

# Notes on Frequency Methods: Gain Margin Phase Margin, Delays, and the Nyquist Map

Dr. Daniel S. Stutts  
Missouri University of Science and Technology  
May 3, 1999<sup>1</sup>

## 1 Introduction to Frequency Methods

Using frequency methods, it is possible to determine a great deal of information from the open-loop transfer function. One of the most important facts about a given system which may be determined via frequency methods is the relative stability of the system. There are two principal measures of system stability determined via frequency methods: *Gain Margin*, and *Phase Margin*. These two measures may be read from Nyquist plots or Bode plots, both of which may be easily constructed using MATLAB. Below, we will first discuss the Nyquist criteria, and work an example, then we will apply MATLAB's Bode plotting capabilities to obtain the same information.

## 2 Nyquist Criteria

The stability of a system may be determined graphically from the open-loop transfer function in yet another way besides the root locus. The open-loop TF,  $F(s) = G(s)H(s)$ , is a function of the complex variable,  $s$ , and may be thought to map a collection of points forming a curve from the  $s$ -plane to the  $F(s)$ -plane. A corresponding curve will be formed in the  $F(s)$ -plane. This concept is illustrated in Figure 1.

The fact that another continuous curve results may be argued based on the continuity of  $F(s)$ . Essentially, two non-singular points (points other than the poles of  $F(s)$ ) in the complex plane, which are within a distance  $\delta$  of one another will map to two new points through the function  $F(s)$  which are within a distance  $\varepsilon$  of each other. This is illustrated in Figure 2.

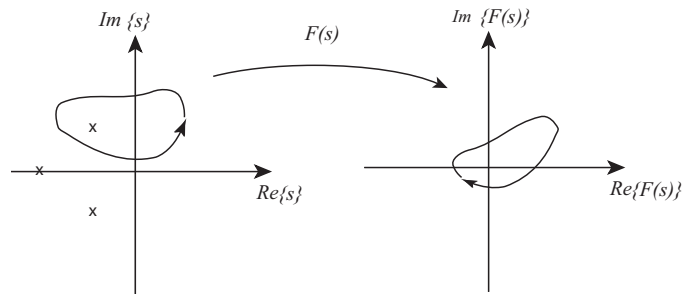


Figure 1. Mapping in the complex plane.

---

<sup>1</sup>Revised December 2, 2005, Second revision December 1, 2008

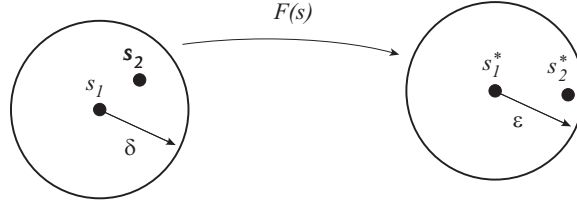


Figure 2. Continuity of curves through a mapping  $F(s)$ .

Hence any two non-singular points in the complex plane (denoted by  $\mathcal{C}$ ) which are within a  $\delta$ -neighborhood of each other will map through a continuous function,  $F(s)$  to the  $F(s)$ -plane to lie within an  $\varepsilon$ -neighborhood of one another. Mathematically:  $F(s)$  is said to be a continuous function, if  $\forall \{s_1, s_2\} \in \mathcal{C}$ , and which are not singular points of  $F(s)$ ,  $\exists \delta$ , and  $\varepsilon$  such that  $|s_2 - s_1| \leq \delta \Rightarrow |F(s_2) - F(s_1)| \leq \varepsilon$ .

Some additional preliminaries are necessary before discussing the Nyquist criteria. First, letting  $G(s) = \frac{N_G}{D_G}$ , and  $H(s) = \frac{N_H}{D_H}$ , so that the open-loop transfer function is given by  $G(s)H(s) = \frac{N_G N_H}{D_G D_H}$ , and the characteristic equation is given by  $1 + G(s)H(s) = 1 + \frac{N_G N_H}{D_G D_H} = \frac{D_G D_H + N_G N_H}{D_G D_H}$ .

Two important facts to note from this exercise are:

1. The poles of  $G(s)H(s)$  (the open-loop poles) are known, and are the poles of  $1 + G(s)H(s)$ .
2. The zeros of  $1 + G(s)H(s)$ , which are not known in general, are the roots of the characteristic equation, and hence, the closed-loop poles.

