Phase Margin Revisited: Phase-Root Locus, Bode Plots, and Phase Shifters

Thomas J. Cavicchi

***Abstract*—In learning undergraduate controls, one of the most abstract and confusing concepts is that of phase margin (PM). One frustration is the difficulty in conceptualizing what physical process could produce the pure phase shift referred to in the definition of PM. Another frustration occurs when simple second-order formulas given for PM do not work in practice. Finally, when resonances occur, there may be multiple gain-crossover frequencies, and which one to use to compute the PM may be unclear. This paper offers visualizations and explanations that students have found helpful in learning about PM-both its evaluation and its application. The phase-root locus (PRL) plot reveals the effect of adding phase in the same way that conventional gain root locus shows the effects of adding gain. The PRL ultimately leads to a definition of PM involving phase shifting that results in physical (real-coefficient) systems, unlike the usual abstract Nyquist plot rotation. This definition of PM suggests a simple, effective compensator design method for improving PM via phase shifters, a solution illustrated by numerical examples. These materials can either be presented in lectures or assigned as supplementary readings and may inspire student projects.**

***Index Terms*—Control engineering, control engineering education, control systems, phase shifters, stability.**

I. Introduction

In THIS paper, several educational tools for undergraduate controls are presented, all concerning phase margin (PM). PM is especially susceptible to confusion and abstractness, yet its importance is equal to that of gain margin (GM) in classical analog or digital controller design. In fact, the most common simple Bode lead/lag controller design algorithms in current textbooks use PM as the fundamental dynamic design parameter. Such basic design and knowledge of classical controls still has widespread use in industrial controls, a major employer of engineering graduates. The phase-root locus (PRL), introduced in [1] (see the Appendix for a simpler, direct PRL calculation method), can be helpful for visualizing effects of system phase shifts in the same way that root locus (RL) is for visualizing effects of changes in gain, as illustrated in Sections II and III and in [1]. The PRL can also help identify PM for systems with resonances (Section IV). In Section V, the accuracies of two commonly cited relations between PM and damping ratio are examined. In Section VI, several definitions/interpretations of PM are reviewed and a definition is offered that avoids phase-shifted systems with complex coefficients. This definition motivates the approximation of an ideal phase shifter using compensating filters in Section VII; examples of such design (which suggest interesting student projects) are presented in Section VIII.

Manuscript received March 9, 2001.

The author is with the Department of Electrical and Computer Engineering, Grove City College, Grove City, PA 16127 USA (e-mail: tjcavicchi@gcc.edu).

Digital Object Identifier 10.1109/TE.2002.808228

II. Vector Plots, Root Loci, and Stability Margins

Let $K_m > 0$ represent the m^{th} value of gain K multiplying the N^{th}-order open-loop plant $G(s)$ in a sequence of negative-unity-feedback designs [replace $G(s)$ by $G(s)H(s)$ for a nonunity feedback transfer function $H(s)$]. The RL is defined as the set of all s, consisting of N branches, for which the angle condition is satisfied

$$\angle G(s) = 180^\circ \ell \ (\ell \text{ odd}) \quad [RL \text{ (angle) condition}] \tag{1}$$

and the PRL is the set of all s, forming one or more contours, for which the magnitude condition is satisfied

$$|K_m G(s) = 1| \quad [PRL \text{ (magnitude) condition}]. \tag{2}$$

See [1] for a complete introductory tutorial on PRL.

The unity-feedback formula shows that the N closed-loop poles $\{s_i(K_m),\ i\ \epsilon[1, N]\}$ occur for $K_m G(s_i(K_m)) = -1$. Thus, for $s_i(K_m)$ to be the closed-loop poles, they must lie not only on the RL but also on the PRL. The RL is the set of paths traced out by the closed-loop poles $\{s_i(K)\}$ as gain K ranges from 0 to ∞ with zero added phase. Let the closed-loop poles, as K is held fixed at the design value K_m but phase θ is added to $K_m G(s)$, be denoted $s_{i,\text{Km}}(\theta)$, $i\ \epsilon[1, N]$. (The practical addition of phase is discussed fully in Sections VI–VIII.) Then the PRL is the set of contours traced out by the closed-loop poles $\{s_{i,\text{Km}}(\theta)\}$ as $K = K_m$ is held fixed and phase θ is added to $G(s)$. The closed-loop poles are located at the intersections of the RL and PRL.

Students have found Fig. 1 helpful for visualizing the relation between RL, PRL, GM, and PM. (All plots in this paper were made using MATLAB [a registered trademark of The MathWorks, Inc.].) The example plant is $G(s) = (s + 10)/\{s(s^2 + 3s + 402.2)\}$. The region of the s-plane depicted includes the $j\omega$-axis and the gain-critical [1] open-loop pole providing oscillatory closed-loop behavior for moderately small increases in gain K_m. The large circular path is the PRL for $K_m = 80$; the large, nearly vertical curve is the RL; the symbol "X" at $s = -1.5$ Np/s$+j20$ rad/s is an open-loop pole; and the symbol "*" is the associated closed-loop pole for $K_m = 80$.

