Control Systems I
Lecture 13: Loop Shaping

Readings: Åstrom and Murray, Chapter 11

Jacopo Tani

Institute for Dynamic Systems and Control
D-MAVT
ETH Zürich

December 14, 2018
Time for feedback!

- By this time in the course we understand the power and importance of feedback.
- Feedback works for every system. We have thought a system as mechanical / electrical / thermal so far.
- Anything that varies in time and has inputs and outputs is a system: including learning.

- Your feedback is critical for improving our ”controller”!
- Only 130 of you filled up the course evaluation so far - there are nearly 400 students enrolled. Take 5 mins to fill up now!
- Additional feedback for Dr. Mousavi lectures: https://goo.gl/forms/ZP9gbdGqJo685MZB2
<table>
<thead>
<tr>
<th>#</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sept. 21</td>
<td>Introduction, Signals and Systems</td>
</tr>
<tr>
<td>2</td>
<td>Sept. 28</td>
<td>Modeling, Linearization</td>
</tr>
<tr>
<td>3</td>
<td>Oct. 5</td>
<td>Analysis 1: Time response, Stability</td>
</tr>
<tr>
<td>4</td>
<td>Oct. 12</td>
<td>Analysis 2: Diagonalization, Modal coordinates</td>
</tr>
<tr>
<td>5</td>
<td>Oct. 19</td>
<td>Transfer functions 1: Definition and properties</td>
</tr>
<tr>
<td>6</td>
<td>Oct. 26</td>
<td>Transfer functions 2: Poles and Zeros</td>
</tr>
<tr>
<td>7</td>
<td>Nov. 2</td>
<td>Analysis of feedback systems: internal stability, root locus</td>
</tr>
<tr>
<td>8</td>
<td>Nov. 9</td>
<td>Frequency response</td>
</tr>
<tr>
<td>9</td>
<td>Nov. 16</td>
<td>Analysis of feedback systems 2: the Nyquist condition</td>
</tr>
<tr>
<td>10</td>
<td>Nov. 23</td>
<td>Frequency response II</td>
</tr>
<tr>
<td>11</td>
<td>Nov. 30</td>
<td>Specifications for feedback systems</td>
</tr>
<tr>
<td>12</td>
<td>Dec. 7</td>
<td>PID control</td>
</tr>
<tr>
<td>13</td>
<td>Dec. 14</td>
<td><strong>Loop Shaping</strong></td>
</tr>
<tr>
<td>14</td>
<td>Dec. 21</td>
<td>State feedback and Luenberger observers</td>
</tr>
</tbody>
</table>
Point of the Situation

- **Analysis**: given a controller, determine properties of the closed loop system from knowledge of open loop transfer function (Root Locus, Bode, Nyquist).
- **Synthesis**: given a plant model (or not?) find a controller that provides:
  - **Stability**: closed loop poles have negative real part;
  - **Performance**: given time domain and/or frequency domain specifications are met (e.g.: steady-state error, rise time, overshoot, crossover frequency, etc.);
  - **Robustness**: margin of gain and margin of phase are satisfactory.

We have talked about PID: the most successful controller of all times.

Today we talk about a more general, albeit less intuitive, approach: loop shaping.
Recap: Proportional-Integral-Derivative Control

- PID control:

\[ u(t) = k_P e(t) + k_I \int_0^t e(\tau) d\tau + k_D \dot{e}(t), \]

\[ C(s) = k_P + \frac{k_I}{s} + k_D s = \frac{k_D s^2 + k_P s + k_I}{s}. \]
**PID Summary**

- **Proportional control**
  - Decrease the steady-state error;
  - Increase the closed-loop bandwidth;
  - Increase sensitivity to noise;
  - Can reduce stability margins for higher-order systems (2nd order or more).

- **Integral control**
  - Eliminates the steady-state error to a step (if the closed-loop is stable);
  - Reduces stability margins, can make a higher-order system unstable.

- **Derivative control**
  - Reduce overshooting, increase damping;
  - Improves stability margins;
  - Increase sensitivity to noise.
PID Tuning

- PID tuning corresponds to choosing the parameters $k_p$, $k_i$ and $k_d$ to reach the feedback control design specifications.

- PID tuning can be done by hand (a.k.a. “synthesis through attempts”, or “tweak until death”) or numerically using MATLAB or other tools (the former does not require a system model).