## 2.1 Geometry of resulting maps

The orientation of the resulting curve in the  $F(s)$ -plane yields a great deal of information about the mapping function,  $F(s)$ . The following are important concepts regarding the orientation of the mapping:

1. If the path of the curve in the  $s$ -plane,  $\Gamma_s$ , encircles a *zero* of  $F(s)$ , then the resulting path in the  $F(s)$ -plane,  $\Gamma_{F(s)}$  will encircle the origin, and preserve the orientation of  $\Gamma_s$  – i.e. clockwise, (CW) or counterclockwise (CCW).
2. If  $\Gamma_s$  encircles a *pole* of  $F(s)$ , then the resulting path,  $\Gamma_{F(s)}$  will encircle the origin, but in the opposite sense of  $\Gamma_s$  – i.e. clockwise, (CW) or counterclockwise (CCW).
3. If  $\Gamma_s$  encircles neither a pole nor a zero of  $F(s)$ , then the resulting path in the  $F(s)$ -plane,  $\Gamma_{F(s)}$ , will not encircle the origin.
4. If  $\Gamma_s$  encircles the same number of poles as zeros, then the net phase change of  $F(s)$  will be zero, and  $\Gamma_{F(s)}$  will not encircle the origin in the  $F(s)$ -plane.

The phase of  $F(s)$  at a given point on  $\Gamma_s$  is determined by the principle of the *Argument* of  $F(s)$ ,  $\angle F(s)$ . That is, the sum of the phase contributions of each vector from the poles and zeros of  $F(s)$  to the curve,  $\Gamma_s$ , contribute to the phase of  $F(s)$  as  $\Gamma_s$  is mapped to  $\Gamma_{F(s)}$ . The angle between the vector  $R$  to  $\Gamma_{F(s)}$  with respect to the real axis is given by  $\Phi = -\phi_1 - \phi_2 - \phi_3$ . This process is illustrated in Figure 3 and in Figure 4, where in case (a) the system is given by  $F(s) = K(s + z_1)$ , and in (b) by  $F(s) = \frac{K}{s + p_1}$ .

Figure 5 illustrates the mapping of a circle of 21 points and unit radius around the pole at  $s = -2$  in the  $s$ -plane to  $F(s) = 2/(s + 2)$ . The the mapped curve encircles the origin in the opposite direction due to the negative phase of  $F(s)$ . The Maple (version 12) code to produce these plots is given in Appendix B.

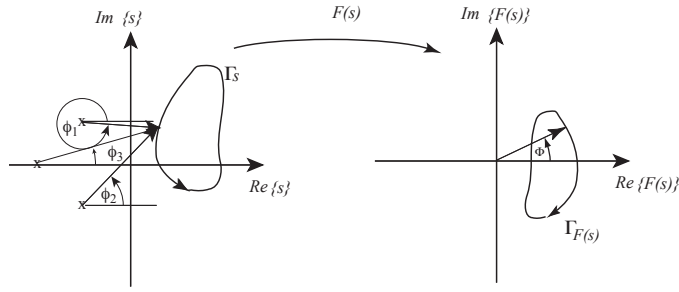


Figure 3. Mapping of  $\Gamma_s$  to  $\Gamma_{F(s)}$  for a system with three poles.

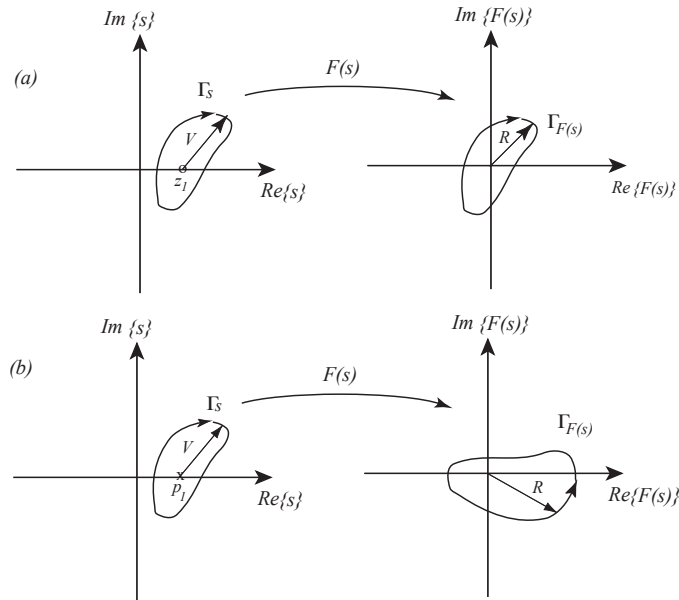


Figure 4. Mapping of  $F(s)$  with a single zero (a) and a single pole (b) encircled.