The vectors in Fig. 1 represent values of $K_m G(s)$. The length of the vector plotted at grid point $s = s_1$ is $|K_m G(s_1)|$. The angle of vector $K_m G(s_1)$ measured from the right-pointing ray originating at s_1 is $\angle G(s_1)(K_m > 0)$. The little circles all have unit radii. (All vector lengths/circles are scaled by 0.4 relative

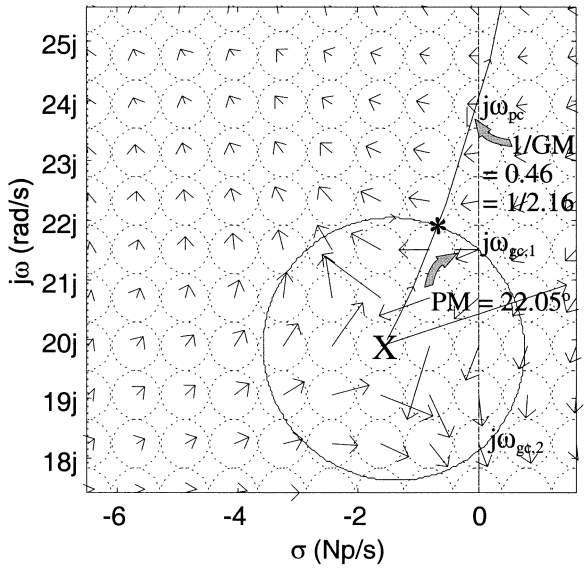


Fig. 1. Vector plot of $K_mG(s) = 80(s+10)/\{s(s^2+3s+402.2)\}$, with RL/PRL superimposed and PM, GM labeled; X = open-loop pole, * = closed-loop pole, arrows = vector representations of $K_mG(s)$ on grid of s values.

to the s-value scale.) On the PRL, where $|K_mG(s)| = 1$, the vector arrowhead exactly meets its unit circle, whatever its direction. On the RL, where $\angle G(s) = 180^\circ\ell$ (ℓ odd), the vector $K_mG(s)$ always points exactly leftwards, whatever its length. At the closed-loop pole * (not a grid point in Fig. 1), the vector has length 1 and points exactly leftwards.

On the $j\omega$ axis, the $K_mG(s)$ vectors represent both the magnitude and phase of the frequency response, $K_mG(j\omega)$. Notice that where the RL intersects the $j\omega$-axis ($s = j\omega_{\rm pc}$, where pc means that at $\omega = \omega_{\rm pc}$, the *p*hase of the frequency response *c*rosses over from greater than to less than -180°), the vector $K_mG(j\omega_{\rm pc})$, whose length is 1/GM, points exactly leftwards [(1) is satisfied] (the vector indicated is not precisely at $j\omega_{\rm pc}$). At the two intersections of the PRL with the $j\omega$ axis ($s = j\omega_{\rm gc}$, where gc means that at $\omega = \omega_{\rm gc}$, the *g*ain of the frequency response *c*rosses over from greater than to less than 1 [or 0 dB]), the vectors $K_mG(j\omega_{\rm gc})$ have exactly unit length [(2) is satisfied]. In Section IV, the PM is shown to be determined in this case from the angle of the vector $K_mG(j\omega_{\rm gc})$ at the indicated $\omega_{\rm gc}$, $\omega_{\rm gc,1}$. Note the weak resonance for ω near 20 rad/s (the open-loop pole oscillation frequency), where the vectors along the $j\omega$-axis are relatively long.

Notice that within the PRL, each $K_mG(s)$ vector extends outside its corresponding unit circle, because $|K_mG(s)| > 1$ due to s being near the open-loop pole. Outside the PRL, no $K_mG(s)$ vector reaches its unit circle because $|K_mG(s)| < 1$ due to s being far from the open-loop pole and thus $|K_mG(s)|$ being small. Similarly, notice how on either side of the RL, the angles of the $K_mG(s)$ vectors are greater than or less than -180°. Thus the RL and PRL are, respectively, the phase- and gain-crossover complex frequencies for *all* vertical contours $\sigma = {\rm Re}\{s\}$, not just for $\sigma = 0$ (the frequency response). The appearance of these vectors over large regions resembles that of fluid flow diagrams with sinks and sources. As the PRL is traversed clockwise, the $K_mG(s)$ vector rotates clockwise $360^\circ\times$(number of enclosed open-loop zeros – number of enclosed open-loop poles).

III. From the s-Plane to the Bode Plot

Letting s_i denote either $s_i(K_m)$ (RL) or $s_{i,{\rm Km}}(\theta)$ (PRL), the s-plane closed-loop stability criterion is

$$\sigma_i = {\rm Re}\{s_i\} < 0,\ i\ \epsilon[1,N] \quad [s-\text{plane stability criterion}] \tag{3}$$

so that the corresponding time-domain natural modes $\exp\{s_it\}$ are exponentially damped. If $K_mG(s)$ is minimum phase, (3) can be translated via RL into an equivalent Bode gain criterion defining GM [2] or via PRL into an equivalent Bode phase criterion defining PM (considered here). (The relationship between GM/PM and the minimum return difference in the Nyquist plot is explored in [3].)

A negative phase increment $\theta = \theta_0$ added to minimum-phase $K_mG(s)$ is generally required to drive the closed-loop system to instability by moving the RL so that it intersects the PRL on the $j\omega$-axis. Thus, if a closed-loop pole $s_{i,{\rm Km}}(\theta_0)$ (on the PRL) is imaginary for $i = i_1$ (and thus $s_{i1,{\rm Km}}(\theta_0) = j\omega_{\rm gc}$), then the system is stable for $|\theta| < |\theta_0|$ (equivalently, $\theta > \theta_0$) and unstable for $|\theta| > |\theta_0|$. (It is assumed that $|\theta_0|$ is the lowest value of $|\theta|$ for which there is an imaginary $s_{i,{\rm Km}}(\theta)$; see Section IV.) Moreover, as $s_{i1,{\rm Km}}(\theta_0) = j\omega_{\rm gc}$ is a closed-loop pole, the phase-modified plant (*not* the original plant) has phase -180° at $s = j\omega_{\rm gc}$, and therefore $\theta_0 = -180^\circ - \angle K_mG(j\omega_{\rm gc})$. Thus for stability, $\theta > -180^\circ - \angle K_mG(j\omega_{\rm gc})$, and the stability requirement on the original $K_mG(j\omega)$ (no phase added: $\theta = 0$) is

$$K_mG(j\omega_{\rm gc}) > -180^\circ \quad [\text{Bode phase stability criterion}]. \tag{4}$$

The amount by which the phase of the original system could be modified if originally closed-loop stable and just avoid closed-loop instability is θ_0. The PM is defined as $-\theta_0$, namely

$$PM(K_m) = \angle K_mG(j\omega_{\rm gc}) + 180^\circ = \angle G(j\omega_{\rm gc}) - (-180^\circ) \tag{5}$$

i.e., $PM(K_m)$ is the amount by which $\angle G(j\omega_{\rm gc})$ exceeds -180°. The dependence of $PM(K_m)$ on K_m is through the dependence of $\omega_{\rm gc}$ on K_m.

