17 CONTROL SYSTEMS - LOOPSHAPING

17.1 Introduction

This section formalizes the notion of loopshaping for linear control system design. The loopshaping approach is inherently two-fold. First, we shape the open-loop transfer function (or matrix) $P(s)C(s)$, to meet performance and robustness specifications. Once this is done, then the compensator must be computed, from from knowing the nominal product $P(s)C(s)$, and the nominal plant $P(s)$.

Most of the analysis here is given for single-input, single-output systems, but the link to multivariable control is not too difficult. In particular, absolute values of transfer functions are replaced with the maximum singular values of transfer matrices. Design based on singular values is the idea of L_2-control, or LQG/LTR, to be presented in the next lectures.

17.2 Roots of Stability - Nyquist Criterion

We consider the SISO feedback system with reference trajectory $r(s)$ and plant output $y(s)$, as given previously. The tracking error signal is defined as $e(s) = r(s) - y(s)$, thus forming the negative feedback loop. The sensitivity function is written as

$$S(s) = \frac{e(s)}{r(s)} = \frac{1}{1 + P(s)C(s)},$$

where $P(s)$ represents the plant transfer function, and $C(s)$ the compensator. The closed-loop *characteristic equation*, whose roots are the poles of the closed-loop system, is $1 + P(s)C(s) = 0$, equivalent to $\underline{P}(s)\underline{C}(s) + \overline{P}(s)\overline{C}(s) = 0$, where the underline and overline denote the denominator and numerator, respectively. The Nyquist criterion allows us to assess the stability properties of a system based on $P(s)C(s)$ only. This method for design involves plotting the complex loci of $P(s)C(s)$ for the range $\omega = [-\infty, \infty]$. There is no explicit calculation of the closed-loop poles, and in this sense the design approach is quite different from the root-locus method (see Ogata).

17.2.1 Mapping Theorem

We impose a reasonable assumption from the outset: The number of poles in $P(s)C(s)$ exceeds the number of zeros. It is a reasonable constraint because otherwise the loop transfer function could pass signals with infinitely high frequency. In the case of a PID controller (two zeros) and a second-order zero-less plant, this constraint can be easily met by adding a high-frequency rolloff to the compensator, the equivalent of low-pass filtering the error signal.

Let $F(s) = 1 + P(s)C(s)$. The heart of the Nyquist analysis is the mapping theorem, which answers the following question: How do paths in the s-plane map into paths in the F-plane? We limit ourselves to *closed, clockwise* (CW) paths in the s-plane, and the remarkable result of the mapping theorem is

Every zero of $F(s)$ enclosed in the s-plane generates exactly one CW encirclement of the origin in the $F(s)$-plane. Conversely, every pole of $F(s)$ enclosed in the s-plane generates exactly one CCW encirclement of the origin in the $F(s)$-plane. Since CW and CCW encirclements of the origin may cancel, the relation is often written $Z - P = CW$.

The trick now is to make the trajectory in the s-plane enclose all unstable poles, i.e., the path encloses the entire right-half plane, moving up the imaginary axis, and then proceeding to the right at an arbitrarily large radius, back to the negative imaginary axis.

Since the zeros of $F(s)$ are in fact the poles of the closed-loop transfer function, e.g., $S(s)$, stability requires that there are *no* zeros of $F(s)$ in the right-half s-plane. This leads to a slightly shorter form of the above relation:

$$P = CCW. \tag{176}$$

In words, stability requires that the number of unstable poles in $F(s)$ is equal to the number of CCW encirclements of the origin, as s sweeps around the entire right-half s-plane.

17.2.2 Nyquist Criterion

The Nyquist criterion now follows from one translation. Namely, encirclements of the origin by $F(s)$ are equivalent to encirclements of the point $(-1+0j)$ by $F(s)-1$, or $P(s)C(s)$. Then the stability criterion can be cast in terms of the *unstable poles of* $P(s)C(s)$, *instead of those of* $F(s)$:

$$P = CCW \longleftrightarrow \text{closed-loop stability} \tag{177}$$

This is in fact the complete Nyquist criterion for stability. It is a necessary and sufficient condition that the number of unstable poles in the loop transfer function $P(s)C(s)$ must be matched by an equal number of CCW encirclements of the critical point $(-1+0j)$.