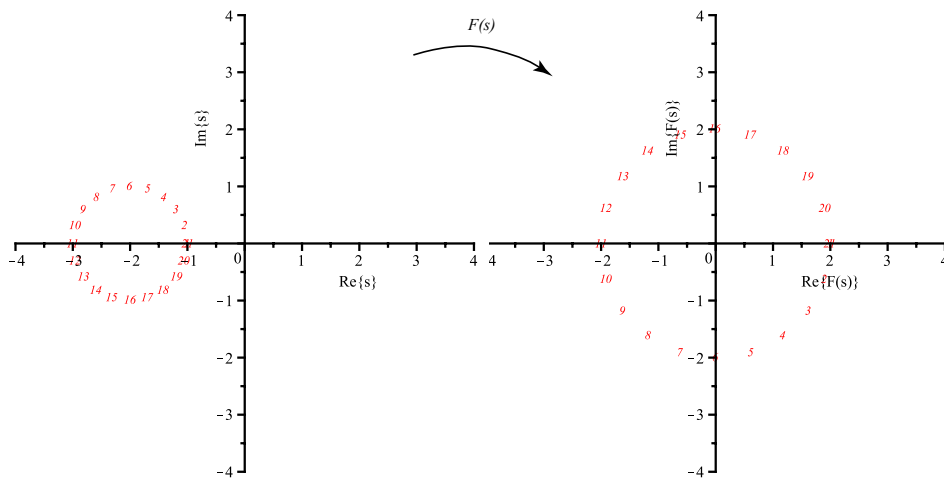


Figure 5. Mapping of  $s$  and  $F(s)$  for  $F(s) = 2/(s+2)$ .

## 2.2 Detection of closed-loop poles in the right half-plane

If we let  $F(s) = 1 + G(s)H(s)$ , and then let  $\Gamma_s$  encircle the entire right half plane in the CW direction without passing through any poles of  $F(s)$ , any CW encirclements of the origin by the mapping  $\Gamma_{F(s)}$  will correspond to the zeros of  $F(s)$ , which are also the system closed-loop poles in the right-half plane – provided there are no open-loop poles or zeros in the right half-plane. Systems with no open-loop poles in the right half-plane are termed *minimum phase systems*. Minimum phase systems are very common, and are amenable to a simplified Nyquist analysis detailed in the next section.

Since the function  $F(s) = 1 + G(s)H(s)$  is merely the open-loop transfer function shifted, we really need only map through  $F(s) = G(s)H(s)$ , and count the encirclements of the point  $s = -1$ . This is how the Nyquist sketch is made in practice.

If  $H(s)G(s)$  contains finite poles in the right half of the complex plane, then the evaluation of the Nyquist mapping becomes slightly more complicated. Letting  $N$  denote the number of encirclements of  $s = -1$ ,  $Z$  denote the number of zeros of  $1 + H(s)G(s)$ , closed-loop poles, in the right half-plane, and  $P$  denote the number of finite open-loop poles in the right half plane. We have

$$N = Z - P. \quad (1)$$

Hence, if  $P \neq 0$ , stability ( $Z = 0$ ) requires that

$$N = -P, \quad (2)$$

which indicates  $P$  CCW encirclements of  $s = -1$ .

It is important to note that while finite open-loop zeros of  $H(s)G(s)$  may correspond to CW encirclements of the origin, it is only CW encirclements of the critical point,  $s = -1$ , that matter in the determination of stability, since these correspond to the zeros of  $1 + H(s)G(s)$ . In addition, Equation (1) may be used not only to determine stability, but also the number of closed-loop poles in the right half-plane.

The Nyquist mapping also reveals the concept of *Gain and Phase Margin*, whereby the *degree* of stability or instability may be quantified. These concepts apply in general to any system, but will be detailed in the following section on so-called minimum phase systems.

## 3 Simplified Nyquist Criteria

Consider the following example:

$$G(s) = \frac{K}{s(s+4)(s+6)}, \quad (3)$$

so that

$$T(s) = \frac{K}{s^3 + 10s^2 + 24s + K}. \quad (4)$$

This system is of the minimal phase variety, having no finite open-loop poles or zeros in the right half-plane.

