Nonlinear Systems

\[ \dot{x} = Ax + Bu \quad \rightarrow \quad \dot{x} = f(x, u) \quad (1) \]

Analysis:

- \( \dot{x} = f(x) \quad f : \mathbb{R}^n \rightarrow \mathbb{R}^n \) time-invariant (autonomous)
- \( \dot{x} = f(t, x) \quad f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n \) time-varying (non-autonomous)

Design:

\[ \dot{x} = f(x, u) \quad u \text{ to be designed as a function of } x. \]

Equilibria

\( x = x^* \) is an equilibrium for \( \dot{x} = f(x) \) if \( f(x^*) = 0. \)

Example: Linear system \( \dot{x} = Ax. \)

- If \( A \) is nonsingular, \( x^* = 0 \) is the unique equilibrium.
- If \( A \) is singular, the nullspace defines a continuum of equilibria.

Example: Logistic growth model in population dynamics

\[ \dot{x} = f(x) = r \left( 1 - \frac{x}{K} \right) x, \quad r > 0, \; K > 0 \quad (2) \]

\( x > 0 \) denotes the population and \( K \) is called the carrying capacity.

For systems with a scalar state variable \( x \in \mathbb{R} \), stability can be determined from the sign of \( f(x) \) around the equilibrium. In this example, \( f(x) > 0 \) for \( x \in (0, K) \), and \( f(x) < 0 \) for \( x > K \); therefore

- \( x = 0 \) unstable equilibrium
- \( x = K \) asymptotically stable.
Linearization

Local stability properties of $x^*$ can be determined by linearizing the vector field $f(x)$ at $x^*$:

$$f(x^* + \tilde{x}) = f(x^*) + \frac{\partial f}{\partial x} \bigg|_{x=x^*} \tilde{x} + \text{higher order terms} = 0 \quad \triangleq A$$

Thus, the linearized model is:

$$\dot{\tilde{x}} = A\tilde{x}.$$  \hspace{1cm} (4)

If $\Re\lambda_i(A) < 0$ for each eigenvalue $\lambda_i$ of $A$, then $x^*$ is asymptotically stable.

If $\Re\lambda_i(A) > 0$ for some eigenvalue $\lambda_i$ of $A$, then $x^*$ is unstable.

Example: Logistic growth model above:

Caveats:

1. Only local properties can be determined from the linearization.
   Example: The logistic growth model linearized at $x = 0$ ($\dot{x} = rx$) would incorrectly predict unbounded growth of $x(t)$. In reality, $x(t) \to K$.

2. If $\Re\lambda_i(A) \leq 0$ with equality for some $i$, then linearization is inconclusive as a stability test. Higher order terms determine stability.
   Example: $f(x) = x^3$ vs. $f(x) = -x^3$.

$f'(0) = 0$ in each case, but one is stable and the other is unstable.
Second order example: Pendulum

\[ \ell m \dot{\theta} = -k \ell \dot{\theta} - mg \sin \theta \]  \hspace{1cm} (5)

Define \( x = \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} \). State space: \( S^1 \times \mathbb{R} \).

\[ \dot{x}_1 = x_2, \quad \dot{x}_2 = -\frac{k}{m} x_2 - \frac{g}{\ell} \sin x_1 \]  \hspace{1cm} (6)

Equilibria: \((0, 0)\) and \((\pi, 0)\)

\[ \frac{\partial f}{\partial x} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{\ell} \cos x_1 & -\frac{k}{\ell} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{\ell} & -\frac{k}{\ell} \end{bmatrix} \]  (stable) at \( x_1 = 0 \)

\[ \begin{bmatrix} 0 & 1 \\ -\frac{g}{\ell} & -\frac{k}{\ell} \end{bmatrix} \]  (unstable) at \( x_1 = \pi \)

Phase portrait: plot of \( x_1(t) \) vs. \( x_2(t) \) for 2nd order systems

Figure 1: Phase portrait of the pendulum for the undamped case \( k = 0 \).

Essentially Nonlinear Phenomena

1. Finite Escape Time

Example: \( \dot{x} = x^2 \)

\[ \frac{d}{dt} x^{-1} = -x^{-2} \dot{x} = -1 \]

\[ \Rightarrow \frac{1}{x(t)} - \frac{1}{x(0)} = -t \]

\[ \Rightarrow x(t) = \frac{1}{\frac{1}{x(0)} - t} \]  \hspace{1cm} (7)

For linear systems, \( x(t) \to \infty \) cannot happen in finite time.
2. Multiple Isolated Equilibria

Linear systems: either unique equilibrium or a continuum

Pendulum: two isolated equilibria (one stable, one unstable)

“Multi-stable” systems: two or more stable equilibria

Example: bistable switch

\[
\begin{align*}
\dot{x}_1 &= -ax_1 + x_2 & x_1 : \text{concentration of protein} \\
\dot{x}_2 &= \frac{x_1^2}{1+x_1^2} - bx_2 & x_2 : \text{concentration of mRNA}
\end{align*}
\]

(8)

\(a > 0, \ b > 0\) are constants. State space: \(\mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}\).

This model describes a positive feedback where the protein encoded by a gene stimulates more transcription via the term \(\frac{x_1^2}{1+x_1^2}\).

Single equilibrium at the origin when \(ab > 0.5\). If \(ab < 0.5\), the line where \(\dot{x}_1 = 0\) intersects the sigmoidal curve where \(\dot{x}_2 = 0\) at two other points, giving rise to a total of three equilibria:
Essentially Nonlinear Phenomena Continued

1. Finite escape time

2. Multiple isolated equilibria

3. Limit cycles: Linear oscillators exhibit a continuum of periodic orbits; e.g., every circle is a periodic orbit for \( \dot{x} = Ax \) where

\[
A = \begin{bmatrix}
0 & -\beta \\
\beta & 0 \\
\end{bmatrix} \quad (\lambda_{1,2} = \mp j\beta).
\]

In contrast, a limit cycle is an isolated periodic orbit and can occur only in nonlinear systems.

Example: van der Pol oscillator

\[
\begin{align*}
C\dot{v}_C &= -i_L + v_C - v_C^3 \\
L\dot{i}_L &= v_C
\end{align*}
\]

Example: Lorenz system (derived by Ed Lorenz in 1963 as a simplified model of convection rolls in the atmosphere):

\[
\begin{align*}
\dot{x} &= \sigma(y - x) \\
\dot{y} &= rx - y - xz \\
\dot{z} &= xy - bz.
\end{align*}
\]

Chaotic behavior with \(\sigma = 10, b = 8/3, r = 28:\)

- For continuous-time, time-invariant systems, \(n \geq 3\) state variables required for chaos.
  - \(n = 1\): \(x(t)\) monotone in \(t\), no oscillations:

- \(n = 2\): Poincaré-Bendixson Theorem (to be studied in Lecture 3) guarantees regular behavior.
- Poincaré-Bendixson does not apply to time-varying systems and \(n \geq 2\) is enough for chaos (Homework problem).
- For discrete-time systems, \(n = 1\) is enough (we will see an example in Lecture 5).

**Planar (Second Order) Dynamical Systems**

**Phase Portraits of Linear Systems:** \(\dot{x} = Ax\)

- Distinct real eigenvalues

\[
T^{-1}AT = \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix}
\]
In \( z = T^{-1}x \) coordinates:

\[
\dot{z}_1 = \lambda_1 z_1, \quad \dot{z}_2 = \lambda_2 z_2.
\]

The equilibrium is called a node when \( \lambda_1 \) and \( \lambda_2 \) have the same sign (stable node when negative and unstable when positive). It is called a saddle point when \( \lambda_1 \) and \( \lambda_2 \) have opposite signs.

- **Complex eigenvalues:** \( \lambda_{1,2} = \alpha \mp j\beta \)

\[
T^{-1}AT = \begin{bmatrix}
\alpha & -\beta \\
\beta & \alpha
\end{bmatrix}
\]

\[
\begin{align*}
\dot{z}_1 &= \alpha z_1 - \beta z_2 \\
\dot{z}_2 &= \alpha z_2 + \beta z_1
\end{align*}
\]

\[
\rightarrow \text{polar coordinates} \quad \rightarrow \begin{cases}
r = \alpha r \\
\theta = \beta
\end{cases}
\]

**Phase Portraits of Nonlinear Systems Near Hyperbolic Equilibria**

Hyperbolic equilibrium: linearization has no eigenvalues on the imaginary axis

Phase portraits of nonlinear systems near hyperbolic equilibria are qualitatively similar to the phase portraits of their linearization. According to the Hartman-Grobman Theorem (below) a “continuous deformation” maps one phase portrait to the other.
Hartman-Grobman Theorem: If $x^*$ is a hyperbolic equilibrium of $\dot{x} = f(x), x \in \mathbb{R}^n$, then there exists a homeomorphism $z = h(x)$ defined in a neighborhood of $x^*$ that maps trajectories of $\dot{x} = f(x)$ to those of $\dot{z} = Az$ where $A \triangleq \frac{\partial f}{\partial x}|_{x=x^*}$.

The hyperbolicity condition can’t be removed:

Example:

$$
\begin{align*}
\dot{x}_1 &= -x_2 + ax_1(x_1^2 + x_2^2) \\
\dot{x}_2 &= x_1 + ax_2(x_1^2 + x_2^2) \\
x^* &= (0, 0)
\end{align*}
$$

$$
A = \left[ \begin{array}{cc}
0 & -1 \\
1 & 0
\end{array} \right]
$$

There is no continuous deformation that maps the phase portrait of the linearization to that of the original nonlinear model:

Periodic Orbits in the Plane

Bendixson’s Theorem: For a time-invariant planar system

$$
\dot{x}_1 = f_1(x_1, x_2) \quad \dot{x}_2 = f_2(x_1, x_2),
$$

if $\nabla \cdot f(x) = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2}$ is not identically zero and does not change sign in a simply connected region $D$, then there are no periodic orbits lying entirely in $D$. 

A continuous map with a continuous inverse.
Proof: By contradiction. Suppose a periodic orbit \( J \) lies in \( D \). Let \( S \) denote the region enclosed by \( J \) and \( n(x) \) the normal vector to \( J \) at \( x \). Then \( f(x) \cdot n(x) = 0 \) for all \( x \in J \). By the Divergence Theorem:

\[
\int_J f(x) \cdot n(x) d\ell = \int_S \nabla \cdot f(x) dx. \\
= 0 \quad \neq 0
\]

Example: \( \dot{x} = Ax, x \in \mathbb{R}^2 \) can have periodic orbits only if

\[ \text{Trace}(A) = 0, \text{ e.g.}, \]

\[
A = \begin{bmatrix}
0 & -\beta \\
\beta & 0
\end{bmatrix}.
\]

Example:

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -\delta x_2 + x_1 - x_1^3 + x_1^2 x_2 \quad \delta > 0
\end{align*}
\]

\[ \nabla \cdot f(x) = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} = x_1^2 - \delta \]

Therefore, no periodic orbit can lie entirely in the region \( x_1 \leq -\sqrt{\delta} \) where \( \nabla \cdot f(x) \geq 0 \), or \( -\sqrt{\delta} \leq x_1 \leq \sqrt{\delta} \) where \( \nabla \cdot f(x) \leq 0 \), or \( x_1 \geq \sqrt{\delta} \) where \( \nabla \cdot f(x) \geq 0 \).
Invarian Sets

Notation: $\phi(t, x_0)$ denotes a trajectory of $\dot{x} = f(x)$ with initial condition $x(0) = x_0$.

Definition: A set $M \subset \mathbb{R}^n$ is positively (negatively) invariant if, for each $x_0 \in M$, $\phi(t, x_0) \in M$ for all $t \geq 0$ ($t \leq 0$).

If $f(x) \cdot n(x) \leq 0$ on the boundary then $M$ is positively invariant.

Example 1: A predator-prey model

\begin{align*}
\dot{x} &= (a - by)x & \text{Prey (exponential growth when } y = 0) \\
\dot{y} &= (cx - d)y & \text{Predator (exponential decay when } x = 0) \\
a, b, c, d > 0
\end{align*}

The nonnegative quadrant is invariant:

Example 2: \[ \begin{align*}
\dot{x}_1 &= x_1 + x_2 - x_1(x_1^2 + x_2^2) \\
\dot{x}_2 &= -2x_1 + x_2 - x_2(x_1^2 + x_2^2)
\end{align*} \]

Show that $B_r \triangleq \{ x | x_1^2 + x_2^2 \leq r^2 \}$ is positively invariant for sufficiently large $r$. 

\[ \text{Licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.} \]
\[ f(x) \cdot n(x) = x_1^2 + x_1x_2 - x_1^2(x_1^2 + x_2^2) - 2x_1x_2 + x_2^2 - x_2^2(x_1^2 + x_2^2) \]
\[ = -x_1x_2 + (x_1^2 + x_2^2) - (x_1^2 + x_2^2)^2 \]
\[-x_1x_2 \leq \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \quad \text{(completion of squares)}\]

Therefore, \( f(x) \cdot n(x) \leq \frac{3}{2}r^2 - r^4 \leq 0 \) if \( r^2 \geq \frac{3}{2} \).

**Periodic Orbits in the Plane Continued**

Two criteria:

1. Bendixson (absence of periodic orbits)
2. Poincaré-Bendixson (existence of periodic orbits)

**Poincaré-Bendixson Theorem:** Suppose \( M \) is compact \( ^4 \) and positively invariant for the planar, time invariant system \( \dot{x} = f(x), x \in \mathbb{R}^2 \). If \( M \) contains no equilibrium points, then it contains a periodic orbit.

**Example 3:** Harmonic Oscillator

\[
A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.
\]

For any \( R > r > 0 \), the ring \( \{x : r^2 \leq x_1^2 + x_2^2 \leq R^2\} \) is compact, invariant and contains no equilibria \( \Rightarrow \) at least one periodic orbit. (We know there are infinitely many in this case.)

The “no equilibrium” condition in the PB theorem can be relaxed as:

“\( M \) contains one equilibrium which is an unstable focus or unstable node”

**Proof sketch:** Since the equilibrium is an unstable focus or node, we can encircle it with a small closed curve on which \( f(x) \) points outward. Then the set obtained from \( M \) by carving out the interior of the closed curve is positively invariant and contains no equilibrium.
Example 2 above: \( B_r \) is positively invariant for \( r \geq \sqrt{\frac{3}{2}} \) but contains the equilibrium \( x = 0 \).

\[
\left. \frac{\partial f}{\partial x} \right|_{x=0} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} \quad \lambda_{1,2} = 1 \mp j\sqrt{2} \quad \text{unstable focus.}
\]

Therefore, \( B_r \) must contain a periodic orbit.

A more general form of the PB Theorem states that, for time invariant, planar systems, bounded trajectories converge to equilibria, periodic orbits, or unions of equilibria connected by trajectories.

\textbf{Corollary:} No chaos for time invariant planar systems.

\textit{Index Theory}

Again, applicable only to planar systems.

\textbf{Definition (index):} The index of a closed curve is \( k \) if, when traversing the curve in one direction, \( f(x) \) rotates by \( 2\pi k \) in the same direction. The index of an equilibrium is defined to be the index of a small curve around it that doesn’t enclose another equilibrium.

<table>
<thead>
<tr>
<th>type of equilibrium or curve</th>
<th>index</th>
</tr>
</thead>
<tbody>
<tr>
<td>node, focus, center</td>
<td>+1</td>
</tr>
<tr>
<td>saddle</td>
<td>-1</td>
</tr>
<tr>
<td>any closed orbit</td>
<td>+1</td>
</tr>
<tr>
<td>a closed curve not encircling any equilibria</td>
<td>0</td>
</tr>
</tbody>
</table>

The last claim (index = 0) follows from the following observations:
- Continuously deforming a closed curve without crossing equilibria leaves its index unchanged.
- A curve not encircling equilibria can be shrunk to an arbitrarily small one, so \( f(x) \) can be considered constant.
**Theorem:** The index of a closed curve is equal to the sum of indices of the equilibria inside.

**Graphical proof:** Shrinking curve $c$ to $c'$ below without crossing equilibria does not change the index. The index of $c'$ is the sum of the indices of the curves encircling the equilibria because the thin "pipes" connecting these curves do not affect the index of $c'$.

The following corollary is useful for ruling out periodic orbits (like Bendixson’s Theorem studied in the previous lecture):

**Corollary:** Inside any periodic orbit there must be at least one equilibrium and the indices of the equilibria enclosed must add up to $+1$.

**Example (from last lecture):**

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -\delta x_2 + x_1 - x_1^3 + x_2^2 + x_2 \quad \delta > 0
\end{align*}
\]

Bendixson’s Criterion: No periodic orbit can lie entirely in one of the regions $x_1 \leq -\sqrt{\delta}$, $-\sqrt{\delta} \leq x_1 \leq +\sqrt{\delta}$, or $x_1 \geq \sqrt{\delta}$.

Now apply the corollary above.

**Equilibria:** $(0,0), (\mp 1,0)$. To find their indices evaluate the Jacobian:

\[
\begin{align*}
\left| \frac{\partial f}{\partial x} \right|_{(x,0)} &= \begin{bmatrix} 0 & 1 \\ 1 & -\delta \end{bmatrix} \quad \lambda^2 + \delta \lambda - 1 < 0 \\
\left| \frac{\partial f}{\partial x} \right|_{(1,0)} &= \begin{bmatrix} 0 & 1 \\ -2 & 1 - \delta \end{bmatrix} \quad \lambda^2 + (\delta - 1) \lambda + 2 > 0
\end{align*}
\]

The eigenvalues are real and have opposite signs, therefore $(0,0)$ is a saddle: index $= -1$.

The eigenvalues are either real with the same sign (node) or complex conjugates (focus or center), therefore $(\mp 1,0)$ each has index $= +1$.

Thus, the corollary above rules out the periodic orbit in the middle plot below. It does not rule out the others, but does not prove their existence either. Bendixson’s Criterion rules out neither of the three.
\[ x_1 = -\sqrt{\delta} \quad \text{not possible} \]

\[ x_1 = \sqrt{\delta} \]

\[ x_1 = -\sqrt{\delta} \quad \text{not possible} \]

\[ x_1 = \sqrt{\delta} \]
Bifurcations

A bifurcation is an abrupt change in qualitative behavior as a parameter is varied. Examples: equilibria or limit cycles appearing/disappearing, becoming stable/unstable.

Fold Bifurcation

Also known as “saddle node” or “blue sky” bifurcation.

Example: \( \dot{x} = \mu - x^2 \)

If \( \mu > 0 \), two equilibria: \( x = \pm \sqrt{\mu} \). If \( \mu < 0 \), no equilibria.

Transcritical Bifurcation

Example: \( \dot{x} = \mu x - x^2 \)

Equilibria: \( x = 0 \) and \( x = \mu \).

\[
\frac{\partial f}{\partial x} = \mu - 2x = \begin{cases} 
\mu & \text{if } x = 0 \\
-\mu & \text{if } x = \mu 
\end{cases}
\]

\( \mu < 0 : x = 0 \) is stable, \( x = \mu \) is unstable

\( \mu > 0 : x = 0 \) is unstable, \( x = \mu \) is stable
Pitchfork Bifurcation

Example: $\dot{x} = \mu x - x^3$
Equilibria: $x = 0$ for all $\mu$, $x = \mp \sqrt{\mu}$ if $\mu > 0$.

$$\frac{\partial f}{\partial x} \bigg|_{x=0} = \mu \quad \text{stable unstable}$$
$$\frac{\partial f}{\partial x} \bigg|_{x=\mp \sqrt{\mu}} = -2\mu \quad \text{N/A stable}$$

"supercritical pitchfork"

Example: $\dot{x} = \mu x + x^3$
Equilibria: $x = 0$ for all $\mu$, $x = \mp \sqrt{-\mu}$ if $\mu < 0$.

$$\frac{\partial f}{\partial x} \bigg|_{x=0} = \mu \quad \text{stable unstable}$$
$$\frac{\partial f}{\partial x} \bigg|_{x=\mp \sqrt{-\mu}} = -2\mu \quad \text{unstable N/A}$$

"subcritical pitchfork"

Example: $\dot{x} = \mu x + x^3 - x^5$

fold $\rightarrow$ subcritical pitchfork
Hysteresis arising from a subcritical pitchfork bifurcation:

Bifurcation and hysteresis in perception:

Figure 1: Observe the transition from a man’s face to a sitting woman as you trace the figures from left to right, starting with the top row. When does the opposite transition happen as you trace back from the end to the beginning? [Fisher, 1967]

Higher Order Systems

Fold, transcritical, and pitchfork are one-dimensional bifurcations, as evident from the first order examples above. They occur in higher order systems too, but are restricted to a one-dimensional manifold.

1D subspace: $c_1^T x = \cdots = c_{n-1}^T x = 0$

1D manifold: $g_1(x) = \cdots = g_{n-1}(x) = 0$

Example 1:

\[
\begin{align*}
\dot{x}_1 &= \mu - x_1^2 \\
\dot{x}_2 &= -x_2
\end{align*}
\]

A fold bifurcation occurs on the invariant $x_2 = 0$ subspace:
Example 2: bistable switch (Lecture 1)

\[
\begin{align*}
\dot{x}_1 &= -ax_1 + x_2 \\
\dot{x}_2 &= \frac{x_2}{1 + x_1^2} - bx_2
\end{align*}
\]

A fold bifurcation occurs at \( \mu \triangleq ab = 0.5 \):

\[
\begin{align*}
\dot{x}_2 &= ax_1 \\
&= \begin{cases} 
  a > 0.5/b \\
  a = 0.5/b \\
  a < 0.5/b 
\end{cases}
\end{align*}
\]

\( x_2 = \frac{x_1^2}{b(1 + x_1^2)} \)

![Graph showing the bistable switch behavior](image)

Characteristic of one-dimensional bifurcations:

\[
\frac{\partial f}{\partial x} \bigg|_{\mu=\mu_c, x=x^*(\mu_c)}
\]

has an eigenvalue at zero

where \( x^*(\mu) \) is the equilibrium point undergoing bifurcation and \( \mu_c \) is the critical value at which the bifurcation occurs.

Example 1 above:

\[
\frac{\partial f}{\partial x} \bigg|_{\mu=0, x=0} = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \rightarrow \lambda_{1,2} = [0, -1]
\]

Example 2 above:

\[
\frac{\partial f}{\partial x} \bigg|_{\mu=\frac{1}{2}, x_1=1, x_2=a} = \begin{bmatrix} -a & 1 \\ \frac{1}{2} & -b \end{bmatrix} \rightarrow \lambda_{1,2} = [0, -(a + b)]
\]

**Hopf Bifurcation**

Two-dimensional bifurcation unlike the one-dimensional types above.

Example: Supercritical Hopf bifurcation

\[
\begin{align*}
\dot{x}_1 &= x_1(\mu - x_1^2 - x_2^2) - x_2 \\
\dot{x}_2 &= x_2(\mu - x_1^2 - x_2^2) + x_1
\end{align*}
\]

In polar coordinates:

\[
\begin{align*}
\dot{r} &= \mu r - r^3 \\
\dot{\theta} &= 1
\end{align*}
\]
Note that a positive equilibrium for the $r$ subsystem means a limit cycle in the $(x_1, x_2)$ plane.

$\mu < 0$: stable equilibrium at $r = 0$

$\mu > 0$: unstable equilibrium at $r = 0$ and stable limit cycle at $r = \sqrt{\mu}$

The origin loses stability at $\mu = 0$ and a stable limit cycle emerges.

**Example:** Subcritical Hopf bifurcation

\[
\begin{align*}
\dot{r} &= \mu r + r^3 - r^5 \\
\dot{\theta} &= 1
\end{align*}
\]

Phase portrait for $\mu < 0$:

Characteristic of the Hopf bifurcation:

\[
\left. \frac{\partial f}{\partial x} \right|_{\mu=\mu^*, x=x^*(\mu^*)} \text{ has complex conjugate eigenvalues on the imaginary axis.}
\]
Center Manifold Theory

\[ \dot{x} = f(x), \quad f(0) = 0 \]  
(1)

Suppose \( A \triangleq \left. \frac{\partial f}{\partial x} \right|_{x=0} \) has \( k \) eigenvalues will zero real parts, and \( m = n - k \) eigenvalues with negative real parts.