There are several details to keep in mind when making Nyquist plots:

> If neither the plant nor the controller have unstable modes, then the loci of $P(s)C(s)$ must not encircle the critical point at all.
>
> Because the path taken in the s-plane includes negative frequencies (i.e., the negative imaginary axis), the loci of $P(s)C(s)$ occur as complex conjugates - the plot is symmetric about the real axis.
>
> The requirement that the number of poles in $P(s)C(s)$ exceeds the number of zeros means that at high frequencies, $P(s)C(s)$ always decays such that the loci go to the origin.
>
> For the multivariable (MIMO) case, the procedure of looking at individual Nyquist plots for each element of a transfer matrix is unreliable and outdated. Referring to the multivariable definition of $S(s)$, we should count the encirclements for the function $[det(I + P(s)C(s)) - 1]$ instead of $P(s)C(s)$. The use of gain and phase margin in design is similar to the SISO case.

17.2.3 Robustness on the Nyquist Plot

The question of robustness in the presence of modelling errors is central to control system design. There are two natural measures of robustness for the Nyquist plot, each having a very clear graphical representation. The loci need to stay away from the critical point; how close the loci come to it can be expressed in terms of magnitude and angle.

> When the angle of $P(s)C(s)$ is -180°, the magnitude $|P(s)C(s)|$ should not be near one.
> When the magnitude $|P(s)C(s)| = 1$, its angle should not be -180°.

These notions lead to definition of the *gain margin* k_g and *phase margin* γ for a design. As the figure shows, the definition of k_g is different for stable and unstable $P(s)C(s)$. Rules of thumb are as follows. For a stable plant, $k_g \geq 2$ and $\gamma \geq 30^\circ$; for an unstable plant, $k_g \leq 0.5$ and $\gamma \geq 30^\circ$.

Stable P(s)C(s)   Unstable P(s)C(s)
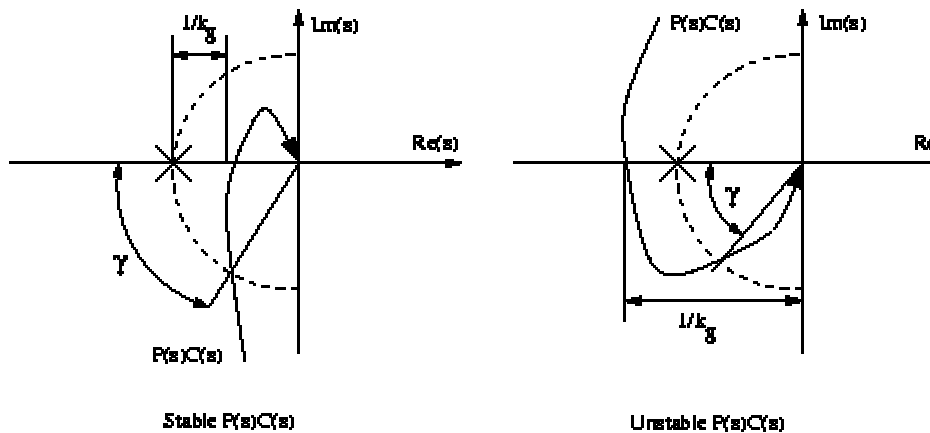
17.3 Design for Nominal Performance