The Routh-Hurwitz criteria predicts stability for  $0 \leq K \leq 240$ , so the critical value of gain is  $K = 240$ . The Nyquist map is constructed for three cases: (1) one-half of critical gain, (2) critical gain, and (3) twice critical gain, and illustrated for each case in Figure 6. These plots were made using the MATLAB function, `nyquist.m` which plots the Nyquist plot from  $\omega = 0$  to  $\omega = +\infty$ , and then simply reflects the

Table 1. Routh Table

$s^3$	1	24
$s^2$	10	$K$
$s$	$\frac{240-K}{10}$	0
$s^0$	$K$	

resulting plot into the upper half of the complex plane to represent the plot from  $\omega = -\infty$  to  $\omega = 0$ . The Nyquist plot will be symmetric if  $\Gamma_s$  is symmetric.

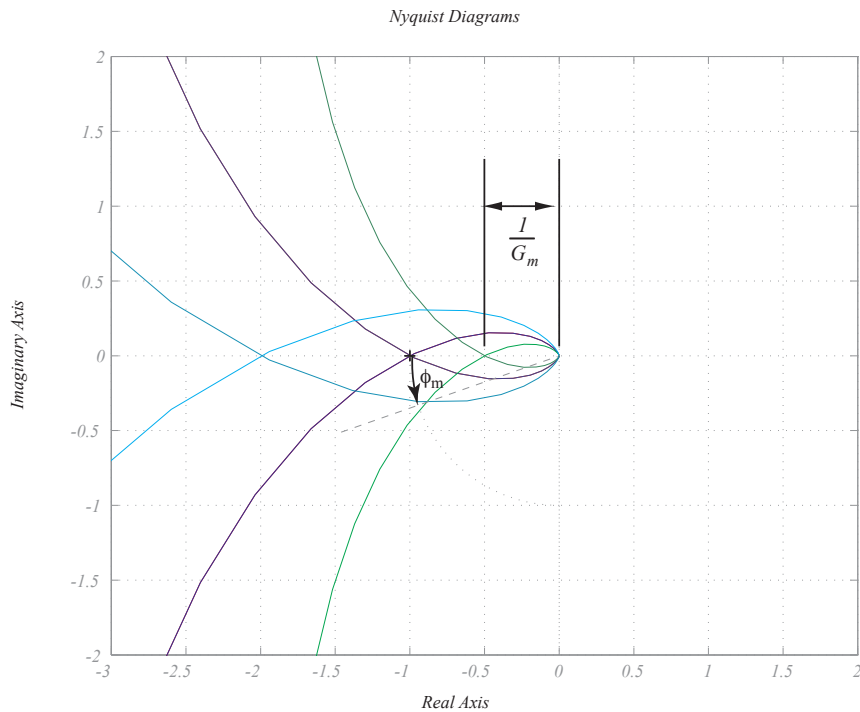


Figure 6. Nyquist plots for  $K = 120, 240$  and  $480$ .

The gain and phase margin is shown for the stable,  $K = 120$  case, but may also be defined for the unstable case as well. In the unstable case, we have negative gain and phase margins. The gain margin represents the amount of gain required to render the system marginally stable. For unstable systems, the negative gain margin, usually expressed in decibels (dB), indicates the amount by which the gain must be divided to reach marginal stability. The phase margin is defined as the angle between a line drawn from the origin through the crossing of the unit circle by the Nyquist mapping,  $\Gamma_{F(s)}$ . The phase margin may also be negative, and is usually expressed in degrees.

### 3.1 Sketching the Nyquist plot by hand

It is possible, with practice, to rapidly sketch the Nyquist mapping. The contour,  $\Gamma_s$  must be chosen so that it does not pass through any of the poles of  $H(s)G(s)$ . This presents no difficulties in principle, but care must be taken to take the contour to the right of any purely imaginary poles. Hence, each  $j\omega$ -axis



point,  $s = \pm\sqrt{24}$ , corresponds to the to the  $j\omega$ -axis crossing for this system on the root locus.

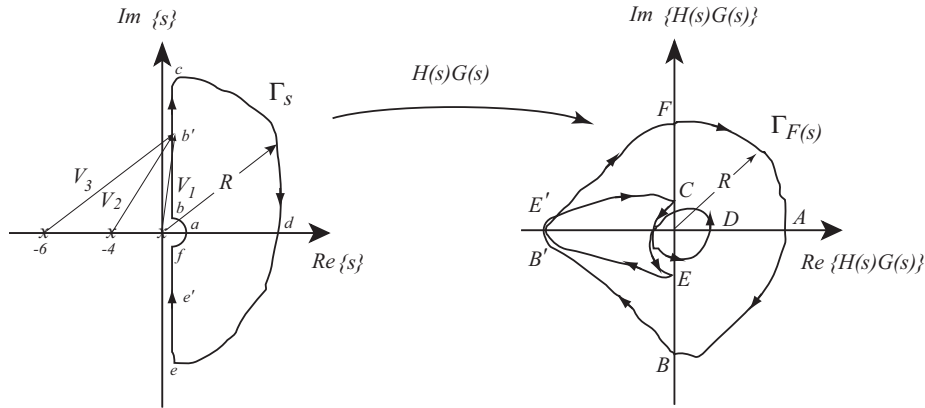


Figure 8. Illustration of the Nyquist mapping process.