Define \( \begin{bmatrix} y \\ z \end{bmatrix} = T x \) such that

\[
T A T^{-1} = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}
\]

where the eigenvalues of \( A_1 \) have zero real parts and the eigenvalues of \( A_2 \) have negative real parts.

Rewrite \( \dot{x} = f(x) \) in the new coordinates:

\[
\begin{align*}
\dot{y} &= A_1 y + g_1(y, z) \\
\dot{z} &= A_2 z + g_2(y, z)
\end{align*}
\]
(2)

\( g_i(0, 0) = 0, \quad \frac{\partial g_i}{\partial y}(0, 0) = 0, \quad \frac{\partial g_i}{\partial z}(0, 0) = 0, \quad i = 1, 2. \)

**Theorem 1**: There exists an invariant manifold \( z = h(y) \) defined in a neighborhood of the origin such that

\[
h(0) = 0, \quad \frac{\partial h}{\partial y}(0) = 0.
\]

**Theorem 2**: If \( y = 0 \) is asymptotically stable (resp., unstable) for the reduced system, then \( x = 0 \) is asymptotically stable (resp., unstable) for the full system \( \dot{x} = f(x) \).
Characterizing the Center Manifold

Define \( w \triangleq z - h(y) \) and note that it satisfies

\[
\dot{w} = A_2 z + g_2(y,z) - \frac{\partial h}{\partial y} \left( A_1 y + g_1(y,h(y)) \right).
\]

The invariance of \( z = h(y) \) means that \( w = 0 \) implies \( \dot{w} = 0 \). Thus, the expression above must vanish when we substitute \( z = h(y) \):

\[
A_2 h(y) + g_2(y,h(y)) - \frac{\partial h}{\partial y} \left( A_1 y + g_1(y,h(y)) \right) = 0.
\]

To find \( h(y) \) solve this differential equation for \( h \) as a function on \( y \).

If the exact solution is unavailable, an approximation is possible. For scalar \( y \), expand \( h(y) \) as

\[
h(y) = h_2 y^2 + \cdots + h_p y^p + o(y^{p+1}).
\]

where \( h_1 = h_0 = 0 \) because \( h(0) = \frac{\partial h}{\partial y}(0) = 0 \).

Example:

\[
\begin{align*}
\dot{y} &= yz \\
\dot{z} &= -z + ay^2 \quad a \neq 0
\end{align*}
\]

This is of the form (2) with \( g_1(y,z) = yz, g_2(y,z) = ay^2, A_2 = -1. \)

Thus \( h(y) \) must satisfy

\[
-h(y) + ay^2 - \frac{\partial h}{\partial y} y h(y) = 0.
\]

Try \( h(y) = h_2 y^2 + o(y^3) \):

\[
0 = -h_2 y^2 + o(y^3) + ay^2 - (2h_2 y + o(y^2)) y (h_2^2 + o(y^3))
\]

\[
= (a - h_2) y^2 + o(y^3)
\]

\[
\implies h_2 = a
\]

Reduced System: \( \dot{y} = y(ay^2 + o(y^3)) = ay^3 + o(y^4). \)

If \( a < 0 \), the full systems is asymptotically stable. If \( a > 0 \) unstable.
Discrete-Time Models and a Chaos Example

CT: \( \dot{x}(t) = f(x(t)) \)
\( f(x^*) = 0 \)

DT: \( x_{n+1} = f(x_n) \) \( n = 0, 1, 2, \ldots \)
\( f(x^*) = x^* \) ("fixed point")

Asymptotic stability criterion:
\[ \Re \lambda_i(A) < 0 \text{ where } A \triangleq \frac{\partial f}{\partial x} \bigg|_{x=x^*} \]
\[ |\lambda_i(A)| < 1 \text{ where } A \triangleq \frac{\partial f}{\partial x} \bigg|_{x=x^*} \]
\[ f'(x^*) < 0 \text{ for first order system} \]

These criteria are inconclusive if the respective inequality is not strict, but for first order systems we can determine stability graphically:

Cobweb Diagrams for First Order Discrete-Time Systems

Example: \( x_{n+1} = \sin(x_n) \) has unique fixed point at 0. Stability test above inconclusive since \( f'(0) = 1 \). However, the "cobweb" diagram below illustrates the convergence of iterations to 0:

In discrete time, even first order systems can exhibit oscillations:
Detecting Cycles Analytically

\[ f(p) = q \quad f(q) = p \quad \implies \quad f(f(p)) = p \quad f(f(q)) = q \]

For the existence of a period-2 cycle, the map \( f(f(\cdot)) \) must have two fixed points in addition to the fixed points of \( f(\cdot) \).

Period-3 cycles: fixed points of \( f(f(f(\cdot))) \).

Chaos in a Discrete Time Logistic Growth Model

\[ x_{n+1} = r(1 - x_n)x_n \quad (3) \]

Range of interest: \( 0 \leq x \leq 1 \) \( (x_n > 1 \implies x_{n+1} < 0) \)

We will study the range \( 0 \leq r \leq 4 \) so that \( f(x) = r(1 - x)x \) maps \([0, 1]\) onto itself.

Fixed points: \( x = r(1 - x)x \) \( \implies \) \( \left\{ \begin{array}{l} x^* = 0 \quad \text{and} \\ x^* = 1 - \frac{1}{r} \quad \text{if } r > 1. \end{array} \right. \)

\( r \leq 1 \): \( x^* = 0 \) unique and stable fixed point

\( r > 1 \): \( x = 0 \) unstable because \( f'(0) = r > 1 \)
Note that a transcritical bifurcation occurred at $r = 1$, creating the new equilibrium

$$x^* = 1 - \frac{1}{r}.$$ 

Evaluate its stability using $f'(x^*) = r(1 - 2x^*) = 2 - r$.

$$r < 3 \Rightarrow |f'(x^*)| < 1 \text{ (stable)}$$
$$r > 3 \Rightarrow |f'(x^*)| > 1 \text{ (unstable)}.$$

At $r = 3$, a period-2 cycle is born:

$$x = f(f(x)) = r(1 - f(x))f(x) = r(1 - r(1 - x))r(1 - x)x = r^2x(1 - x)(1 - r + rx - rx^2)$$

$$0 = r^2x(1 - x)(1 - r + rx - rx^2) - x$$

Factor out $x$ and $(x - 1 + \frac{1}{r})$, find the roots of the quotient:

$$p, q = \frac{r + 1 \mp \sqrt{r - 3}(r + 1)}{2r}$$

This period-2 cycle is stable when $r < 1 + \sqrt{6} = 3.4494$:

$$\left.\frac{d}{dx}f(f(x))\right|_{x=p} = f'(f(p))f'(p) = f'(p)f'(q) = 4 + 2r - r^2$$

$$|4 + 2r - r^2| < 1 \Rightarrow 3 < r < 1 + \sqrt{6} = 3.4494$$

At $r = 3.4494$, a period-4 cycle is born!

"period doubling bifurcations"
\[ r_1 = 3 \quad \text{period-2 cycle born} \]
\[ r_2 = 3.4494 \quad \text{period-4 cycle born} \]
\[ r_3 = 3.544 \quad \text{period-8 cycle born} \]
\[ r_4 = 3.564 \quad \text{period-16 cycle born} \]
\[ \vdots \]
\[ r_\infty = 3.5699 \]

After \( r > r_\infty \), chaotic behavior for a window of \( r \), followed by windows of periodic behavior (e.g., period-3 cycle around \( r = 3.83 \)).

Below is the cobweb diagram for \( r = 3.9 \) which is in the chaotic regime:
Mathematical Background

\[ \dot{x} = f(t, x) \quad x(t_0) = x_0 \quad (1) \]

Do solutions exist? Are they unique? Do they depend continuously on \( x_0 \)?

- If \( f(t, x) \) is continuous (\( C^0 \)) then a solution exists, but \( C^0 \) is not sufficient for uniqueness.
  
  Example: \( \dot{x} = x^{1/3} \) with \( x(0) = 0 \)

\[ x(t) \equiv 0, \quad x(t) = \left( \frac{2}{3} t \right)^{3/2} \] are both solutions

- Sufficient condition for uniqueness: “Lipschitz continuity” (more restrictive than \( C^0 \))

\[ |f(t, x) - f(t, y)| \leq L|x - y| \quad (2) \]

**Definition:** \( f(t, x) \) is locally Lipschitz if every point \( x^0 \) has a neighborhood where (2) holds for all \( x, y \) in this neighborhood and for all \( t \) for some \( L \).

**Example:** \( x^{1/3} \) is NOT locally Lipschitz (due to \( \infty \) slope)

\( x^3 \) is locally Lipschitz:

\[ x^3 - y^3 = (x^2 + xy + y^2)(x - y) \]

in any nbhd of \( x^0 \), we can find \( L \) to upper bound this

\[ \implies |x^3 - y^3| \leq L|x - y| \]
• If \( f(x) \) is continuously differentiable (\( C^1 \)), then it is locally Lipschitz.

Examples: \( x^3, x^2, e^x \), etc.

The converse is not true: local Lipschitz \( \not\Rightarrow C^1 \)

Example:

Not differentiable at \( x = \pm 1 \), but locally Lipschitz:

\[
| \text{sat}(x) - \text{sat}(y) | \leq | x - y | \quad (L = 1).
\]

Definition continued: \( f(t, x) \) is globally Lipschitz if (2) holds \( \forall x, y \in \mathbb{R}^n \) (i.e., the same \( L \) works everywhere).

Examples: sat\((x)\) is globally Lipschitz. \( x^3 \) is not globally Lipschitz:

• Suppose \( f(x) \) is \( C^1 \). Then it is globally Lipschitz iff \( \frac{\partial f}{\partial t} \) is bounded.

\[
L = \sup_x |f'(x)|
\]
Preview of existence theorems:
1. \( f(t,x) \) is \( C^0 \) \( \implies \) existence of solution \( x(t) \) on finite interval \([0,t_f]\).
2. \( f(t,x) \) locally Lipschitz \( \implies \) existence and uniqueness of \([0,t_f]\).
3. \( f(t,x) \) globally Lipschitz \( \implies \) existence and uniqueness on \([0,\infty)\).

Examples:
- \( \dot{x} = x^2 \) (locally Lipschitz) admits unique solution on \([0,t_f]\), but \( t_f < \infty \) from Lecture 1 (finite escape).
- \( \dot{x} = A(t)x \) where \( \|A(t)\| \leq L \ \forall t \). Globally Lipschitz because
  \[
  |A(t)x - A(t)y| \leq L|x - y|.
  \]
  \( \implies \) Linear systems do not exhibit finite escape time.

**Normed Linear Spaces**

Definition: \( \mathbb{X} \) is a normed linear space if there exists a real-valued norm \( |\cdot| \) satisfying:
1. \( |x| \geq 0 \ \forall x \in \mathbb{X}, \ |x| = 0 \ \text{iff} \ x = 0 \).
2. \( |x + y| \leq |x| + |y| \ \forall x, y \in \mathbb{X} \) (triangle inequality)
3. \( |\alpha x| = |\alpha| \cdot |x| \ \forall \alpha \in \mathbb{R} \) and \( x \in \mathbb{X} \).

Definition: A sequence \( \{x_k\} \) in \( \mathbb{X} \) is said to be a Cauchy sequence if
\[
|x_k - x_m| \to 0 \ \text{as} \ k, m \to \infty.
\]

Every convergent sequence is Cauchy. The converse is not true.

Definition: \( \mathbb{X} \) is a Banach space if every Cauchy sequence is convergent.

All Euclidean spaces are Banach spaces.

Example:
\( C^n[a,b] \): the set of all continuous functions \([a,b] \to \mathbb{R}^n\) with norm:
\[
|x|_C = \max_{t \in [a,b]} |x(t)|
\]
1. \( |x|_C \geq 0 \) and \( |x|_C = 0 \iff x(t) \equiv 0 \).
2. \( |x + y|_C = \max_{t \in [a,b]} |x(t) + y(t)| \leq \max_{t \in [a,b]} \{|x(t)| + |y(t)|\} \leq |x|_C + |y|_C \)
3. \( |\alpha \cdot x|_C = \max_{t \in [a,b]} |\alpha| \cdot |x(t)| = |\alpha| \cdot |x|_C \)

It can be shown that \( C^n[a,b] \) is a Banach space.
Fixed Point Theorems

\[ T(x) = x \]  \hspace{1cm} (4)

**Brouwer’s Theorem** (Euclidean spaces):
If \( U \) is a closed bounded subset of a Euclidean space and \( T : U \rightarrow U \) is continuous, then \( T \) has a fixed point in \( U \).

**Schauder’s Theorem** (Brouwer’s Thm → Banach spaces):
If \( U \) is a closed bounded convex subset of a Banach space \( X \) and \( T : U \rightarrow U \) is completely continuous* 2, then \( T \) has a fixed point in \( U \).

**Contraction Mapping Theorem:**
If \( U \) is a closed subset of a Banach space and \( T : U \rightarrow U \) is such that
\[
|T(x) - T(y)| \leq \rho |x - y| \quad \rho < 1 \quad \forall x, y \in U
\]
then \( T \) has a unique fixed point in \( U \) and the solutions of \( x_{n+1} = T(x_n) \) converge to this fixed point from any \( x_0 \in U \).

**Example:** The logistic map (Lecture 5)
\[ T(x) = rx(1 - x) \]  \hspace{1cm} (5)
with \( 0 \leq r \leq 4 \) maps \( U = [0, 1] \) to \( U \). \( |T'(x)| \leq r \quad \forall x \in [0, 1] \), so the contraction property holds with \( \rho = r \).

If \( r < 1 \), the contraction mapping theorem predicts a unique fixed point that attracts all solutions starting in \([0, 1]\).

Proof steps for the Contraction Mapping Thm:

1. Show that \( \{x_n\} \) formed by \( x_{n+1} = T(x_n) \) is a Cauchy sequence. Since we are in a Banach space, this implies a limit \( x^* \) exists.
2. Show that \( x^* = T(x^*) \).

3. Show that \( x^* \) is unique.

Details of each step:

1. For all \( n \geq 1 \),

\[
|x_{n+1} - x_n| = |T(x_n) - T(x_{n-1})| \leq \rho |x_n - x_{n-1}|
\]

\[
\leq \rho^2 |x_{n-1} - x_{n-2}|
\]

\[
\vdots
\]

\[
\leq \rho^n |x_1 - x_0|.
\]

Since \( \rho^n \rightarrow 0 \) as \( n \rightarrow \infty \), we have \( |x_{n+r} - x_n| \rightarrow 0 \) as \( n \rightarrow \infty \).

2. \( |x^* - T(x^*)| = |x^* - x_n + T(x_{n-1}) - T(x^*)| \)

\[
\leq |x^* - x_n| + |T(x_{n-1}) - T(x^*)|
\]

\[
\leq |x^* - x_n| + \rho |x^* - x_{n-1}|.
\]

Since \( \{x_n\} \) converges to \( x^* \), we can make this upper bound arbitrarily small by choosing \( n \) sufficiently large. This means that \( |x^* - T(x^*)| = 0 \), hence \( x^* = T(x^*) \).

3. Suppose \( y^* = T(y^*) \) \( y^* \neq x^* \).

\[
|x^* - y^*| = |T(x^*) - T(y^*)| \leq \rho |x^* - y^*| \implies x^* = y^*.
\]

Thus we have a contradiction.
Existence and Uniqueness Theorems for ODEs

\[ \dot{x} = f(t, x) \quad x(0) = x_0 \quad (1) \]

Theorem 1: If \( f(t, x) \) locally Lipschitz in \( x \) and continuous in \( t \)
\[ \implies \text{existence and uniqueness on some finite interval } [0, \delta]. \]

Sketch of the proof: From the local Lipschitz assumption, we can find \( r > 0 \)
and \( L > 0 \) such that
\[ |f(t, x) - f(t, y)| \leq L|x - y| \quad \forall x, y \in \{x \in \mathbb{R}^n : |x - x_0| \leq r\} \]
\[ \triangleq B(x_0, r) \]

If \( x(t) \) is a solution, then:
\[ x(t) = x_0 + \int_0^t f(\tau, x(\tau))d\tau. \]
\[ \triangleq T(x(t)) \]

To apply the Contraction Mapping Theorem:

1. Choose \( \delta \) small enough that \( T(\cdot) \) maps the following subset of \( C^n[0, \delta] \) to itself:
\[ U \triangleq \{x(t) \in C^n[0, \delta] : |x(t) - x_0|_C \leq r\}, \]
i.e.
\[ \max_{t \in [0, \delta]} |x(t) - x_0| \leq r \implies \max_{t \in [0, \delta]} |T(x(t)) - x_0| \leq r \]

\[ T(x(t)) - x_0 = \int_0^t f(\tau, x(\tau))d\tau = \int_0^t \left(f(\tau, x(\tau)) - f(\tau, x_0) + f(\tau, x_0)\right)d\tau \]
\[ |T(x(t)) - x_0| \leq \int_0^\delta |f(\tau, x(\tau)) - f(\tau, x_0)|d\tau + \int_0^\delta |f(\tau, x_0)|d\tau \]
\[ \quad \leq \int_0^\delta L|x(\tau) - x_0|d\tau + \int_0^\delta h d\tau \quad \text{where } h \text{ is a bound on } |f(\tau, x_0)| \]
\[ \quad \leq (Lr + h)\delta. \]

Thus by choosing \( \delta \leq \frac{r}{Lr + h} \) we can ensure that
\[ \max_{t \in [0, \delta]} |T(x(t)) - x_0| \leq (Lr + h)\delta \leq r. \]
2. Show that $T(\cdot)$ is a contraction in $U$, i.e., there exists $\rho < 1$ s.t. 

$$ x(\cdot), y(\cdot) \in U \implies |T(x(t)) - T(y(t))| \leq \rho |x(t) - y(t)|.$$ 

We see that 

$$ |T(x(t)) - T(y(t))| = \int_0^t |f(\tau, x(\tau)) - f(\tau, y(\tau))| d\tau \leq L \int_0^t |x(\tau) - y(\tau)| d\tau. $$ 

Therefore 

$$ \max_{t \in [0,\delta]} |T(x(t)) - T(y(t))| \leq L\delta \max_{t \in [0,\delta]} |x(t) - y(t)| $$ 

and $\rho < 1$ if $\delta \leq \frac{r}{L r + h}$ as prescribed above.

**Theorem 2**: $f(t, x)$ globally Lipschitz in $x$ and continuous in $t \implies$ existence and uniqueness on $[0, \infty)$.

**Proof**: Choose a $\delta$ that doesn’t depend on $x_0$ and apply Theorem 1 repeatedly to cover $[0,\infty)$. This is possible because $L$ works everywhere and we can pick $r$ as large as we wish. Indeed, for any $\delta < \frac{1}{L}$, we can choose $r$ large enough that $\delta \leq \frac{r}{L r + h}$.

Q: Why can’t we do this in Theorem 1?

A: $\delta$ depends on $x_0$ (no universal $L$) and $x_0$ changes at the next iteration. We can’t use the same $\delta$ in every iteration:

- The theorems above are sufficient only, and can be conservative:

  Example: $\dot{x} = -x^3$ is not globally Lipschitz but 

  $$ x(t) = \text{sgn}(x_0) \sqrt{\frac{x_0^2}{1 + 2tx_0^2}} $$

  is defined on $[0, \infty)$.

**Theorem 3**: (Continuous dependence on initial conditions) Let $x(t), y(t)$ be two solutions of $\dot{x} = f(t, x)$ starting from $x_0$ and $y_0$, and remaining in a set with Lipschitz constant $L$ on $[0, \tau]$. Then, for any $\epsilon > 0$, there exists $\delta(\epsilon, \tau) > 0$ such that 

$$ |x_0 - y_0| \leq \delta \implies |x(t) - y(t)| \leq \epsilon \ \forall t \in [0, \tau]. $$

- This conclusion does not hold on infinite time intervals (even if $f$ is globally Lipschitz).
Example: bistable system

If $\epsilon$ is smaller than the distance between the two stable equilibria, no choice of $\delta$ guarantees $|x(t) - y(t)| \leq \epsilon \ \forall t \geq 0$.

• Theorem 3 also shows continuous dependence on parameter $\mu$ in $f(t, x, \mu)$ with the following trick:

$$
\dot{x} = f(t, x, \mu) \\
\dot{\mu} = 0
$$

$$
X = \begin{bmatrix} x \\ \mu \end{bmatrix} \\
\dot{X} = F(t, X) \triangleq \begin{bmatrix} f(t, x, \mu) \\ 0 \end{bmatrix}
$$

Q: How do you reconcile bifurcations with continuous dependence on parameters? We could pick two values of the bifurcation parameter arbitrarily close, but one below and one above the critical value, thereby expecting a drastic difference in the solutions.

A: The two solutions are close in the short term (Theorem 3 holds on finite time intervals); the drastic difference builds up over time.

The following lemma is used in the proof of Theorem 3:

**Bellman-Gronwall Lemma:** Let $a(t)$, $u(t)$ be continuous functions on $[0, T]$ and let $a(t) \geq 0$. If a continuous function $z(t)$ satisfies:

$$
z(t) \leq u(t) + \int_0^t a(\tau)z(\tau)d\tau \ \forall t \in [0, T]
$$

then:

$$
z(t) \leq u(t) + \int_0^t a(\tau)u(\tau)e^{\int_\tau^t a(\sigma)d\sigma}d\tau \ \forall t \in [0, T].
$$

If $u(t) \equiv u$ is a constant then:

$$
z(t) \leq ue^{\int_0^t a(\sigma)d\sigma}.
$$

If, in addition, $a(t) \equiv a \geq 0$ is a constant, then

$$
z(t) \leq ue^{at}.
$$
Proof of Theorem 3:

\[ x(t) - y(t) = x_0 - y_0 + \int_0^t \left( f(\tau, x(\tau)) - f(\tau, y(\tau)) \right) d\tau \]

\[ |x(t) - y(t)| \leq |x_0 - y_0| + \int_0^t \underbrace{|f(\tau, x(\tau)) - f(\tau, y(\tau))|}_{L|x(\tau) - y(\tau)|} d\tau \]

Apply BG Lemma with \( z(t) = |x(t) - y(t)|, u \equiv |x_0 - y_0|, a \equiv L: \)

\[ |x(t) - y(t)| \leq |x_0 - y_0| e^{Lt} \]
\[ \leq e^{LT} |x_0 - y_0| \quad \forall t \in [0, T]. \]

Choose \( \delta \leq \frac{\epsilon}{eLT}. \)
Lyapunov Stability Theory

Consider a time invariant system
\[ \dot{x} = f(x) \]
and assume equilibrium at \( x = 0 \), i.e. \( f(0) = 0 \). If the equilibrium of interest is \( x^* \neq 0 \), let \( \tilde{x} = x - x^* \):
\[ \dot{\tilde{x}} = f(x) = f(\tilde{x} + x^*) \triangleq \tilde{f}(\tilde{x}) \implies \tilde{f}(0) = 0. \]

Definition: The equilibrium \( x = 0 \) is stable if for each \( \varepsilon > 0 \), there exists \( \delta > 0 \) such that
\[ |x(0)| \leq \delta \implies |x(t)| \leq \varepsilon \quad \forall t \geq 0. \]

It is unstable if not stable.

Asymptotically stable if stable and \( x(t) \to 0 \) for all \( x(0) \) in a neighborhood of \( x = 0 \).

Globally asymptotically stable if stable and \( x(t) \to 0 \) for every \( x(0) \).

Note that \( x(t) \to 0 \) does not necessarily imply stability: one can construct an example where trajectories converge to the origin, but only after a large detour that violates the stability definition.
Lyapunov’s Second (Indirect) Method

1. Let \( D \) be an open, connected subset of \( \mathbb{R}^n \) that includes \( x = 0 \). If there exists a \( C^1 \) function \( V : D \to \mathbb{R} \) such that
   \[
   V(0) = 0 \quad \text{and} \quad V(x) > 0 \quad \forall x \in D - \{0\} \quad \text{(positive definite)}
   \]
   and
   \[
   V(x) \triangleq \nabla V(x)^T \cdot f(x) \leq 0 \quad \forall x \in D \quad \text{(negative semidefinite)}
   \]
   then \( x = 0 \) is stable.