- There exist heuristic methods to tune a PID controller without a model of the plant $P(s)$, e.g. the tuning rules proposed by Ziegler and Nichols.

- Regardless of chosen method, one can always think of a PID as:

$$C(s) = k_{RL} \frac{(s - z_1)(s - z_2)}{s}$$

i.e., as two zeros and one pole at the origin. Decide where you want these zeros (in the complex plane, or in terms of natural frequency and damping ratio on the Bode plot), and what you want the (root-locus) gain to be. Finally, compute the corresponding $k_P$, $k_I$, $k_D$. 
Ziegler Nichols Tuning Rules

- Assumption: Plant can be approximated by the transfer function

\[ P(s) = \frac{k}{\tau s + 1} e^{-Ts} \]

with \( T/(T + \tau) \) small.

- Apply the controller \( C(s) = k_p \) to the system starting at \( k_p = 0 \) and increase \( k_p \) until the system is in a steady-state oscillation, then note the "critical \( k_p \)" called \( k_p^* \) and the corresponding critical oscillation period \( T^* \).

- Use \( k_p^* \) and \( T^* \) to calculate the control gains:

<table>
<thead>
<tr>
<th>type</th>
<th>( k_p )</th>
<th>( T_i )</th>
<th>( T_d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>( 0.5 \cdot k_p^* )</td>
<td>( \infty \cdot T^* )</td>
<td>( 0 \cdot T^* )</td>
</tr>
<tr>
<td>PI</td>
<td>( 0.45 \cdot k_p^* )</td>
<td>( 0.85 \cdot T^* )</td>
<td>( 0 \cdot T^* )</td>
</tr>
<tr>
<td>PD</td>
<td>( 0.55 \cdot k_p^* )</td>
<td>( \infty \cdot T^* )</td>
<td>( 0.15 \cdot T^* )</td>
</tr>
<tr>
<td>PID</td>
<td>( 0.6 \cdot k_p^* )</td>
<td>( 0.5 \cdot T^* )</td>
<td>( 0.125 \cdot T^* )</td>
</tr>
</tbody>
</table>
Towards loop shaping
Command tracking/disturbance rejection vs. noise rejection

- Sensitivity function $S(s) = \frac{e(s)}{r(s)} = \frac{e(s)}{d(s)} = \frac{1}{1 + L(s)} = \frac{1}{1 + P(s)C(s)}$
  - If we want small errors to reference inputs, and want to reject disturbances, we need $S(s)$ to be “small.”

- Complementary sensitivity $T(s) = \frac{y(s)}{n(s)} = \frac{L(s)}{1 + L(s)}$.
  - If we do not want the effect of the noise to be observed at the output, then we need $T(s)$ to be “small.”

So if we want to reject both disturbances and noise, and want to follow commands well, we need both $T(s)$ and $S(s)$ to be “small”.

But $T(s) + S(s) = 1$ so they cannot be “small” at the same time!
Recap: Frequency-domain specifications

- Typically commands \((u)\) and disturbances \((d)\) act at “low” frequency, e.g., no more than 10 Hz.

- Noise \((n)\) is typically a high-frequency phenomenon, e.g., more than 100 Hz.

- So we can reconcile both command tracking/disturbance rejection AND noise rejection by separating them frequency-wise!

  - Make \(|S(j\omega)| << 1\) (hence \(|T(j\omega)| \approx 1\)) at low frequencies.
    e.g., “ensure that commands are tracked with max 10% error up to a frequency of 10Hz.

  - Make \(|T(j\omega)| << 1\) at high frequencies.
    e.g., “ensure that noise is reduced by a factor of 10 at the output at frequencies higher than 100 Hz.
Frequency-domain specifications on the Bode plot

- Frequency-domain specifications are usually expressed in terms of closed-loop frequency response. Can we write them in terms of the open loop frequency response? Yes we can.

- For good command tracking / disturbance rejection, we want $|S(j\omega)| = \frac{1}{|1 + L(j\omega)|}$ to be small at low frequencies, i.e., we want $|L(j\omega)|$ to be large at low frequencies.

