Constraints on the Sensitivity and Complementary Sensitivity Functions

A. Overview

Given a single-input, single-output (SISO) system with unity feedback and loop gain (return ratio) $L(s)$, the closed-loop sensitivity $S(s)$ and complementary sensitivity $T(s)$ transfer functions are

$$S(s) = \frac{1}{1+L(s)}, \qquad T(s) = \frac{L(s)}{1+L(s)} \tag{1}$$

Letting the transforms of the reference input, output disturbance, and measurement noise be $R(s)$, $D(s)$, and $M(s)$, respectively, the expression for the transform of the system output is

$$Y(s) = S(s)D(s) + T(s)\left[R(s) - M(s)\right] \tag{2}$$

and the expression for the tracking error is

$$E(s) = R(s) - Y(s) = S(s)\left[R(s) - D(s)\right] + T(s)M(s) \tag{3}$$

Thus, the sensitivity and complementary sensitivity functions are closed-loop transfer functions between specified inputs and outputs in the system. Specifically,

$$\begin{aligned} S(s) &= \left.\frac{Y(s)}{D(s)}\right|_{\substack{R(s)=0\\M(s)=0}} = \left.\frac{E(s)}{R(s)}\right|_{\substack{D(s)=0\\M(s)=0}} = \left.\frac{-E(s)}{D(s)}\right|_{\substack{R(s)=0\\M(s)=0}} \\ T(s) &= \left.\frac{Y(s)}{R(s)}\right|_{\substack{D(s)=0\\M(s)=0}} = \left.\frac{-Y(s)}{M(s)}\right|_{\substack{D(s)=0\\R(s)=0}} = \left.\frac{E(s)}{M(s)}\right|_{\substack{D(s)=0\\R(s)=0}} \end{aligned} \tag{4}$$

It can easily be seen from (1) that there is a fundamental constraint on $S(s)$ and $T(s)$, namely

$$\begin{aligned} S(s) + T(s) &= 1 \\ |S(s) + T(s)| &= |1| = 1 \\ |S(s) + T(s)| &\leq |S(s)| + |T(s)| \end{aligned} \tag{5}$$

Even though $S(s)$ and $T(s)$ are functions of the complex variable s, and their frequency responses are functions of $j\omega$, the sum of the two functions is a constant. Closed-loop stability and performance issues will impose other constraints as well. In these notes, we will look at the constraints imposed by the relative degree of the loop gain $L(s)$ and by the presence of right-half plane zeros in $L(s)$.

We can see from (2) and (3) that in order to have good tracking of the reference signal and good rejection of the disturbance signal, we need the conditions that $S(s) \approx 0, T(s) \approx 1$. Both of these conditions can be satisfied by having $|L(s)| \gg 1$. However, in order to prevent propagation of measurement noise to the error and output signals, we need $T(s) \approx 0 \Rightarrow S(s) \approx 1$. These conditions are only satisfied by $|L(s)| \ll 1$. In order to achieve all of these objectives, there must be a frequency separation between the reference and disturbance signals on the one hand and the measurement noise signal on the other. In many real situations, the reference and disturbance signals contain mostly low frequencies, and the measurement noise contains a wide band of frequencies, including high frequencies. Therefore in general, $|L(j\omega)| \gg 1$ is required at low frequencies and $|L(j\omega)| \ll 1$ is required at high frequencies.

In addition to a large loop gain at low frequencies, steady-state error constraints might require that $|L(j0)|$ be infinitely large, with the slope of the magnitude curve as $\omega \to 0$ depending on the specific steady-state error characteristics that are desired. For example, if zero steady-state error is required for a constant reference signal or disturbance signal, then the low-frequency slope of $|L(j\omega)|$ must be at least -20 db/decade (or -1 on a log-log plot). If zero steady-state error is required for a linearly increasing reference or disturbance signal, then the low-frequency slope of $|L(j\omega)|$ must be at least -40 db/decade (or -2 on a log-log plot).

These notes are lecture notes prepared by Prof. Guy Beale for presentation in a course on multivariable and robust control in the Electrical and Computer Engineering Department, George Mason University, Fairfax, VA. Additional notes can be found at the following website: http://teal.gmu.edu/~gbeale/examples.html.