2. If \( \dot{V}(x) < 0 \quad \forall x \in D - \{0\} \quad \text{(negative definite)} \)
   then \( x = 0 \) is asymptotically stable.

3. If, in addition, \( D = \mathbb{R}^n \) and
   \[
   |x| \to \infty \implies V(x) \to \infty \quad \text{(radially unbounded)}
   \]
   then \( x = 0 \) is globally asymptotically stable.

Sketch of the proof:
The sets \( \Omega_c \triangleq \{ x : V(x) \leq c \} \) for constants \( c \) are called level sets of \( V \) and are positively invariant because \( \nabla V(x)^T f(x) \leq 0 \).

Stability follows from this property: choose a level set inside the ball of radius \( \varepsilon \), and a ball of radius \( \delta \) inside this level set. Trajectories starting in \( B_\delta \) can’t leave \( B_\varepsilon \) since they remain inside the level set.
Asymptotic stability:
Since $V(x(t))$ is decreasing and bounded below by 0, we conclude
\[ V(x(t)) \to c \geq 0. \]
We will show $c = 0$ (i.e., $x(t) \to 0$) by contradiction. Suppose $c \neq 0$:

Let 
\[ \gamma \triangleq \max_{\{x \mid c \leq V(x) \leq V(x_0)\}} -\dot{V}(x) > 0 \]
where the maximum exists because it is evaluated over a bounded set, and is positive because $\dot{V}(x) < 0$ away from $x = 0$. Then,
\[ \dot{V}(x) \leq -\gamma \implies V(x(t)) \leq V(x_0) - \gamma t, \]
which implies $V(x(t)) < 0$ for $t > \frac{V(x_0)}{\gamma}$ — a contradiction because $V \geq 0$. Therefore, $c = 0$ which implies $x(t) \to 0$.

Global asymptotic stability:
Why do we need radial unboundedness?
Example:
\[ V(x) = \frac{x_1^2}{1 + x_1^2} + x_2^2 \] (2)
Set $x_2 = 0$, let $x_1 \to \infty$: $V(x) \to 1$ (not radially unbounded). Then $\Omega_c$ is not a bounded set for $c \geq 1$:

Therefore, $x_1(t)$ may grow unbounded while $V(x(t))$ is decreasing.
Finding Lyapunov Functions

Example: \[ \dot{x} = -g(x) \quad x \in \mathbb{R}, \quad xg(x) > 0 \quad \forall x \neq 0 \] (3)

\[ V(x) = \frac{1}{2}x^2 \] is positive definite and radially unbounded.

\[ \dot{V}(x) = -xg(x) \] is negative definite. Therefore \( x = 0 \) is globally asymptotically stable.

If \( xg(x) > 0 \) only in \((-b,c) - \{0\}\), then take \( D = (-b,c) \)

\( \implies x = 0 \) is locally asymptotically stable.

There are other equilibria where \( g(x) = 0 \), so we know global asymptotic stability is not possible.

Example:

\[ \begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -ax_2 - g(x_1) \quad a \geq 0, \quad xg(x) > 0 \quad \forall x \in (-b,c) - \{0\}
\end{align*} \] (4)

The choice \( V(x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \) doesn’t work because \( \dot{V}(x) \) is sign indefinite (show this).

The function

\[ V(x) = \int_0^{x_1} g(y) dy + \frac{1}{2} x_2^2 \]

is positive definite on \( D = (-b,c) - \{0\} \) and

\[ \dot{V}(x) = g(x_1)x_2 - ax_2^2 - x_2g(x_1) = -ax_2^2 \]

is negative semidefinite \( \implies \) stable.

If \( a = 0 \), no asymptotic stability because \( \dot{V}(x) = 0 \implies V(x(t)) = V(x(0)) \).

If \( a > 0 \), (4) is asymptotically stable but the Lyapunov function above doesn’t allow us to reach that conclusion. We need either another \( V(x) \) with negative definite \( \dot{V}(x) \), or the Lasalle-Krasovskii Invariance Principle to be discussed in the next lecture.
LaSalle-Krasovskii Invariance Principle

- Applicable to time-invariant systems.
- Allows us to conclude asymptotic stability from $\dot{V}(x) \leq 0$ if additional conditions hold:

Suppose $\Omega_c = \{ x : V(x) \leq c \}$ is bounded and $\dot{V}(x) \leq 0$ in $\Omega_c$. Define $S = \{ x \in \Omega_c : \dot{V}(x) = 0 \}$ and let $M$ be the largest invariant set in $S$. Then, for every $x(0) \in \Omega_c$, $x(t) \to M$.

Corollary: If no solution other than $x(t) \equiv 0$ can stay identically in $S$ then $M = \{ 0 \}$ and we conclude asymptotic stability.

Example (from last lecture):
\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -ax_2 - g(x_1) & a > 0, xg(x) > 0 \forall x \neq 0
\end{align*}
\] (1)

\[ V(x) = \int_0^{x_1} g(y)dy + \frac{1}{2}x_2^2 \implies \dot{V}(x) = -ax_2^2 \]

$S = \{ x \in \Omega_c | x_2 = 0 \}$

If $x(t)$ stays identically in $S$, then $x_2(t) \equiv 0 \implies \dot{x}_2(t) \equiv 0 \implies g(x_1(t)) \equiv 0 \implies x_1(t) \equiv 0 \implies$ asymptotic stability from Corollary.

Example (linear system): Same system above with $g(x_1) = bx_1$:
\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -ax_2 - bx_1 & a > 0, b > 0
\end{align*}
\] (2)

\[ V(x) = \frac{b}{2}x_1^2 + \frac{1}{2}x_2^2 \implies \dot{V}(x) = -ax_2^2 \implies \text{Invariance Principle works as in the example above.} \]

Alternatively, construct another Lyapunov function with negative definite $\dot{V}(x)$. Try $V(x) = x^TPx$ where $P = P^T > 0$ is to be selected.

\[ \dot{V}(x) = x^TP\dot{x} + \dot{P}x = x^T(A^TP + PA)x \text{ where } A = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} \]

Let $P = \frac{1}{\epsilon} \begin{bmatrix} b & \epsilon \\ \epsilon & 1 \end{bmatrix}$, that is $V(x) = \frac{b}{2}x_1^2 + \epsilon x_1 x_2 + \frac{1}{2}x_2^2$.

Note that $P > 0$ if $\epsilon^2 < b$. 

\[ A^T P + PA = \begin{bmatrix} -eb & -ea/2 \\ -ea/2 & \epsilon - a \end{bmatrix} \leq 0 \text{ if } \epsilon = 0 \\
< 0 \text{ if } 0 < \epsilon < a \text{ and } eb(a - \epsilon) > \frac{\epsilon^2 a^2}{4} \]

\[ 0 < \epsilon < \frac{ba}{b + \frac{\epsilon}{2}} \]

**Linear Systems**

\[
\dot{x} = Ax \quad x \in \mathbb{R}^n
\]

(3)

\[ x = 0 \text{ is stable if } \Re\{\lambda_i(A)\} \leq 0 \text{ for all } i = 1, \cdots, n \text{ and eigenvalues on the imaginary axis have Jordan blocks of order one.}^2 \]

\[ \text{It is asymptotically stable if } \Re\{\lambda_i(A)\} < 0 \text{ for all } i, \text{i.e., } A \text{ is "Hurwitz."} \]

**Example:**

\[
A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ is unstable: } \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = 0 \end{cases} \]

\[
x_1(t) = x_1(0) + x_2(0)t
\]

\[ A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ is stable.} \]

\[ \text{Lyapunov Functions for Linear Systems} \]

\[ V(x) = x^T P x \quad P = P^T > 0 \]

(4)

\[ V(x) = x^T (A^T P + PA) x \]

If \( \exists P = P^T > 0 \) such that \( A^T P + PA = -Q < 0 \), then \( A \) is Hurwitz.

The converse is also true:

**Theorem:** \( A \) is Hurwitz if and only if for any \( Q = Q^T > 0 \), there exists \( P = P^T > 0 \) such that

\[ A^T P + PA = -Q. \]

(5)

Moreover, the solution \( P \) is unique.

**Proof:**

(if) From (4) above, the Lyapunov function \( V(x) = x^T P x \) proves asymptotic stability which means \( A \) is Hurwitz.

(only if) Assume \( \Re\{\lambda_i(A)\} < 0 \forall i \). Show \( \exists P = P^T > 0 \) such that \( A^T P + PA = -Q \).

**Candidate:**

\[ P = \int_0^\infty e^{A^T t} Q e^{A t} dt. \]

(6)
• The integral exists because \( \|e^{At}\| \leq \kappa e^{-\alpha t} \).

• \( P = P^T \)

• \( P > 0 \) because \( x^T P x = \int_0^\infty (e^{At} x)^T Q (e^{At} x) \, dt \geq 0 \) and
  \[ x^T P x = 0 \implies \phi(t, x) = 0 \implies x = 0 \] because \( e^{At} \) is nonsingular.

• \( A^T P + PA = \int_0^\infty \left( A^T e^{At} Q e^{At} + e^{At} Q e^{At} A \right) \, dt \)
  \[ = \frac{d}{dt} \left( e^{At} Q e^{At} \right) \bigg|_0^\infty = 0 - Q = -Q \]

Uniqueness:
Suppose there is another \( \hat{P} = \hat{P}^T > 0 \) satisfying \( \hat{P} \neq P \), and \( A^T \hat{P} + \hat{P} A = -Q \).
\[ \implies (P - \hat{P}) A + A^T (P - \hat{P}) = 0 \]
Define \( W(x) = x^T (P - \hat{P}) x \).
\[ \frac{d}{dt} W(x(t)) = 0 \implies W(x(t)) = W(x(0)) \quad \forall t. \]
Since \( A \) is Hurwitz, \( x(t) \to 0 \) and \( W(x(t)) \to 0 \).
Combining the two statements above, we conclude \( W(x(0)) = 0 \) for any \( x(0) \). This is possible only if \( P - \hat{P} = 0 \) which contradicts \( \hat{P} \neq P \).

**Invariance Principle Applied to Linear Systems**

\[ A^T P + PA = -Q \leq 0 \quad (7) \]
Can we conclude that \( A \) is Hurwitz if \( Q \) is only semidefinite?
Decompose \( Q \) as \( Q = C^T C \) where \( C \in \mathbb{R}^{r \times n} \), \( r \) is the rank of \( Q \).
\[ \dot{V}(x) = -x^T Q x = -x^T C^T C x = -y^T y \]
where \( y \triangleq C x \). The invariance principle guarantees asymptotic stability if
\[ y(t) = C x(t) \equiv 0 \implies x(t) \equiv 0. \]
This implications is true if the pair \((C, A)\) is observable.

**Example** (beginning of the lecture):
\[ A = \begin{bmatrix} 0 & 1 \\ -b & a \end{bmatrix} \quad Q = \begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix} \quad \implies C = \begin{bmatrix} 0 & \sqrt{a} \end{bmatrix} \]
\((C, A)\) is observable if \( b \neq 0 \).
Lyapunov’s Direct Method

\[ \dot{x} = f(x) \quad f(0) = 0 \]

Define \( A = \frac{\partial f(x)}{\partial x} \bigg|_{x=0} \) and decompose \( f(x) \) as

\[ f(x) = Ax + g(x) \quad \text{where} \quad \frac{|g(x)|}{|x|} \rightarrow 0 \quad \text{as} \quad |x| \rightarrow 0. \]

**Theorem:** The origin is asymptotically stable if \( A \) is Hurwitz and unstable if \( \Re\{\lambda_i(A)\} > 0 \) for at least one eigenvalue.

**Note:** We can conclude only local asymptotic stability from this linearization. Inconclusive if \( A \) has eigenvalues on the imaginary axis.

**Proof:** Find \( P = P^T > 0 \) such that \( A^T P + PA = -Q < 0 \). Use \( V(x) = x^T P x \) as a Lyapunov function for the nonlinear system \( \dot{x} = Ax + g(x) \).

\[
V(x) = x^T P(Ax + g(x)) + (Ax + g(x))^T Px \\
= x^T (PA + A^T P)x + 2x^T Pg(x) \\
\leq -x^T Qx + 2|x|\|P||g(x)|
\]

\[
\lambda_{\min}(Q)|x|^2 \leq x^T Qx \leq \lambda_{\max}(Q)|x|^2
\]

\[ \dot{V}(x) \leq -\lambda_{\min}(Q)|x|^2 + 2\|P\||x||g(x)| \]

Since \( \frac{|g(x)|}{|x|} \rightarrow 0 \), for any \( \gamma > 0 \) we can find \( r > 0 \) such that \( |x| \leq r \implies |g(x)| \leq \gamma|x| \).

Choose \( \gamma < \frac{\lambda_{\min}(Q)}{2\|P\|} \). Then \( \dot{V}(x) \) is negative definite in a ball with radius \( r(\gamma) \) around the origin, \( B_{r(\gamma)}(0) \). We appeal to Lyapunov’s indirect method to conclude (local) asymptotic stability.
Region of Attraction

\[ R_A = \{ x : \phi(t, x) \to 0 \} \]  

(1)

“Quantifies” local asymptotic stability. Global asymptotic stability: \( R_A = \mathbb{R}^n \).

Proposition: If \( x = 0 \) is asymptotically stable, then its region of attraction is an open, connected, invariant set. Moreover, the boundary is formed by trajectories.

Example: van der Pol system in reverse time:

\[
\begin{align*}
\dot{x}_1 &= -x_2 \\
\dot{x}_2 &= x_1 - x_2 + x_2^3 
\end{align*}
\]  

(2)

The boundary is the (unstable) limit cycle. Trajectories starting within the limit cycle converge to the origin.

Estimating the Region of Attraction with a Lyapunov Function

Suppose \( \dot{V}(x) < 0 \) in \( D - \{0\} \). The level sets of \( V \) inside \( D \) are invariant and trajectories starting in them converge to the origin.
Therefore we can use the largest level set of $V$ that fits into $D$ as an (under)approximation of the region of attraction.

$$D \uparrow \{x : V(x) \leq c\} \subset R_A$$

This estimate depends on the choice of Lyapunov function. A simple (but often conservative) choice is: $V(x) = x^T P x$ where $P$ is selected for the linearization (see p.1).

**Time-Varying Systems**

$$\dot{x} = f(t, x) \quad f(t, 0) \equiv 0 \quad (4)$$

To simplify the definitions of stability and asymptotic stability for the equilibrium $x = 0$, we first define a class of functions known as "comparison functions."

**Comparison Functions**

**Definition:** A continuous function $\alpha : [0, \infty) \rightarrow [0, \infty)$ is class-$K$ if it is zero at zero and strictly increasing. It is class-$K_\infty$ if, in addition, $\alpha(r) \rightarrow \infty$ as $r \rightarrow \infty$.

A continuous function $\beta : [0, \infty) \times [0, \infty) \rightarrow [0, \infty)$ is class-$KL$ if:

1. $\beta(\cdot, s)$ is class-$K$ for every fixed $s$,
2. $\beta(r, \cdot)$ is decreasing and $\beta(r, s) \rightarrow 0$ as $s \rightarrow \infty$, for every fixed $r$.

**Example:** $\alpha(r) = \tan^{-1}(r)$ is class-$K$, $\alpha(r) = r^c, c > 0$ is class-$K_\infty$, $\beta(r, s) = r^c e^{-s}$ is class-$KL$.

**Proposition:** If $V(\cdot)$ is positive definite, then we can find class-$K$ functions $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ such that

$$\alpha_1(|x|) \leq V(x) \leq \alpha_2(|x|). \quad (5)$$

If $V(\cdot)$ is radially unbounded, we can choose $\alpha_1(\cdot)$ to be class-$K_\infty$.

**Example:** $V(x) = x^T P x \quad P = P^T > 0$

$$\alpha_1(|x|) = \lambda_{\min}(P)|x|^2 \quad \alpha_2(|x|) = \lambda_{\max}(P)|x|^2.$$

Khalil (Sec. 4.5), Sastry (Sec. 5.2)
Stability Definitions

**Definition:** $x = 0$ is stable if for every $\epsilon > 0$ and $t_0$, there exists $\delta > 0$ such that

$$|x(t_0)| \leq \delta(t_0, \epsilon) \implies |x(t)| \leq \epsilon \quad \forall t \geq t_0.$$ 

If the same $\delta$ works for all $t_0$, i.e. $\delta = \delta(\epsilon)$, then $x = 0$ is uniformly stable.

It is easier to define uniform stability and uniform asymptotic stability using comparison functions:

- $x = 0$ is uniformly stable if there exists a class-$K$ function $\alpha(\cdot)$ and a constant $c > 0$ such that
  $$|x(t)| \leq \alpha(|x(t_0)|)$$
  for all $t \geq t_0$ and for every initial condition such that $|x(t_0)| \leq c$.

- uniformly asymptotically stable if there exists a class-$K\mathcal{L}$ $\beta(\cdot, \cdot)$ s.t.
  $$|x(t)| \leq \beta(|x(t_0)|, t - t_0)$$
  for all $t \geq t_0$ and for every initial condition such that $|x(t_0)| \leq c$.

- globally uniformly asymptotically stable if $c = \infty$.

- uniformly exponentially stable if $\beta(r, s) = k e^{-\lambda s}$ for some $k, \lambda > 0$:
  $$|x(t)| \leq k|x(t_0)|e^{-\lambda(t-t_0)}$$
  for all $t \geq t_0$ and for every initial condition such that $|x(t_0)| \leq c$.

**Example:** Consider the following system, defined for $t > -1$:

$$\dot{x} = \frac{-x}{1+t} \quad (6)$$

$$x(t) = x(t_0)e^{\int_{t_0}^{t} \frac{1}{1+s} \, ds} = x(t_0)e^{\log(1+t)|_{t_0}}$$

$$= x(t_0)e^{\log \frac{1+t}{1+t_0}} = x(t_0) \frac{1+t_0}{1+t}$$

$$|x(t)| \leq |x(t_0)| \implies \text{the origin is uniformly stable with } \alpha(r) = r.$$ 

The origin is also asymptotically stable, but not uniformly, because the convergence rate depends on $t_0$:

$$x(t) = x(t_0) \frac{1+t_0}{1+t_0 + (t-t_0)} = \frac{x(t_0)}{1 + \frac{t-t_0}{1+t_0}}.$$
Time-Varying Systems Continued

Uniform stability: There exists a class $\mathcal{K}$ function $\alpha(\cdot)$ and a constant $c > 0$, both independent of $t_0$, such that
\begin{equation}
|x(t)| \leq \alpha(|x(t_0)|) \quad \forall t \geq t_0 \quad \text{when} \quad |x(t_0)| \leq c.
\end{equation}

Uniform asymptotic stability: There exists a class $\mathcal{KL}$ function $\beta(\cdot, \cdot)$ and a constant $c > 0$ such that
\begin{equation}
|x(t)| \leq \beta(|x(t_0)|, |t - t_0|) \quad \forall t \geq t_0 \quad \text{when} \quad |x(t_0)| \leq c.
\end{equation}

Uniform exponential stability: There exist constants $k, \lambda, c > 0$ s.t.
\begin{equation}
|x(t)| \leq k|x(t_0)|e^{-\lambda(t-t_0)} \quad \forall t \geq t_0 \quad \text{when} \quad |x(t_0)| \leq c,
\end{equation}
that is $\beta(r, s) = kre^{-\lambda s}$.

$k > 1$ allows for overshoot:

\[
\begin{aligned}
&k|x(t_0)| \quad \text{arrow} \quad |x(t)| \quad \text{arrow} \quad k|x(t_0)|e^{-\lambda(t-t_0)} \\
&|x(t_0)| \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \\
&t_0 \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \quad \text{arrow} \\
&k|x(t_0)| \quad \text{arrow} \quad |x(t)| \quad \text{arrow} \quad k|x(t_0)|e^{-\lambda(t-t_0)}
\end{aligned}
\]

Example:

\[
\dot{x} = -x^3 \implies x(t) = \text{sgn}(x(t_0))\sqrt{\frac{x_0^2}{1 + 2(t-t_0)x_0}}
\]

$x = 0$ is asymptotically stable but not exponentially stable.

Proposition: $x = 0$ is exponentially stable for $\dot{x} = f(x)$, $f(0) = 0$, if and only if $A \triangleq \frac{df}{dx}|_{x=0}$ is Hurwitz, that is $\Re \lambda_i(A) < 0 \ \forall i$.

Although strict inequality in $\Re \lambda_i(A) < 0$ is not necessary for asymptotic stability (see example above where $A = 0$), it is necessary for exponential stability.
Lyapunov’s Indirect Method for Time-Varying Systems

1. If \( W_1(x) \leq V(t, x) \leq W_2(x) \) and \( \dot{V}(t, x) = \frac{\partial V}{\partial x} + \frac{\partial W}{\partial x} f(t, x) \leq 0 \) for some positive definite functions \( W_1(\cdot), W_2(\cdot) \) on a domain \( D \) that includes the origin, then \( x = 0 \) is uniformly stable.

2. If, further, \( \dot{V}(t, x) \leq -W_3(x) \) \( \forall x \in D \) for some positive definite \( W_3(\cdot) \), then \( x = 0 \) is uniformly asymptotically stable.

3. If \( D = \mathbb{R}^n \) and \( W_1(\cdot) \) is radially unbounded, then \( x = 0 \) is globally uniformly asymptotically stable.

4. If \( W_i(x) = k_i|\alpha|^a, i = 1, 2, 3, \) for some constants \( k_1, k_2, k_3, a > 0 \), then \( x = 0 \) is uniformly exponentially stable.

Proof:

1. \( \alpha_1(|\alpha|) \leq W_1(x) \leq V(t, x) \leq W_2(x) \leq \alpha_2(|\alpha|) \)
\[
\dot{V} \leq 0 \implies V(x(t), t) \leq V(x(t_0), t_0) \\
\implies \alpha_1(|\alpha(t)|) \leq \alpha_2(|\alpha(t_0)|) \\
\implies |\alpha(t)| \leq \alpha(|\alpha(t_0)|) \triangleq (\alpha^{-1}_1 \circ \alpha_2)(|\alpha(t_0)|).
\]

Note: The inverse of a class-\( K \) function is well defined locally (globally if \( K_{\infty} \)) and is class-\( K \). The composition of two class-\( K \) functions is also class-\( K \).

2. \( \dot{V} \leq -W_3(x) \leq -\alpha_3(|\alpha|) \leq -\alpha_3(\alpha^{-1}_2(V)) \triangleq -\gamma(V) \)
\[
\frac{d}{dt}V(t, x(t)) \leq -\gamma(V(t, x(t)))
\]
Let \( y(t) \) be the solution of \( \dot{y} = -\gamma(y), y(t_0) = V(t_0, x(t_0)) \). Then,
\[
V(t, x(t)) \leq y(t).
\]

Since \( \dot{y} = -\gamma(y) \) is a first order differential equation and \( -\gamma(y) < 0 \) when \( y > 0 \), we conclude monotone convergence of \( y(t) \) to 0:
\[
y(t) = \beta(y(t_0), t - t_0) \implies V(t, x(t)) \leq \beta(V(t_0, x(t_0)), t - t_0) \\
\leq \beta_k(|x(t)|)
\]
\[
\implies \alpha_1(|\alpha(t)|) \leq \beta_k(|\alpha(t_0)|), t - t_0 \\
\implies |\alpha(t)| \leq \tilde{\beta}(\beta_k(|\alpha(t_0)|), t - t_0) \triangleq \alpha^{-1}_1(\beta_k(|\alpha(t_0)|), t - t_0)
\]
3. If $a_1(\cdot)$ is class $\mathcal{K}_\infty$ then $a_1^{-1}(\cdot)$ exists globally above.

