APPPHYS217 Thursday 6 May 2010

A bit more on Nyquist plots and stability

Let's start by recalling the scenario of interest. In a standard unity-feedback setup we have a given plant transfer function $P(s)$ and controller transfer function $C(s)$. If we know how to write these as rational functions,

$$P(s) = \frac{n_p}{d_p}, \quad C(s) = \frac{n_c}{d_c},$$

then it is simple to read off their poles (roots of d_p, d_c) and zeros (roots of n_p, n_c). Then for the 'loop transfer function'

$$L(s) \equiv C(s)P(s) = \frac{n_p n_c}{d_p d_c},$$

we know that the set of poles of $L(s)$ is the union of the sets of poles of $P(s)$ and $C(s)$, likewise the set of zeros of $L(s)$ is the union of the sets of zeros of $P(s)$ and $C(s)$. This follows simply from the fact that series interconnection corresponds to multiplication of transfer functions.

What we want to know for feedback analysis and design purposes is how to predict the stability of the closed-loop transfer function

$$H(s) = \frac{L(s)}{1 + L(s)} = \frac{n_p n_c}{d_p d_c + n_p n_c}.$$

Without the use of a computer algebra package, it generally isn't easy to determine the roots of $d_p d_c + n_p n_c$. Luckily, we can often proceed with analysis/design by making use of the Nyquist theorem,

$$Z = N + P,$$

where Z is the number of right half-plane zeros of $1 + L(s)$ (that is, zeros of $1 + L(s)$ that have positive real part), P is the number of right half-plane poles of $L(s)$, and N is the number of clockwise encirclements of the -1 point on a Nyquist plot of $L(s)$. In order for the closed-loop system to be stable we want $Z = 0$, and we usually know P from factorizations of d_p and d_c.

Last time we described two practical strategies for generating a Nyquist plot of $L(s)$. One is to use the MATLAB routine; the other is to sketch the Nyquist plot from a Bode plot of $L(s)$. To start today's lecture we'll talk more about the sketching method, and review relationships among the frequency response, Nyquist plot and transfer function. We'll discuss some subtleties along the way that were only mentioned in passing last time.

The Nyquist plot is a plot of $L(s)$ as s follows the Nyquist "D" curve, which runs up along the imaginary axis (from negative $j\omega$ to positive $j\omega$, with semicircular detours of radius r around any poles of $L(s)$ on the imaginary axis) and then closes with a semicircle of radius R. We work in the limits $R \to \infty$ and $r \to 0$. The part of the contour along the imaginary axis coincides of course with the frequency response, and thus can in principle be read off from a Bode plot. It is usually the case that $L(s) \to 0$ as $|s| \to \infty$ (we say that $L(s)$ is 'strictly proper' in this case, which holds for any linear

state-space model with $D = 0$), which means that the infinite-semicirclular part of the D curve usually maps to the origin. Hence the main thing we should discuss here is generating Nyquist curve from Bode plot.

We begin by noting two important symmetries of the frequency response under the mapping $\omega \mapsto -\omega$. For $L(s)$ a rational function,

$$L(s) = \frac{n_l}{d_l} = \frac{a_p s^p + a_{p-1} s^{p-1} + a_{p-2} s^{p-2} + \cdots + a_1 s + a_0}{b_r s^r + b_{r-1} s^{r-1} + b_{r-2} s^{r-2} + \cdots + b_1 s + b_0},$$

where p is the order of the numerator polynomial and r is the order of the denominator polynomial. For a strictly proper transfer function, $r > p$. For $s = j\omega$, the even powers of s will all be purely real and the odd powers will be purely imaginary. This means that we can write

$$\mathrm{Re}[n_l] = a_0 - a_2 \omega^2 + a_4 \omega^4 + \cdots, \quad \mathrm{Im}[n_l] = a_1 \omega - a_3 \omega^3 + a_5 \omega^5 + \cdots,$$

$$\mathrm{Re}[d_l] = b_0 - b_2 \omega^2 + b_4 \omega^4 + \cdots, \quad \mathrm{Im}[d_l] = b_1 \omega - b_3 \omega^3 + b_5 \omega^5 + \cdots.$$