It should be noted that the MATLAB function, `nyquist.m`, only plots the portion of the plot between points  $b$  and  $c$ , and then reflects the resulting plot into the upper half of the complex plane due to symmetry. Hence, the resulting contour,  $\Gamma_{F(s)}$ , produced by MATLAB consists of the segment  $BC$  and its reflection. The numerically produced plot, one of those shown in Figure 6, is obviously much easier to obtain, and much cleaner!

### 3.2 Gain and phase margin from Bode plots

The information in the polar Nyquist plot may also be found in the Bode Plot, where magnitude and phase are separated. The same example given in the previous section is displayed in Figures 9 through 11, which were obtained using the MATLAB version 5.0 (and, more recently, via version 7.01 in the case of Figure 11) command, `margin.m`.

The gain and phase margins are given in terms of how much either may be increased before instability results. When these metrics are negative, the system is already unstable, and the values given represent the amount by which either must be reduced to achieve stability. For the gain margin, since the metric are given in dB, a negative gain margin implies that the gain must be divided by a factor,  $G_m$ , where  $G_m$  is related to the gain margin in decibels as shown in Equation (7).

$$G_m = 10^{\frac{G_{m\text{dB}}}{20}} \quad (7)$$

The gain margin is determined by locating the phase crossing of the  $-180^\circ$ -line, and drawing a vertical line up to the corresponding magnitude plot. The distance between the 0-dB line and the magnitude plot is the gain margin. The frequency at which this crossing occurs is called the gain margin frequency, or  $\omega_G$  ( $\omega_{cg}$  in MATLAB).

The the phase margin is determined in a similar way. A vertical line is drawn down from the 0-dB crossing at  $\omega_\Phi$  ( $\omega_{cp}$  in MATLAB) of the magnitude plot through the corresponding phase plot. The distance between the phase plot and the  $-180^\circ$ -line is the phase margin.

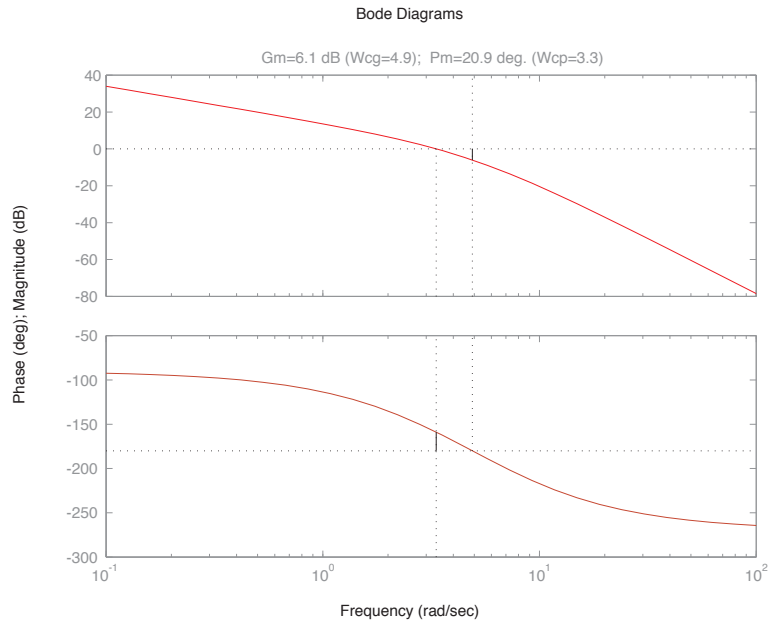


Figure 9. Bode magnitude and phase plots for  $K = 120$ .