4. $a_3(|x|) = k_3|x|^a$, $a_2(|x|) = k_2|x|^a$

\[ \implies \gamma(V) = a_3(a_2^{-1}(V)) = k_3 \left( \frac{V}{k_2} \right)^{\frac{1}{a}} = \frac{k_3}{k_2} V \]

\[ \dot{y} = -\frac{k_3}{k_2} y \implies y(t) = y(t_0) e^{-(k_2/k_2)(t-t_0)} \]

\[ \beta(r,s) = r e^{-(k_3/k_2)s} \implies \tilde{\beta}(r,s) = \left( \frac{k_2}{k_1} r e^{-(k_3/k_2)s} \right)^{\frac{1}{a}} = \left( \frac{k_2}{k_1} \right)^{\frac{1}{a}} r^{\frac{k_3}{k_2} s} \]

**Example:**

\[ \dot{x} = -g(t)x^3 \quad \text{where} \quad g(t) \geq 1 \quad \text{for all} \quad t \]

\[ V(x) = \frac{1}{2} x^2 \implies \dot{V}(t,x) = -g(t)x^4 \leq -x^4 \triangleq W_3(x) \]

Globally uniformly asymptotically stable but not exponentially stable. Take $g(t) \equiv 1$ as a special case:

\[ \dot{x} = -x^3 \implies x(t) = \text{sgn}(x(t_0)) \sqrt{\frac{x_0^2}{1 + 2(t-t_0)x_0^2}} \]

which converges slower than exponentially.

**Example:** $\dot{x} = A(t)x$. Take $V(x) = x^T P(t)x$:

\[ \dot{V}(x) = x^T \dot{P}(t)x + \dot{x}^T P(t)x + x^T P(t) \dot{x} \]

\[ = x^T (\dot{P} + A^T P + PA)x \triangleq Q(t) \]

If $k_1 I \leq P(t) \leq k_2 I$ and $k_3 I \leq Q(t)$, $k_1, k_2, k_3 > 0$, then

\[ k_1 |x|^2 \leq V(t,x) \leq k_2 |x|^2 \quad \text{and} \quad \dot{V}(t,x) \leq -k_3 |x|^2 \]

\[ \implies \text{global uniform exponential stability.} \]

What if $W_3(\cdot)$ is only semidefinite? \hfill Khalil, Section 8.3

Lasalle-Krasovskii Invariance Principle is **not** applicable to time-varying systems. Instead, use the following (weaker) result:

**Theorem:** Suppose $W_1(x) \leq V(t,x) \leq W_2(x)$

\[ \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t,x) \leq -W_3(x), \]

where $W_1(\cdot), W_2(\cdot)$ are positive definite and $W_3(\cdot)$ is positive semidefinite. Suppose, further, $W_1(\cdot)$ is radially unbounded, $f(t,x)$ is locally Lipschitz in $x$ and bounded in $t$, and $W_3(\cdot)$ is $C^1$. Then

\[ W_3(x(t)) \to 0 \quad \text{as} \quad t \to \infty. \]
**Note:** This proves convergence to \( S = \{ x : W_3(x) = 0 \} \) whereas the Invariance Principle, when applicable, guarantees convergence to the largest invariant set within \( S \).

**Example:**

\[
\begin{align*}
\dot{x}_1 &= -x_1 + w(t)x_2 \\
\dot{x}_2 &= -w(t)x_1
\end{align*}
\] *(5)*

\( V(t, x) = \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 \implies V(t, x) = -x_1^2 \). Thus, \( x_1(t) \to 0 \) as \( t \to \infty \), but no guarantee about the convergence of \( x_2(t) \) to zero.

If \( w(t) \equiv w \neq 0 \), then we can use the Invariance Principle and conclude \( x_2(t) \to 0 \) from the observability of \( x_2 \) from \( x_1 \).

**Next time:** a class of functions \( w(t) \) that guarantee \( x_2(t) \to 0 \) using observability notions for linear time varying systems.

**Barbalat’s Lemma** (used in proving the theorem above):

If \( \lim_{t \to \infty} \int_0^t \phi(\tau)d\tau \) exists and is finite, and \( \phi(\cdot) \) is uniformly continuous\(^2\) then \( \phi(t) \to 0 \) as \( t \to \infty \).

Uniform continuity in Barbalat’s Lemma can’t be relaxed:

Example: Let \( \phi(t) \) be a sequence of pulses centered at \( k = 1, 2, 3, \ldots \) with amplitude \( k \), width \( \frac{1}{k^3} \), then

\[
\int_0^\infty \phi(t)dt = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty \quad \text{but} \quad \phi(t) \not\to 0.
\]

Proof of the theorem:

\[
\begin{align*}
\alpha_1(|x|) &\leq V(t, x) \leq \alpha_2(|x|) \quad \alpha_1 \in K_{\infty} \\
\implies |x(t)| &\leq \alpha_1^{-1}(\alpha_2(|x(t_0)|))
\end{align*}
\]

\( x(t) \) bounded \( \implies \dot{x} = f(t, x) \) is bounded \( \implies x(t) \) is uniformly continuous.

\[
\begin{align*}
V(t, x) &\leq -W_3(x(t)) \\
\implies V(x(T)) - V(x(t_0), t_0) &\leq -\int_{t_0}^{T} W_3(x(t))dt \\
\implies \int_{t_0}^{\infty} W_3(x(t))dt &\leq V(x(t_0), t_0) < \infty.
\end{align*}
\]

Since \( W_3(\cdot) \) is \( C^1 \), it is uniformly continuous on the bounded domain where \( x(t) \) resides. So, by Barbalat’s Lemma, \( W_3(x(t)) \to 0 \) as \( t \to \infty \).
Linear Time-Varying Systems

\[ \dot{x} = A(t)x \quad x(t) = \Phi(t, t_0)x(t_0) \]  (1)

where \( \Phi(t, t_0) \) is the state transition matrix.

- No eigenvalue test for stability in the time-varying case:
  
  \[
  A(t) = \begin{bmatrix}
  -1 + 1.5 \cos^2 t & 1 - 1.5 \sin t \cos t \\
  -1 - 1.5 \sin t \cos t & -1 + 1.5 \sin^2 t
  \end{bmatrix}
  \]

  eigenvalues: \(-0.25 \pm i0.25\sqrt{7}\) for all \(t\), but unstable:

  \[
  \Phi(t, 0) = \begin{bmatrix}
  e^{0.5t} \cos t & e^{-t} \sin t \\
  e^{-0.5t} \sin t & e^{-t} \cos t
  \end{bmatrix}
  \]

- For linear systems uniform asymptotic stability is equivalent to uniform exponential stability:

  \[ \text{Theorem}^2: \ x = 0 \text{ is uniformly asymptotically stable if and only if} \]

  \[ \|\Phi(t, t_0)\| \leq ke^{-\lambda(t-t_0)} \text{ for some } k > 0, \lambda > 0. \]

- Last lecture: \( V(x) = x^T P(t)x \) proves uniform exp. stability if

  \[\begin{align*}
  (i) & \quad P(t) + A^T(t)P(t) + P(t)A(t) = -Q(t) \\
  (ii) & \quad 0 < k_1 I \leq P(t) \leq k_2 I \\
  (iii) & \quad 0 < k_3 I \leq Q(t) \text{ for all } t.
  \end{align*}\]

  The converse is also true:

  \[ \text{Theorem:} \sup \text{p} x = 0 \text{ is uniformly exponentially stable, } A(t) \text{ is continuous and bounded, } Q(t) \text{ is continuous and symmetric, and there exist } k_3, k_4 > 0 \text{ such that} \]

  \[ 0 < k_3 I \leq Q(t) \leq k_4 I \text{ for all } t. \]

  Then, there exists a symmetric \( P(t) \) satisfying (i)–(ii) above.

  \[ \text{Proof:} \]

  time-invariant: \( P = \int_0^\infty e^{A^T \tau} Q e^{A\tau} d\tau \)

  time-varying: \( P(t) = \int_t^\infty \Phi^T(\tau, t)Q(\tau)\Phi(\tau, t)d\tau \)
From the Leibniz rule:
\[
\dot{P}(t) = \int_t^\infty \left( \frac{\partial}{\partial t} \Phi^T(\tau,t)Q(\tau)\Phi(\tau,t) + \Phi^T(\tau,t)Q(\tau) \frac{\partial}{\partial t} \Phi(\tau,t) \right) d\tau \\
- \Phi^T(t,t)Q(t)\Phi(t,t) \\
= -P(t)A(t) - A^T(t)P(t) - Q(t).
\]

• Suppose we find \(P(t)\) satisfying (i), (ii) above but, instead of (iii),
\[ Q(t) = C^T(t)C(t) \geq 0 \]
where \(C(t)\) is a fat matrix. Can we conclude asymptotic stability?

**Definition:** The pair \((A(t),C(t))\) is uniformly observable if there exist \(\alpha > 0\) and \(\delta > 0\) (both independent of \(t_0\)) such that, for all \(t_0\), the observability Gramian
\[
W(t_0, t_0 + \delta) = \int_{t_0}^{t_0+\delta} \Phi^T(\tau, t_0)C^T(\tau)C(\tau)\Phi(\tau, t_0) d\tau \geq \alpha I.
\]

**Theorem:** If \((A(t), C(t))\) is uniformly observable, then \(x = 0\) is uniformly asymptotically stable.

**Proof:**
\[ V(t, x) = -x(t)^T C^T(t) C(t) x(t) \]
Substitute \(x(t) = \Phi(t, t_0)x(t_0)\) in this equation and integrate both sides from \(t_0\) to \(t_0 + \delta\):
\[
V(t_0 + \delta, x(t_0 + \delta)) - V(t_0, x(t_0)) = -x^T(t_0)W(t_0, t_0 + \delta)x(t_0) \\
\leq -\alpha |x(t_0)|^2 \leq -\frac{\alpha}{k^2} V(t_0, x(t_0))
\]

\(V(t, x(t))\) is nonincreasing because \(V \leq 0\), and \(V(t, x) \to 0\) because of exponential decay of samples at \(t_0 + k\delta, k = 1, 2, 3, \ldots\)
Example: Gradient Algorithm for Parameter Estimation

\[ y(t) = \phi^T(t)\theta^* \]  

(Parametric Model)

\( y(t) \): observed output, \( \theta^* \): vector of unknown parameters, 
\( \phi(t) \) is called the "regressor."

Let \( \theta(t) \) be the estimate for \( \theta^* \) at time \( t \), and define the estimation error:

\[ e(t) \triangleq y(t) - \phi^T(t)\theta(t) \]

Gradient Law: update \( \theta(t) \) by

\[ \dot{\theta}(t) = \phi(t)e(t) \]

Does \( \theta(t) \rightarrow \theta^* \) as \( t \rightarrow \infty \)?

Let \( \tilde{\theta}(t) \equiv \theta(t) - \theta^* \) and note that \( e(t) = \phi^T(t)\theta^* - \phi^T(t)\theta = -\phi^T(t)\tilde{\theta} \).

\[ \dot{\tilde{\theta}} = \dot{\theta} = \phi(t)e(t) = -\phi(t)\phi^T(t)\tilde{\theta} \]

\[ P = \frac{1}{2}I \Rightarrow A(t)^TP + PA(t) = -\phi(t)\phi^T(t) \]

Applying the theorem with \( C(t) = \phi^T(t) \), we conclude \( \theta(t) \rightarrow \theta^* \) if \( (A(t), C(t)) \) is uniformly observable.

This observability condition is difficult to check because the Gramian \( W(t_0, t_0 + \delta) \) depends on the unknown state transition matrix. The following proposition helps us circumvent this difficulty:

Proposition: If \( (A(t), C(t)) \) is uniformly observable, then so is \( (A(t) + K(t)C(t), C(t)) \) for any bounded \( K(t) \).

In our problem, \( A(t) = -\phi(t)\phi^T(t) \) and \( C(t) = \phi^T(t) \). If we choose \( K(t) = \phi(t) \), then \( A(t) + K(t)C(t) = 0 \), which means \( \Phi(\tau, t_0) = I \).

Uniform observability condition:

\[ W(t_0, t_0 + \delta) = \int_{t_0}^{t_0+\delta} \phi(\tau)\phi^T(\tau)d\tau \geq \alpha I \]  

(2)

\( \phi(t) \) is called “persistently exciting” (PE) if it satisfies (2) for some \( \alpha > 0 \) and \( \delta > 0 \) that are independent of \( t_0 \).

Examples:

- \( \phi(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \) is not PE: \( \phi\phi^T \) is constant and \( W(t_0, t_0 + \delta) = \delta\phi\phi^T \) is singular no matter how we choose \( \delta \).
• $\phi(t) = \begin{bmatrix} 1 \\ \sin t \end{bmatrix}$ is PE because $\phi(t)\phi^T(t) = \begin{bmatrix} 1 & \sin t \\ \sin t & \sin^2 t \end{bmatrix}$ and, if we choose $\delta = 2\pi$, then:

$$
\int_{t_0}^{t_0+2\pi} \phi(\tau)\phi^T(\tau)d\tau = \begin{bmatrix} 2\pi & 0 \\ 0 & \pi \end{bmatrix} \geq \pi I.
$$

• $\phi(t) = e^{-t}$ is not PE: $\int_{t_0}^{t_0+\delta} e^{-2\tau}d\tau > 0$ but no lower bound that is independent of $t_0$. 

Review of convergence results for time-varying systems

1. \( \dot{x} = f(t,x) \) \( f(t,0) \equiv 0 \), \( f(t,x) \) Lipschitz in \( x \), bounded in \( t \)
   \( W_1(x) \leq V(t,x) \leq W_2(x) \) \( W_1,W_2 \) pos. definite and radially unbdd
   \( \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(t,x) \leq -W_3(x) \) \( W_3 \) positive semidefinite and \( C^1 \)
   \( \Rightarrow \) uniform stability and \( W_3(x(t)) \to 0. \)

2. Sharper result for the linear time-varying case \( \dot{x} = A(t)x \): If
   \( \dot{P}(t) + P(t)A(t) + A^T(t)P(t) = -C^T(t)C(t) \) \( k_1 I \leq P(t) \leq k_2 I \)
   and \( (A(t),C(t)) \) uniformly observable \( \Rightarrow \) unif. asymptotic stability.

   \[ \int_{t_0}^{t_0+\delta} \Phi^T(t,t_0)C^T(t)C(t)\Phi(t,t_0)dt \geq \alpha I \]
   for some \( \delta, \alpha > 0 \) independent of \( t_0. \)

Examples:

1. Parameter estimation: \( \dot{\theta} = -\phi(t)\phi^T(t) \dot{\theta} \)
   \( =A(t) \)
   \( P = \frac{1}{2} I \) \( \Rightarrow \) \( A^T P + PA = -\phi(t)\phi^T(t) \)

   \( A(t) \to A(t) + \phi(t)C(t) \equiv 0 \) leaves unif. observability unchanged.
   Uniform observability of \( (0,C(t)) = (0,\phi^T(t)) \):

   \[ \int_{t_0}^{t_0+\delta} \phi(t)\phi^T(t)dt \geq \alpha I. \]
   If this property holds, \( \phi(t) \) is called persistently exciting (PE).

2. An example from Lecture 11 generalized:
   \[
   \begin{align*}
   \dot{x}_1 &= -ax_1 + w^T(t)x_2 & a > 0 \\
   \dot{x}_2 &= -w(t)x_1 & x_1 \in \mathbb{R}, x_2 \in \mathbb{R}^{n-1}
   \end{align*}
   \]
   (\( n = 2 \) in Lecture 11)

   \[
   A(t) = \begin{bmatrix}
   -a & w^T(t) \\
   -w(t) & 0
   \end{bmatrix}
   \]
\[ P = \frac{1}{2a} I \Rightarrow PA(t) + A^T(t)P = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} = -C^T C \]

where \( C = \begin{bmatrix} 1 & 0 \end{bmatrix} \)

**Claim:** \((A(t), C)\) is uniformly observable if \(w(t)\) is PE.

Thus, if \(w(t)\) is PE, then \(x = 0\) is uniformly asymptotically stable.

**Proof of the claim:**

Recall that \( A(t) \rightarrow A(t) + K(t)C(t) \) does not change uniform observability. The choice \( K(t) = \begin{bmatrix} a \\ w(t) \end{bmatrix} \) yields

\[
A(t) + K(t)C = \begin{bmatrix} 0 & w^T(t) \\ 0 & 0 \end{bmatrix}
\]

whose sparsity simplifies the observability analysis:

\[
\begin{align*}
\dot{x}_1 &= w^T(t)x_2 \\
\dot{x}_2 &= 0 \\
y &= x_1.
\end{align*}
\]

Adding an integrator at the output does not change observability, therefore the system above is unif. observable if the following is:

\[
\begin{align*}
\dot{x}_2 &= 0 \\
y_0 &= w^T(t)x_2 \\
A_0(t) &\equiv 0 \\
C_0(t) &= w^T(t)
\end{align*}
\]

\((A_0(t), C_0)\) is uniformly observable if \(w(t)\) is PE.

**Lyapunov-based Feedback Design Examples**

**Model Reference Adaptive Control**

Illustrated on a first order system:

\[ \dot{y} = a^* y + u \]  \hspace{1cm} (2)

where \(a^*\) is unknown.

Reference model:

\[ \dot{y}_m = -ay_m + r(t) \quad a > 0, \ r(t) : \text{reference signal.} \]  \hspace{1cm} (3)

**Goal:** Design a controller that guarantees \( y(t) - y_m(t) \rightarrow 0 \) without the knowledge of \(a^*\).
If we knew $a^*$, we would choose:

$$u = -(a^* + a)y + r(t) \quad \triangleq k,$$

$$\Rightarrow \quad \dot{y} = -ay + r(t).$$

The tracking error $\dot{e}(t) \triangleq y(t) - y_m(t)$ satisfies:

$$\dot{e} = -ae \Rightarrow e(t) \to 0 \text{ exponentially.}$$

**Adaptive design** when $a^*$ (therefore, $k^*$) is unknown:

$$u = -ky + r(t)$$

$$k = \text{to be designed.}$$

Then:  

$$\dot{e} = -ae - (k - k^*)y.$$ 

Use the Lyapunov function: $V = \frac{1}{2}e^2 + \frac{1}{2}k^2$:

$$\dot{V} = -ae^2 - k\dot{y} + \dot{k}k$$

$$= -ae^2 + \ddot{k}(k - ey).$$

Choose $\dot{k} = ey$ so that $\dot{V} = -ae^2$.

If $r(t)$ is bounded, then $e(t) \to 0$ from Result 1 (page 1).

Whether $\ddot{k}(t) \to 0$ ($k(t) \to k^*$) depends on the reference signal $r(t)$:

$$\dot{e} = -ae - \ddot{k}y(t)$$

$$\dot{k} = ey(t).$$

From Example 2 above, $\ddot{k}(t) \to 0$ if $y(t)$ is persistently exciting.

**Homework:** Extend the adaptive design above to the system:

$$\dot{y} = a^*y + b^*u, \quad a^*, b^* \text{unknown,}$$

but $b^* \neq 0$ and its sign is known.

**Backstepping**

**Feedback stabilization:** Given the system

$$\dot{x} = f(x) + g(x)u \quad (4)$$

with input $u$, design a control law $u = a(x)$ such that $x = 0$ is asymptotically stable for the closed-loop system:

$$\dot{x} = f(x) + g(x)a(x).$$
Backstepping is a technique that simplifies this task for a class of systems.

Suppose a stabilizing feedback \( u = \alpha(X) \) is available for:

\[
\dot{X} = F(X) + G(X)u \quad X \in \mathbb{R}^n, u \in \mathbb{R}
\]

and suppose the closed-loop system admits a Lyapunov function \( V(X) \) such that

\[
\frac{\partial V}{\partial X} \left( F(X) + G(X)\alpha(X) \right) \leq -W(X) < 0 \quad \forall X \neq 0.
\]

Can we modify \( \alpha(X) \) to stabilize the augmented system below?

\[
\dot{X} = F(X) + G(X)x \\
\dot{x} = u.
\]

Define the error variable \( z = x - \alpha(X) \) and change variables: \((X, x) \to (X, z)\):

\[
\dot{X} = F(X) + G(X)\alpha(X) + G(X)z \\
\dot{z} = u - \dot{\alpha}(X, z)
\]

where \( \dot{\alpha}(X, z) = \frac{\partial \alpha}{\partial X} \left( F(X) + G(X)\alpha(X) + G(X)z \right) \). Take the new Lyapunov function:

\[
V_+(X, z) = V(X) + \frac{1}{2}z^2.
\]

\[
\dot{V}_+ = \frac{\partial V}{\partial X} \left( F(X) + G(X)\alpha(X) \right) + \frac{\partial V}{\partial X} G(X)z + z(u - \dot{\alpha})
\leq -W(X) + z \left( u - \dot{\alpha} + \frac{\partial V}{\partial X} G(X) \right)
\]

Let: \( u = \dot{\alpha} - \frac{\partial V}{\partial X} G(X) - kz, \quad k > 0 \).

Then, \( \dot{V}_+ \leq -W(X) - k z^2 \Rightarrow (X, z) = 0 \) is asymptotically stable.
Backstepping (continued)

Last time we discussed backstepping over a pure integrator. The main idea generalizes trivially to:

\[
\begin{align*}
X &= F(X) + G(X)x \\
\dot{x} &= f(X, x) + g(X, x)u
\end{align*}
\]

where \( X \in \mathbb{R}^n, x \in \mathbb{R} \), and \( g(X, x) \neq 0 \) for all \( (X, x) \in \mathbb{R}^{n+1} \).

With the preliminary feedback

\[
u = \frac{1}{g(X, x)}(-f(X, x) + v)
\]

the \( x \)-subsystem becomes a pure integrator: \( \dot{x} = v \). Substituting the backstepping control law from last time:

\[
v = \dot{\alpha} - \frac{\partial V}{\partial X} G(X) - k z, \quad z \triangleq x - \alpha(X), \quad k > 0
\]

into (1), we get:

\[
u = \frac{1}{g(X, x)} \left( -f(X, x) + \dot{\alpha} - \frac{\partial V}{\partial X} G(X) - k z \right).
\]

Example 1:

\[
\begin{align*}
\dot{x}_1 &= x_1^2 + x_2 \\
\dot{x}_2 &= u.
\end{align*}
\]

Treat \( x_2 \) as “virtual” control input for the \( x_1 \)-subsystem:

\[
\begin{align*}
\alpha(x_1) &= -k_1 x_1 - x_1^2, \quad k_1 > 0 \\
V_1(x_1) &= \frac{1}{2} x_1^2.
\end{align*}
\]

Apply backstepping:

\[
\begin{align*}
z_2 &= x_2 - \alpha(x_1) = x_2 + k_1 x_1 + x_1^2 \\
\dot{z}_2 &= u - \dot{\alpha} \\
u &= \dot{\alpha} - \frac{\partial V_1}{\partial x_1} z_2, \quad k_2 > 0 \\
&= -(k_1 + 2x_1)(x_1^2 + x_2) - \frac{x_1}{x_1} - k_2(x_2 + k_1 x_1 + x_1^2).
\end{align*}
\]
• Backstepping can be applied recursively to systems of the form:

\[ \begin{align*}
\dot{x}_1 &= f_1(x_1) + g_1(x_1)x_2 \\
\dot{x}_2 &= f_2(x_1, x_2) + g_2(x_1, x_2)x_3 \\
\dot{x}_3 &= f_3(x_1, x_2, x_3) + g_3(x_1, x_2, x_3)x_4 \\
&\vdots \\
\dot{x}_n &= f_n(x) + g_n(x)u
\end{align*} \]  

Systems of this form are called “strict feedback systems.”

\[ \begin{bmatrix}
\dot{x}_1 \\
\dot{x}_2 \\
\vdots \\
\dot{x}_n
\end{bmatrix} = \begin{bmatrix}
f_1(x_1) + g_1(x_1)x_2 \\
f_2(x_1, x_2) + g_2(x_1, x_2)x_3 \\
\vdots \\
f_n(x) + g_n(x)u
\end{bmatrix} 
\]

where \( g_i(x_1, \ldots, x_i) \neq 0 \) for all \( x \in \mathbb{R}^n, i = 1, 2, \ldots, n \).