- Typically this is written as $|S(j\omega)| \cdot |W_1(j\omega)| < 1$ for some function $|W_1(j\omega)|$ that is large at low frequency. This translates to $|1 + L(j\omega)| > |W_1(j\omega)|$, which is approximated as

  $$|L(j\omega)| > |W_1(j\omega)|.$$  

- This can be seen as a “low frequency obstacle” on the magnitude Bode plot.
For good noise rejection, we want \( |T(j\omega)| = \frac{|L(j\omega)|}{|1 + L(j\omega)|} \) to be small at high frequencies.

If \( |T(j\omega)| \) is small, then \( |L(j\omega)| \) has to be small, and \( |T(j\omega)| \approx |L(j\omega)| \) (at high frequencies).

Typically this is written as \( |T(j\omega)| \cdot |W_2(j\omega)| < 1 \) for some function \( |W_2(j\omega)| \) that is large at high frequency. This translates to \( |L(j\omega)| < |W_2(j\omega)|^{-1} \).

This can be seen as a “high-frequency obstacle” on the Bode plot.
Closed-loop bandwidth and (open-loop) crossover

- The bandwidth of the closed-loop system is defined as the maximum frequency $\omega$ for which $|T(j\omega)| > 1/\sqrt{2}$, i.e., the output ($y$) can track the reference commands ($r$) to within a factor of $\approx 0.7$.

- Let $\omega_c$ be the crossover frequency, such that $|L(j\omega_c)| = 1$. If we assume that the phase margin is about $90^\circ$, then $L(j\omega_c) = -j$, and $|T(j\omega_c)| = 1/\sqrt{2}$.

- In other words, the (open-loop) crossover frequency is approximately equal to the bandwidth of the closed-loop system.
Bode-plot “obstacle course”

Load disturbance attenuation

Robustness

$\omega_{gc}$

$W_{2}^{-1}$

$W_1$

High frequency measurement noise

$\log |L(i\omega)|$

$\log \omega$
Loop Shaping

- How can we “steer” the open-loop frequency response through the Bode obstacle course?

- **Proportional (static) compensation:** Choose a proportional controller with transfer function \( C(s) = k \).

- **Dynamic compensation,** i.e., by choosing a controller (compensator) with transfer function \( C(s) \) so that \( L(s) = P(s)C(s) \) satisfies the requirements.

- **Idea:** the basic approach to designing a dynamic compensator is to choose one or more of the following elements (call the lead/lag \( R_i(s) \)):

  - Gain: \( k \),
  - Lead: \( \frac{s/a + 1}{s/b + 1} \) (with \( 0 < a < b \)),
  - Lag: \( \frac{s/a + 1}{s/b + 1} \) (with \( 0 < b < a \)).

- The final controller to satisfy specifications on the steady-state error, cross-over frequency and margin of phase will look like:

  \[
  C(s) = R_1(s) \cdots R_q(s) \frac{k_C}{s^h}.
  \]
Loop Shaping: proportional control

- The effects of proportional control (i.e., a simple positive gain $k$) are to shift the magnitude plot of the loop transfer function up and down. The phase plot is not affected.
  - affects the crossover frequency, i.e., bandwidth of the closed loop system,
  - reduces the steady state error to ramps of orders equal to the system type (move above low-frequency Bode obstacle).

- If the system is open-loop stable, we know that small enough gains ($k \to 0$) yield stable closed-loops. (Remember root locus: closed loop poles are the same as open loop when $k = 0$)

- Increasing the gain too much however might cause not meet the other constraints (crossover/bandwidth, or command tracking/disturbance rejection), without compromising stability.
Loop shaping: Lead compensator

\[ C_{\text{lead}} = \frac{s/a + 1}{s/b + 1} = \frac{b}{a} \frac{s + a}{s + b}, \quad 0 < a < b \]

- **Lead compensator effects:**
  - Increase the magnitude at high frequencies, by \( b/a \); magnitude at low frequencies is not affected.
  - Increase the slope of the magnitude at frequencies between \( a \) and \( b \) by 20 dB/decade.
  - Increase the phase around \( \sqrt{ab} \) (i.e., the mid point between \( a \) and \( b \) on the Bode plot, by up to 90 degrees.

- A lead compensator approximates PD control as \( a \to 0 \) (and \( k/a \) remains finite).
Using a lead compensator

- The typical use of a lead compensator is to **increase the phase margin**.

  - Pick $\sqrt{ab}$ at the desired crossover frequency.