These basic ideas are clarified further in the series of Figs. 2 and 3 concerning the plant $G(s) = (s+30)(s+0.4)/[s^2(s+50)(s+10)(s^2+s+10)]$ to which various amounts of gain and phase are, respectively, added. In each figure is shown the RL and PRL (horizontal plane) and the Bode plot (vertical plane). Students find the series of plots in Figs. 2 and 3 helpful for linking two fundamental linear control system analysis and design tools: the RL/PRL plot and the Bode plot. Figs. 2 and 3 facilitate visualization of the analytical steps earlier that are used to derive stability requirements on the Bode diagram from the stability requirement (3) on the RL/PRL plot. The (linear) ω axis in the Bode plot numerically matches the $j\omega$ axis in the RL/PRL plot. The RL is the set of contours that originate on open-loop poles (denoted "X") and end on open-loop zeros (denoted "O") as gain K is increased from 0 to ∞. The PRL loops encircle open-loop poles (and possibly zeros), and in the upper half-plane the closed-loop poles move clockwise along them as

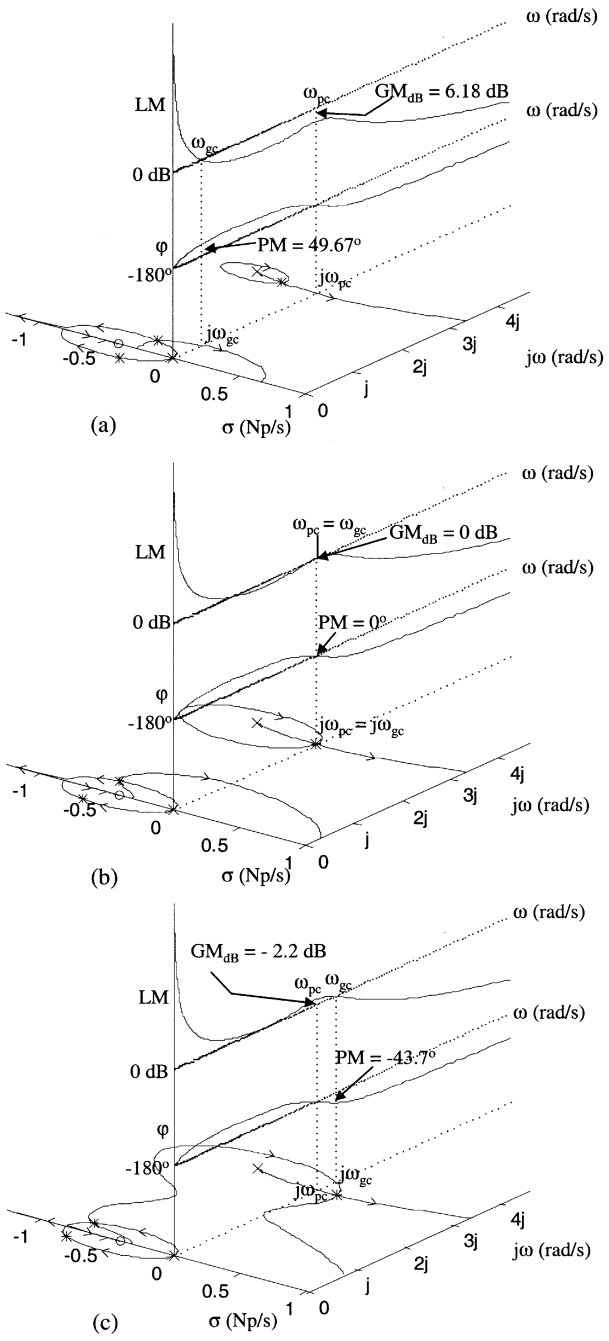


Fig. 2. RL, PRL, and Bode plots of $K_mG(s) = K_m(s+30)(s+0.4)/[s^2(s+50)(s+10)(s^2+s+10)]$ for K_m = (a) 80. (b) $163 = K_0$. (c) 210. X = open-loop pole, * = closed-loop pole, O = open-loop zero. RL begins on X; PRL encloses X.

negative phase is added and K is fixed at K_m (clockwise only if the PRL encloses at least as many open-loop poles as open-loop zeros, which is typically the case). Again, the intersections of the RL and PRL (denoted "*") are the closed-loop poles.

In Fig. 2(a)–(c), the gain is increased respectively from $K_m = 80$ ($< K_0 =$ marginal stability gain) to $K_0 = 163$ and to 210 ($> K_0$); the PRLs grow progressively larger, and the closed-loop poles are thus moved along the RL. The closed-loop pole for the RL branch having higher frequency (ω) values is seen to move from the left-half plane, to the $j\omega$ axis, and into the right-half plane. On the Bode diagrams, the phase curve remains unaltered. The magnitude curve is progressively raised by $\Delta\{20\log_{10}(K_m)\}$, causing $\omega_{\rm gc}$ to migrate from below $\omega_{\rm pc}$, to coincident with $\omega_{\rm pc}$, and to above $\omega_{\rm pc}$. A dotted line is drawn from $\omega_{\rm pc}$ in the Bode plot to $j\omega_{\rm pc}$ on the RL/PRL plot, and also from $\omega_{\rm gc}$ in the Bode plot to $j\omega_{\rm gc}$ in the RL/PRL plot. For each K_m value, the corresponding $GM_{\rm dB} = 20\log_{10}(GM)$ value and PM value are indicated.