**Example 2:**
\[ \begin{align*}
\dot{x}_1 &= (x_1x_2 - 1)x_1^3 + (x_1x_2 + x_3^2 - 1)x_1 \\
\dot{x}_2 &= x_3 \\
\dot{x}_3 &= u.
\end{align*} \]

Not in strict feedback form because \( x_3 \) appears too soon. In fact, this system is not globally stabilizable because the set \( x_1x_2 \geq 2 \) is positively invariant regardless of \( u \):

To see this, note that
\[ n(x) \cdot f(x, u) = [(x_1x_2 - 1)x_1^3 + (x_1x_2 + x_3^2 - 1)x_1]x_2 + x_3x_1 \]
and substitute \( x_1x_2 = 2 \):

\[ \begin{align*}
&= \left( x_1^3 + (1 + x_3^2)x_1 \right)x_2 + x_3x_1 \\
&= \left( x_1^2 + (1 + x_3^2) \right)x_1x_2 + x_3x_1 \\
&= 2x_1^2 + 2(1 + x_3^2) + x_3x_1 \\
&\geq 0
\end{align*} \]

• The condition \( g_i(x_1, \ldots, x_i) \neq 0 \) in (3) can be relaxed in some cases:

**Example 3:**
\[ \begin{align*}
\dot{x}_1 &= x_1^2x_2 \\
\dot{x}_2 &= u
\end{align*} \]
Treat $x_2$ as virtual control and let $\alpha_1(x_1) = -x_1$ which stabilizes the $x_1$-subsystem, as verified with Lyapunov function $V_1(x_1) = \frac{1}{2}x_1^2$.

Then $z_2 \triangleq x_2 - \alpha_1(x_1)$ satisfies $\dot{z}_2 = u - \dot{\alpha}_1$, and

$$u = \dot{\alpha}_1 - \frac{\partial V_1}{\partial x_1}x_1^2 - k_2 z_2 = -x_1^2 x_2 - x_1^3 - k_2(x_2 + x_1)$$

achieves global asymptotic stability:

$$V = \frac{1}{2}x_1^2 + \frac{1}{2}z_2^2 \Rightarrow \dot{V} = -x_1^4 - k_2 z_2^2.$$ 

Note that we can’t conclude exponential stability due to the quartic term $x_1^4$ above (recall the Lyapunov sufficient condition for exponential stability in Lecture 11, p.2). In fact, the linearization of the closed-loop system proves the lack of exponential stability:

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -k_2 \end{bmatrix} \rightarrow \lambda_{1,2} = 0, -k_2. $$

**Design example:** Active suspension

Define state variables: $x_1 = x_s, x_2 = \dot{x}_s, x_3 = x_a, x_4 = Q$:

$$\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -\frac{k_a}{M_b} (x_1 - x_3) - \frac{c_a}{M_b} (x_2 - \frac{1}{A} x_4) \\
\dot{x}_3 &= \frac{1}{A} x_4 \\
\dot{x}_4 &= -c_f x_4 + k_f u.
\end{align*}$$
This system is not in strict feedback form due to the \( x_4 \) term in \( \dot{x}_2 \). To overcome this problem define:

\[
\bar{x}_3 \triangleq \frac{k_a}{M_b} x_3 + \frac{c_a}{M_b A} x_4 \\
\bar{\xi} \triangleq x_3
\]

and change variables to \((x_1, x_2, \bar{x}_3, \bar{\xi})\):

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= -\frac{k_a}{M_b} x_1 - \frac{c_a}{M_b} x_2 + \bar{x}_3 \\
\dot{\bar{x}}_3 &= \frac{k_a - c_d c_f}{M_b A} x_4 + \frac{c_d k_f}{M_b A} \bar{u}.
\end{align*}
\]

Two steps of backstepping starting with the virtual control law:

\[
a_1(x_1) = -c_1 x_1 - k_1 x_1^3
\]

will stabilize the \((x_1, x_2, \bar{x}_3)\) subsystem. Full \((x_1, x_2, \bar{x}_3, \bar{\xi})\) system:

\[
\begin{array}{ccc}
(x_1, x_2, \bar{x}_3) & \xrightarrow{\bar{x}_3} & \bar{\xi} \\
\text{subsystem} & & \dot{\bar{\xi}} = -\frac{k_f}{M_b A} \bar{\xi} + \frac{1}{A} \bar{x}_3
\end{array}
\]

The \( \bar{\xi} \)-subsystem is an asymptotically stable linear system driven by \( \bar{x}_3 \); therefore the full system is stabilized.
Input-to-State Stability

\[ \dot{x} = f(x, u) \]  \( u \): exogenous input

For linear systems, asymptotic stability of the zero-input model \( \dot{x} = Ax \) implies BIBO\(^2\) stability for \( \dot{x} = Ax + Bu \):

\[
\begin{align*}
x(t) &= e^{At}x_0 + \int_0^t e^{A(t-\tau)}Bu(\tau)d\tau \\
\implies |x(t)| &\leq \|e^{At}\||x_0| + \int_0^t \|e^{A(t-\tau)}\||\|B|||u(\tau)|d\tau \\
&\leq \kappa e^{-\alpha t}|x_0| + \|B\| \sup_{0 \leq \tau \leq t} |u(\tau)| \int_0^t \kappa e^{-\alpha \tau}d\tau \\
&\leq \underbrace{\kappa e^{-\alpha t}|x_0|}_{\text{effect of initial condition}} + \underbrace{\frac{\kappa \|B\| \sup_{0 \leq \tau \leq t} |u(\tau)|}{\alpha}}_{\text{effect of input}}.
\end{align*}
\]

For nonlinear systems, asymptotic stability of the origin for \( \dot{x} = f(x, 0) \) does not imply BIBO stability for \( \dot{x} = f(x, u) \):

**Example 1:** \( \dot{x} = -x + xu \)

\( u(t) \equiv \text{constant} > 1 \implies \text{exponential growth of } x(t). \)

A precise formulation of BIBO stability for nonlinear systems:

**Definition:** The system \( \dot{x} = f(x, u) \), \( f(0, 0) = 0 \) is said to be input-to-state stable (ISS) if:

\[
|x(t)| \leq \beta(|x(0)|, t) + \gamma \left( \sup_{0 \leq \tau \leq t} |u(\tau)| \right)
\]

for some class-\( \mathcal{K}_L \) function \( \beta \) and class-\( \mathcal{K} \) function \( \gamma \), called an ISS gain function.

**Example:** For the linear system above, \( \gamma(s) = \frac{s}{\alpha} \|B\|s. \)

Implications of ISS

1. \( \dot{x} = f(x, u) \) ISS \( \implies \dot{x} = f(x, 0) \) globally asymptotically stable

   **Proof:**
   Substitute \( u(t) \equiv 0 \) in the definition above: \( |x(t)| \leq \beta(|x(0)|, t). \)
2. \( u(t) \to 0 \) as \( t \to \infty \) \( \Rightarrow \) \( x(t) \to 0 \) as \( t \to \infty \).

**Proof:**

Need to show that for any \( \epsilon > 0 \), there exists \( T \) such that

\[ |x(t)| \leq \epsilon \quad \forall t \geq T. \]

Since \( u(t) \to 0 \), we can find \( T_1 \) such that \( \gamma(|u(t)|) \leq \epsilon/2 \) for all \( t \geq T_1 \). Choose \( t_0 = T_1 \) and apply ISS definition:

\[ |x(t)| \leq \beta(|x(T_1)|, t - T_1) + \epsilon/2 \quad \forall t \geq T_1. \]

Choose \( T_2 \) such that

\[ \beta(|x(T_1)|, T_2) \leq \epsilon/2. \]

Then, \( |x(t)| \leq \epsilon \) for all \( t \geq T_1 + T_2 \triangleq T \).

**A Lyapunov Characterization of ISS**

The system \( \dot{x} = f(x, u) \) is ISS if there exist class-\( \mathcal{K}_\infty \) functions \( \alpha_i \), \( i = 1, 2, 3, 4 \), and a \( C^1 \) function \( V \) such that

\[
\begin{align*}
\alpha_1(|x|) &\leq V(x) \leq \alpha_2(|x|) \\
\frac{\partial V}{\partial x} f(x, u) &\leq -\alpha_3(|x|) + \alpha_4(|u|).
\end{align*}
\]

\( V \) is called an “ISS Lyapunov function.”

**Sketch of the proof:**

Let \( \bar{u} \triangleq \sup_{\tau \geq 0} |u(\tau)|. \) Then:

\[
|x| \geq r \triangleq \alpha_3^{-1}(\alpha_4(\bar{u})) \quad \Rightarrow \quad \frac{\partial V}{\partial x} f(x, u(t)) \leq 0 \quad \forall t \geq 0.
\]

This implies that the level set \( \{ x : V(x) \leq \alpha_2(r) \} \) is invariant and attractive. Thus, all trajectories converge to this level set which is enclosed in the outer ball \( |x| \leq R \triangleq \alpha_1^{-1}(\alpha_2(r)) \).

![Lyapunov Diagram](image-url)
Example 2: $\dot{x} = -x^r + x^s u$, $r$: odd integer, is ISS if $r > s$. Take:

$$V(x) = \frac{1}{2} x^2$$
$$\dot{V}(x) = -x^{r+1} + x^{s+1} u.$$  

Young’s inequality (recall from homework):

$$yz \leq \frac{\lambda^p}{p} |y|^p + \frac{1}{q \lambda^q} |z|^q$$

for any $\lambda > 0$, and $p > 1, q > 1$ satisfying $(p - 1)(q - 1) = 1$. Apply to:

$$x^{s+1} u \leq \frac{\lambda^p}{p} |x|^{(s+1)p} + \frac{1}{q \lambda^q} |u|^q$$

and choose

$$p = \frac{r+1}{s+1} \quad \text{and} \quad \lambda \text{ such that } \frac{\lambda^p}{p} = \frac{1}{2}$$

$$\Rightarrow \quad x^{s+1} u \leq \frac{1}{2} |x|^{r+1} + \frac{1}{q \lambda^q} |u|^q$$
$$\Rightarrow \quad \dot{V}(x) \leq -|x|^{r+1} + \frac{1}{2} |x|^{r+1} + \frac{1}{q \lambda^q} |u|^q$$
$$\leq -\alpha_3(|x|) + \frac{1}{q \lambda^q} |u|^q.$$

Note:

• $\dot{x} = -x + xu$ ($r = s = 1$) is not ISS as shown in Example 1.
• $\dot{x} = -x + x^2 u$ ($r = 1, s = 2$) is not ISS: it exhibits finite time escape for $u(t) \equiv \text{constant} \neq 0$, even with an exponentially decaying $u(t)$.
• $\dot{x} = -x^3 + u$ ($r = 3, s = 0$) is ISS.

Example 3:

$$\begin{align*}
\dot{x}_1 &= -x_1 + x_2^2 \\
\dot{x}_2 &= -x_2 + u.
\end{align*}$$

Let $V(x) = \frac{1}{2} x_1^2 + \frac{a}{4} x_2^4$, $a > 0$ to be determined.\(^3\)

$$V(x) = -x_1^2 + x_1 x_2^2 + a(-x_2^4 + x_2^3 u)$$

Apply the Young Inequalities:

$$\begin{align*}
x_1 x_2^2 &\leq \frac{1}{2} x_1^2 + \frac{1}{2} x_2^4 \\
x_2^3 u &\leq \frac{\lambda^{4/3}}{4 \lambda^{1/3}} x_2^4 + \frac{1}{4 \lambda^4} u^4.
\end{align*}$$
Choose \( \lambda \) such that \( \frac{4/3}{4/3} = \frac{1}{2} \).

\[
V(x) \leq -\frac{1}{2} x_1^2 + \frac{1}{2} x_4^4 + a \left( -\frac{1}{2} x_2^4 + \frac{1}{4\lambda^4} u^4 \right)
\]

Let \( a = 2 \):

\[
\dot{V}(x) \leq -\frac{1}{2} x_1^2 - \frac{1}{2} x_4^4 + \frac{1}{2\lambda^4} u^4 \leq \alpha_3(|x|) = \alpha_4(|u|)
\]

for an appropriate choice of \( \alpha_3 \). Thus, the system is ISS.

**Stability of Series Interconnections**

\[
\begin{align*}
\dot{x}_1 &= f_1(x_1, x_2) & x_1 &\in \mathbb{R}^{n_1} \\
\dot{x}_2 &= f_2(x_2) & x_2 &\in \mathbb{R}^{n_2}
\end{align*}
\]

Suppose \( x_2 = 0 \) is globally asymptotically stable for \( \dot{x}_2 = f_2(x_2) \) and \( x_1 = 0 \) is globally asymptotically stable for \( \dot{x}_1 = f_1(x_1, 0) \). Is \( (x_1, x_2) = 0 \) globally asymptotically stable for the interconnection?

**Answer:** No.

**Example 4:**

\[
\begin{align*}
\dot{x}_1 &= -x_1 + x_1^2 x_2 \\
\dot{x}_2 &= -x_2
\end{align*}
\]

exhibits finite time escape.

**Proposition:** Consider the series interconnection:

\[
\begin{align*}
\dot{x}_1 &= f_1(x_1, x_2) \\
\dot{x}_2 &= f_2(x_2, u).
\end{align*}
\]

If the \( x_1 \) subsystem is ISS with \( x_2 \) viewed as an input, and the \( x_2 \) subsystem is ISS with input \( u \), then the interconnection is ISS.

**Example 3 revisited:**

\[
\begin{align*}
\dot{x}_1 &= -x_1 + x_2^2 & \text{is ISS with respect to } x_2 \\
\dot{x}_2 &= -x_2 + u & \text{is ISS with input } u
\end{align*}
\]

\( \Rightarrow \) the interconnection is ISS — an alternative to the proof in Ex. 3.

**Corollary:** \( (x_1, x_2) = 0 \) is globally asymptotically stable when \( u \equiv 0 \).

Note that Example 4 fails the ISS condition for the \( x_1 \) subsystem.

GAS \( \Rightarrow \) ISS \( \equiv \) GAS
Example: Active suspension design example in Lecture 14:

\[
\begin{align*}
(x_1, x_2, \dot{x}_3) & \quad \text{subsystem} \\
\dot{x}_3 & \quad \xi = -\frac{k_4}{M_b A} \xi + \frac{1}{A} \dot{x}_3
\end{align*}
\]

The \((x_1, x_2, \dot{x}_3)\)-subsystem globally asymptotically stabilized by backstepping. The \(\xi\)-subsystem is an asymptotically stable linear system, therefore ISS with respect to the input \(\dot{x}_3\).
Reachable Sets and Safety Certification

Reachable sets with unit peak inputs

\[ R_T \triangleq \{ x(T) \mid \dot{x} = f(x,u), \ x(0) = 0, \ |u| \leq 1 \} \]  

(1)

The set of points that can be reached from \( x(0) = 0 \) with inputs not exceeding unit magnitude. Difficult to find exactly, but methods exist to find overapproximations.

ISS gives a very conservative bound:

\[ |x(T)| \leq \beta(|x(0)|,T) + \gamma \left( \sup_{0 \leq t \leq T} |u(t)| \right) \leq \gamma(1). \]

A less conservative estimate with level sets:

Find positive definite \( V(\cdot) \) and a constant \( c > 0 \) such that

\[ |u| \leq 1 \quad \text{and} \quad V(x) \geq c \quad \Rightarrow \quad \nabla V(x) \cdot f(x,u) \leq 0. \]

Then, the level set \( \Omega_c \triangleq \{ x : V(x) \leq c \} \) contains the reachable set:

\[ R_T \subset \Omega_c \quad \forall T \geq 0. \]

Example: Linear system \( \dot{x} = Ax + Bu \). Use \( V(x) = x^T P x \). If there exists \( P = P^T > 0 \) such that

\[ u^T u \leq 1 \quad \text{and} \quad x^T P x \geq 1 \quad \Rightarrow \quad x^T (A^T P + PA)x + x^T PBu + u^T B^T P x \leq 0 \]

then the ellipsoid \( \{ x : x^T P x \leq 1 \} \) is an overapproximation of \( R_T \).

Rewrite the above implication as:

\[ \left\{ \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} + 1 \geq 0 \right\} \land \left\{ \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} - 1 \geq 0 \right\} \]

\[ \Rightarrow \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} A^T P + PA & PB \\ B^T P & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \leq 0. \]

Note that this statement is verified if we can find \( \alpha \geq 0, \beta \geq 0 \) such
that:

\[
\begin{bmatrix}
  x \\
  u
\end{bmatrix}^T
\begin{bmatrix}
  A^TP + PA & PB \\
  B^TP & 0
\end{bmatrix}
\begin{bmatrix}
  x \\
  u
\end{bmatrix} + \alpha \left( \begin{bmatrix}
  x \\
  u
\end{bmatrix}^T
\begin{bmatrix}
  P & 0 \\
  0 & 0
\end{bmatrix}
\begin{bmatrix}
  x \\
  u
\end{bmatrix} - 1 \right)
\]
\[+ \beta \left( \begin{bmatrix}
  x \\
  u
\end{bmatrix}^T
\begin{bmatrix}
  0 & 0 \\
  0 & -1
\end{bmatrix}
\begin{bmatrix}
  x \\
  u
\end{bmatrix} + 1 \right) \leq 0
\]

or, equivalently:

\[
\begin{bmatrix}
  x \\
  u \\
  1
\end{bmatrix}^T
\begin{bmatrix}
  A^TP + PA + \alpha P & PB & 0 \\
  B^TP & -\beta I & 0 \\
  0 & 0 & \beta - \alpha
\end{bmatrix}
\begin{bmatrix}
  x \\
  u \\
  1
\end{bmatrix} \leq 0.
\]

Next, using the following property (prove this as an exercise):

\[
\begin{bmatrix}
  y \\
  1
\end{bmatrix}^T M \begin{bmatrix}
  y \\
  1
\end{bmatrix} \geq 0 \text{ for all } y \text{ if and only if } M \geq 0,
\]

we conclude

\[
\begin{bmatrix}
  x \\
  u \\
  1
\end{bmatrix}^T
\begin{bmatrix}
  A^TP + PA + \alpha P & PB \\
  B^TP & -\beta I \\
  0 & 0 & \beta - \alpha
\end{bmatrix}
\begin{bmatrix}
  x \\
  u \\
  1
\end{bmatrix} \leq 0
\]
\[\iff \begin{bmatrix}
  A^TP + PA + \alpha P & PB \\
  B^TP & -\beta I \\
  \beta - \alpha
\end{bmatrix} \leq 0
\]

Choose \( \beta = \alpha \) which is the best choice to ensure the block matrix above is nonpositive:

\[
\begin{bmatrix}
  A^TP + PA + \alpha P & PB \\
  B^TP & -\alpha I
\end{bmatrix} \leq 0.
\]

\[\text{(3)}\]

**Summary: procedure to overapproximate the reachable set**

Look for \( P = P^T > 0 \) and \( \alpha > 0 \) satisfying the matrix inequality (3). This is not a linear matrix inequality (LMI) in \( \alpha \) and \( P \), but it is an LMI in \( P \) if \( \alpha \) is fixed. The resulting ellipsoid \( \{ x : x^TPx \leq 1 \} \) is a superset of \( R_T \).

Additional objectives can be incorporated, such as minimizing the volume of the ellipsoid, which is proportional to \( \sqrt{\det P^{-1}} \):

\[
\text{minimize } \log(\det P^{-1}) \text{ which is convex in } P.
\]
S-procedure

The principle used to obtain (2) is known as the S-procedure in control theory. To show that:

\[ q_0(\xi) \geq 0 \quad \text{whenever} \quad q_i(\xi) \geq 0 \quad i = 1, 2, \ldots, p \]

look for \( \tau_1, \tau_2, \ldots, \tau_p \geq 0 \) such that

\[ q_0(\xi) - \sum_{i=1}^{p} \tau_i q_i(\xi) \geq 0. \]

In (2), \( q_i(\cdot), i = 0, 1, 2, \) are quadratic functions of \( \xi = \begin{bmatrix} x \\ u \end{bmatrix} \).

Reachable sets with unit energy inputs

\[ R_T \triangleq \{ x(T) \mid \dot{x} = f(x, u), \; x(0) = 0, \; \int_0^T u^T(t)u(t)dt \leq 1 \} \tag{4} \]

For an overapproximation, find positive definite \( V(\cdot) \) such that

\[ \nabla V(x) \cdot f(x, u) \leq u^T u. \]

\[ \frac{d}{dt} V(x(t)) \leq u^T u \Rightarrow V(x(T)) - V(x(0)) \leq \int_0^T u^T(t)u(t)dt \leq 1 \]

\[ \Rightarrow V(x(T)) \leq 1. \]

Therefore, \( x \in R_T \) implies \( V(x) \leq 1 \), i.e., the level set contains the reachable set:

\[ R_T \subset \{ x : V(x) \leq 1 \}. \]

Example:

\[ \dot{x} = Ax + Bu \quad V(x) = x^T P x. \]

Find \( P = P^T > 0 \) such that

\[ x^T (A^T P + PA)x + x^T PBu + u^T B^T P x \leq u^T u \]

or, written more compactly:

\[
\begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} A^T P + PA & PB \\ B^T P & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \leq \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}.
\]

This means

\[
\begin{bmatrix} A^T P + PA & PB \\ B^T P & -I \end{bmatrix} \leq 0
\]

which is a LMI in \( P \).
Safety Certification

Given an “unsafe” set $U$, show that

$$RT \cap U = \emptyset.$$ 

The level set overapproximations above can be used to prove safety:

Look for a $V$ with the additional property that $x \in U \Rightarrow V(x) > 1$. Such functions $V$ are sometimes called “barrier functions.”

**Example:** Suppose the unsafe set is the half-space:

$$U = \{ x : a^Tx > 1 \}.$$ 

Let $V(x) = x^TPx$. From the S-procedure, if there exists $\tau > 0$ such that

$$(x^TPx - 1) - \tau(a^Tx - 1) \geq 0,$$ 

then $x \in U \Rightarrow V(x) > 1$.

**Exercise:** Show that (5) is equivalent to: $P \geq aa^T$.

Thus, the LMIs in the previous examples can be augmented with this additional constraint to certify safety.
Sum of Squares Programming

Establishing nonnegativity of functions is critical in nonlinear system analysis, e.g., a Lyapunov function $V$ for $\dot{x} = f(x)$ must satisfy

$$V(x) > 0 \quad \forall x \neq 0 \quad (1)$$

$$-\nabla V(x)^T f(x) \geq 0 \quad \forall x. \quad (2)$$

For $f(x) = Ax$ and $V(x) = x^T P x$, the conditions above are simple matrix inequalities:

$$P > 0, \quad -A^T P - PA \geq 0.$$ 

How can we check nonnegativity when $f$ and $V$ are more general polynomials?

Sum of Squares (SOS) Polynomials

A monomial is a product of powers of variables (e.g., $m(x) = x_1^2 x_2$) and its degree is the sum of its exponents (e.g., 3 for $m(x) = x_1^2 x_2$).

A polynomial is a finite linear combination of monomials and its degree is the maximum degree of these monomials.

Example 1: The polynomial

$$q(x_1, x_2) = x_1^2 - 2x_1 x_2^2 + 2x_1^4 + 2x_1^3 x_2 - x_1^2 x_2^2 + 6x_2^4 \quad (3)$$

has degree 4.

Definition: A polynomial $p$ is a sum of squares (SOS) if there exist polynomials $g_1, \cdots, g_r$ such that

$$p = \sum_{i=1}^{r} g_i^2. \quad (4)$$

A SOS polynomial $p(x)$ is nonnegative for all $x$. The converse is not true: there exist nonnegative polynomials that are not SOS.

The polynomial $q(x_1, x_2)$ in (3) is SOS because it can be rewritten as:

$$(x_1 - x_2^2)^2 + \frac{1}{2} \left( 2x_1^2 + x_1 x_2 - 3x_2^2 \right)^2 + \frac{1}{2} \left( 3x_1 x_2 + x_2^2 \right)^2. \quad (5)$$

You can verify the equivalence of (3) and (5) by multiplying out terms in (5) and matching them to those in (3).

How a SOS decomposition like (5) can be obtained is discussed next.
**SOS Decomposition**

Let \( z(x) \) be the vector of all monomials of degree \( \leq d \) in \( n \) variables:
\[
z(x) \triangleq [1, x_1, x_2, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_n^d]^T.
\]
Then any polynomial with degree \( \leq 2d \) can be rewritten as
\[
p(x) = z(x)^T Q z(x)
\]  \( (6) \)
where \( Q \) is a symmetric matrix.