Under the mapping $\omega \mapsto -\omega$, we see that $\mathrm{Re}[n_l]$ and $\mathrm{Re}[d_l]$ are unchanged while $\mathrm{Im}[n_l]$ and $\mathrm{Im}[d_l]$ are simply multiplied by -1. Generally speaking,

$$\begin{aligned}
|L(j\omega)| &= \left| \frac{\mathrm{Re}[n_l] + i\,\mathrm{Im}[n_l]}{\mathrm{Re}[d_l] + i\,\mathrm{Im}[d_l]} \right| = \frac{|(\mathrm{Re}[n_l] + i\,\mathrm{Im}[n_l])(\mathrm{Re}[d_l] - i\,\mathrm{Im}[d_l])|}{(\mathrm{Re}[d_l])^2 + (\mathrm{Im}[d_l])^2} \\
&= \frac{|\mathrm{Re}[n_l]\,\mathrm{Re}[d_l] + i\,\mathrm{Re}[d_l]\,\mathrm{Im}[n_l] - i\,\mathrm{Re}[n_l]\,\mathrm{Im}[d_l] + \mathrm{Im}[n_l]\,\mathrm{Im}[d_l]|}{(\mathrm{Re}[d_l])^2 + (\mathrm{Im}[d_l])^2} \\
&= \frac{\sqrt{(\mathrm{Re}[n_l]\,\mathrm{Re}[d_l] + \mathrm{Im}[n_l]\,\mathrm{Im}[d_l])^2 + (\mathrm{Re}[d_l]\,\mathrm{Im}[n_l] - \mathrm{Re}[n_l]\,\mathrm{Im}[d_l])^2}}{(\mathrm{Re}[d_l])^2 + (\mathrm{Im}[d_l])^2}.
\end{aligned}$$

Likewise,

$$\angle L(j\omega) = \tan^{-1}\left(\frac{\mathrm{Re}[d_l]\,\mathrm{Im}[n_l] - \mathrm{Re}[n_l]\,\mathrm{Im}[d_l]}{\mathrm{Re}[n_l]\,\mathrm{Re}[d_l] + \mathrm{Im}[n_l]\,\mathrm{Im}[d_l]} \right).$$

Hence we see by inspection that under $\omega \mapsto -\omega$, the frequency response of a rational transfer function transforms as

$$|L(s)| \mapsto |L(s)|, \quad \angle L(s) \mapsto -\angle L(s).$$

This in turn means that when we are sketching the Nyquist curve, we can first draw the part corresponding to $\omega \in (0, \infty)$ and the the part corresponding to $\omega \in (-\infty, 0)$ will simply be its mirror image (reflected through the real axis).

Of course, the frequency response for $\omega \in (0, \infty)$ is in principle what we see in a Bode plot. Since the "D" curve runs upwards along the imaginary axis, we should think about plotting the positive frequency part of the Nyquist curve from $\omega = 0$ to $\omega = +\infty$. If we have a rational $L(s)$ and the lowest power of s appearing in the numerator is equal to or higher than the lowest power of s appearing in the denominator, then $\lim_{\omega \to 0^+}|L(j\omega)|$ is finite and we can locate it somewhere on the finite complex plane. It will be the point whose radius (in polar coordinates) is the magnitude of the frequency response at $\omega = 0$ and whose angle (in polar coordinate) is the phase of the frequency response at $\omega = 0$; these can be directly read off of a Bode plot.

If the lowest power of s appearing in the denominator is higher than in the numerator, however, then $\lim_{\omega \to 0^+}|L(j\omega)|$ diverges. Note that under these assumptions,

$L(s)$ has at least a simple pole at the origin. Remember that in this case the "D" curve takes a small semicircular detour around the origin. Consider $L(j\eta)$ for some very small value η. We know that $|L(j\eta)|$ will be large, and from the rational function form of $L(s)$ we can determine $\angle L(j\eta)$. Likewise we can find $|L(-j\eta)|$ and $\angle L(-j\eta)$ straightforwardly. If we can evaluate one more point, $L(\varepsilon)$ for ε small and real, this usually suffices to determine how the image of the semicircle closes the Nyquist curve. It's easiest to see this with an example.
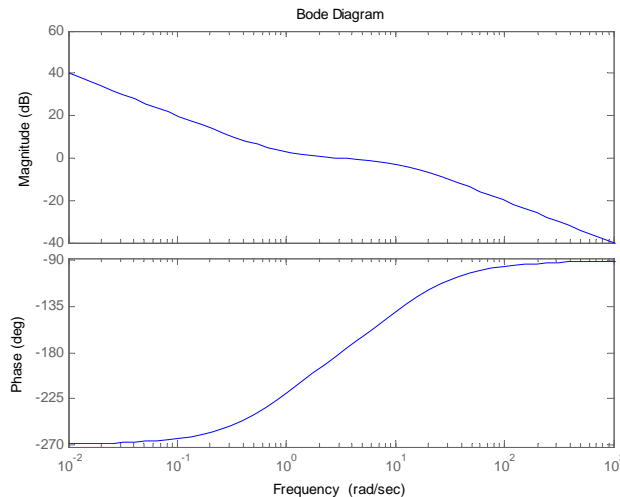
Consider