In Figs. 2(a) and 3(a) and (b), nonpositive phase θ is added via $e^{j\theta}$ to $K_mG(s)$ for $K_m = 80$ from, respectively, 0° ($< |\theta_0| = |$marginal stability added phase$|$) to $\theta_0 = -49.7°$, and to $-74.5°$ (greater in magnitude than $|\theta_0|$). The root loci of the phase-modified plant are swung around relative to those of $G(s)$ so that $\theta + \angle K_mG(s) = -180°$ along them (see Fig. 1 in [1]), and the closed-loop poles move along the unmodified PRL. On the Bode diagrams, the magnitude curve remains unchanged. The phase curve is lowered uniformly by θ, causing $\omega_{\rm pc}$ to decrease. When θ falls below θ_0 (Fig. 3(b)), the entire phase curve lies below $-180°$; there is no $\omega_{\rm pc}$ and no stabilizing value of K_m for such θ, because the phase-altered RL branch beginning at $s = 0$ is seen to lie entirely in the right half-plane. (Analogously, a Bode magnitude curve entirely below 0 dB ["$PM = \infty$" in MATLAB] corresponds to the PRL being entirely in the left half-plane.) The lack of conjugate symmetry of RL branches in Fig. 3 is addressed fully in Section VI.

IV. PM When Resonances Occur

The PRL can also help explain the determination of PM in the case of open-loop resonances, such as occurs in Fig. 1. For example, Fig. 4(a) shows the Bode plot of the plant in Fig. 1. Three gain-crossover frequencies exist; therefore, the question is which one is used to determine the PM. Fig. 4(b) shows (without vectors) a larger view of the RL and PRL than in Fig. 1. The three gain-crossover frequencies correspond to the three $j\omega$-axis crossings of the PRL $s_n = j\omega_{{\rm gc},n}, n \in [1, 3]$. Evaluation of $K_mG(s_n)$ indicates that for s_n to be closed-loop poles, the respective phases that must be added to $K_mG(s_n)$ are $-22.05°$, $-114.6°$, and $-100.7°$.

If increasing negative phase is added to $K_mG(s)$ with $K_m = 80$ and thus the closed-loop poles slide along the PRL in Fig. 4(b), then $s_1 = j\omega_{{\rm gc},1}$ is reached long before s_2 (on the same PRL loop) or s_3 (on the lower-frequency PRL loop). Specifically, when $-22.05° = -PM(80)$ is added, the closed-loop pole on the upper PRL has moved to s_1, while that along the lower PRL has moved from the negative real axis (at zero added phase) by an angle of only 18.8° toward the imaginary axis. The PRL clearly shows that to get to $s_2 = j\omega_{{\rm gc},2}$, the closed-loop pole at s_1 has already been passed through on the way, and since then, that closed-loop pole has been in the right-half-plane (unstable). Similarly, the lower pole reaches the RHP before the upper pole re-enters the LHP. This PRL analysis proves that $\omega_{{\rm gc},1}$ is the correct gain-crossover frequency for evaluation of PM because $s_1 = j\omega_{{\rm gc},1}$ is the PRL $j\omega$-axis crossing reached first as negative phase is added to $K_mG(s)$. (Resonances do not always significantly affect stability margins; see [2].)

Finally, notice that for Fig. 2(b), $PM = GM_{\rm dB} = 0$ as shown, for there is an imaginary closed-loop pole,

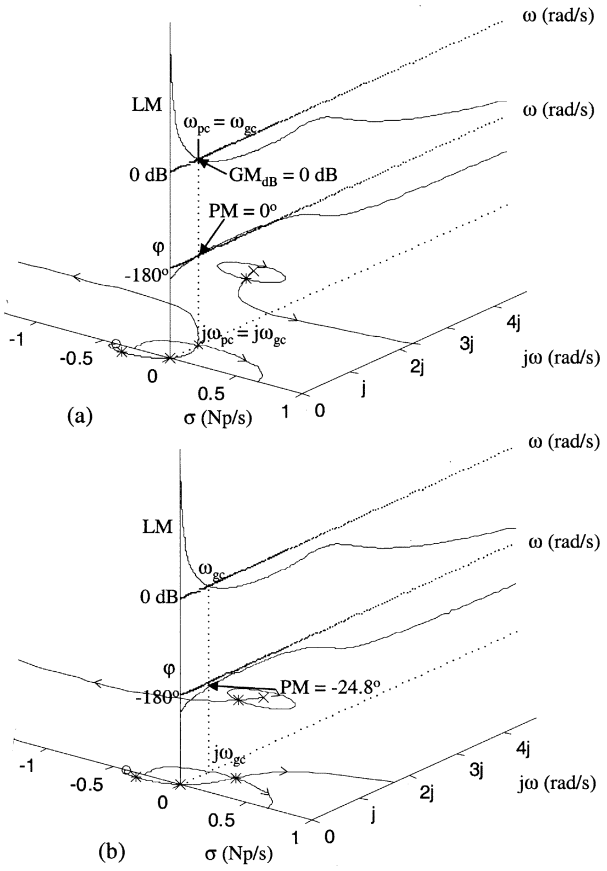


Fig. 3. RL, PRL, and Bode plots of $e^{j\theta}K_mG(s)$ [$G(s)$ as in Fig. 2; $K_m = 80$] for θ = (a) $-49.7^\circ = \theta_0$. (b) -74.5°.

yet MATLAB's "margin" command gives the large value $PM = 57.64^o$ with K_m even infinitesimally (or more) lower than K_0, e.g., via entering a rounded value of K_0. Students often find it confusing that they obtain from MATLAB one stability margin significantly nonzero when the other margin is "zero." The confusion is easily resolved by reference to Fig. 2(b); if K_m is even slightly less than K_0, the upper PRL/Bode resonance peak no longer intersect the $j\omega$ axis/0-dB line; and, thus, PM is determined from the lower-PRL (and now the only existing) $\omega_{\rm gc}$ value; PM can be a discontinuous function of K_m.