**Example 2:** Let \( p(x_1, x_2) = 2x_1^2 x_2^2 \) which has degree 4. With \( n = 2 \) and \( d = 2 \),
\[
z(x) = [1, x_1, x_2, x_1^2, x_1 x_2, x_2^2]^T,
\]  \( (7) \)
and (6) holds with either
\[
Q_1 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]
or
\[
Q_2 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

Thus, the choice of \( Q \) is not unique.

**Theorem:** A polynomial \( p \) with degree \( \leq 2d \) is SOS if and only if there exists \( Q = Q^T \geq 0 \) satisfying (6).

**Proof: (only if)** If \( p \) is SOS then, by definition, \( p = \sum_{i=1}^r g_i^2 \) for some polynomials \( g_i, i = 1, \ldots, r \). Write \( g_i \) as:
\[
g_i(x) = C_i z(x)
\]  \( (8) \)
where \( C_i \) is a row vector of coefficients. Then \( g_i^2 = z^T C_i^T C_i z \) and
\[
p = \sum_{i=1}^r g_i^2 = z^T \left( \sum_{i=1}^r C_i^T C_i \right) z.
\]

**(if)** Given \( Q = Q^T \geq 0 \) satisfying (6), decompose \( Q \) as \( Q = C^T C \) where \( C \) has as many rows as the rank of \( Q \), say \( r \). Then,
\[
Q = C^T C = \sum_{i=1}^r C_i^T C_i
\]
where \( C_i \) is the \( i \)th row. If we define \( g_i \) as in (8), then
\[
z^T Q z = \sum_{i=1}^r g_i^2.
\]
Since \( Q \) is not unique, not all \( Q \) satisfying (6) will certify SOS. In Example 2 above, \( Q_1 \geq 0 \) but \( Q_2 \) is indefinite. We need to characterize the set of all \( Q \) satisfying (6) and search for a \( Q \geq 0 \) in this set.
Parameterization of all matrices $Q$ satisfying (6):

Find a particular solution $Q_0$ such that

$$p(x) = z(x)^T Q_0 z(x),$$

and find a basis of symmetric matrices $N_j$, $j = 1, 2, \cdots, K$, such that

$$z(x)^T N_j z(x) = 0 \quad \text{for all } x. \quad (9)$$

Then we can parameterize the set of all $Q$ satisfying (6) as

$$Q = Q_0 + \sum_{j=1}^{K} \lambda_j N_j \quad \lambda_j \in \mathbb{R},$$

and $p$ is SOS if and only if there exist $\lambda_1, \cdots, \lambda_K$ such that $Q \geq 0$.

For $n = d = 2$, $z(x)$ is as defined in (7) and a basis as in (9) is:

$N_1 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$

$N_2 = \begin{bmatrix}
0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$

$N_3 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0
\end{bmatrix}$

$N_4 = \begin{bmatrix}
0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$

$N_5 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0
\end{bmatrix}$

$N_6 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0
\end{bmatrix}$

Example 1 revisited: For $q(x_1, x_2)$ in (3), a suitable choice for $Q_0$ is

$$Q_0 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 1 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 \\
0 & -1 & 0 & 0 & 0 & 6
\end{bmatrix}.$$
Note that $Q_0 \preceq 0$, but $Q_0 + 6N_0 \succeq 0$. Moreover, $Q_0 + 6N_0$ can be decomposed as
\[
\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 2 & 1 & -3 \\
0 & 0 & 0 & 3 & 1
\end{bmatrix}
\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & -1 \\
0 & 1/2 & 0 & 0 & 0 & 2 & 1 & -3 \\
0 & 0 & 1/2 & 0 & 0 & 0 & 3 & 1
\end{bmatrix}
\]
which explains how the SOS form (5) was obtained.

*Synthesizing SOS Polynomials*

With the method above we can numerically check whether a given polynomial function $V$ satisfies (1)-(2). However, in practice, it is more important to be able to search for a $V$ satisfying (1)-(2). This is accomplished by synthesizing $V$ as a weighted sum of basis polynomials with weights left as decision variables.

This leads to the following SOS synthesis problem:

*Given basis polynomials $p_i$, $i = 0, 1, \cdots, m$, each with degree $\leq 2d$, find parameters $a_1, \cdots, a_m$ such that $p_0 + a_1 p_1 + \cdots + a_m p_m$ is SOS.*

To solve this problem, find a matrix $Q_i$ satisfying $p_i = z^T Q_i z$ for each $i = 0, 1, \cdots, m$. Then search for $a_1, \cdots, a_m$ and $\lambda_1, \cdots, \lambda_K$ satisfying

$$Q_0 + \sum_{i=1}^m a_i Q_i + \sum_{j=1}^K \lambda_j N_j \succeq 0. \quad (10)$$

This is a linear matrix inequality (LMI) and can be solved numerically with standard semidefinite program (SDP) solvers.

There are also software packages\(^2\) that follow the procedures above to automatically convert SOS programs to LMIs, such as (10).\(^\text{e.g., SOSOPT}\)
Review of Sum of Squares (SOS) Polynomials

Checking whether a polynomial is SOS: A polynomial $p$ with degree $\leq 2d$ is a sum of squares if and only if there exists $Q = Q^T \succeq 0$ s.t.

$$p(x) = z(x)^T Q z(x)$$

where $z(x)$ is the vector of all monomials of degree $\leq d$:

$$z(x) \triangleq [1, x_1, x_2, \ldots, x_n, x_1^2, x_1 x_2, \ldots, x_n^d]^T.$$

Find a particular solution $Q_0$ such that

$$p(x) = z(x)^T Q_0 z(x),$$

and find a basis of symmetric matrices $N_j, j = 1, 2, \cdots, K,$ such that

$$z(x)^T N_j z(x) = 0 \text{ for all } x.$$

Then $p$ is SOS if and only if there exist reals $\lambda_1, \cdots, \lambda_K$ such that

$$Q = Q_0 + \sum_{j=1}^{K} \lambda_j N_j \succeq 0.$$  

This is a linear matrix inequality (LMI) and can be solved numerically with standard semidefinite program (SDP) solvers.

Synthesizing SOS Polynomials: Given $p_i, i = 0, 1, \cdots, m$, each with degree $\leq 2d$, find reals $a_1, \cdots, a_m$ s.t. $p_0 + a_1 p_1 + \cdots + a_m p_m$ is SOS.

Find a particular $Q_i$ satisfying $p_i = z^T Q_i z$ for each $i = 0, 1, \cdots, m$.

Then search for $a_1, \cdots, a_m$ and $\lambda_1, \cdots, \lambda_K$ satisfying the LMI

$$Q_0 + \sum_{i=1}^{m} a_i Q_i + \sum_{j=1}^{K} \lambda_j N_j \succeq 0.$$  

Applications

Searching for a Lyapunov Function

Given $\dot{x} = f(x), f(0) = 0$, where $f$ is a vector of polynomials, search for a Lyapunov function of the form

$$V(x) = p_0(x) + a_1 p_1(x) + \cdots + a_m p_m(x)$$

(5)
where $p_i, i = 0, 1, \cdots, m$ are basis polynomials selected ahead of time, and $a_i, i = 1, \cdots, m$ are weights to be determined.

To ensure $V$ is positive definite, pick a positive definite polynomial $\ell$ (e.g., $\ell(x) = \varepsilon x^T x$ for some small $\varepsilon$) and impose the constraint:

$$V(x) - \ell(x) \text{ is SOS.} \quad (6)$$

To ensure $\nabla V(x)^T f(x)$ is negative semidef., impose the constraint:

$$-\nabla V(x)^T f(x) \text{ is SOS.} \quad (7)$$

Constraints (6) and (7) can be brought to the LMI form (4) and feasible $a_i, i = 1, \cdots, m$ can be determined numerically (if they exist).

**Overapproximating Reachable Sets**

Recall from Lecture 16 that

$$R_T \triangleq \{ x(T) | \dot{x} = f(x,u), x(0) = 0, \int_0^T u^T(t)u(t)dt \leq 1 \} \quad (8)$$

defines the reachable set from $x(0) = 0$ under unit energy inputs and, if we can find a positive definite $V$ such that

$$\nabla V(x)^T f(x,u) \leq u^T u, \quad (9)$$

then we can overapproximate $R_T$ by:

$$R_T \subset \{ x : V(x) \leq 1 \}.$$  

This follows because, from (9),

$$\frac{d}{dt} V(x(t)) \leq u^T u \Rightarrow V(x(T)) - V(x(0)) \leq \int_0^T u^T(t)u(t)dt \leq 1 \Rightarrow V(x(T)) \leq 1.$$  

If $f(x,u)$ is a vector of polynomials in $x$ and $u$, we can search for a polynomial $V$ of the form (5), and encode (9) with the constraint:

$$-\nabla V(x)^T f(x,u) + u^T u \text{ is SOS in } x \text{ and } u. \quad (10)$$

This can then be combined with (6) and brought to the LMI form (4).

**Certifying Safety**

If unsafe set $U$ does not intersect the overapproximation above, then it can’t intersect the actual reachable set. Thus, we can certify safety by proving the implication:

$$x \in U \Rightarrow V(x) \geq 1 + \varepsilon \quad (11)$$
for some $\varepsilon > 0$.

Suppose the unsafe set can be expressed as

$$U = \{ x : q_i(x) \geq 0, i = 1, \ldots, p \}$$

where $q_i$ are polynomials. Then we can encode (11) with the constraints:

$$V(x) - (1 + \varepsilon) - \sum_{i=1}^{p} s_i(x)q_i(x) \text{ is SOS} \quad (12)$$

$$s_i(x), i = 1, \cdots, p \text{ are SOS.} \quad (13)$$

We can parameterize the search space for $s_i$ as we did for $V$ in (5), and combine (6), (10), (12)-(13) into a LMI.

Above we implicitly used a generalization of the S-procedure from Lecture 16. Specifically, to prove that

$$q_0(x) \geq 0 \quad \text{whenever} \quad q_i(x) \geq 0, i = 1, 2, \ldots, p$$

we look for nonnegative functions $s_1, s_2, \ldots, s_p$ (rather than constants as in Lecture 16) such that

$$q_0(x) - \sum_{i=1}^{p} s_i(x)q_i(x) \geq 0.$$

**Underapproximating the Region of Attraction**

Given system $\dot{x} = f(x)$ with asymptotically stable equilibrium at the origin $x = 0$, the region of attraction, denoted $R_A$, is the set of initial conditions from which the trajectories converge to the origin.

Recall from Lecture 10 that, if $V$ is positive definite and

$$\nabla V(x)^T f(x) < 0 \quad \text{whenever} \quad x \neq 0 \text{ and } V(x) \leq \gamma \quad (14)$$

then $\Omega_\gamma \triangleq \{ x : V(x) \leq \gamma \} \subset R_A$.

Let $\ell$ be a positive definite polynomial. If there exists a SOS polynomial $s$ such that

$$- [\ell(x) + \nabla V(x)^T f(x)] - s(x) [\gamma - V(x)] \text{ is SOS,} \quad (15)$$

then $V(x) \leq \gamma$ implies $\nabla V(x)^T f(x) \leq -\ell(x)$ as stipulated in (14).

To obtain a LMI from (15), one option is to fix the Lyapunov function $^2 V$ and to parameterize the search space for $s$. We can further maximize $\gamma$ subject to (15) by incrementing $\gamma$ until the the resulting LMI is infeasible.

Alternatively $s$ can be fixed and $V$ parameterized. If we parameterize both $s$ and $V$, however, (15) is no longer affine in the parameters because the term $s(x)V(x)$ contains the products of these parameters.

[^2]: choose, e.g., a quadratic Lyapunov function for the linearized model at $x = 0$.\]
Below is a procedure that alternates between first fixing $V$, varying $s$, and next fixing $s$, varying $V$. When a new $V$ is obtained, however, the shape of the level set changes and it may be ambiguous whether the new one is bigger. To remove this ambiguity we define a "shape function" $p$ and use its level sets to judge the size of the region of attraction estimate.

**Step 1:** Let $V_0(x)$ be an initial choice for a Lyapunov function, e.g., a quadratic function for the linearized model at the origin. Find $\gamma^* := \max \gamma$ s.t. $\nabla V_0(x)^T f(x) < 0$ whenever $x \neq 0$ and $V_0(x) \leq \gamma$.

To satisfy the constraint look for a SOS multiplier $s_1(x)$ that satisfies

$$-[\ell(x) + \nabla V_0(x)^T f(x)] - s_1(x)[\gamma - V_0(x)] \text{ is SOS}$$

where $\ell$ is positive definite, e.g., $\ell(x) := \epsilon(x_1^2 + x_2^2)$ for some $\epsilon > 0$.

**Step 2:** Let $p(x)$ be some fixed, positive definite convex polynomial (e.g., $p(x) = x_1^2 + x_2^2$), and let $V_0(x)$ and $\gamma^*$ be as in Step 1. Find $\beta^* := \max \beta$ s.t. $V_0(x) \leq \gamma^*$ whenever $p(x) \leq \beta$.

To satisfy the constraint look for a SOS multiplier $s_2(x)$ such that

$$[\gamma^* - V_0(x)] - s_2(x)[\beta - p(x)] \text{ is SOS}.$$ 

This means that $\{x : p(x) \leq \beta\}$ is contained in $\{x : V_0(x) \leq \gamma^*\}$.

**Step 3:** Given $\gamma^*, s_1(x)$ from Step 1 and $p(x), s_2(x)$ from Step 2, search for $V(x)$ to solve:

$$\max_{\beta > 0} \beta$$

subject to $4$th-order $V(x)$

$$V(x) - \ell(x) \text{ is SOS}$$

$$- [\ell(x) + V(x)^T f(x)] - s_1(x)[\gamma^* - V(x)] \text{ is SOS}$$

$$[\gamma^* - V(x)] - s_2(x)[\beta - p(x)] \text{ is SOS}.$$ 

The first constraint ensures $V$ is positive definite. The second implies that the level set $\{x : V(x) \leq \gamma^*\}$ is invariant, hence a valid approximation for the region of attraction. The third constraint and the maximization of $\beta$ ensure that $V$ is selected such that the level set $\{x : V(x) \leq \gamma^*\}$ is as large as possible, as measured by function $p$.

To proceed, replace $V_0(x)$ in Step 1 with the function $V(x)$ from Step 3, and repeat the steps above for several iterations, until the change in $\beta^*$ in Step 2 is sufficiently small. The final approximation of the ROA is the set where $V(x) \leq \gamma^*$. 
Feedback Linearization

Consider the single-input single-output (SISO) nonlinear system:

\[
\dot{x} = f(x) + g(x)u \\
y = h(x).
\]

(1)

Relative degree (informal definition): Number of times we need to take the time derivative of the output to see the input:

\[
\dot{y} = \frac{\partial h}{\partial x} f(x) + \frac{\partial h}{\partial x} g(x) u \\
\triangleq L_f h(x) \triangleq L_g h(x)
\]

If \(L_g h(x) \neq 0\) in an open set containing the equilibrium, then the relative degree is equal to 1. If \(L_g h(x) \equiv 0\), continue taking Lie derivatives:

\[
\dot{y} = L_f L_f h(x) + L_g L_f h(x) u. \\
\triangleq L^2_f h(x)
\]

If \(L_g L_f h(x) \neq 0\), then relative degree is 2. If \(L_g L_f h(x) \equiv 0\), continue...

Definition: The system (1) has relative degree \(r\) if, in a neighbourhood of the equilibrium,

\[
L_g L_f^{r-1} h(x) = 0 \quad i = 1, 2, \ldots, r - 1 \\
L_g L_f^{r-1} h(x) \neq 0.
\]

(2)

Examples:

1. \(\dot{x}_1 = x_2 \)
   \(\dot{x}_2 = -x_3^3 + u \)
   \(y = x_1\)

   has relative degree = 2.

2. SISO linear system:

   \(\dot{x} = Ax + Bu \quad y = Cx\)
\[ L_g h(x) = CB, \quad L_g L_f h(x) = CAB, \quad \ldots, \quad L_g L_f^{r-1} = CA^{r-1}B. \]
\[ CB \neq 0 \Rightarrow \text{relative degree} = 1 \]
\[ CB = 0, \quad CAB \neq 0 \Rightarrow \text{relative degree} = 2 \]
\[ CB = \cdots = CA^{r-2}B = 0, \quad CA^{r-1}B \neq 0 \Rightarrow \text{relative degree} = r \]

The parameters \( CA^i B \quad i = 1, 2, 3, \ldots \) are called Markov parameters and are invariant under similarity transformations.

3. \[
\begin{align*}
\dot{x}_1 &= x_2 + x_3^3 \\
\dot{x}_2 &= x_3 \\
\dot{x}_3 &= u
\end{align*}
\]
\[ y = x_1, \quad \dot{y} = \dot{x}_1 = x_2 + x_3^3, \quad \dot{y} = x_2 + 3x_3^2x_3 = x_3 + 3x_3^2u \]
\[ L_g L_f h(x) = 3x_3^2 = 0 \quad \text{when} \quad x_3 = 0, \quad \text{and} \quad \neq 0 \quad \text{elsewhere}. \]
Thus, this system does not have a well-defined relative degree around \( x = 0 \).

**Input-Output Linearization**

If a system has a well-defined relative degree then it is input-output linearizable:

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

Apply preliminary feedback:

\[
\boxed{u = \frac{1}{L_g L_f^{r-1} h(x)} \left( -L_f h(x) + v \right)} \tag{4}
\]

where \( v \) is a new input to be designed. Then, \( y^{(r)} = v \) is a linear system in the form of an integrator chain:

\[
\begin{align*}
\dot{\zeta}_1 &= \zeta_2 \\
\dot{\zeta}_2 &= \zeta_3 \\
&\vdots \\
\dot{\zeta}_r &= v
\end{align*}
\]

where \( \zeta_1 \triangleq y = h(x), \quad \zeta_2 \triangleq \dot{y} = L_f h(x), \quad \ldots, \quad \zeta_r \triangleq y^{(r-1)} = L_f^{r-1} h(x). \)
To ensure \( y(t) \to 0 \) as \( t \to \infty \), apply the feedback:

\[
\boxed{v = -k_1 \zeta_1 - k_2 \zeta_2 - \cdots - k_r \zeta_r} \tag{5}
\]

where \( k_1, \ldots, k_r \) are such that \( s^r + k_1 s^{r-1} + \cdots + k_r s + k_1 \) has all roots in the open left half-plane.
Does the controller (4)-(5) achieve asymptotic stability of \( x = 0 \)?
Not necessarily! It renders the \((n - r)\)-dimensional manifold:
\[
h(x) = L_f h(x) = \cdots = L_f^{r-1} h(x) = 0
\]
invariant and attractive. The dynamics restricted to this manifold are called zero dynamics and determine whether or not \( x = 0 \) is stable.
If the origin of the zero dynamics is asymptotically stable, the system is called minimum phase. If unstable, it is called nonminimum phase.

Example: \( n = 3, \ r = 1 \)

Finding the Zero Dynamics

Set \( y = \dot{y} = \cdots = y^{(r-1)} = 0 \) and substitute (4) with \( v = 0 \), that is:
\[
u^* = \frac{-L_f h(x)}{L_f^{r-1} h(x)}.
\]
The remaining dynamical equations describe the zero dynamics.
Example:
\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= \alpha x_3 + u \\
\dot{x}_3 &= \beta x_3 - u \\
y &= x_1
\end{align*}
\]
This system has relative degree 2. With \( x_1 = x_2 = 0 \) and \( u^* = -\alpha x_3 \), the remaining dynamical equation is
\[
\dot{x}_3 = (\alpha + \beta) x_3.
\]
Thus this system is minimum phase if \( \alpha + \beta < 0 \).

For a linear SISO system, relative degree is the difference between the degrees of the denominator and the numerator of the transfer function, and zeros are the roots of the numerator. The definitions of relative degree and zero dynamics above generalize these concepts to nonlinear systems. As an example, the transfer function for (6) is
\[
\frac{s - (\alpha + \beta)}{s^2(s - \beta)},
\]
which has relative degree two and a zero at \( s = \alpha + \beta \) as expected.
Example: Cart/Pole

\[
\dot{y} = \frac{1}{M/m + \sin^2 \theta} \left( \frac{u}{m} + \dot{\theta}^2 \ell \sin \theta - g \sin \theta \cos \theta \right)
\]

\[
\ddot{\theta} = \frac{1}{\ell(M/m + \sin^2 \theta)} \left( - \frac{u}{m} \cos \theta - \dot{\theta}^2 \ell \cos \theta \sin \theta + \frac{M + m}{m} g \sin \theta \right)
\] (7)

Relative degree = 2.

To find the zero dynamics, substitute \( y = \dot{y} = 0 \), and

\[ u^* = -m(\dot{\theta}^2 \ell \sin \theta - g \sin \theta \cos \theta) \]

in the \( \ddot{\theta} \) equation:

\[ \ddot{\theta} = \frac{g}{\ell} \sin \theta. \]

Same as the dynamics of the pole when the cart is held still:

Nonminimum phase because \( \theta = 0 \) is unstable for the zero dynamics.
Feedback Linearization (continued)

Nonlinear Changes of Variables

$T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is called a diffeomorphism if its inverse $T^{-1}$ exists, and both $T$ and $T^{-1}$ are continuously differentiable ($C^1$).

Examples:

1. $\xi = Tx$ is a diffeomorphism if $T$ is a nonsingular matrix
2. $\xi = \sin x$ is a local diffeomorphism around $x = 0$, but not global

\[ \begin{array}{c}
\xi \\
\hline
x
\end{array} \]

3. $\xi = x^3$ is not a diffeomorphism because $T^{-1}(\cdot)$ is not $C^1$ at $\xi = 0$

\[ \begin{array}{c}
\xi \\
\hline
x
\end{array} \]

slope = 0

How to check if $\xi = T(x)$ is a local diffeomorphism?

Implicit Function Theorem

Suppose $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is $C^1$ and there exists $x_0 \in \mathbb{R}^n, \xi_0 \in \mathbb{R}^m$ such that

$f(x_0, \xi_0) = 0$.

If $\frac{\partial f}{\partial x}(x_0, \xi_0)$ is nonsingular, then in a neighborhood of $(x_0, \xi_0)$,

$f(x, \xi) = 0$

has a unique solution $x = g(\xi)$ where $g$ is $C^1$ at $\xi = \xi_0$.

Corollary: Let $f(x, \xi) = T(x) - \xi$. If $\frac{\partial T}{\partial x}$ is nonsingular at $x_0$, then $T(\cdot)$ is a local diffeomorphism around $x_0$. 
A "Normal Form" that Explicitly Displays the Zero Dynamics

**Theorem:** If \( \dot{x} = f(x) + g(x)u, \ y = h(x) \) has a well-defined relative degree \( r \leq n \), then there exist a diffeomorphism \( T : x \rightarrow [z \ \zeta]^T \), \( z \in \mathbb{R}^{n-r}, \ \zeta \in \mathbb{R}^r \), that transforms the system to the form:

\[
\begin{align*}
\dot{z} &= f_0(z, \zeta) \\
\dot{\zeta}_1 &= \zeta_2 \\
& \vdots \\
\dot{\zeta}_r &= b(z, \zeta) + a(z, \zeta)u, \ \ y = \zeta_1.
\end{align*}
\]

In particular, \( \dot{z} = f_0(z, 0) \) represents the zero dynamics. \( \square \)

To obtain this form, let \( \zeta = [h(x) \ \ L_fh(x) \ \ldots \ \ L_f^{r-1}h(x)]^T \), and find \( n-r \) independent variables \( z \) such that \( \dot{z} \) does not contain \( u \).

Note that the terms \( b(z, \zeta) \) and \( a(z, \zeta) \) correspond to \( L_f^r(x) \) and \( L_f^{r-1}h(x) \) in the original coordinates.

