq k_1.\]

On the sliding patch, \(x_{1,\text{avg}} \approx 0\) and \(\dot{x}_{1,\text{avg}} \approx 0\)

We then have: \(\dot{x}_{2,\text{avg}} \approx -\frac{k_2}{k_1} x_{2,\text{avg}} + Y(x_{2,\text{avg}})\)

What about sensor noise? \(\text{sgn}(\tilde{x}_1 + v(t)) = 0\)

One possible adaptation is to add a \textbf{dead zone}:  

Nonlinear Compensators

Observers seldom work in isolation.

Consider the system:

\[
\begin{align*}
\dot{x} &= f(x(t), u(t)) \\
y &= h(x(t)) \\
\ddot{x} &= g(\dot{x}(t), y(t), u(t))
\end{align*}
\]

For a linear system, this can be analyzed and it can be shown that the eigenvalues decouple. For a nonlinear system, there are only a few things that you can say about local behavior.

Theorem (Vidyasagar, IEEE TAC)


Assume \( f \) is continuously differentiable with \( f(0,0)=0 \), \( h \) is continuous and \( h(0,0)=0 \).

If:

a. \( u(t)=m(x(t)) \) is a uniformly asymptotically stable control law to \( x = 0 \) for the combined system \((*) + (**))\)

b. The system \((**)\) is an observer for system \((*)\), that is:

\[
\|\hat{x}(t) - x(t)\| \to 0 \text{ uniformly and asymptotically}
\]

Then:
\((x = 0, \dot{x} = 0)\) is a uniformly asymptotically stable equilibrium point of \((*) + (**)\) under the control law \(u(t) = m(x(t))\).

Note: this is a local result.

**Nonlinear Compensator Design**

Problems can occur due to peaking (large errors during the transient period).

Due to \(\dot{x}(0) \neq 0\), \(\tilde{x} = x - \hat{x} \implies u(\hat{x}) \neq u(x)\) during the transient period.

**Design Method**

a. The observer is designed for “good” state estimation, \(\hat{x}(t) \rightarrow x(t)\)

b. The controller is designed on the observed state estimate, that is \(\hat{x}(t) \rightarrow x_d(t)\)

c. Implement the resulting control law on the physical model

Note:

\(x = \tilde{x} + \hat{x}\)

If \(\hat{x} \rightarrow 0\), then \(x \rightarrow \tilde{x}\) and \(\hat{x} \rightarrow x_d\), so \(x \rightarrow x_d\)

**Second-Order Example**

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= f(x_1, x_2) + u \\
z &= y = x_1
\end{align*}
\]

GOAL: \(x_1 \rightarrow x_{1d}(t)\)

a. Design the asymptotic observer

\[
\begin{align*}
\dot{\hat{x}}_1 &= \hat{x}_2 + l_1(x_1 - \hat{x}_1) = \hat{x}_2 + l_1(\tilde{x}_1) \\
\dot{\hat{x}}_2 &= f(\hat{x}_1, \hat{x}_2) + u + l_2(\tilde{x}_1) \\
\end{align*}
\]

\[
\begin{align*}
\dot{\hat{x}}_1 &= \tilde{f} - l_1(\tilde{x}_1) \\
\dot{\hat{x}}_2 &= \tilde{f} - l_2(\tilde{x}_1)
\end{align*}
\]

Assuming that \(f\) is Lipschitz and we can find \(l_1\) and \(l_2\), \(\|\tilde{x}\| \rightarrow 0\)
b. Design the control law

Let $S = \dot{x}_2 - \dot{x}_{id} + \lambda \left[ \dot{x}_1 - x_{id} \right]

(Note: we don’t use $\dot{x}_1$ in the control surface, or we would have to differentiate the measurement, which is bad because it introduces a lot of noise).

$$\dot{S} = f(\dot{x}_1, \dot{x}_2) + u + l_2 \ddot{x}_1 - \dot{x}_{id} + \lambda \left[ \ddot{x}_2 + l_1 \ddot{x}_1 - \dot{x}_{id} \right]$$

We choose $u$ such that:

$$\dot{S} = -k_i S$$

$$u = -f(\dot{x}_1, \dot{x}_2) - l_2 \ddot{x}_1 + \dot{x}_{id} - \lambda \left[ \ddot{x}_2 + l_1 \ddot{x}_1 - \dot{x}_{id} \right] - k_i S$$

The terms containing $l_1$ and $l_2$ are due to the design of the control law being based on the observer.

So the control law knows $x_1 \neq \dot{x}_1$, and in steady-state these terms go to zero.