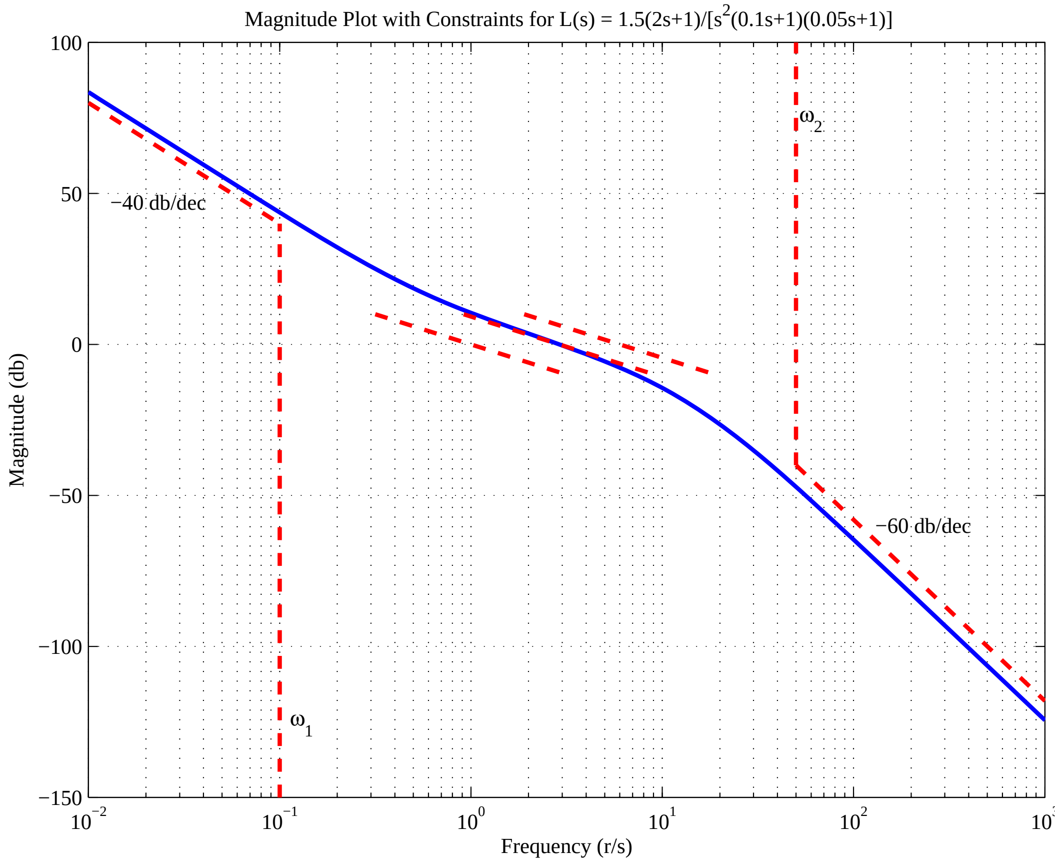

Fig. 1. Typical open-loop Bode plots for a system with constraints on transient and steady-state performance.

In addition to providing high loop gain at low frequencies and low loop gain at high frequencies, closed-loop stability is also an issue that affects the shape of $|L(j\omega)|$. In order to provide adequate phase margin, the slope of the magnitude curve at the gain crossover frequency ω_c, should be -20 db/decade. An example of typical constraints imposed on $|L(j\omega)|$ are shown in Fig. 1, giving the Bode magnitude plot of the transfer function

$$L(s) = \frac{1.5\,(2s+1)}{s^2\,(0.1s+1)\,(0.05s+1)} \tag{6}$$

$|L(j\omega)|$ must lie above the dashed lines at low frequencies, which requires $|L(j\omega)| \geq 40$ db for $\omega \leq \omega_1 = 0.1$ rad/s. Thus, the system must have an open-loop gain of at least 100 for $\omega \leq \omega_1$. The constraint also requires a low-frequency slope of -40 db/decade (or more negative). This slope will produce zero steady-state error when the reference or disturbance signals are either constant or linearly increasing with time. At high frequencies, $|L(j\omega)|$ must lie below the dashed lines shown. The magnitude must be less than or equal to -40 db (attenuation by at least a factor of 100) for all frequencies $\omega \geq \omega_2 = 50$ rad/s. The high-frequency slope of the magnitude curve must be -60 db/decade (or more negative), so measurement noise will be rapidly attenuated above the bandwidth frequency. The sloped dashed lines in the middle of the plot indicate that gain crossover should occur in the interval $1 \leq \omega_c \leq 6$ rad/s with a slope of -20 db/dec. The middle dashed line shows that the actual gain crossover occurs at $\omega \approx 2.9$ rad/s. The slope of the magnitude curve at that frequency is very nearly -20 db/decade, so good stability margins should be achieved. The gain margin for this system is 9.25 (19.3 db), and the phase margin is $55.8^{\cdot}$.

The following sections will investigate some relations between the loop gain $L(s)$ and the closed-loop transfer functions $S(s)$ and $T(s)$. The effects of the relative degree of the loop gain $L(s)$ is discussed first, and then the effect of right-half plane zeros.

B. Relative Degree of $L(s)$

Assume that there are no open-loop poles or zeros in the open right-half of the s-plane. Let n and m indicate the total number of poles and the total number of zeros in $L(s)$, respectively. Then the relative degree of $L(s)$ is defined to be $n - m$. We will show that if the closed-loop system is stable and $n - m > 1$, then the polar plot of $L(j\omega)$ must enter the unit circle centered at -1 in the complex $L(s)$ plane, which means that the sensitivity function magnitude must obey $|S(j\omega)| > 1$ over some frequency interval.