**Example:**

\[
\begin{align*}
\dot{x}_1 &= x_2 \\
\dot{x}_2 &= ax_3 + u \\
\dot{x}_3 &= \beta x_3 - u \\
y &= x_1.
\end{align*}
\]

Let \( \zeta_1 = x_1, \ \zeta_2 = x_2 \), and note that \( z = x_2 + x_3 \) is independent of \( \zeta_1, \zeta_2 \), and \( \dot{z} \) does not contain \( u \). Thus, the normal form is:

\[
\begin{align*}
\dot{z} &= (a + \beta)x_3 = (a + \beta)z - (a + \beta)\zeta_2 \\
\dot{\zeta}_1 &= \zeta_2 \\
\dot{\zeta}_2 &= ax_3 + u = az - a\zeta_2 + u.
\end{align*}
\]

**I/O Linearizing Controller** in the new coordinates (1):

1. \( u = \frac{1}{a(z, \zeta)} \left( -b(z, \zeta) + v \right) \) \hspace{1cm} (2)
2. \( v = -k_1\zeta_1 \cdots - k_r\zeta_r \) \hspace{1cm} (3)

where \( k_1, \ldots, k_r \) are such that all roots of \( s^r + k_r s^{r-1} + \cdots + k_2 s + k_1 \) have negative real parts.

**Theorem:** If \( z = 0 \) is locally exponentially stable for the zero dynamics \( \dot{z} = f_0(z, 0) \), then (2)–(3) locally exponentially stabilizes \( x = 0 \).

**Proof:** Closed-loop system:

\[
\begin{align*}
\dot{z} &= f_0(z, \zeta) \\
\dot{\zeta} &= A\zeta
\end{align*}
\]
where

\[
A = \begin{bmatrix}
0 & 1 & 0 & \ldots \\
0 & 0 & 1 & \ldots \\
\vdots & & & \\
-k_1 & -k_2 & -k_3 & \ldots & -k_r
\end{bmatrix}
\]

is Hurwitz. The Jacobian linearization at \((z, \xi) = 0\) is:

\[
J = \begin{bmatrix}
\frac{\partial f_0}{\partial z}(0,0) & \frac{\partial f_0}{\partial \xi}(0,0) \\
0 & A
\end{bmatrix}
\]

where \(\frac{\partial f_0}{\partial z}(0,0)\) is Hurwitz since \(\dot{z} = f_0(z,0)\) is exponentially stable by the proposition in Lecture 11, page 1. Since \(A\) is also Hurwitz, all eigenvalues of \(J\) have negative real parts \(\Rightarrow\) exponential stability.

Global asymptotic stability can be guaranteed with additional assumptions on the zero dynamics, such as ISS of \(\dot{z} = f_0(z,\xi)\) with respect to the input \(\xi\):

\[
\dot{y} = f_0(z,\xi)\]

Example: \(\dot{z} = -z + z^2\xi, \ \ \dot{\xi} = -k\xi\)

\((z, \xi) = 0\) is locally exponentially stable, but not globally: solutions escape in finite time for large \(z(0)\).

**I/O Linearizing Controller for Tracking**

For the output \(y(t)\) to track a reference signal\(^2\) \(y_d(t)\), replace (3) with:

\[
v = -k_1(\xi_1 - y_d(t)) - k_2(\xi_2 - \dot{y}_d(t)) \ldots - k_r(\xi_r - y_d^{(r-1)}(t)) + y_d^{(r)}(t)
\]

Let \(e_1 \triangleq \xi_1 - y_d(t), e_2 \triangleq \xi_2 - \dot{y}_d(t), \ldots, e_r \triangleq \xi_r - y_d^{(r-1)}(t)\). Then:

\[
\begin{align*}
\dot{e}_1 &= e_2 \\
\dot{e}_2 &= e_3 \\
& \vdots \\
\dot{e}_r &= v - y_d^{(r)}(t) = -k_1e_1 - \cdots - k_re_r
\end{align*}
\]

Thus \(e(t) \rightarrow 0\), that is \(y(t) - y_d(t) \rightarrow 0\).

If \(y_d(t)\) and its derivatives are bounded, then \(\xi(t)\) is bounded. If the zero dynamics \(\dot{z} = f_0(z,\xi)\) is ISS with respect to \(\xi\), then \(z(t)\) is also bounded. Thus, all internal signals are bounded.
Full-State Feedback Linearization

The system \( \dot{x} = f(x) + g(x)u, x \in \mathbb{R}^n, u \in \mathbb{R} \), is (full state) feedback linearizable if a function \( h(x) \) exists such that the relative degree from \( u \) to \( y = h(x) \) is \( n \).

Since \( r = n \), the normal form (1) has no zero dynamics and
\[
x \to \begin{bmatrix}
\xi_1 \\
\xi_2 \\
\vdots \\
\xi_n
\end{bmatrix} = \begin{bmatrix}
h(x) \\
L_f h(x) \\
\vdots \\
L_f^{n-1} h(x)
\end{bmatrix}
\]
is a diffeomorphism that transforms the system to the form:
\[
\begin{align*}
\dot{\xi}_1 &= \xi_2 \\
\dot{\xi}_2 &= \xi_3 \\
&\quad \vdots \\
\dot{\xi}_n &= L_f^n h(x) + L_g L_f^{n-1} h(x) u.
\end{align*}
\tag{4}
\]
Then, (2)-(3) with \( r = n \) is a feedback linearizing controller.

Closed-loop system in the new coordinates:
\[
\dot{\xi} = A\xi \quad \text{where} \quad A = \begin{bmatrix}
0 & 1 & 0 & \ldots \\
0 & 0 & 1 & \ldots \\
& & & & 1 \\
-k_1 & -k_2 & -k_3 & \ldots & -k_r
\end{bmatrix}.
\]

Example:
\[
\begin{align*}
\dot{x}_1 &= x_2 + 2x_1^2 \\
\dot{x}_2 &= x_3 + u \\
\dot{x}_3 &= x_1 - x_3
\end{align*}
\tag{5}
\]
The choice \( y = x_3 \) gives relative degree \( r = n = 3 \).

Let \( \xi_1 = x_3, \xi_2 = x_3 = x_1 - x_3, \xi_3 = \dot{x}_3 = \dot{x}_1 - \dot{x}_3 = x_2 + 2x_1^2 - x_1 + x_3 \).
\[
\begin{align*}
\dot{\xi}_1 &= \xi_2 \\
\dot{\xi}_2 &= \xi_3 \\
\dot{\xi}_3 &= (4x_1 - 1)(x_2 + 2x_1^2) + x_1 + u
\end{align*}
\]
Feedback linearizing controller:
\[
u = -(4x_1 - 1)(x_2 + 2x_1^2) - x_1 - k_1 \xi_1 - k_2 \xi_2 - k_3 \xi_3 \\
= -(4x_1 - 1)(x_2 + 2x_1^2) - x_1 - k_1 x_3 - k_2 (x_1 - x_3) \\
- k_3 (x_2 + 2x_1^2 - x_1 + x_3).
Feedback Linearization (continued)

Summary so far:

I/O linearization:
- suitable for tracking and set point stabilization
- output $y$ is a specific variable of physical interest
- relies on minimum phase property

Full state linearization:
- set point stabilization only
- output serves only to define a change of variables
- output selected such that $r = n$; no zero dynamics

Today: When is a system feedback linearizable, i.e., how do we know whether a relative degree $r = n$ output exists?

Basic Definitions from Differential Geometry

Definition: The Lie bracket of two vector fields $f(x)$ and $g(x)$ is a new vector field defined as:

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

Note:
1. $[f, g] = -[g, f]$,
2. $[f, f] = 0$,
3. If $f, g$ are constant then $[f, g] = 0$.

Notation for repeated applications:

$$[f, [f, g]] = \text{ad}_f^2 g, \quad [f, [f, [f, g]]] = \text{ad}_f^3 g, \quad \cdots$$

$$\text{ad}_f^0 g(x) \triangleq g(x), \quad \text{ad}_f^k g \triangleq [f, \text{ad}_f^{k-1} g] \quad k = 1, 2, 3, \ldots$$

Definition: Given vector fields $f_1(x), \ldots, f_k(x)$, a distribution $\Delta$ is defined as $\Delta(x) = \text{span}\{f_1(x), \ldots, f_k(x)\}$.

$f(x) \in \Delta(x)$ means that there exist scalar functions $a_i(x)$ such that

$$f(x) = a_1(x)f_1(x) + \cdots + a_k(x)f_k(x).$$

Definition: $\Delta(x)$ is said to be nonsingular if $f_1(x), \ldots, f_k(x)$ are linearly independent for all $x$. 
Definition: \( \Delta(x) \) is said to be involutive if
\[
g_1 \in \Delta, g_2 \in \Delta \implies [g_1, g_2] \in \Delta
\]
that is, \( \Delta \) is closed under the Lie bracket operation.

Proposition: \( \Delta(x) = \text{span}\{f_1(x), \ldots, f_k(x)\} \) is involutive if and only if
\[
[f_i, f_j] \in \Delta \quad 1 \leq i, j \leq k.
\]

Example 1: \( \Delta = \text{span}\{f_1, \ldots, f_k\} \) where \( f_1, \ldots, f_k \) are constant vectors

Example 2: single vector field \( f(x) \) is involutive since \([f, f] = 0 \in \Delta\)

Definition: A nonsingular \( k \)-dimensional distribution
\[
\Delta(x) = \text{span}\{f_1(x), \ldots, f_k(x)\} \quad x \in \mathbb{R}^n
\]
is said to be completely integrable if there exist \( n-k \) functions
\[
\phi_1(x), \ldots, \phi_{n-k}(x)
\]
such that
\[
\frac{\partial \phi_i}{\partial x} f_j(x) = 0 \quad i = 1, \ldots, n-k, \quad j = 1, \ldots, k
\]
and \( d\Phi_i(x) \triangleq \frac{\partial \phi_i}{\partial x}, \quad i = 1, \ldots, n-k, \) are linearly independent.

Example 3: If \( f_1, \ldots, f_k \) are linearly independent constant vectors, then we can find \( n-k \) independent row vectors \( T_1, \ldots, T_{n-k} \) s.t.
\[
T_i[f_1 \ldots f_k] = 0.
\]

Therefore, \( \Delta = \text{span}\{f_1, \ldots, f_k\} \) is completely integrable and
\[
\phi_i(x) = T_i x, \quad i = 1, \ldots, n-k.
\]

Frobenius Theorem: A nonsingular distribution is completely integrable if and only if it is involutive.

Example 3 above is a special case since \( \Delta \) is involutive by Example 1.

Back to (Full State) Feedback Linearization

Recall: \( \dot{x} = f(x) + g(x)u, x \in \mathbb{R}^n, u \in \mathbb{R} \) is feedback linearizable if we can find an output \( y = h(x) \) such that relative degree \( r = n \).

How do we determine if a relative degree \( r = n \) output exists?
\[
L_y h(x) = L_y L_f h(x) = \cdots = L_y L_f^{n-2} h(x) = 0 \quad \text{in a nbhd of } x_0 \quad (1)
\]
\[
L_y L_f^{n-1} h(x_0) \neq 0. \quad (2)
\]
Given \( C \):

\[
L_g h(x) = L_{\text{ad}_g} h(x) = \cdots = L_{\text{ad}_g^{n-2}} h(x) = 0 \text{ in a nbhd of } x_0 \tag{3}
\]

\[
L_{\text{ad}_g^{n-1}} h(x_0) \neq 0. \tag{4}
\]

The advantage of (3) over (1) is that it has the form:

\[
\frac{\partial h}{\partial x}[g(x) \ \text{ad}_f g(x) \ \cdots \ \text{ad}_f^{n-2} g(x)] = 0
\]

which is amenable to the Frobenius Theorem.

**Theorem:** \( \dot{x} = f(x) + g(x)u \) is feedback linearizable around \( x_0 \) if and only if the following two conditions hold:

C1) \( [g(x_0) \ \text{ad}_f g(x_0) \ \cdots \ \text{ad}_f^{n-1} g(x_0)] \) has rank \( n \)

C2) \( \Delta(x) = \text{span}\{g(x), \text{ad}_f g(x), \ldots, \text{ad}_f^{n-2} g(x)\} \) is involutive in a neighborhood of \( x_0 \).

**Proof:** (if) Given C1 and C2 show that there exists \( h(x) \) satisfying (3)-(4).

\( \Delta(x) \) is nonsingular by C1 and involutive by C2. Thus, by the Frobenius Theorem, there exists \( h(x) \) satisfying (3) and \( dh(x) \neq 0 \).

To prove (4) suppose, to the contrary, \( L_{\text{ad}_g^{n-1}} h(x_0) = 0 \). This implies

\[
dh(x_0)[g(x_0) \ \text{ad}_f g(x_0) \ \cdots \ \text{ad}_f^{n-1} g(x_0)] = 0.
\]

Thus \( dh(x_0) = 0 \), a contradiction.

(only if) Given that \( y = h(x) \) with \( r = n \) exists, that is (3)-(4) hold, show that C1 and C2 are true.

We will use the following fact\(^3\) which holds when \( r = n \):

\[
L_{\text{ad}_g} L_f^i h(x) = \begin{cases} 0 & \text{if } i + j \leq n - 2 \\ (-1)^{n-1-i} L_g L_f^{n-1} h(x) \neq 0 & \text{if } i + j = n - 1. \end{cases} \tag{5}
\]

Define the matrix

\[
M = \begin{bmatrix}
    dh \\
    dL_f h \\
    \vdots \\
    dL_f^{n-1} h
\end{bmatrix}
\begin{bmatrix}
    g \\
    -\text{ad}_f g \\
    \text{ad}_f^2 g \\
    \ldots \\
    (-1)^{n-1} \text{ad}_f^{n-1} g
\end{bmatrix} \tag{6}
\]

and note that the \((k, \ell)\) entry is:

\[
M_{k\ell} = dL_f^{k-1} h(x)(-1)^{\ell-1} \text{ad}_f^{\ell-1} g(x) = (-1)^{\ell-1} L_{\text{ad}_f^{\ell-1}} h(x).
\]
Then, from (5):

\[
M_{k\ell} = \begin{cases} 
0 & \ell + k \leq n \\
\neq 0 & \ell + k = n + 1.
\end{cases}
\]

Since the diagonal entries are nonzero, \( M \) has rank \( n \) and thus the factor

\[
\begin{bmatrix}
g & -\text{ad}_f g & \text{ad}_f^2 g & \ldots & (-1)^{n-1} \text{ad}_f^{n-1} g 
\end{bmatrix}
\]

in (6) must have rank \( n \) as well. Thus \( C_1 \) follows.

This also implies \( \Delta(x) \) is nonsingular; thus, by the Frobenius Thm,

\[
\text{complete integrability } \equiv \text{ involutivity.}
\]

\( \Delta(x) \) is completely integrable since \( h(x) \) satisfying (3) exists by assumption; thus, we conclude involutivity (C2).

\[ \square \]

Example:

\[
\begin{align*}
\dot{x}_1 &= x_2 + 2x_1^2 \\
\dot{x}_2 &= x_3 + u \\
\dot{x}_3 &= x_1 - x_3 
\end{align*}
\]

(7)

Feedback linearizability shown in previous lecture by inspection:

\( y = x_3 \) gives relative degree = 3. Verify with the theorem above:

\[
\begin{bmatrix}
f(x) = \\
g(x) = \\
[f, g](x) = \\
[f, [f, g]](x) =
\end{bmatrix}
\begin{bmatrix}
\begin{pmatrix} x_2 + 2x_1 \\ x_3 \\ x_1 - x_3 \end{pmatrix} \\
\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
\begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \\
\begin{pmatrix} 4x_1 \\ 0 \\ 1 \end{pmatrix}
\end{bmatrix}
\]

Conditions of the theorem:

1. \[
\begin{bmatrix}
0 & -1 & 4x_1 \\
1 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\] full rank

2. \( \Delta = \text{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \right\} \) involutive

\[
\frac{\partial h}{\partial x} = 
\begin{bmatrix}
0 & -1 \\
1 & 0 \\
0 & 0
\end{bmatrix}
\] satisfied by \( h(x) = x_3 \).
**Dissipativity Theory**

The notion of *dissipativity* characterizes dynamical systems by how their inputs and outputs correlate. This correlation is described by a scalar valued *supply rate* $s(u, y)$ the choice of which distinguishes the type of dissipativity.

**Definition:** The system below, where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $y \in \mathbb{R}^p$,

\[
\begin{align*}
\dot{x} &= f(x, u) \quad f(0,0) = 0 \\
y &= h(x, u) \quad h(0,0) = 0,
\end{align*}
\]

is said to be *dissipative* with respect to a *supply rate* $s(u, y)$ if there exists a $C^1$ function $V : \mathbb{R}^n \mapsto \mathbb{R} \geq 0$ such that $V(0) = 0$ and

\[
\nabla V(x)^T f(x, u) \leq s(u, h(x, u)) \quad \forall x \in \mathbb{R}^n, \forall u \in \mathbb{R}^m.
\]

$V$ is called a *storage function*.

Noting that the left hand side of (3) equals $\frac{d}{dt}V(x(t))$, and integrating from $t = 0$ to $\tau$, we get

\[
V(x(\tau)) - V(x(0)) \leq \int_0^\tau s(u(t), y(t))dt.
\]

Since $V(x(\tau)) \geq 0$, (4) implies

\[
-V(x(0)) \leq \int_0^\tau s(u(t), y(t))dt.
\]

Thus, the integral of the supply rate $s(u(t), y(t))$ along the trajectories is nonnegative when $x(0) = 0$. This means that the system favors a positive sign for $s(u(t), y(t))$ when averaged over time.

Important special cases of dissipativity are discussed below:

- **Finite $L_2$-gain:** $s(u, y) = \gamma^2 |u|^2 - |y|^2 \quad \gamma > 0$

The $L_2$ norm of a signal $u(t)$ is defined as

\[
\lim_{T \to \infty} \sqrt{\int_0^T |u(t)|^2 dt}
\]

when the limit exists. Note from (5) that

\[
-V(x(0)) \leq \gamma^2 \int_0^\tau |u(t)|^2 dt - \int_0^\tau |y(t)|^2 dt
\]
Dissipativity characterizes a dynamical system with a supply rate \( s(u, y) \) that describes how the inputs and outputs correlate, and an accompanying storage function \( V \).

\[
\Rightarrow \quad \int_0^\tau |y(t)|^2 dt \leq \gamma^2 \int_0^\tau |u(t)|^2 dt + V(x(0)).
\]

Taking square roots of both sides and applying the inequality \( \sqrt{a^2 + b^2} \leq |a| + |b| \) to the right-hand side we get

\[
\sqrt{\int_0^\tau |y(t)|^2 dt} \leq \gamma \sqrt{\int_0^\tau |u(t)|^2 dt} + \sqrt{V(x(0))}.
\]

This means that the \( L_2 \) norm of \( y(t) \) is bounded by that of \( u(t) \) multiplied by \( \gamma \), plus an offset term due to initial conditions. Thus \( \gamma \) serves as an \( L_2 \) gain for the system.

- **Passivity:** \( s(u, y) = u^T y \)

  With this choice of supply rate, (5) implies

  \[
  \int_0^\tau u(t)^T y(t) dt \geq -V(x(0))
  \]

  which favors a positive sign for the inner product of \( u(t) \) and \( y(t) \).

  Periods of time when \( u(t)^T y(t) < 0 \) must be outweighed by those when \( u(t)^T y(t) > 0 \).

- **Output Strict Passivity:** \( s(u, y) = u^T y - \varepsilon |y|^2 \quad \varepsilon > 0 \)

  This supply rate implies passivity since \( s(u, y) \leq u^T y \), but is more stringent than (6):

  \[
  \int_0^\tau u(t)^T y(t) dt \geq -V(x(0)) + \varepsilon \int_0^\tau |y(t)|^2 dt \geq 0
  \]

  Output strict passivity also implies an \( L_2 \) gain of \( \gamma = 1/\varepsilon \) because a completion of squares argument gives

  \[
  u^T y - \frac{1}{\gamma} y^T y \leq \frac{\gamma}{2} u^T u - \frac{1}{2\gamma} y^T y
  \]

  where the right-hand side is equal to

  \[
  \frac{1}{2\gamma} (\gamma^2 |u|^2 - |y|^2).
  \]

  Thus we use \( 2\gamma V \) as a storage function and conclude dissipativity with the \( L_2 \) gain supply rate \( \gamma^2 |u|^2 - |y|^2 \).
Graphical Interpretation

For a memoryless system
\[ y(t) = h(u(t)) \]
we take the storage function in (3) to be zero and interpret dissipativity as the static inequality
\[ s(u, h(u)) \geq 0 \quad \forall u \in \mathbb{R}^m. \quad (8) \]
This inequality characterizes a class of maps \( h(\cdot) \) that are dissipative with respect to the supply rate \( s(\cdot, \cdot) \). For example, a scalar function \( h(\cdot) \) is passive if \( uh(u) \geq 0 \) for all \( u \), which means that the graph of \( h(\cdot) \) lies in the first and third quadrants as in Figure 2 (left). Likewise, the sector in the middle represents the output strict passivity supply rate \( s(u, y) = uy - \varepsilon y^2, \varepsilon > 0 \), and the sector on the right represents the finite gain supply rate \( s(u, y) = \gamma^2 u^2 - y^2 \).

Examples of Passive Systems

Suppose we wish to prove passivity of the system
\[
\begin{align*}
\dot{x} &= f_0(x) + g(x)u \\
y &= h(x)
\end{align*}
\]
which is a special case of (1)-(2) with \( f(x, u) = f_0(x) + g(x)u \) affine in \( u \) and \( h(x, u) = h(x) \) independent of \( u \). Then (3) becomes
\[ \nabla V(x)^T f_0(x) + \nabla V(x)^T g(x)u \leq h(x)^T u \quad \forall x \in \mathbb{R}^n, \forall u \in \mathbb{R}^m \quad (9) \]
which is equivalent to
\[ \nabla V(x)^T f_0(x) \leq 0 \quad \nabla V(x)^T g(x) = h^T(x) \quad \forall x \in \mathbb{R}^n. \quad (10) \]
The inequality in (10) follows by setting \( u = 0 \) in (9). To see how the equality follows suppose there exists an \( x \) for which \( \nabla V(x)^T g(x) -
$h^T(x) \neq 0$. Then we can select a $u$ such that $(\nabla V(x)^T g(x) - h^T(x))u$ is positive and large enough to contradict (9).

Similar arguments show that output strict passivity is equivalent to

$$\nabla V(x)^T f_0(x) \leq -\epsilon h(x)^T h(x) \quad \nabla V(x)^T g(x) = h^T(x) \quad \forall x \in \mathbb{R}^n. \quad (11)$$

**Example 1:** Consider the scalar system

$$\dot{x} = f_0(x) + u, \quad y = h(x), \quad u, x, y \in \mathbb{R} \quad (12)$$

where $xh(x) \geq 0$ for all $x$, as in Figure 2 (left). For this system the equality in (11) is

$$\frac{dV(x)}{dx} = h(x)$$

whose solution subject to $V(0) = 0$ is

$$V(x) = \int_0^x h(z) dz. \quad (13)$$

Furthermore $V(x) \geq 0$ because $h(z)$ and $dz$ have equal signs (positive when the limit of integration is $x > 0$ and negative when $x < 0$).

The inequality condition in (11) is then

$$h(x)(f_0(x) + \epsilon h(x)) \leq 0$$

which is equivalent to

$$x(f_0(x) + \epsilon h(x)) \leq 0 \quad (14)$$

since $xh(x) \geq 0$. Thus, we conclude passivity when (14) holds with $\epsilon = 0$ and output strict passivity when (14) holds with $\epsilon > 0$.

Note that an integrator, where $f_0(x) \equiv 0$, is passive since (14) holds with $\epsilon = 0$, but not output strictly passive since (14) with $\epsilon > 0$ contradicts the assumption $xh(x) \geq 0$.

**Example 2:** Consider the second order model

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -kx_2 - \phi'(x_1) + u$$
$$y = x_2$$

where $\phi'(\cdot)$ is the derivative of a continuously differentiable and nonnegative function $\phi(\cdot)$ satisfying $\phi(0) = 0$. We interpret $x_1$ as position, $x_2$ as velocity, $u$ as force, $k \geq 0$ as damping coefficient, and $\phi(x_1)$ as potential energy of a mechanical system.