V. Validity of PM-Damping Ratio Relations

The previous discussions have addressed the exact evaluation of PM using either Bode plots or the PRL. PM is the basis of simple Bode-based compensator design algorithms (e.g., [4]), with the intention of achieving a specified closed-loop overshoot via the damping ratio-PM link. However, the major quantitative limitations of this link are overlooked in textbooks. For the canonical second-order plant $G_2(s) = \omega_n^2/[s(s + 2\zeta\omega_n)]$ in a negative-unity-feedback system, PM is determined exactly in terms of the parameter ζ in $G_2(s)$ by finding $\omega = \omega_{\rm gc}$ for which $|G_2(j\omega)| = 1$, evaluating $\angle G_2(j\omega_{\rm gc})$, and then using (5), yielding [4]–[7]

$$PM = \tan^{-1}\left\{\frac{2\zeta}{[(4\zeta^4+1)^{1/2} - 2\zeta^2]^{1/2}}\right\} \qquad (6)$$

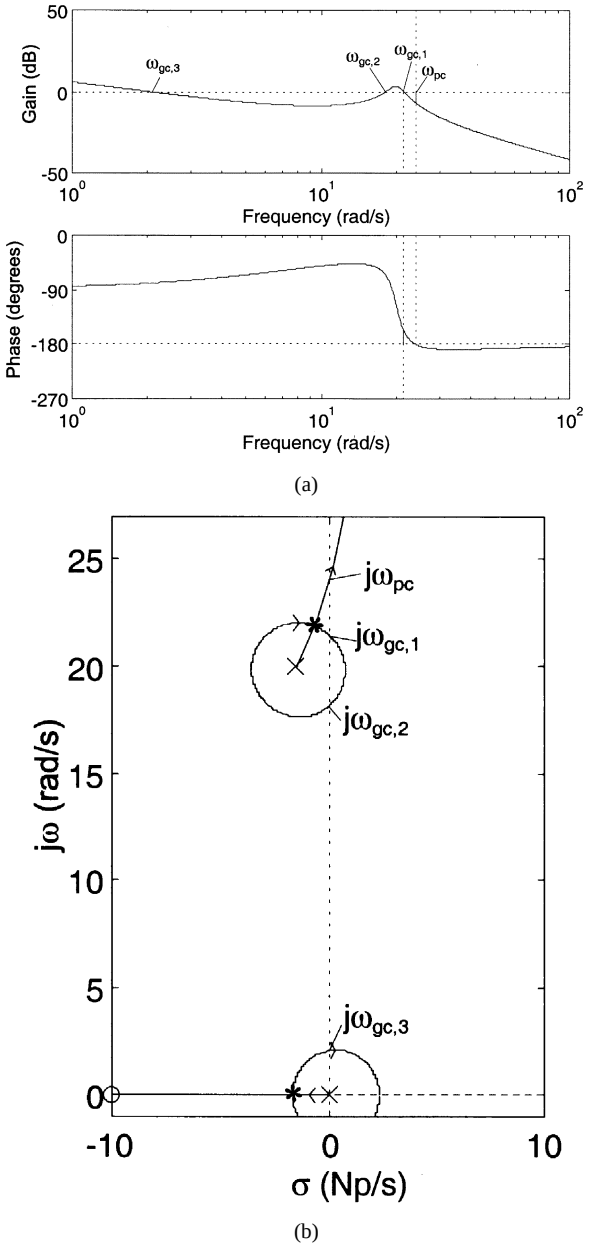


Fig. 4. (a) Bode and (b) RL/PRL plots for $K_mG(s) = 80(s+10)/\{s(s^2 + 3s + 402.2)\}$ (same system as in Fig. 1).

and its inverse $\zeta = \sin(PM)/[2(\cos(PM))^{1/2}]$; ζ is in turn closely related to closed-loop overshoot for the plant $G_2(s)$. Equation (6) is well approximated by [5],[6]

$$PM = 100\zeta \qquad (7)$$

for $0 < \zeta < 0.7$. Equation (7) can be visualized using the PRL, as shown in [2]. However, Fig. 5 (for $\zeta = 0.4$, $\omega_n = 1$ rad/s) demonstrates the extreme unreliability of either formula even for $G(s) = G_2(s)$ if $K_m \neq 1$.

Given that the derivation of (6) is based on solving for ω such that $|G_2(j\omega)| = 1$ and not $|K_mG_2(j\omega)| = 1$, obviously (6) relates PM and ζ in $G_2(s)$ only for $K_m = 1$, as may be verified in Fig. 5–the only point of agreement. The true PM (Fig. 5) traverses the entire range $90^\circ > PM(K_m) > 0^\circ$ as K_m ranges from 0.001 to 10000, with 74% of that variation on [0.1, 10].

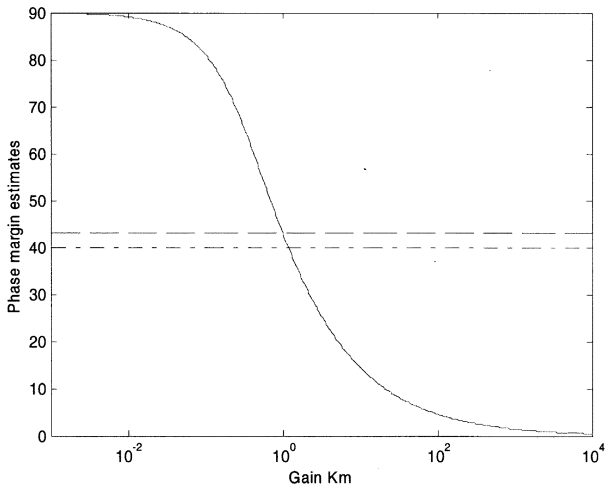


Fig. 5. PM versus K_m for $K_mG_2(s) = K_m\omega_n^2/[s(s+2\zeta\omega_n)]$ for $\zeta = 0.4$, $\omega_n = 1$ rad/s; solid= exact via "margin" or canonical relation using $\zeta_{\rm CL} = \zeta/\{K_m\}^{1/2}$ instead of $\zeta = 0.4$ in (6), dashed = (6) using $\zeta = 0.4$; dash dot = 100ζ.