$$L(s) = \frac{9(s+1)}{s(s-10)}.$$

This has a left half-plane zero at $s = -1$ and poles at $s = 0$ and $s = 10$. For $s = j\eta$, with η small but positive, we can infer

$$L(s) = \frac{9s+9}{s^2-10s} \rightarrow -\frac{0.9}{s} = -\frac{0.9}{j\eta}.$$

Thus $|L(s)| \sim 1/\eta$ and $\angle L(s) \sim -270°$. The Bode plot for intermediate frequencies looks like

Hence the magnitude decreases monotonically as $\omega \rightarrow +\infty$ while the phase wraps from -270 to -90. By plotting a few points and connecting the dots, one can obtain the parts of the Nyquist curve corresponding to $\omega \in (0, +\infty)$ and $\omega \in (-\infty, 0)$:
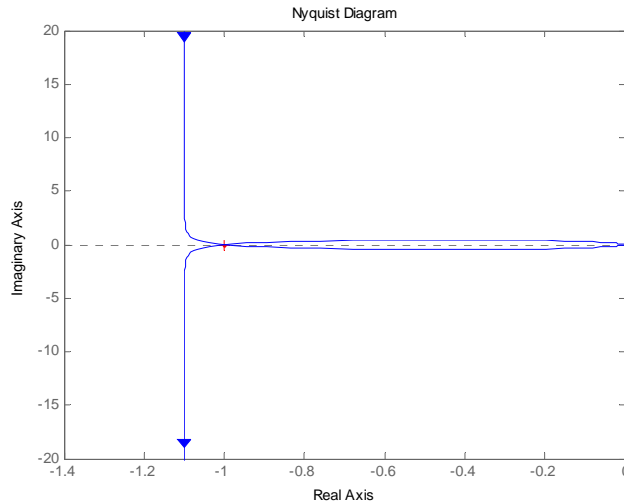
Now we need to be careful about evaluating the response near the -1 point because the Nyquist curve comes very close! Analytically, we can solve

$$L(j\omega) \mapsto \frac{9j\omega + 9}{-\omega^2 - 10j\omega} = -\frac{9 + 9\omega j}{\omega^2 + 10\omega j} \frac{\omega^2 - 10\omega j}{\omega^2 - 10\omega j}$$

$$= -\frac{(9 + 9\omega j)(\omega^2 - 10\omega j)}{\omega^4 + 100\omega^2} = -\frac{9\omega^2 - 90\omega j + 9\omega^3 j + 90\omega^2}{\omega^4 + 100\omega^2}$$

$$= \frac{-j99\omega + (90 - 9\omega^2)j}{\omega^3 + 100\omega}.$$

Hence the Nyquist curve crosses the real axis at $\omega = \sqrt{10}$, and that point the magnitude is

$$\left| L\left(j\sqrt{10}\right) \right| = \frac{99}{\omega^2 + 100} < 1.$$

Hence, $N = 0$ or $N = +1$ depending on whether the Nyquist curve closes 'to the right' or 'to the left,' which we can settle by evaluating $L(\varepsilon + 0j)$:

$$L(s) = \frac{9s + 9}{s^2 - 10s} \rightarrow -\frac{0.9}{s} = -\frac{0.9}{\varepsilon}.$$

Hence we see that the Nyquist curve closes in a way such that $\angle L(s)$ passes through $-180°$, and looking at the orientation of the contour we conclude $N = +1$. According to the Nyquist theorem this means that $Z = N + P$ should be *two*, since $P = 1$. Indeed,

$$H(s) = \frac{L(s)}{1 + L(s)} = \frac{9s + 9}{s^2 - 10s + 9s + 9} = \frac{9(s + 1)}{s^2 - s + 9}.$$

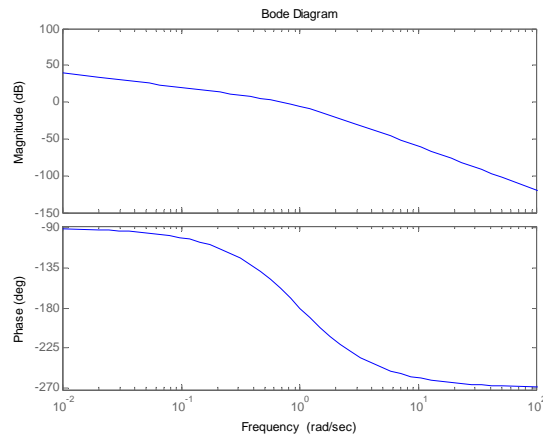This has a single left half-plane zero and a conjugate pair of RHP poles.