The phase shift of $L(j\omega)$ as $\omega \to \infty$ is $-90^\circ\,(n-m)$. With a relative degree $n-m>1$, the phase shift will be at least -180° at high frequencies, so the polar plot of $L(s)$ must enter the third quadrant of the complex plane. Since the Nyquist stability criterion states that the -1 point must not be encircled in a clockwise direction in order for the closed-loop system to be stable, the polar plot cannot go completely around the unit circle centered at -1. Therefore, the plot must enter that circle if the closed-loop system is stable. Since the distance from the -1 point to a point on the $L(j\omega)$ curve is $|1+L(j\omega)|$, and $|S(j\omega)| = 1/|1+L(j\omega)|$, whenever the polar plot of $L(j\omega)$ is inside that circle, $|S(j\omega)| > 1$. Relative stability measures, such as gain and phase margins, are in general inversely related to the peak value of $|S(j\omega)|$, so large peak values are to be avoided in order to achieve good transient performance.

To illustrate this, consider the following two loop gains and the corresponding sensitivity and complementary sensitivity functions.

$$L_1(s) = \frac{2}{s\,(s+1)}\ , \qquad S_1(s) = \frac{s\,(s+1)}{s^2+s+2}\ , \qquad T_1(s) = \frac{2}{s^2+s+2} \tag{7}$$

$$L_2(s) = \frac{2\,(s+0.5)}{s}\cdot L_1(s) = \frac{4\,(s+0.5)}{s^2\,(s+1)}\ , \qquad S_2(s) = \frac{s^2\,(s+1)}{s^3+s^2+4s+2}\ , \qquad T_2(s) = \frac{4\,(s+0.5)}{s^3+s^2+4s+2} \tag{8}$$

The loop gain $L_1(s)$ might correspond to an uncompensated system model. Then the loop gain $L_2(s)$ could correspond to the system model in series with a Proportional + Integral (PI) controller. This controller might be chosen to provide zero-steady-state error for reference or disturbance signals that increase linearly with time. The relative degree for each of these systems is $n-m=2$, so $|S(j\omega)|_{\max} = \|S(s)\|_\infty > 1$ for each of them. Figure 2 shows the Bode plots and polar plots for $L_1\,(j\omega)$ and $L_2\,(j\omega)$. The polar plots clearly show that the graphs enter the unit circle centered at -1. The effects of adding the open-loop pole at $s=0$ are also clear from the reduced phase margin of $L_2(j\omega)$ and the smaller distance from the -1 point to the graph. This latter fact means that $\|S_2(s)\|_\infty > \|S_1(s)\|_\infty$. The top two plots in Fig. 3 show that this is indeed the case, with $|S_2(j\omega)|_{\max} = |S_1(j\omega)|_{\max} + 8.1$ db (the peak magnitude of $|S_2(j\omega)|$ is 2.53 times the peak magnitude of $|S_1(j\omega)|$).

The effect of the reduced relative stability can be seen in the step response plots in Fig. 3. The closed-loop system with loop gain $L_2(s)$ has much larger overshoot, larger oscillations, and longer settling time than the system with loop gain $L_1(s)$. One advantage of the second pole at $s=0$ in $L_2(s)$ is seen in the ramp response plots. That second integrator forces the steady-state error for the ramp input to be zero. The price that is paid for that is the increased overshoot and settling time in the step response.

Although the graphs in Figs. 2 and 3 are for particular systems, there are some general comments that can be made. As previously mentioned, any time the relative degree $n-m$ of the loop gain $L(s)$ is two or more, then the polar plot of $L(s)$ will enter the unit circle centered at -1, assuming that the closed-loop system is stable. That means that $\|S(s)\|_\infty > 1$. In general, the larger that peak value, the smaller the phase margin and the worse the transient response of the closed-loop system.

In order to obtain adequate phase margin and step response characteristics, limits should be imposed on $\|S(s)\|_\infty$. This can be done through the use of a frequency selective weighting function $w_P(s)$. The limit on $\|S(s)\|_\infty$ is achieved through the following constraints[1].

$$|w_P\,(j\omega)\,S(j\omega)| < 1\ \forall\omega \Leftrightarrow \|w_P\,(s)\,S(s)\|_\infty < 1 \Leftrightarrow |S(j\omega)| < \frac{1}{|w_P\,(j\omega)|}\ \forall\omega \tag{9}$$

Simple first-order and second-order transfer functions for $w_P(s)$ that can be used to limit the peak value of $|S(j\omega)|$ to $1/|w_P\,(j\omega)|$ are given by

$$w_P(s) = \frac{s/M+\omega_B}{s+\omega_B A}\ , \qquad w_P(s) = \frac{\left(s/\sqrt{M}+\omega_B\right)^2}{\left(s+\omega_B\sqrt{A}\right)^2} \tag{10}$$

where A is the bound on $|S(j\omega)|$ at zero frequency, M is the bound on $|S(j\omega)|$ at high frequency, and ω_B is the desired closed-loop bandwidth defined in terms of $|S(j\omega)|$.