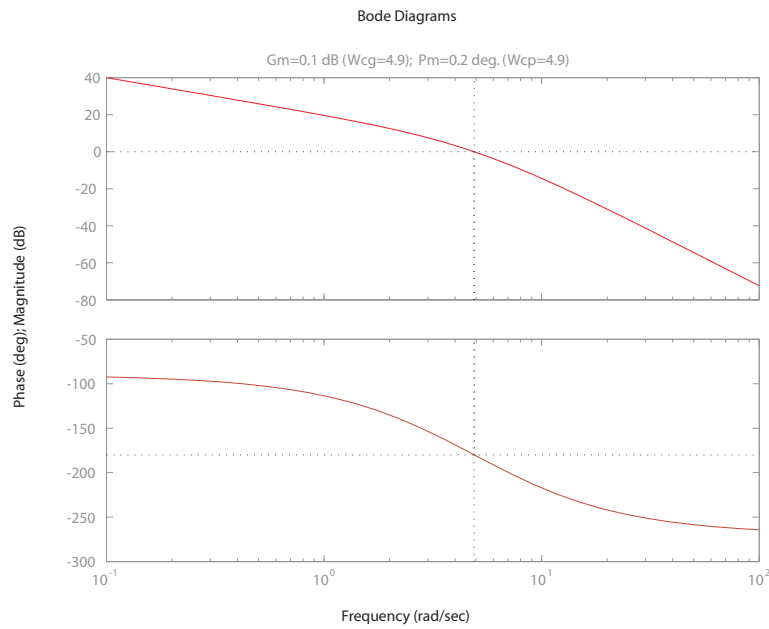


Figure 10. Bode magnitude and phase plots for  $K = 240$ .



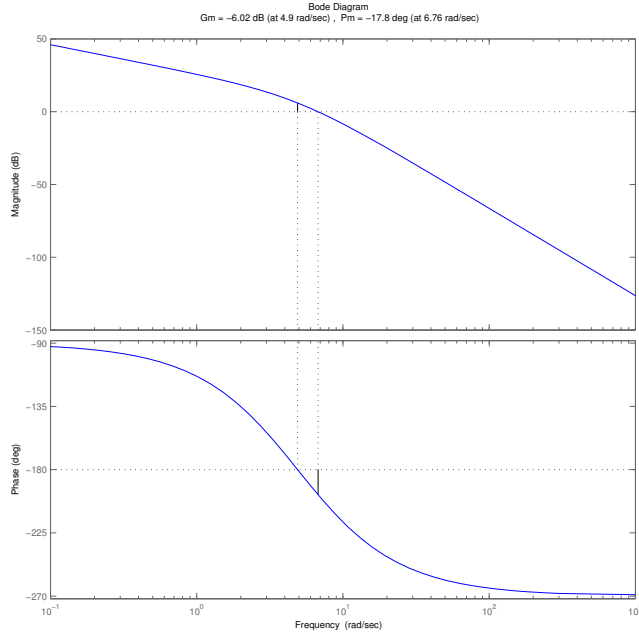


Figure 11. Bode magnitude and phase plots for  $K = 480$ .

#### 4 Nyquist Plots of Systems with Delays

The simplified Nyquist criteria may be used to graphically determine the  $G_m$  and  $\Phi_m$  for systems with delays. Such systems have an open-loop transfer function of the form

$$F(s) = KG(s)H(s)e^{-sT} \quad (8)$$

where  $T$  denotes the time delay in seconds between the input and the measured response. Because the exponential function is irrational, the Routh-Hurwitz method will not work, nor is it possible to plot the root-locus. Fortunately, the simplified Nyquist method incorporating the delay is easily implemented in MATLAB. For  $s = j\omega$ , the delay function,  $e^{-sT}$  is a pure clockwise rotation of the points mapped to through  $KG(j\omega)H(j\omega)$ . Hence, a stable system such as the system given by (1) with  $K = 120$  will eventually become unstable given sufficient response lag,  $T$ . The delay has the effect of rotating the resultant spiral obtained by plotting the imaginary versus real parts of  $KG(j\omega)$ , so that it tends to encircle the  $s = -1$  point. This effect is illustrated in Figure 12.

It can be seen that the system is just barely stable with a delay of  $T = 0.1$  seconds. The MATLAB function which was used to produce Figure 12 is given in the MATLAB M-Files section.

#### 5 Conditions for Marginal Stability for Systems with Delays: A Second-Order Example

A relationship between the delay, gain, and forcing frequency may be obtained by expanding Equation (8) evaluated at  $s = j\omega$ . Setting

$$F(j\omega) = KG(j\omega)H(j\omega)e^{-j\omega T} = KG(j\omega)H(j\omega) (\cos \omega T - j \sin \omega T) \quad (9)$$