  - Pick $b/a$ depending on the desired phase increase (the larger $b/a$ the larger the phase increase, but no more than 90 degrees). Options on how to compute $a$, $b$:
    - by iteration (start with an educated guess, test, reiterate until satisfaction. Quick with computer aid.)
    - by using tables: see next slide

  - Adjust $k$ to put the crossover at the desired frequency.

- Possible side effect: increase magnitude at high frequencies, and hence noise sensitivity
Lead compensation example

- Plant$^1$: $P(s) = \frac{1}{s^2 + 2s + 2}$
- Desired bandwidth: $\omega_b^* = 5$ rad/s.
- Desired phase margin: $m_\phi^* > 45^\circ$.

⇒ $C(s) = \frac{52s + 2.5}{s + 10}$

Check phase margin at 5 rad/s ($20^\circ$)

$\sqrt{ab} = 5$

If $b/a = 10 \rightarrow \approx 90^\circ$, $b/a = 0 \rightarrow 0^\circ$, We need $+25^\circ$, ballpark $b/a = 4 \rightarrow R(s) = \frac{10}{2.5} \frac{s+2.5}{s+10}$.

Choose $k_C$ such that $\omega_b = \omega_b^*$.

$C(s) = k_C R(s)$.

---

$^1$Quick plotting tool: http://mathlets.org/mathlets/bode-and-nyquist-plots/
Lead compensator tables

- Think:
  \[ R_{lead}(s) = \frac{1 + \tau + s}{1 + \frac{1}{m+} s}, \quad m_+ > 1. \]
  \[ R_{lag}(s) = \frac{1 + \frac{1}{m-} s}{1 + \tau - s}, \quad m_- > 1 \]

- Pick \( m \) from the tables (e.g., \( m = 4 \)).

- Tables are expressed in \( \omega_N = \omega \tau \) scale.

- Let \( \omega = \omega_b^* \) (the desired crossover frequency), \( \omega_N = \omega_N^* \) (the \( \omega_N \) you pick on the tables based on design, e.g., \( \omega_N^* = 2 \) for \( +35^\circ \) with \( m = 4 \) network).

- Obtain \( \tau = \frac{\omega_N^*}{\omega_b^*} \).
Loop shaping: Lag compensator

\[ C_{\text{lag}} = \frac{s/a + 1}{s/b + 1} = \frac{b}{a} \frac{s + a}{s + b}, \quad 0 < b < a \]

- A lag compensator has the following main effects on the frequency response:
  - Decrease the magnitude at high frequencies, by \( b/a \); magnitude at low frequencies is not affected.
  - Decrease the slope of the magnitude at frequencies between \( a \) and \( b \), by 20 dB/decade.
  - Decrease the phase around \( \sqrt{ab} \) (i.e., the mid point between \( a \) and \( b \) on the Bode plot, by up to 90 degrees.

- A lag compensator approximates PI control as \( b \to 0 \) (and \( kb \) remains finite).
Lag compensator Bode plot

\[ C(s) = \frac{s/10+1}{s/0.1+1} \]
Using a lag compensator

- The typical use of a lag compensator is to **improve command tracking/disturbance rejection**
  - Pick $a/b$ as the desired increase in magnitude at low frequencies.
  - Pick $a$ so that it is sufficiently smaller than the crossover frequency, not to affect crossover frequency and phase margin.
  - Increase the gain $k$ by $a/b$

- Possible side effect: phase lag at low frequencies, potential reduction of phase margin.
Lag compensation example

- Plant: \( P(s) = 52 \frac{s + 2.5}{s + 10} \cdot \frac{1}{s^2 + 2s + 2} \)

- Previous specs + desired stead-state error to a unit step: less than 1%.

\[ \Rightarrow C(s) = \frac{s + 0.1}{s + 0.1/15.3} \]
A general procedure for open-loop stable systems

Proceed from “the left,” i.e., from low frequencies to high frequencies.

1. Figure out how many integrators are needed in $C(s)$. This depends on the order of the ramp that must be tracked with zero steady-state error.

2. Fix the gain in such a way that the low-frequency asymptote clears the command-tracking/disturbance specification (the low-frequency Bode obstacle).