Figure 4 shows the Bode magnitude plots for the sensitivity functions $S_1(j\omega)$ and $S_2(j\omega)$. The steeper roll-off at low frequencies and the larger peak value for $|S_2(j\omega)|$ are clearly seen. The dashed line in the figure is the inverse magnitude of the weighting function

$$w_P(s) = \frac{0.5(s+1)^2}{s^2} \tag{11}$$

[1] *Multivariable Feedback Control: Analysis and Design*, S. Skogestad and I. Postlethwaite, John Wiley & Sons, Chichester, UK, 1996.

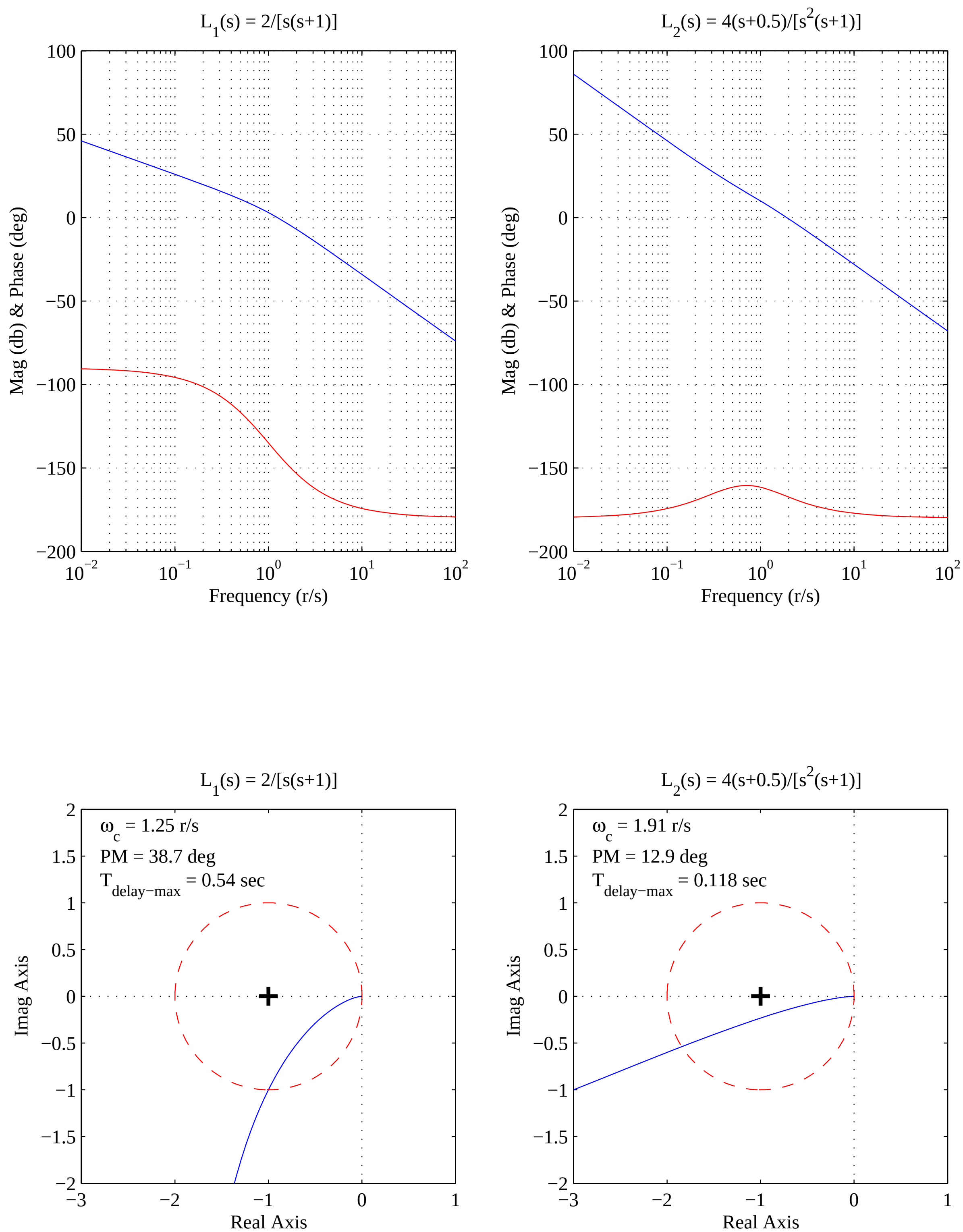

Fig. 2. Bode plots and polar plots for $L_1(j\omega)$ and $L_2(j\omega)$.