Performance requirements of a feedback controller, using the nominal plant model, can be cast in terms of the Nyquist plot. We restrict the discussion to the scalar case in the following sections. Since the sensitivity function maps reference input $r(s)$ to tracking error $e(s)$, we know that $|S(s)|$ should be small at low frequencies. For example, if one-percent tracking is to be maintained for all frequencies below $\omega = \lambda$, then $|S(s)| < 0.01, \forall \omega < \lambda$. This can be formalized by writing

$$|W_1(s)S(s)| < 1, \tag{178}$$

where $W_1(s)$ is a stable weighting function of frequency. To force $S(s)$ to be small at low ω, $W_1(s)$ should be large in the same range. The requirement $|W_1(s)S(s)| < 1$ is equivalent to $|W_1(s)| < |1 + P(s)C(s)|$, and this latter condition can be interpreted as: The loci of $P(s)C(s)$ must stay outside the disk of radius $W_1(s)$, which is to be centered on the critical point $(-1 + 0j)$. The disk is to be quite large, possibly infinitely large, at the lower frequencies.

17.4 Design for Robustness

It is a ubiquitous observation that models of plants degrade with increasing frequency. For example, the DC gain and slow, lightly-damped modes or zeros are easy to observe, but higher-frequency components in the response may be hard to capture or even excite repeatably. Higher-frequency behavior may have more nonlinear properties as well.

The effects of modeling uncertainty can be considered to enter the nominal feedback system as a disturbance at the plant output, d_y. One of the most useful descriptions of model uncertainty is the multiplicative uncertainty:

$$\tilde{P}(s) = (1 + \Delta(s)W_2(s))P(s). \tag{179}$$

Here, $P(s)$ represents the nominal plant model used in the design of the control loop, and $\tilde{P}(s)$ is the actual, perturbed plant. The perturbation is of the multiplicative type, $\Delta(s)W_2(s)P(s)$, where $\Delta(s)$ is an *unknown but stable* function of frequency for which $|\Delta(s)| \leq 1$. The weighting function $W_2(s)$ scales $\Delta(s)$ with frequency; $W_2(s)$ should be growing with increasing frequency, since the uncertainty grows. However, $W_2(s)$ should not grow any faster than necessary, since it will turn out to be at the cost of nominal performance.

In the scalar case, the weight can be estimated as follows: since $\tilde{P}/P - 1 = \Delta W_2$, it will suffice to let $|\tilde{P}/P - 1| < |W_2|$.

Example: Let $\tilde{P} = k/(s - 1)$, where k is in the range 2-5. We need to create a nominal model $P = k_0/(s - 1)$, with the smallest possible value of W_2, which will not vary with frequency in this case. Two equations can be written using the above estimate, for the two extreme values of k, yielding $k_0 = 7/2$, and $W_2 = 3/7$.

For constructing the Nyquist plot, we observe that $\tilde{P}(s)C(s) = (1 + \Delta(s)W_2(s))P(s)C(s)$. The path of the perturbed plant could be anywhere on a disk of radius $|W_2(s)P(s)C(s)|$, centered on the nominal loci $P(s)C(s)$. The robustness condition is that this disk should not intersect the critical point. This can be written as

$$|1 + PC| > |W_2 PC| \longleftrightarrow$$
$$1 > \frac{|W_2 PC|}{|1 + PC|} \longleftrightarrow$$
$$1 > |W_2 T|, \tag{180}$$

where T is the complementary sensitivity function. The last inequality is thus a condition for robust stability in the presence of multiplicative uncertainty parametrized with W_2.

17.5 Robust Performance

The condition for good performance with plant uncertainty is a combination of the above two conditions. Graphically, the disk at the critical point, with radius $|W_1|$, should not intersect the disk of radius $|W_2 PC|$, centered on the nominal locus PC. This is met if

$$|W_1 S| + |W_2 T| < 1. \tag{181}$$

The robust performance requirement is related to the magnitude $|PC|$ at different frequencies, as follows:

1. At low frequency, $|W_1 S| \simeq |W_1 / PC|$, since $|PC|$ is large. This leads directly to the performance condition $|PC| > |W_1|$ in this range.
2. At high frequency, $|W_2 T| \simeq |W_2 PC|$, since $|PC|$ is small. We must therefore have $|PC| < 1/|W_2|$, for robustness.