3. Add terms of the form $(\tau s + 1)$ at the numerator or denominator in such a way that the Bode magnitude plot intersects the 0dB line with a slope of about 20 dB/s (this would give you a 90-degree phase margin. Poles steer “down”, zeros steer “up.” Note that normalizing the zero-order term to 1 makes it so that these do not affect the Bode plot on the left of the pole/zero.

4. Past the crossover, add poles as needed to clear the high-frequency Bode obstacle (noise reduction/uncertainty).
A general design example

Consider the system with transfer function

\[ P(s) = 0.01 \frac{(s + 10)^2}{s(s^2 + 2s + 2)}; \]

Design a feedback control system such that

- The closed-loop is stable;
- The steady-state error to a unit step is zero.
- Commands up to 0.05 rad/s are followed with at most 1% error.
- The phase margin is no less than 40 degrees;
- Noise at frequencies higher than 100 rad/s is attenuated by at least 40 dB.
Why does PID work so well?

- The form of an “implementable” PID is the following:

\[
PID(s) = k \frac{(s/z_1 + 1)(s/z_2 + 1)}{s(s/p + 1)},
\]

where \( p \gg 1 \) refers to an additional “fast” pole introduced to make \( PID(s) \) a proper transfer function.

- This can also be interpreted as follows:

\[
PID(s) = k \cdot \frac{s/z_1 + 1}{s + 0} \cdot \frac{s/z_2 + 1}{s/p + 1},
\]

i.e., a PID corresponds to an “extreme” lead-lag compensator, with one pole at \( s = 0 \), and one pole at \( -s = p \gg 1 \).
Loop shaping for non-minimum-phase/unstable systems

- Can we extend the techniques learned so far to systems with poles/zeros in the right half plane? We can — but please always double check with Nyquist and/or the root locus.

- That said... factor the plant transfer function as follows:

\[ P(s) = P_{mp}(s)D(s), \]

where

- \( P_{mp}(s) \) is obtained by replacing all poles/zeros of \( P(s) \) in the right half plane with their mirror image w.r.t. the imaginary axis.

- \( D(s) \) contains all the poles/zeros of \( P(s) \) in the right half plane, times the inverse all the mirror images introduced in \( P_{mp}(s) \). Note that \( |D(j\omega)| = 1 \) for all \( \omega \), i.e., \( D(s) \) is an all-pass filter. Choose the sign of \( D(s) \) so that the phase of \( D(j\omega) \) is negative.

- Example: if \( P(s) = \frac{s-z}{s-p} \), then \( P_{mp}(s) = \frac{s+z}{s+p} \), and \( D(s) = \frac{z-s}{s+z} \cdot \frac{s+p}{s-p} \).
Pretend that you have to stabilize a “nice” plant $P_{mp}(s)$, but a little “Bode demon” placed $D(s)$ inside your control system.

$D(s)$ does not alter the magnitude plot of $L(s)$, but messes up its phase plot—by introducing a possibly massive phase lag, and hence reducing the phase margin.
Loop shaping for non-minimum-phase system

\[ P(s) = \frac{1 - s}{s^2 + 2s + 2} \]

- We have that \( D(s) = -\frac{s - 1}{s + 1} \). Note that \( \angle D(0) = 0 \), \( \angle D(j) = -90^\circ \), and \( \lim_{\omega \to \infty} \angle D(j\omega) = -180^\circ \).

- NMP zeros limit the crossover frequency and the feedback gain.
Loop shaping for open-loop unstable system

\[ P(s) = \frac{1}{(s + 2)(s - 1)} \]

- We have that \( D(s) = \frac{s+1}{s-1} \). Note that \( \angle D(0) = -180^\circ \), \( \angle D(j) = -90^\circ \), and \( \lim_{\omega \to \infty} \angle D(j\omega) = 0^\circ \).

- Unstable open-loop poles force the crossover frequency to be large.
In this lecture, we covered the following material

- Specifications on the Bode plot;
- Loop-shaping concept for control design
- Use standard elements such as
  - Gain: increase crossover frequency, reduce steady-state error, etc.
  - Lead: increase the phase margin at crossover frequency.
  - Lag: decrease the steady-state error.
- ... or use poles/zeros to “steer” the Bode plot through the obstacle course.
- Non-minimum-phase zeros limit the crossover frequency (closed-loop bandwidth)
- Open loop unstable poles require the crossover frequency to be higher.