Approximations (6) and (7) are nearly equal, and clearly depend on only $\zeta = 0.4$, not on K_m. For $K_mG(s) = K_mG_2(s)$ with $K_m \neq 1$, the closed-loop transfer function can be put in the canonical "$K_m = 1$" form by defining $\omega_n' = \omega_n\{K_m\}^{1/2}$ and $\zeta' = \zeta_{\rm CL} = \zeta/\{K_m\}^{1/2} \neq \zeta$ and then (6) applies exactly for ζ replaced by ζ'. Thus (6) is useful for the plant $G(s) = \omega_n^2/[s(s+2\zeta\omega_n)]$ with applied gain $K_m \neq 1$ only if in (6) "ζ" is not the parameter ζ in $G(s)$ but rather $\zeta_{\rm CL} = \zeta/\{K_m\}^{1/2}$, which is actually desirable for design as a valid link between PM and $\zeta_{\rm CL}$.

However, in general, $K_mG(s)$ does not even have the form of $K_mG_2(s)$, and there is no simple relation between $\zeta_{\rm CL}$ and the parameter ζ in (6) or (7). Addition of a compensator $G_c(s)$ generally guarantees that $K_mG_c(s)G(s) \neq \alpha G_2(s)$ for any α, let alone $\alpha = 1$. One might attempt to design PM for $K_mG_c(s)G(s)$ using (6) in hopes of achieving a desired $\zeta_{\rm CL}$, yet (6) applies as a link between PM and $\zeta_{\rm CL}$ *only* if $K_mG_c(s)G(s) = 1 \cdot \omega_n^2/[s(s+2\zeta_{\rm CL}\omega_n)]$. One may thus see the substantial limitation on the applicability of (6) and (7).

A further example is $K_mG(s) = 12\,000/[(s+50)(s^2+20s+109)]$, which has a dc gain of 2.2, for which the unity-feedback closed-loop poles are -55.72 Np/s and -7.14 Np/s $\pm$ $j16.19$ rad/s. Note that the real closed-loop pole is 7.8 times as far into the left half-plane as are the complex closed-loop poles, so the closed-loop system is second-order dominant. Now $\zeta_{\rm CL} = 7.14/(7.14^2 + 16.19^2)^{1/2} = 0.4035$. The true PM is 70.55°, but the PMs predicted by (6) and (7) for "ζ" $= \zeta_{\rm CL}$ are, respectively, 43.45° and 40.35°—completely useless approximations in this case. This discussion should serve as a caution to the student not to apply simple formulas found in the literature unless the conditions for their validity have been verified. One may conclude that PM has relevance as a robustness metric and as a qualitative, but not necessarily quantitative, indicator of overshoot.

VI. PM Defined for Physical Systems

A previous attempt to connect PM with RL ideas is presented in [8]. A complex gain K is used to show how the angles of closed-loop poles, $\angle s_i(K)$, vary with $\angle K$. A disadvantage of [8] is that a complex gain is nonphysical, so the results may be unacceptable. Also, it is easier for the student to interpret one graphical display of the movement of all the closed-loop poles in the s-plane as added phase is varied (PRL) than to interpret N plots [8] of only the angles of closed-loop poles $\angle s_i(K)$ versus $\angle K$. Moreover, for the complex gain used in [8], the RL becomes unsymmetrical, implying complex differential equation coefficients, which are nonphysical. The complex gain in [8] moves closed-loop poles in the opposite direction along the PRL in the lower half-plane relative to that in the upper half-plane (e.g., if an upper-half-plane closed-loop pole moves toward the $j\omega$-axis, the corresponding conjugate lower-half-plane closed-loop pole moves away from it). The physical relevance of the resulting transfer function is in question.

The traditional definition of $PM(K_m)$ is the clockwise angle through which the Nyquist plot can be uniformly rotated before it passes through the -1 point in the $K_mG(s)$-plane [6], [7], [9], [10]. This definition suffers from the same abstractness as that in [8]: How can the gain of a physical system be complex, and what is the physical meaning of the rotation?

In an attempt to define PM on physically meaningful grounds, [11] associates PM with the time delay τ the system can tolerate before becoming unstable. The statements in [11] are true, but the correspondence of delay with PM is not the dual of that of K_m with GM; the true dual of uniform-in-s K_m is uniform conjugate-symmetric phase shift (UCSPS). A UCSPS is equal to θ everywhere in the upper half-plane, and to $-\theta$ in the lower half-plane. Delay (e.g., on a conveyor belt) adds pure phase ($-\omega\tau$) onto $G(s)$ only on the $j\omega$ axis, and $e^{-s\tau}$ is uniform nowhere in the s-plane (not even on the $j\omega$ axis). $PM(K_m)$ as defined below specifies a UCSPS (which, other than conjugate symmetry, is independent of s) that when added to $K_mG(s)$ will move a complex closed-loop pole and its conjugate pair to the $j\omega$-axis along the PRL [1]. *Any* allpass modification (phase-only for $s = j\omega$) contributing a phase $\theta = -PM$ at $\omega_{\rm gc}$, e.g., a delay $\tau = PM/\omega_{\rm gc}$ for $PM > 0$, may be viewed as a PM annihilator. However, although delay does reduce PM, it does not produce the uniform rotation of the polar plot that is universally and definitively associated with PM, as does UCSPS (albeit opposite rotation for $\omega < 0$). The following PM definition involving UCSPS for all s is the true dual of GM because PRL is the true dual of RL [1].