17.6 Implications of Bode's Integral

The loop transfer function PC cannot roll off too rapidly in the crossover region. The simple reason is that a steep slope induces a large phase loss, which in turn degrades the phase margin. To see this requires a short foray into Bode's integral. For a transfer function $H(s)$, the crucial relation is

$$angle(H(j\omega_0)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{d}{d\nu}(\ln|H(j\omega)|) \cdot \ln(\coth(|\nu|/2)) d\nu, \qquad (182)$$

where $\nu = \ln(\omega/\omega_0)$. The integral is hence taken over the log of a frequency normalized with ω_0. It is not hard to see how the integral controls the angle: the function $\ln(\coth(|\nu|/2))$ is nonzero only near $\nu = 0$, implying that the angle depends only on the local slope $d(\ln|H|)/d\nu$. Thus, if the slope is large, the angle is large.

Example: Suppose $H(s) = \omega_0^n / s^n$, i.e., it is a simple function with n poles at the origin, and no zeros; ω_0 is a fixed constant. It follows that $|H| = \omega_0^n/\omega^n$, and $\ln|H| = -n\ln(\omega/\omega_0)$, so that $d(\ln|H|)/d\nu = -n$. Then we have just

$$angle(H) = -\frac{n}{\pi} \int_{-\infty}^{\infty} \ln(\coth(|\nu|/2)) d\nu = -\frac{n\pi}{2}. \qquad (183)$$

This integral is trivial to look up or compute. Each pole at the origin clearly induces $90°$ of phase loss. In the general case, each pole not at the origin induces $90°$ of phase loss for frequencies above the pole. Each zero at the origin adds $90°$ phase lead, while zeros not at the origin at $90°$ of phase lead for frequencies above the zero. In the immediate neighborhood of these poles and zeros, the phase may vary significantly with frequency.

The Nyquist loci are clearly susceptible to these variations is phase, and the phase margin can be easily lost if the slope of PC at crossover (where the magnitude is unity) is too steep. The slope can safely be first-order ($-20dB/decade$, equivalent to a single pole), and may be second-order ($-40dB/decade$) if an adequate phase angle can be maintained near crossover.

17.7 The Recipe for Loopshaping

In the above analysis, we have extensively described what the open loop transfer function PC should look like, to meet robustness and performance specifications. We have said very little about how to get the compensator C, the critical component. For clarity, let the designed loop transfer function be renamed, $L = PC$. We will use concepts from optimal linear control for the MIMO case, but in the scalar case, it suffices to just pick

$$C = L/P. \qquad (184)$$

This extraordinarily simple step involves a plant inversion.

The overall idea is to first shape L as a stable transfer function meeting the requirements of stability and robustness, and then divide through by the plant.

When the plant is stable and has stable zeros (minimum-phase), the division can be made directly.

One caveat for the well-behaved plant is that lightly-damped poles or zeros should not be cancelled verbatim by the compensator, because the closed-loop response will be sensitive to any slight change in the resonant frequency. The usual procedure is to widen the notch or pole in the compensator, through a higher damping ratio.

Non-minimum phase or unstable behavior in the plant can usually be handled by performing the loopshaping for the closest stable model, and then explicitly considering the effects of adding the unstable parts. In the case of unstable zeros, we find that they impose

an unavoidable frequency limit for the crossover. In general, the troublesome zeros must be *faster* than the closed-loop frequency response.

In the case of unstable poles, the converse is true: The feedback system must be faster than the corresponding frequency of the unstable mode.

When a control system involves multiple inputs and outputs, the ideas from scalar loopshaping can be adapted using the singular value. We list below some basic properties of the singular value decomposition, which is analogous to an eigenvector, or modal, analysis. Useful properties and relations for the singular value are found in the section *MATH FACTS*.
The condition for MIMO robust performance can be written in many ways, including a direct extension of our scalar condition

$$\bar{\sigma}(W_1 S) + \bar{\sigma}(W_2 T) < 1. \tag{185}$$

The open-loop transfer matrix L should be shaped accordingly. In the following sections, we use the properties of optimal state estimation and control to perform the plant inversion for MIMO systems.