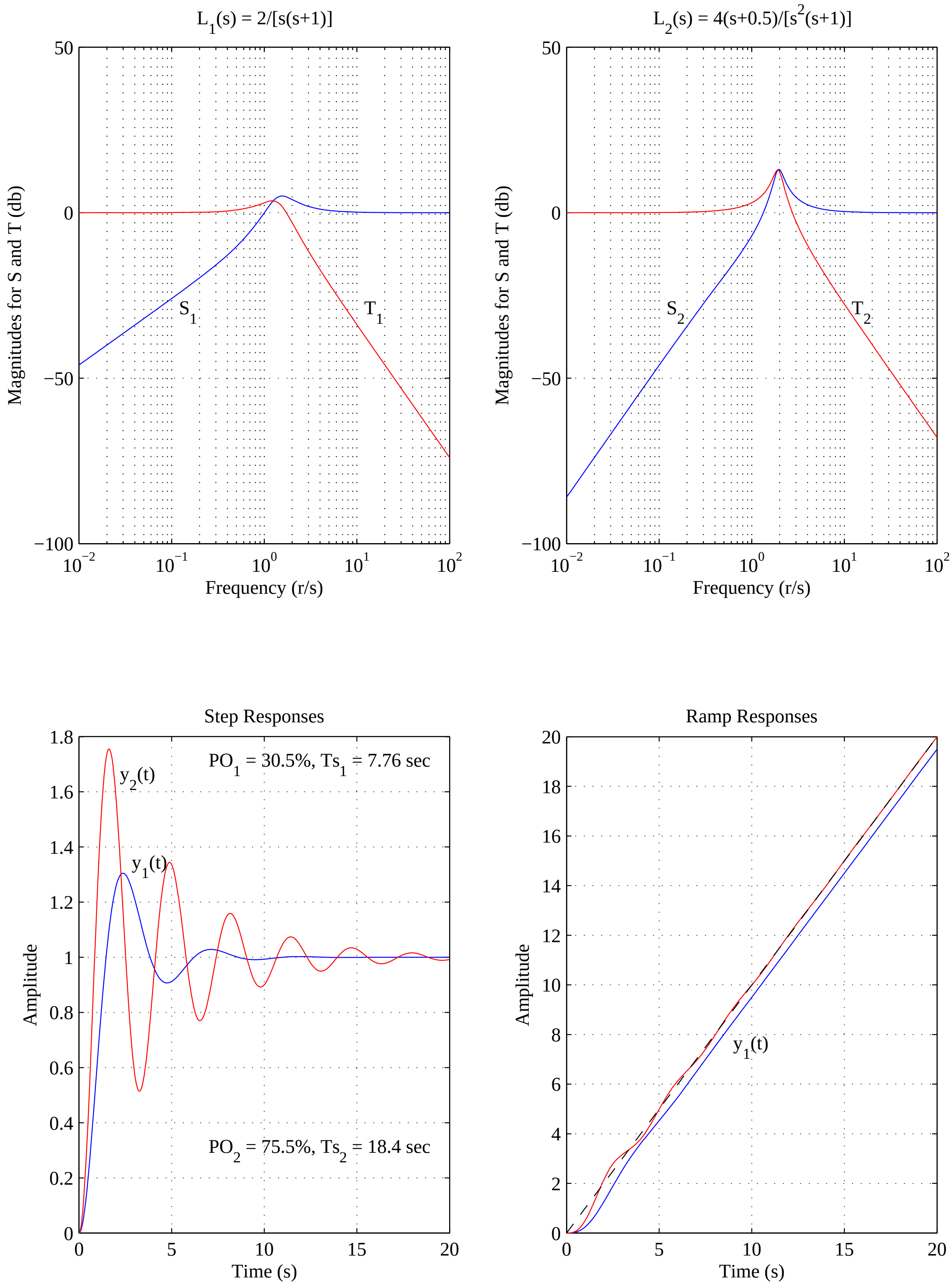

Fig. 3. Sensitivity functions and time responses for the two systems.