To avoid ending up with nonsymmetric, nonphysical transfer functions, phase margin is defined as minus the largest negative [for $PM(K_m) > 0$] angle that can be added *uniformly in the upper half of the s-plane (including the real axis) to the open-loop transfer function $K_mG(s)$, and simultaneously the positive phase of the same magnitude added in the lower half-plane to $K_mG(s)$* and still have the resulting closed-loop system be at least marginally stable. This definition, consistent with (5), corresponds to symmetrical movement of closed-loop conjugate-pair poles along the PRL just as added gain corresponds to symmetrical movement of closed-loop conjugate-pair poles along the conventional RL.

When pure phase is added conjugate-symmetrically, the resulting transfer function has the same open-loop poles/zeros as before, but a new PM due to closed-loop pole movement along the *(unmodified)* PRL. The closed-loop poles occur at

the intersections of the RL (which *does* move when phase is added) and the PRL; if complex, they occur in complex conjugate pairs, as required for physical systems. The path to instability is, respectively, along the RL or PRL as gain is added or as negative/positive phase is added in the upper/lower half-plane (see [2] for details about the effects of UCSPS on real closed-loop poles).

Note that one may view plant parameter variations or modeling errors as producing phase and gain changes on $K_mG(s)$, which may occur simultaneously and/or depend on s. Contrary to the intuitive but erroneous notion that gain variations will produce variations in only GM and not in PM, and to the analogous statement for phase variations, GM and PM are interdependent (e.g., see Fig. 3(b) in [3]). Although gain and phase are orthogonal polar coordinates of $K_mG(j\omega)$, this independence is not retained in the corresponding GM and PM. It nevertheless *is* true that a multiplicative change in gain of the open-loop transfer function translates directly to the same multiplicative change in 1/GM and a change in phase translates to the same change in PM; the dynamics of $K_mG(s)$ then determine the change in the other margin not directly altered.

VII. Approximate Phase-Shifting Compensators

What sort of filtering operation could produce a UCSPS? In communication theory, the Hilbert transformer $h(t) = 1/(\pi t)$ for all t performs a -90° phase shift on sinusoids (i.e., $-/+$ 90° is added for ω greater than/less than zero). The shift occurs because the bilateral Fourier transform (FT) of $h(t)$ is [12]

$$H(j\omega) = -j \cdot \mathrm{sgn}(\omega) = 1\angle -\left(\frac{\pi}{2}\right) \cdot \mathrm{sgn}(\omega). \tag{8}$$

A general UCSPS *for* $s = j\omega$ can be added to $G(j\omega)$ by a phase shifter $H_{\mathrm{ps}}(j\omega)$ producing $G_{\mathrm{ps}}(j\omega)$, where $H_{\mathrm{ps}}(j\omega)$ is obtained from $H(j\omega)$ in (8) for A and B satisfying $\{A^2 + B^2\}^{1/2} = 1$ as follows:

$$\begin{aligned} H_{\mathrm{ps}}(j\omega) =& \frac{G_{\mathrm{ps}}(j\omega)}{G(j\omega)} = A + BH(j\omega) \\ =& A - jB \cdot \mathrm{sgn}(\omega) = 1\angle\theta \cdot \mathrm{sgn}(\omega) = 1\angle\theta \cdot \frac{\omega}{|\omega|} \end{aligned} \tag{9}$$

where $\theta = -\tan^{-1}(B/A)$, giving $A = \cos(\theta)$, $B = -\sin(\theta)$. The impulse response of the phase shifter is

$$h_{\mathrm{ps}}(t) = A\delta(t) + Bh(t) = \cos(\theta)\delta(t) - \frac{\sin(\theta)}{(\pi t)}. \tag{10}$$

In controls, $H_{\mathrm{ps}}(j\omega)$ increases the PM of the plant $G(s)$ on which it operates by θ, in a manner consistent with the above definition of PM (but restricted to the $j\omega$ axis). Assuming distinct poles and $\omega_\ell = 0$ for real poles and denoting the inverse Laplace transform operator as LT^{-1}, the impulse response of the plant causal $g(t)$ is

$$g(t) = LT^{-1}\{G(s)\} = \sum_{\ell=1}^{N} A_\ell e^{-\sigma_\ell t}\cos(\omega_\ell t + \varphi_\ell),\ t \geq 0 \tag{11a}$$

where A_ℓ and φ_ℓ are real-valued. The impulse response $g_{\mathrm{ps}}(t)$ of the phase-shifted system is

$$g_{\mathrm{ps}}(t) = g(t) * h_{\mathrm{ps}}(t) \tag{11b}$$

$$\neq \sum_{\ell=1}^{N} A_\ell e^{-\sigma_\ell t}\cos(\omega_\ell t + \varphi_\ell + \theta),\ t \geq 0 \tag{11c}$$

where the intuitive (11c) holds only approximately, and then only for a certain class of plants (see later).