yields the two criteria for instability

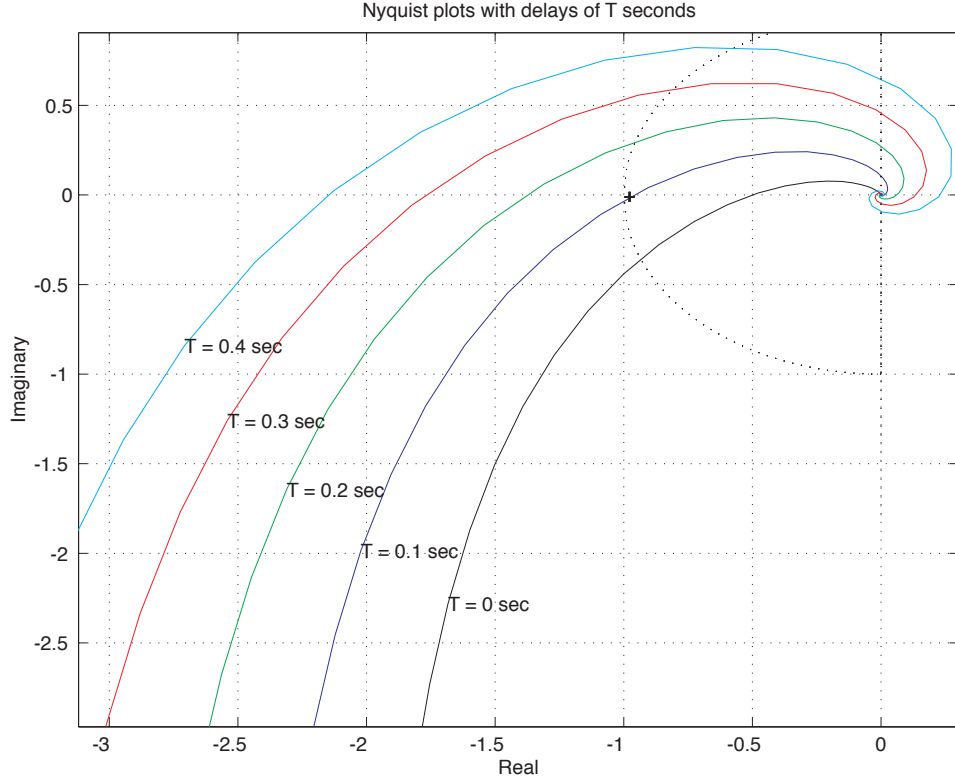


Figure 12. Nyquist plot for  $K = 120$ , but with various delays denoted by  $T$ .

$$\text{Im} \{F(j\omega)\} = 0 \quad (10)$$

and

$$|\text{Re} \{F(j\omega)\}| > 1 \quad (11)$$

For example, consider a second order system with open-loop transfer function

$$G(s) = \frac{K e^{-sT}}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (12)$$

at  $s = j\omega$ , Equations (10) and (11) yield the following equations at marginal stability

$$\tan \omega T = \frac{2\zeta\omega_n}{\omega} \quad (13)$$

and

$$K = \omega \sqrt{\omega^2 + 4\zeta^2\omega_n^2} \quad (14)$$

Equations (13) and (14) supply the necessary, and combined, the sufficient conditions for marginal stability. Hence, Equations (13) and (14) may be recast as

$$\omega = \sqrt{\sqrt{4\zeta^4\omega_n^4 + K^2} - 2\zeta^2\omega_n^2} \quad (15)$$

and

$$\omega = \frac{1}{T} \arctan \left( \frac{2\zeta\omega_n}{\omega} \right) \quad (16)$$

and combined to yield the relationships between gain,  $K$ , forcing frequency,  $\omega$ , and delay,  $T$  for marginal stability in a second order unity feedback system:

$$\omega = \frac{2\zeta\omega_n}{\arctan\left(T\sqrt{\sqrt{4\zeta^4\omega_n^4 + K^2} - 2\zeta^2\omega_n^2}\right)} \quad (17)$$

or

$$T = \frac{\tan\left(\frac{2\zeta\omega_n}{\omega}\right)}{\sqrt{\sqrt{4\zeta^4\omega_n^4 + K^2} - 2\zeta^2\omega_n^2}} \quad (18)$$

or

$$K = \frac{\arctan\left(\frac{2\zeta\omega_n}{\omega}\right)}{T^2} \sqrt{\arctan^2\left(\frac{2\zeta\omega_n}{\omega}\right) + 4\zeta^2\omega_n^2 T^2} \quad (19)$$