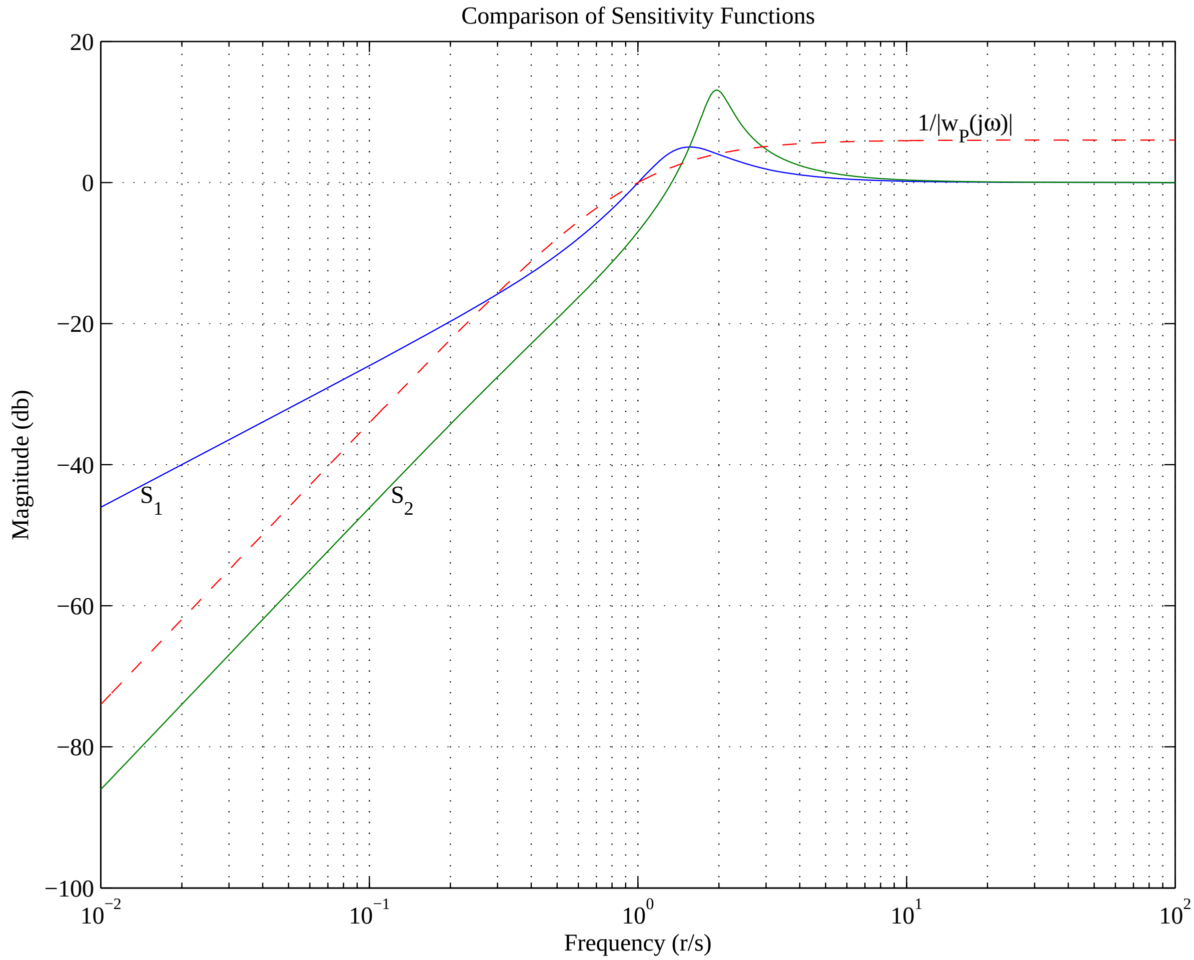

Fig. 4. Comparison of the two sensitivity functions.

Neither of the sensitivity functions satisfies the constraint imposed by $w_P(s)$ at all frequencies, but $|S_2(j\omega)|$ does have the right low frequency behavior, while $|S_1(j\omega)|$ has approximately the correct bandwidth. Note that when $|S_2(j\omega)|$ is "pushed down" at low frequencies, it "pops up" at the higher frequencies. This is known as the waterbed effect, arising from Bode's sensitivity integral. For stable open-loop systems with $n - m > 1$, the Bode sensitivity integral is

$$\int_0^\infty \ln |S(j\omega)|\, d\omega = 0 \tag{12}$$

which says that any area of sensitivity reduction must be balanced exactly by an equal area of sensitivity increase. For unstable open-loop systems, the area of sensitivity increase is larger than the area of sensitivity reduction. Since virtually any real system will have a pole excess of at least two, it is important to recognize this waterbed effect and take steps to limit $\|S(s)\|_\infty$ through careful design.

If the open-loop transfer function $L(s)$ is stable and $n - m = 1$, the Bode integral becomes[2]

$$\int_0^\infty \ln |S(j\omega)|\, d\omega = -\frac{\pi}{2} \cdot \lim_{s\to\infty} [sL(s)] = -\frac{\pi}{2} \cdot \frac{b_{n-1}}{a_n} \tag{13}$$

where b_{n-1} and a_n are the leading coefficients in the numerator and denominator polynomials of $L(s)$, respectively.

[2] *Control System Design*, G.C. Graham, S.F. Graebe, and M.E. Salgado, Prentice Hall, Upper Saddle River, NJ, 2001.

C. *Right-Half Plane Zeros in $L(s)$*

Now we will look at the effect of right-half plane (RHP) zeros on the sensitivity function. A zero in the RHP has the phase shift of a pole—creating negative phase shift—but the magnitude of a zero. Therefore, the phase becomes more negative without the corresponding attenuation of the magnitude. Since the magnitude must be less than 1 at the frequency where the phase shift is $-180°$ in order to have closed-loop stability, RHP zeros impose another constraint on system performance.

Consider the following two loop gains and the corresponding sensitivity functions in a unity feedback system.

$$\begin{aligned} L_1(s) &= \frac{2(0.4s+1)}{s(10s+1)}, \qquad S_1(s) = \frac{s(10s+1)}{(10s^2+1.8s+2)} = \frac{s(s+0.1)}{(s^2+0.18s+0.2)} \\ L_2(s) &= \frac{2(-0.4s+1)}{s(10s+1)}, \qquad S_2(s) = \frac{s(10s+1)}{(10s^2+0.2s+2)} = \frac{s(s+0.1)}{(s^2+0.02s+0.2)} \end{aligned} \tag{14}$$

The two loop gains are identical except for the fact that the zero of $L_2(s)$ is in the right-half plane rather than the left-half plane. The effects of this are seen in the Bode and polar plots of Fig. 5. The Bode magnitude plots are identical, but the phase plot of $L_2(j\omega)$ has $180°$ of additional negative phase shift as $\omega \to \infty$ due to the RHP zero. The polar plots show that $L_2(j\omega)$ comes very close to the -1 point, so it should be expected that $\|S_2(s)\|_\infty$ is large. Since the gain crossover frequencies of the two systems are the same, and $L_2(j\omega)$ has more negative phase shift, the phase margin for $L_2(j\omega)$ is much less than for $L_1(j\omega)$. The second system is nearly unstable.