Another example:

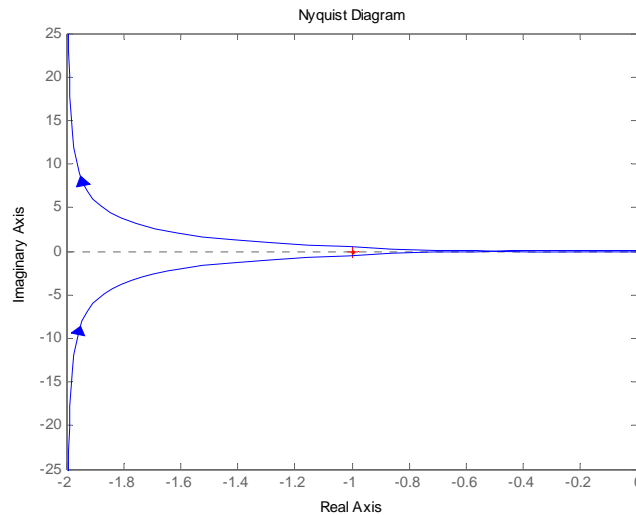$$L(s) = \frac{1}{s(s + 1)^2} = \frac{1}{s^3 + 2s^2 + s},$$

which has a simple pole at the origin. For $s = j\eta$, we have

$$L(s) \rightarrow \frac{1}{s} = \frac{1}{j\eta},$$

4

so the phase is -90. For intermediate positive frequencies, the Bode plot looks like

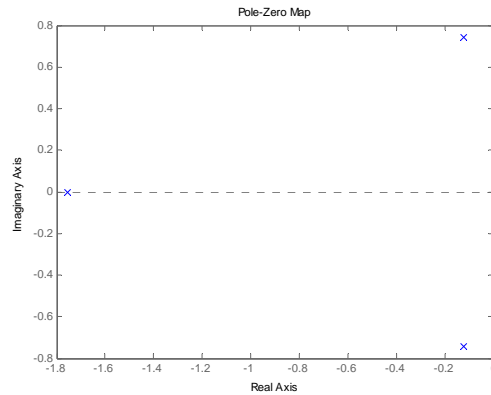and the corresponding branches of the Nyquist curve:

Now we check

$$L(\varepsilon + 0j) \rightarrow \frac{1}{\varepsilon},$$

which has phase 0. Hence the Nyquist curve closes to the right, and $N = 0$. Hence the Nyquist theorem tells us $Z = N + P = 0$ and we should be stable in closed-loop. Indeed,

$$H(s) = \frac{L(s)}{1 + L(s)} = \frac{1}{s^3 + 2s^2 + s + 1},$$
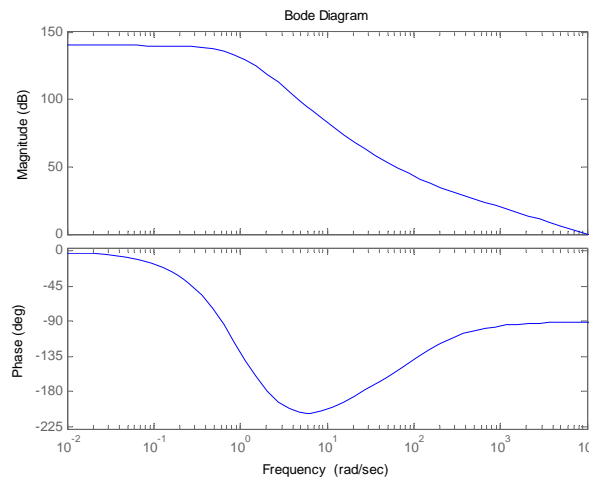
which has pole-zero map:

Pole-Zero Map

Systems with more than $180°$ phase lag where $|L(j\omega)| \gg 1$

Consider the following loop transfer function:

$$L(s) = \frac{10000(s+10)(s+100)}{(s+1)^3} = 10000\frac{s^2 + 110s + 1000}{s^3 + 3s^2 + 3s + 1},$$

which has Bode plot

Bode Diagram

This has phase $\sim -210°$ in a region where $|L(j\omega)| > 90$ dB. Is this a problem for closed-loop stability?
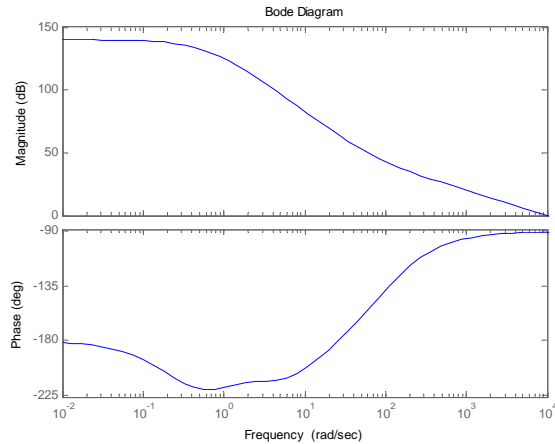
We can simply compute,

$$H(s) = \frac{L(s)}{1 + L(s)} = 10000\frac{s^2 + 110s + 1000}{s^3 + 3s^2 + 3s + 1 + 10000(s^2 + 110s + 1000)}$$

$$= 10000\frac{s^2 + 110s + 1000}{s^3 + 10\,003s^2 + 1\,100\,003s + 10\,000\,001},$$

which turns out to have only left half-plane poles.

Alternatively, let us look at

6

$$L(s) = \frac{10000(s + 10)(s + 100)}{(s - 1)(s + 1)^2} = 10000 \frac{s^2 + 110s + 1000}{s^3 + s^2 - s - 1},$$

which has Bode plot

Like the previous example, this has phase $< -180°$ in some region where the gain is large. Looking at the closed-loop transfer function,
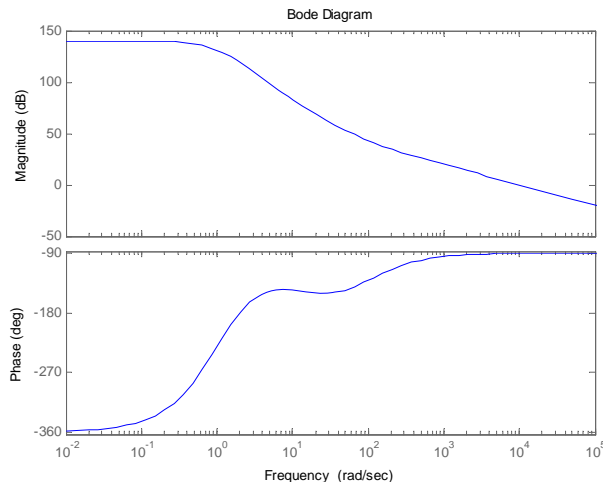
$$H(s) = \frac{L(s)}{1 + L(s)} = 10000 \frac{s^2 + 110s + 1000}{s^3 + s^2 - s - 1 + 10000(s^2 + 110s + 1000)}$$
$$= 10000 \frac{s^2 + 110s + 1000}{s^3 + 10001s^2 + 1099\,999s + 9999\,999},$$

which again turns out to have only left half-plane poles.

Finally, consider

$$L(s) = \frac{10000(s - 10)(s + 100)}{(s - 1)^3} = 10000 \frac{s^2 + 90s - 1000}{s^3 - 3s^2 + 3s - 1},$$

which has Bode plot

In this case,

$$H(s) = 10000 \frac{s^2 + 90s - 1000}{s^3 - 3s^2 + 3s - 1 + 10000(s^2 + 90s - 1000)}$$

$$= 10000 \frac{s^2 + 90s - 1000}{s^3 + 9997s^2 + 900\,003s - 10\,000\,001}.$$

This now does have a right half-plane pole...

Exercises: analyze the stability of the examples in this section using Nyquist plots and the Nyquist theorem.

Time delays and Padé approximation

Suppose we want to represent the frequency response (transfer function) of a time delay. Then for example,

$$u = \sin(\omega t) \mapsto y = \sin(\omega(t - \tau)),$$

(note the minus sign) so the magnitude should be one and the phase should be $-\omega\tau$.