For this system the equality condition $\nabla V(x)^T g(x) = h^T(x)$ becomes:

$$\frac{\partial V(x_1, x_2)}{\partial x_2} = x_2.$$
Thus we restrict the storage function to be of the form:

\[ V(x_1, x_2) = V_1(x_1) + \frac{1}{2} x_2^2 \]

and examine the inequality condition \( \nabla V(x)^T f_0(x) \leq 0 \). We have

\[
\nabla V(x)^T f_0(x) = \frac{dV_1(x_1)}{dx_1} x_2 + x_2 (-kx_2 - \phi'(x_1)) \\
= -kx_2^2 + x_2 \left( \frac{dV_1(x_1)}{dx_1} - \phi'(x_1) \right).
\]

The choice \( V_1(x_1) = \phi(x_1) \) ensures \( \nabla V(x)^T f_0(x) = -kx_2^2 = -kh(x)^2 \) which proves passivity when \( k = 0 \) and output strict passivity when \( k > 0 \).

The resulting storage function \( V(x_1, x_2) = \phi(x_1) + \frac{1}{2} x_2^2 \) is the sum of potential and kinetic energy terms, and \( u(t)y(t) = \text{force} \cdot \text{velocity} \) may be interpreted as the power supplied to the system. The definition of dissipativity (4) is thus consistent with the physical notion of energy storage, and dissipation when damping is present.
Stability of Interconnected Systems

Consider the interconnected system in Figure 1 where each subsystem $G_i, i = 1, \cdots, N$, is described by

$$\dot{x}_i = f_i(x_i, u_i)$$

$$y_i = h_i(x_i, u_i)$$

with $x_i \in \mathbb{R}^{n_i}, u_i \in \mathbb{R}^{m_i}, y_i \in \mathbb{R}^{p_i}, f_i(0,0) = 0, h_i(0,0) = 0$.

The constant matrix $M$ defines the coupling of these subsystems by

$$u = My$$

where $u = [u_1^T \cdots u_N^T]^T$ and $y = [y_1^T \cdots y_N^T]^T$.

For example,

$$M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

describes a negative feedback interconnection of two subsystems as in the block diagram on the right where $u_1 = -y_2$ and $u_2 = y_1$.

We assume that the interconnection is well-posed: when we substitute $y_i = h_i(x_i, u_i)$, the equation (3) admits a unique solution for $u$ as a function $x$.

To derive a stability test for the interconnection we assume each subsystem is dissipative with a positive definite storage function $V_i(x_i)$ and a quadratic supply rate:

$$s_i(u_i, y_i) = \begin{bmatrix} u_i \\ y_i \end{bmatrix}^T X_i \begin{bmatrix} u_i \\ y_i \end{bmatrix} = \begin{bmatrix} u_i \\ y_i \end{bmatrix}^T \begin{bmatrix} X_{i11} & X_{i12} \\ X_{i21} & X_{i22} \end{bmatrix} \begin{bmatrix} u_i \\ y_i \end{bmatrix}$$

where $X_{jk}^i, j, k \in \{1, 2\}$, are conformal block partitions of $X_i$. 

Figure 1: An interconnection of subsystems $G_1, \cdots, G_N$. The inputs depend on the outputs of other subsystems by $u = My$ where $M$ is a constant matrix.
Then we search for a weighted sum of storage functions

\[ V(x) = p_1 V_1(x_1) + \cdots + p_N V_N(x_N) \quad p_i > 0, \ i = 1, \cdots, N \]  

(6)

that serves as a Lyapunov function for the interconnection. This means that the right hand side of the inequality

\[ \sum_{i=1}^{N} p_i \nabla V(x_i)^T f_i(x_i, u_i) \leq \sum_{i=1}^{N} p_i \begin{bmatrix} u_i \\ y_i \end{bmatrix}^T X_i \begin{bmatrix} u_i \\ y_i \end{bmatrix} \]  

(7)

must be negative semidefinite in \( y \) when \( u \) is eliminated with the substitution \( u = My \). Rewriting the right hand side of (7) as

\[ \begin{bmatrix} u_1 \\ \vdots \\ u_N \\ y_1 \\ \vdots \\ y_N \end{bmatrix}^T \begin{bmatrix} p_1 X_{11}^{11} & \cdots & p_1 X_{12}^{12} \\ \cdots & \ddots & \cdots \\ p_N X_{N1}^{11} & \cdots & p_N X_{N2}^{12} \\ p_1 X_{11}^{21} & \cdots & p_1 X_{12}^{21} \\ \cdots & \ddots & \cdots \\ p_N X_{N1}^{21} & \cdots & p_N X_{N2}^{22} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_N \\ y_1 \\ \vdots \\ y_N \end{bmatrix} \]

\( \triangleq X(p_1 X_1, \cdots, p_N X_N) \)

\[ = y^T \begin{bmatrix} M & I \end{bmatrix}^T X(p_1 X_1, \cdots, p_N X_N) \begin{bmatrix} M & I \end{bmatrix} y \]

we obtain the following stability criterion:

**Theorem:** If there exist \( p_i > 0, \ i = 1, \cdots, N, \) such that

\[ \begin{bmatrix} M & I \end{bmatrix}^T X(p_1 X_1, \cdots, p_N X_N) \begin{bmatrix} M & I \end{bmatrix} \leq 0 \]  

(9)

where \( X(p_1 X_1, \cdots, p_N X_N) \) is as defined in (8), then \( x = 0 \) is stable for the composite system (1)-(3) and (6) is a Lyapunov function.

Note that (9) is a linear matrix inequality (LMI) and the search for \( p_i > 0 \) satisfying this inequality can be performed numerically.

For memoryless subsystems of the form \( y_i = h_i(u_i) \) we take the corresponding storage function in (6) to be zero.

Asymptotic stability requires additional assumptions; e.g., the Invariance Principle is applicable if the inequality (9) is strict and \( x(t) = 0 \) is the only solution satisfying \( h_i(x_i(t), 0) = 0, \ i = 1, \cdots, N \) for all \( t \).

We next specialize the LMI (9) to particular types of dissipativity. This allows us to derive analytical LMI feasibility conditions for special interconnection matrices \( M \).
Small Gain Theorem

Suppose each subsystem is single-input-single-output and possesses a finite $L_2$ gain; that is the supply rate is as in (5) with

$$X_i = \begin{bmatrix} \gamma_i^2 & 0 \\ 0 & -1 \end{bmatrix}.$$ 

Defining $P \triangleq \text{diag}(p_1, \ldots, p_N)$ and $\Gamma \triangleq \text{diag}(\gamma_1, \ldots, \gamma_N)$ we get

$$X(p_1X_1, \ldots, p_NX_N) = \begin{bmatrix} \Gamma P & 0 \\ 0 & -P \end{bmatrix}$$

and (9) becomes

$$(\Gamma M)^T P (\Gamma M) - P \leq 0. \tag{10}$$

Thus a diagonal matrix $P > 0$ satisfying this LMI certifies the stability of the interconnection.

When $M$ is as in (4), the LMI (10) becomes

$$\begin{bmatrix} p_2\gamma_2^2 & 0 \\ 0 & p_1\gamma_1^2 \end{bmatrix} - \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} \leq 0$$

which consists of two simultaneous inequalities, $p_2\gamma_2^2 \leq p_1$ and $p_1\gamma_1^2 \leq p_2$. We rewrite them as

$$\gamma_2^2 \leq \frac{p_1}{p_2} \leq \frac{1}{\gamma_1^2}$$

and note that such $p_1$ and $p_2$ exist if and only if $\gamma_2^2 \leq \frac{1}{\gamma_1^2}$, that is

$$\gamma_1\gamma_2 \leq 1.$$ 

This is known as the “small gain” condition, as it restricts the feedback loop gain by one.

The derivation above yields the same condition, $\gamma_1\gamma_2 \leq 1$, when adapted to the positive feedback interconnection where

$$M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$ 

Thus the small gain criterion is oblivious to the feedback sign.
Stability of Interconnected Systems Continued

Summary: Suppose each subsystem $G_i, i = 1, \cdots, N$, in the block diagram below is dissipative with quadratic supply rate:

$$s_i(u_i, y_i) = \begin{bmatrix} u_i \\ y_i \end{bmatrix}^T X_i \begin{bmatrix} u_i \\ y_i \end{bmatrix} = \begin{bmatrix} u_i \\ y_i \end{bmatrix}^T \begin{bmatrix} X_{11}^i & X_{12}^i \\ X_{21}^i & X_{22}^i \end{bmatrix} \begin{bmatrix} u_i \\ y_i \end{bmatrix} \quad (1)$$

and a positive definite storage function $V_i(x_i)$. \footnote{For static blocks we interpret the storage function to be zero.}

If there exist $p_i > 0, i = 1, \cdots, N$, such that

$$\begin{bmatrix} M & I \\ \end{bmatrix}^T \begin{bmatrix} p_1 X_{11}^1 & \cdots & p_1 X_{12}^1 \\ \cdots & \cdots & \cdots \\ p_N X_{11}^N & \cdots & p_N X_{12}^N \\ p_1 X_{21}^1 & \cdots & p_1 X_{22}^1 \\ \cdots & \cdots & \cdots \\ p_N X_{21}^N & \cdots & p_N X_{22}^N \end{bmatrix} \begin{bmatrix} M & I \end{bmatrix} \leq 0 \quad (2)$$

then the following is a Lyapunov function for the interconnection:

$$V(x) = p_1 V_1(x_1) + \cdots + p_N V_N(x_N) \quad (3)$$

Passivity Theorem

Specialize (2) to the case of passivity where

$$X_i = \begin{bmatrix} 0 & 1/2 \\ 1/2 & -\epsilon_i \end{bmatrix} \quad \epsilon_i \geq 0.$$
With \( P \triangleq \text{diag}(p_1, \ldots, p_N) \) and \( E \triangleq \text{diag}(\varepsilon_1, \ldots, \varepsilon_N) \) we get
\[
X(p_1X_1, \ldots, p_NX_N) = \frac{1}{2} \begin{bmatrix} 0 & P \\ P & -2PE \end{bmatrix}
\]
which means that (2) is equivalent to
\[
P(M - E) + (M - E)^T P \leq 0. \tag{4}
\]
Thus, a diagonal matrix \( P > 0 \) satisfying this LMI certifies the stability of the interconnected system. Below we exhibit classes of interconnection structures for which (4) admits a diagonal solution \( P > 0 \).

**Skew Symmetric Interconnections (\( M = -M^T \))**

When \( M + M^T = 0 \) and \( E \geq 0 \), (4) holds trivially with \( P = I \).

Thus stability is inherent to skew symmetric interconnections of passive subsystems. A simple example of a skew symmetric interconnection is the negative feedback interconnection of two subsystems:
\[
M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{5}
\]
The stability of this interconnection with passive subsystems is a classical result known as the Passivity Theorem.

**Example:** Consider the SISO linear system:
\[
G_1 : \quad \dot{x} = Ax + Bu_1 \quad y_1 =Cx
\]
in negative feedback with a static nonlinearity:
\[
G_2 : \quad y_2 = h(u_2)
\]
satisfying \( uh(u) \geq 0 \ \forall u \in \mathbb{R} \), which means that \( G_2 \) is passive. Thus, if \( G_1 \) is also passive\(^3\) we conclude that the origin \( x = 0 \) is stable for the composite system:
\[
\dot{x} = Ax - Bh(Cx).
\]
Since \( G_2 \) is static, we take its storage function \( V_2 \) to be zero. The Lyapunov function is then the storage function \( V_1 \) of \( G_1 \).

**Negative Feedback Cyclic Interconnection**

Next consider a negative feedback loop of \( N \) subsystems where the interconnection matrix is
\[
M = \begin{bmatrix} 0 & \cdots & 0 & \delta_1 \\ \delta_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \delta_N & 0 \end{bmatrix} \quad \text{with} \quad \prod_{i=1}^{N} \delta_i = -1. \tag{6}
\]

\(^3\) If there exists \( Q = Q^T > 0 \) s.t. \( A^TQ + QA \leq 0 \) and \( QB = C^T \), then \( V_1 = \frac{1}{2} x^TQx \) is a pos.def. storage function verifying passivity.
Theorem: When $M$ is as in (6), the LMI (4) admits a diagonal solution $P > 0$ if and only if
\[ \varepsilon_1 \cdots \varepsilon_N \geq \cos^N(\pi/N). \] (7)
Moreover, (4) holds with strict inequality if and only if (7) is strict.

For $N = 2$ the condition (7) recovers the classical Passivity Theorem: $\cos(\pi/2) = 0$ and passivity ($\varepsilon_i \geq 0$) guarantees stability. For $N \geq 3$, $\cos(\pi/N) > 0$ and (7) demands output strict passivity ($\varepsilon_i > 0$).

To compare (7) with the Small Gain Theorem (Lecture 23), recall that output strict passivity implies an $L_2$ gain of $\gamma_i = 1/\varepsilon_i$ and rewrite (7) as
\[ \gamma_1 \cdots \gamma_N \leq \sec^N(\pi/N) \] (8)
where $\sec(\cdot) = 1/\cos(\cdot)$. Unlike the Small Gain Theorem which restricts the feedback loop gain by one, the “secant condition” (8) offers the relaxed bound $\sec^N(\pi/N)$ which is equal to 8 when $N = 3$, and decreases towards one as $N \to \infty$. This sharper bound is due to the output strict passivity assumption which restricts the subsystems further than an $L_2$ gain bound.

Example: Consider the following model for a ring oscillator circuit (Figure 2) that consists of a feedback loop of three inverters:
\[ \begin{align*}
\tau_1 \dot{x}_1 &= -x_1 - h_3(x_3) \\
\tau_2 \dot{x}_2 &= -x_2 - h_1(x_1) \\
\tau_3 x_3 &= -x_3 - h_2(x_2)
\end{align*} \] (9)
where $\tau_i = R_i C_i > 0$, $i = 1, 2, 3$, and $x_i$ represent voltages. The functions $h_i(\cdot)$ depend on the inverter characteristics and satisfy
\[ h_i(0) = 0, \quad x h_i(x) > 0 \quad \forall x \neq 0, \] (10)
as in the commonly used model
\[ h_i(x) = \alpha_i \tanh(\beta_i x) \quad \alpha_i > 0, \beta_i > 0. \] (11)
We decompose (9) into the subsystems
\[ G_i : \quad \tau_i \dot{x}_i = -x_i + u_i \quad y_i = h_i(x_i) \]
interconnected according to $u = My$ where $M \in \mathbb{R}^{3 \times 3}$ is as in (6) with $\delta_1 = \delta_2 = \delta_3 = -1$. 

Using the storage function

$$V_i(x_i) = \tau_i \int_0^{x_i} h_i(z) dz,$$  \hspace{1cm} (12)

we conclude that each $G_i$ is output strictly passive if there exist $\epsilon_i > 0$ s.t.

$$x h_i(x) \geq \epsilon_i h_i(x)^2.$$  

This inequality restricts the graph of $h_i(\cdot)$ to the sector depicted in the figure on the right with slope $\gamma_i = 1/\epsilon_i$. An example of such a function is (11) where $\gamma_i = \alpha_i \beta_i$.

Then, an application of (8) with $N = 3$ shows that the equilibrium of the interconnection $x = 0$ is stable when:

$$\gamma_1 \gamma_2 \gamma_3 \leq 8$$  \hspace{1cm} (13)

and a weighted sum of storage functions, each constructed as in (12), serves as a Lyapunov function:

$$V(x) = \sum_{i=1}^{3} p_i \tau_i \int_0^{x_i} h_i(z) dz.$$  

The weights $p_i > 0$ are to be obtained from the LMI (4) which is guaranteed to have a diagonal solution $P > 0$ by (13).

When the inequality (13) is strict we conclude asymptotic stability because (4) is also strict which means that $\dot{V}$ is upper bounded by a negative definite quadratic function of $y$. Since, further, $y_i = h_i(x_i) = 0 \Rightarrow x_i = 0$ by (10), we conclude $\dot{V}$ is a negative definite function of $x$.

When $\tau_1 = \tau_2 = \tau_3$, the secant condition (13) is also necessary for stability. Once the loop gain exceeds 8, the equilibrium loses its stability and a limit cycle emerges, hence the term "ring oscillator."
Case Study: A Vehicle Platoon

Consider a platoon where the velocity of each vehicle is governed by
\[ \dot{v}_i = -v_i + v_0^i + u_i, \quad i = 1, \ldots, N \] (1)
in which \( u_i \) is a coordination feedback to be designed and \( v_0^i \) is the (constant) nominal velocity of vehicle \( i \) in the absence of such feedback. The position of vehicle \( i \) is then obtained from
\[ \dot{x}_i = v_i. \]

We will design feedback laws that depend on relative positions with respect to a subset of other vehicles, typically nearest neighbors.

We introduce an undirected graph where the vertices represent the vehicles and an edge between vertices \( i \) and \( j \) means that vehicles \( i \) and \( j \) have access to the relative position measurement \( x_i - x_j \). Next we assign an orientation to each edge by selecting one end to be the head and the other to be the tail. Then the incidence matrix
\[ D_{il} = \begin{cases} 
1 & \text{if vertex } i \text{ is the head of edge } l \\
-1 & \text{if vertex } i \text{ is the tail of edge } l \\
0 & \text{otherwise}
\end{cases} \] (2)
generates a vector of relative positions \( z_l \) for the edges \( l = 1, \ldots, L \) by
\[ z = D^T x. \] (3)

As an illustration, in Figure 1,
\[ D = \begin{bmatrix} 
1 & 0 \\
-1 & 1 \\
0 & -1
\end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 
z_1 \\
z_2
\end{bmatrix} = D^T x = \begin{bmatrix} 
x_1 - x_2 \\
x_2 - x_3
\end{bmatrix}. \]

We propose the feedback law
\[ u = -D \begin{bmatrix} 
h_1(z_1) \\
\vdots \\
h_L(z_L)
\end{bmatrix} \] (4)
where each function \( h_l : \mathbb{R} \rightarrow \mathbb{R} \) is increasing and onto. This means that vehicle \( i \) applies the input
\[
u_i = -\sum_{l=1}^{L} D_{il} h_l(z_l)
\] (5)

which depends on locally available measurements because \( D_{il} \neq 0 \) when vertex \( i \) is the head or tail of edge \( l \). In the case of Figure 1,
\[
u_1 = -h_1(z_1) \quad \nu_2 = h_1(z_1) - h_2(z_2) \quad \nu_3 = h_2(z_2)
\]

where we may interpret \( h_1(z_1) \) and \( h_2(z_2) \) as virtual spring forces between vehicles 1 and 2, and 2 and 3 respectively.

Note from (3) that
\[
\dot{z} = D^T v \triangleq w
\] (6)

where we interpret \( w \) as an input and define the output
\[
y \triangleq \begin{bmatrix}
h_1(z_1) \\
\vdots \\
h_L(z_L)
\end{bmatrix}
\] (7)

Then the closed-loop system is as in Figure 2 (left) where the feed-forward blocks \( u_i \rightarrow v_i \) represent the velocity dynamics (1) and the feedback blocks \( w_l \rightarrow y_l \) represent the \( l \)th subsystem of the relative position dynamics (6)-(7).

This block diagram is equivalent to the one in Figure 2 (right) which is of the standard form in Lecture 23 with the interconnection matrix
\[
M = \begin{bmatrix}
0 & -D \\
D^T & 0
\end{bmatrix}
\] (8)

The skew symmetry of \( M \) will allow us to conclude stability from the passivity properties of the subsystems.

Figure 2: A block diagram for the platoon dynamics. Left: the feed-forward blocks \( u_i \rightarrow v_i \) represent the velocity dynamics (1) and the feedback blocks \( w_l \rightarrow y_l \) represent the \( l \)th subsystem of the relative position dynamics (6)-(7). Right: the diagram on the left brought to the standard form in Lecture 23 with the interconnection matrix (8).
Determining the Equilibrium

At equilibrium the right hand side of (6) must vanish, that is

$$D^T v^* = 0.$$  \hfill (9)

By the definition (2) above, the null space of $D^T$ includes the vector of ones: $D^T \mathbb{1} = 0$. In addition, if the graph is connected then the span of $\mathbb{1}$ constitutes the entire null space: there is no solution to (9) other than $v^* = \alpha \mathbb{1}$ where $\alpha$ is a common platoon velocity.

Setting the right hand side of (1) to zero, we see that the equilibrium value of the inputs $u_i$ must compensate for the variations in the nominal velocities $v_0^i$ so that a common velocity $\alpha$ can be maintained:

$$-\alpha + v_0^i + u_i^* = 0 \quad i = 1, \ldots, N.$$  \hfill (10)

Note that $\sum_{i=1}^N u_i = \mathbb{1}^T u = 0$, which follows from (4) and $\mathbb{1}^T D = 0$.

Thus, if we add the equation (10) for $i = 1$ to $i = N$ we get

$$-N\alpha + \sum_{i=1}^N v_0^i = 0$$

which shows that the common velocity $\alpha$ must be the average $\frac{1}{N} \sum_{i=1}^N v_0^i$.

Substituting this average for $\alpha$ and (5) for $u_i^*$ back in (10) we obtain the following equations for $z_l^*$:

$$v_0^i - \frac{1}{N} \sum_{i=1}^N v_0^i = \sum_{i=1}^L D_i h_1(z_i^*) \quad i = 1, \ldots, N.$$  

These equations are particularly transparent for a line graph as in Figure 1 where the head and tail of edge $l$ are vertices $l$ and $l+1$:

$$v_0^1 - \frac{1}{N} \sum_{i=1}^N v_0^i = h_1(z_1^*)$$

$$v_0^i - \frac{1}{N} \sum_{i=1}^N v_0^i = -h_{i-1}(z_{i-1}^*) + h_i(z_i^*) \quad i = 2, \ldots, N - 1$$

$$v_0^N - \frac{1}{N} \sum_{i=1}^N v_0^i = -h_{N-1}(z_{N-1}^*).$$

Adding equations $i = 1$ to $l$ we get a new equation that depends only on $h_l(z_l^*)$. Then a solution $z_l^*$ exists since $h_l(\cdot)$ is onto, and is unique since $h_l(\cdot)$ is increasing. A similar argument may be developed for other acyclic graphs.
Stability Analysis

To analyze the stability of the equilibrium characterized above, we define the shifted state variables

\[ \tilde{v}_i \triangleq v_i - \alpha \quad \tilde{z}_l \triangleq z_l - z_{l}^* \]

so that, at equilibrium \( \tilde{v}_i = 0 \) and \( \tilde{z}_l = 0 \).

From (1) and (10),

\[ \dot{\tilde{v}}_i = -v_i + v_i^0 + u_i = -\tilde{v}_i + \tilde{u}_i \]

which is output strictly passive with input \( \tilde{u}_i \triangleq u_i - u_i^* \) and output \( \tilde{v}_i \), since the storage function

\[ V_i(\tilde{v}_i) = \frac{1}{2} \tilde{v}_i^2 \]

satisfies

\[ \dot{V}_i = -\tilde{v}_i^2 + \tilde{v}_i \tilde{u}_i. \]

Likewise, from (6)-(7),

\[ \dot{\tilde{z}}_l = w_l \]

which is passive with input \( w_l \) and output:

\[ \tilde{y}_l \triangleq h_l(z_l) - h_l(z_l^*). \]

To see this, take the storage function

\[ W_l(\tilde{z}_l) = \int_{0}^{\tilde{z}_l} [h_l(z_l^* + \sigma) - h_l(z_l^*)]d\sigma \]

which satisfies

\[ \dot{W}_l = [h_l(z_l) - h_l(z_l^*)]w_l = \tilde{y}_l w_l. \]

It follows from the skew symmetry of the interconnection matrix \( M \) and the passivity of the subsystems (see Lecture 24) that the origin \( \tilde{v}_i = 0 \) and \( \tilde{z}_l = 0 \) is stable and a Lyapunov function is

\[ \sum_{i=1}^{N} V_i(\tilde{v}_i) + \sum_{l=1}^{L} W_l(\tilde{z}_l). \]