The sensitivity magnitudes and the step response plots for the two systems are shown in Fig. 6. The larger peak in $|S_2(j\omega)|$ can be seen, with $|S_2(j\omega)|_{\max} = |S_1(j\omega)|_{\max} + 18.9$ db (a factor of 8.81 larger). The overshoot in the step response for the second system is nearly 100%, and the settling time is approximately 400 seconds, approximately 10 times that of the first system. This illustrates the very negative effects of having an open-loop zero in the right-half plane.

D. *Concluding Remarks*

These two examples have shown some of the problems that must be taken into consideration when designing a control system. The peak of the sensitivity magnitude $|S(j\omega)|$ plays an important role in closed-loop system stability and performance. The waterbed effect that occurs whenever $n - m > 1$ shows that there will always be a trade-off in the design. If we want to improve the loop shape of $|S(j\omega)|$ in some frequency range, then there will be a worsening of the loop shape in some other frequency range This will be unavoidable.

Open-loop zeros in the right-half plane also cause problems for $|S(j\omega)|$, particularly when they are near the $j\omega$ axis. The negative phase shift produced by the zero, without a corresponding reduction in magnitude, reduces the phase margin and increases the value of $\|S(s)\|_\infty = |S\,(j\omega)|_{\max}$.

Additional constraints on the achievable bandwidth of $|S(j\omega)|$ are imposed by time delays in the system and RHP zeros. The use of weighting factors, such as $w_P(j\omega)$, and the H_∞ design methodology can be used to design a control system that will provide a sensitivity function that provides the best stability and performance characteristics that can be achieved within the given constraints, at least in terms of using the infinity norm as a performance measure.