A simple way to represent this as a transfer function is

$$P(s) = e^{-s\tau},$$

which clearly has

$$|P(j\omega)| = |e^{-j\omega\tau}| = 1, \quad \angle P(j\omega) = -\omega\tau.$$

Time delays are very common in practical scenarios, although they are often small enough to be neglected. When they are large enough to be significant, they can cause major problems! As we'll see in later lectures, one particularly problematic feature of time delays is that they cause phase lags without adding any 'slope' to the magnitude part of the Bode plot - this means you can't correct the lags via derivative control.

From an analysis perspective, time delays are inconvenient because the natural representation of the transfer function isn't a rational function. We could try expanding it as a power series,

$$\exp(-s\tau) \approx 1 - s\tau + \frac{1}{2}(s\tau)^2 + \frac{1}{3!}(s\tau)^3 + \cdots,$$

but it turns out that one can do better by using a rational function of the form

$$\exp(-s\tau) \approx \frac{a_0 + a_1 s + a_2 s^2 + \cdots}{b_0 + b_1 s + b_2 s^2 + \cdots},$$

called a Padé approximant. For example, if we use two terms each in the numerator and denominator,

$$\exp(-s\tau) \approx \frac{a_0 + a_1 s}{1 + b_1 s},$$

where b_0 is set to 1 without loss of generality, we can find the coefficients by comparison with the power series expansion of the exponential:

$$1 - s\tau + \frac{1}{2}(s\tau)^2 + \frac{1}{3!}(s\tau)^3 + \cdots \leftrightarrow \frac{a_0 + a_1 s}{1 + b_1 s},$$

$$(1 + b_1 s)\left(1 - s\tau + \frac{1}{2}(s\tau)^2 + \cdots\right) = a_0 + a_1 s,$$

and equating powers of s we find (keeping enough terms on the left-hand side to generate three equations)
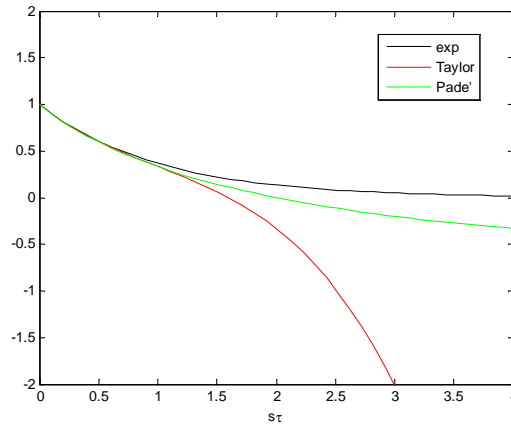
$$a_0 = 1, \quad b_1 - \tau = a_1, \quad \frac{1}{2}\tau - b_1 = 0 \quad \Rightarrow \quad \exp(-s\tau) \approx \frac{1 - s\tau/2}{1 + s\tau/2}.$$

This has essentially three adjustable coefficients, so we should compare it with a third-order power series,

$$\exp(-s\tau) \approx 1 - s\tau + \frac{1}{2}(s\tau)^2 + \frac{1}{3!}(s\tau)^3.$$

Comparing all three functions:

General expressions for higher-order Padé approximants of $\exp(-s\tau)$ can be found, *e.g.*, on Mathworld. For example,

$$\exp(-s\tau) \approx \frac{(s\tau)^2 - 6s\tau + 12}{(s\tau)^2 + 6s\tau + 12},$$

$$\exp(-s\tau) \approx \frac{-(s\tau)^3 + 12(s\tau)^2 - 60s\tau + 120}{(s\tau)^3 + 12(s\tau)^2 + 60s\tau + 120}.$$

Matlab has a `pade` function as well.

One thing to watch out for with Padé approximants is the fact that they tend to introduce right half-plane zeros, which could possibly cause unstable (in fact, inaccurate) cancellations of right-half plane poles in some other component of the transfer function. The exact location of these zeros depends on the numerator order of the Padé approximation. For example,

$$\exp(-s\tau) \approx \frac{1 - s\tau/2}{1 + s\tau/2} \to z = 2/\tau,$$

whereas

$$\exp(-s\tau) \approx \frac{(s\tau)^2 - 6s\tau + 12}{(s\tau)^2 + 6s\tau + 12} \to z = \left(3 \pm i\sqrt{3}\right)/\tau.$$

Clearly the locations of these poles is not completely accurate at any finite order of approximation. Consider the loop transfer function

$$L(s) = \frac{1}{s - 2/\tau}\exp(-s\tau).$$

For the Nyquist criterion, is P zero or one?