Equations (17), through (19) establish the upper bound on forcing frequency, gain, and delay for second-order system stability in terms of the system parameters and each other. Using Equations (17), through (19), limiting cases may be easily investigated. For example, for  $\omega \gg 2\zeta\omega_n$ , we have, from (19)

$$K \approx \frac{4\zeta^2\omega_n^2}{\omega T} \quad (20)$$

## 6 Appendix A: MATLAB M-Files

```
% Function: nyqdel.m
%
% Author: Dr. D. S. Stutts
%         Associate Professor of Mechanical Engineering
%         and Engineering Mechanics
%         University of Missouri-Rolla
%
% Version: 1.0b 4-26-99
%
%
% Purpose: plots the Nyquist graph of a system with a delay of T seconds
%         defined by the vectors num for numerator, and den for
%         denominator polynomials describing the open-loop transfer
%         function. Useful for systems with no open-loop poles
%         or zeros in the right half of the complex plane.
%
% num: numerator polynomial coefficients.
% den: denominator polynomial coefficients.
% T: Time delay in seconds.
% omegaMax: maximum frequency in radians per second.
% mxpts: Number of points to plot.
% color: character for desired color of plot.
%        Must be inclosed in single quotes, 'b' for blue, for
%        example. See 'plot' for color options.
%
function niq(num,den,T,Maxomega,mxpts,color)
label=num2str(T)
```

```

w=logspace(0,Maxomega,mxpts);
n=length(w)
for i=1:n,
g(i)=exp(-T*j*w(i))*polyval(num,j*w(i))/polyval(den,j*w(i));
end
maxre=max(real(g))
minre=min(real(g))
maxim=max(imag(g))
minim=min(imag(g))
op1=strcat(color,'-')
legend=strcat('T = ',label,' sec')
plot(g,op1)
axis([minre-.1*abs(minre) maxre+.1*abs(maxre) minim-.1*abs(minim) maxim+.1*abs(maxim)])
grid on
sgrid(0,1)
text(-1,0,'+')
text(real(g(5)),imag(g(5)),legend)
xlabel('Real')
ylabel('Imaginary')
title('Nyquist plot with delay of T seconds')

```

## 7 Appendix B: Maple 10 Mapping Code

In the following code all output is suppressed via the colon to save space.

```

#
# The following is a short example of mapping in the complex plane.
# It illustrates the geometric effect of pole factors on the
# orientation of the resulting curve.
#
# Note that this text should be cut and pasted into
# a "Code Edit" region in Maple (Version 12+),
# and then executed via the right mouse button options.
#
# DSS: 11/21/05, Revised 12-1-08:DSS
#
# Define mapping function:
#
F := unapply(2/(z+2), z):
# Define mapping to construct a circle of unit radius #around the pole at s = -2:
x := unapply(-2+cos(theta), theta):
y := unapply(sin(theta), theta):
s := unapply(x(theta)+I*y(theta), theta):
plot([Re(s(theta)), Im(s(theta)), theta = 0 .. 2*Pi], -4 .. 4, -4 .. 4,
  labels = ["Re{s}", "Im{s}"], labeldirections = [horizontal, vertical]);
# Generate an array of 21 angles around the circle at which to plot a point:
thetapoints := [seq(1/10*(i-1)*Pi, i = 1 .. 21)]:
#Load the plots library
with(plots):

```

```

# Calculate the points on the circle in the s-plane around the pole:
spoints := [seq([Re(s(thetapoints[i])), Im(s(thetapoints[i]))], i = 1 .. 21)]:
numbers := [seq(i, i = 1 .. 21)]:
# Define axis labels
labelxs:= "Re{s}":
labelys:= "Im{s}":
labelxF := "Re{F(s)}":
labelyF := "Im{F(s)}":
# Plot the points and note the order in the counter clockwise or + direction.
# Note also that I've appended the last points [-4, -4, " "], and [4, 4, " "]
# to trick the textplot function into scaling the plots the way I wanted them.
textplot([seq([Re(s(thetapoints[i])), Im(s(thetapoints[i]))],
numbers[i]], i = 1 .. 21),[-4, -4, " "], [4, 4, " "]], labels = [labelxs, labelys],
labeldirections = [horizontal, vertical], font = [times, italic, 8], color=red);
# Create a list of mapped points and their respective label
# numbers to plot using textplot():
Ftext := [seq([Re(F(s(thetapoints[i]))), Im(F(s(thetapoints[i]))),
numbers[i]], i = 1 .. 21),[-4, -4, " "], [4, 4, " "]]:
# Plot the resulting mapping and note that the orientation of the
# encirclement of the origin is in the clockwise
# or negative direction:
textplot(Ftext, labels = [labelxF, labelyF], labeldirections = [horizontal, vertical],
font = [times, italic, 8], color=red);

```