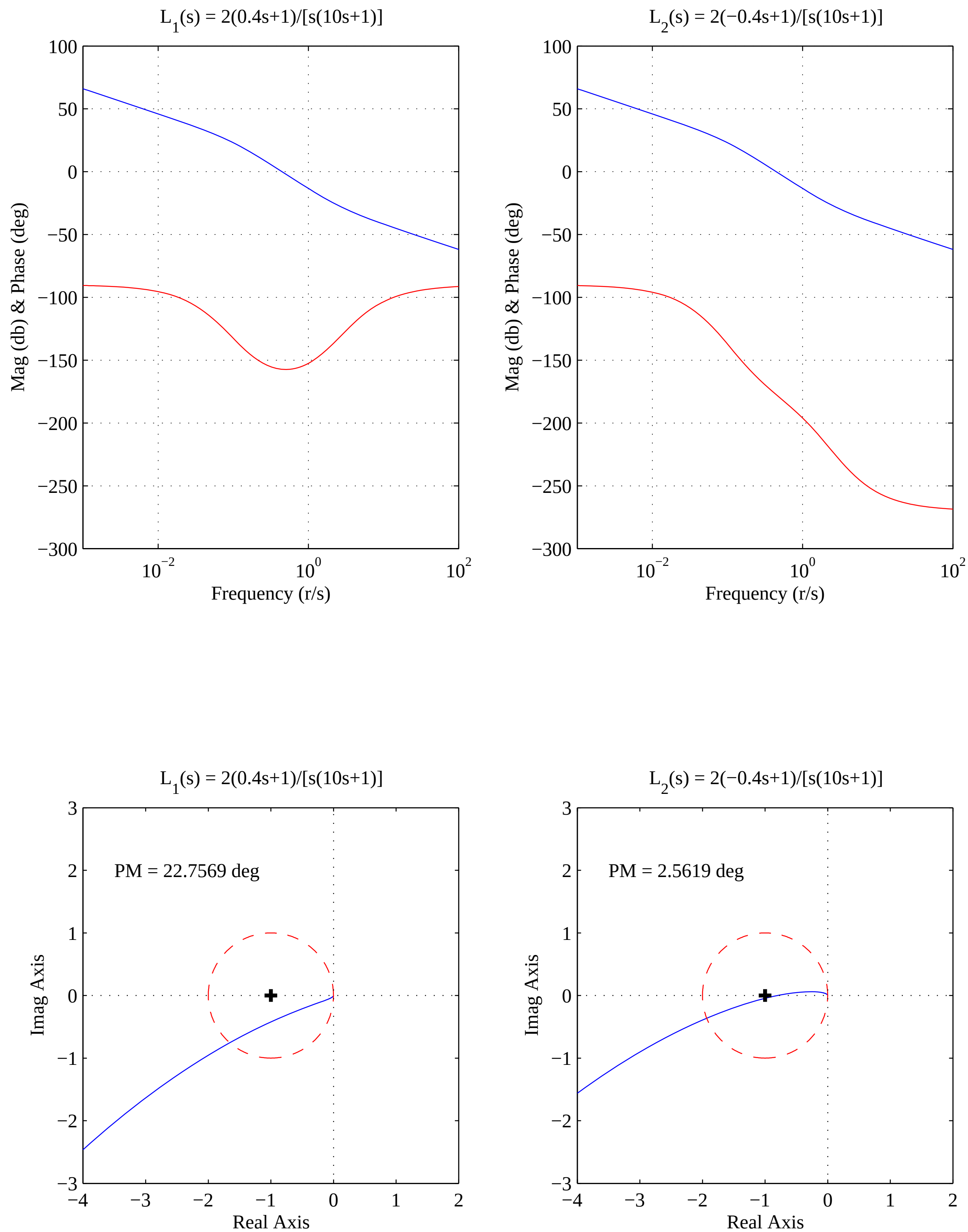

Fig. 5. Bode plots and polar plots showing the effects of a RHP zero.

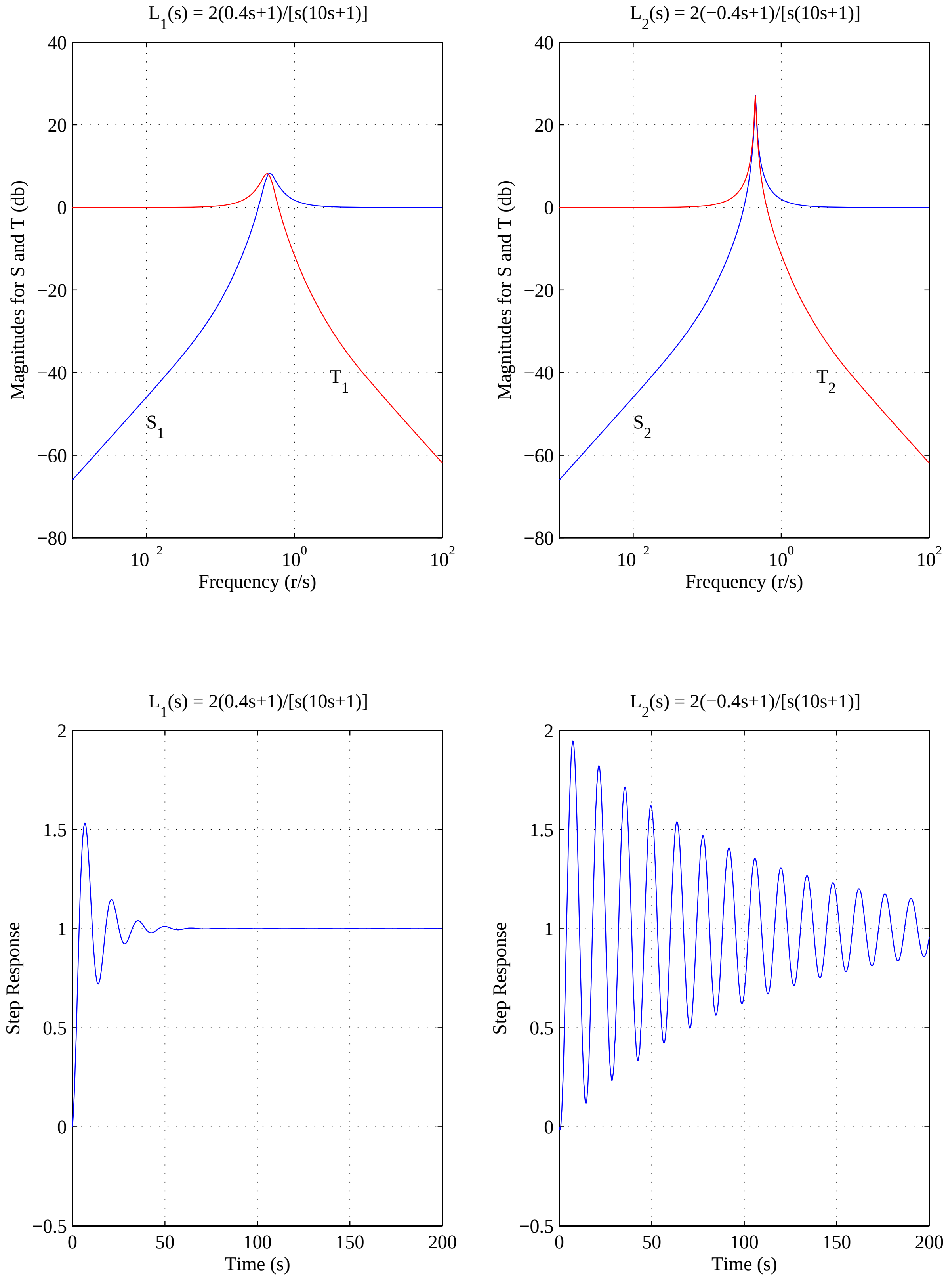

Fig. 6. Sensitivity functions and step responses with a RHP zero.