Specifically, the complex open-loop pole terms in the partial fraction expansion (PFE) $(1/2)A_\ell \exp(j\varphi_\ell)/(s - s_\ell)$ and $(1/2)A_\ell \exp(-j\varphi_\ell)/(s - s_\ell^*)$ where $s_\ell = -\sigma_\ell + j\omega_\ell$ are in (11c) multiplied by, respectively, $e^{j\theta}$ and $e^{-j\theta}$ for all s, but the polarity of the UCSPS is opposite for $\omega > 0$ and $\omega < 0$. Thus, if s_ℓ is in the upper half-plane, $(1/2)A_\ell \exp(j\varphi_\ell)/(s - s_\ell)$ is nonzero in the lower half-plane [there denoted the "tail" of $(1/2)A_\ell \exp(j\varphi_\ell)/(s - s_\ell)$ as its magnitude there may be small] and for UCSPS should be multiplied in the lower half-plane by $e^{-j\theta}$, whereas (11c) has it multiplied there (as everywhere) by $e^{j\theta}$. The true UCSP shifter $1\angle\theta \cdot \omega/|\omega|$ (defined for all $s = \sigma + j\omega$; use $1\angle\theta$ for $\omega = 0$) is nonanalytic because of the phase discontinuity across the real axis which occurs with nonzero (unit) magnitude. Thus, it has no rational transfer function representation, and the residue theorem cannot be applied to evaluate the LT^{-1} to obtain the terms in (11c); see [2] for expressions for $g_{\mathrm{ps}}(t)$ that hold generally.

In (10), one may view the UCSPS for $s = j\omega$ as a Hilbert-transform "leakage" [no leakage/phase change would be $\theta = 0$, giving $h_{\mathrm{ps}}(t) = \delta(t)$]. Only this sort of phase shifting makes sense in the context of PM: A shift in PM is caused by a single phase-angle θ added to each of the open-loop real, bicausal Fourier sinusoidal components of $g(t)$ [as opposed to the causal modes in (11a)] whose sum constitutes the inverse FT, $g_{\mathrm{ps}}(t)$ (see [2]). Although $h_{\mathrm{ps}}(t)$ in (10) produces the symmetric version of the Nyquist rotation, $h(t)$ and thus $h_{\mathrm{ps}}(t)$ are bicausal. Moreover, $LT\{h(t)\}$ does not even exist except for $s = j\omega$, and there only as a principal value [12] [and the same for $LT\{h_{\mathrm{ps}}(t)\}$]. Nevertheless, $h_{\mathrm{ps}}(t)$ lends insight to the topics of PM and UCSPS for $s = j\omega$.

Although the ideal UCSP shifter is unrealizable, (11c) *can* be used to obtain a simple approximation of UCSPS for certain plants. The ℓ^{th} term in (11a) can be written $B_\ell \exp(s_\ell t) + B_\ell^* \exp(s_\ell^* t)$ where $B_\ell = (1/2)A_\ell \exp(j\varphi_\ell)$. Replacing B_ℓ by $B_\ell e^{j\theta}$, one obtains (11c). The fewer real poles $G(s)$ has and the closer to the resonance region one looks (i.e., the vicinity of the dominant s_ℓ), the better the approximation of a UCSPS holding for all s. Near the real axis, the incorrectly multiplied tails may dominate, and the relative error be large; the absolute error may be large or small depending on $G(s)$. For values of s far from the dominant pole, the relative errors may be large, but the absolute error is small because the (modified) PFE terms are small there. However, it is the resonance region where dynamic compensation is often performed and transient performance determined; so, (11c) can be a useful and simple model of how PM may be changed fairly directly. Practical communication system phase shifters have similar limitations, but they approximately UCSP-shift signals within a portion of the $j\omega$ axis; here

UCSP-shifting transfer functions are approximated in a region of the s-plane.

This approximate UCSP-shifting suggests a compensation method for *increasing* PM; e.g., let the compensator be $G_c(s)$ = phase-shifted numerator/plant numerator. (Time delays, which have been used to explain PM, can only decrease PM, not increase it.) Additional (real) poles may have to be added to $G_c(s)$ and to the approximately UCSP-shifted transfer function $G'(s)$ to make a realizable compensator, because the numerator order of $G'(s)$ may exceed that of the plant $G(s)$ and $G_c(s) = G'(s)/G(s)$. Denoting common denominator formation of modified conjugate terms by PFE^{-1}, the approximately UCSP-shifted transfer function is

$$G'(s) = PFE^{-1}\{e^{j\theta} \cdot (B_\ell \text{ of } PFE\{G(s)\})\} \quad \text{(plus added poles)}. \quad (12)$$

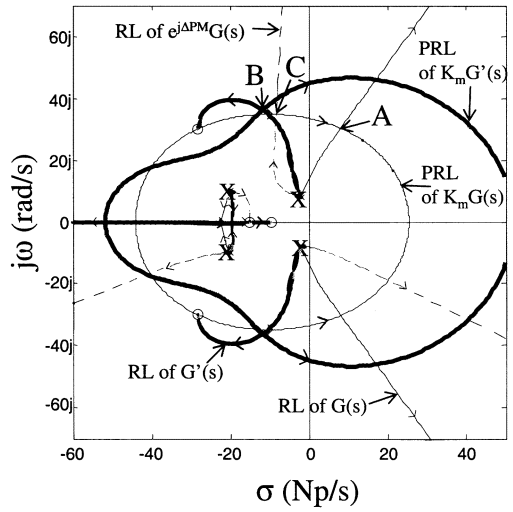
VIII. Design Examples

Initial studies suggest that this design method effectively increases PM. Steady-state error $e_{\rm ss}$ specification is also incorporated into the design by solving for K_m that yields the required error constant, just as in Bode design; enforcement of unity dc gain for the compensator $G_c(s)$ then preserves $e_{\rm ss}$. A design example is presented in Fig. 6. The type 0 plant, $G(s) = 3200(s+15)/[(s^2+40s+500)(s^2+4s+68)]$, has poles at $-2\pm j8$, $-20\pm j10$; the specified $e_{\rm ss,step} = 0.05$ gives $K_m = 13.46$. The RL and PRL of $K_mG(s)$ and $K_mG'(s)$ are shown in Fig. 6(a), where $G'(s)$ is obtained in (12) for a requested UCSPS of $\theta = 30^\circ$ and with required additional poles chosen at $s = -\omega_A = 10\omega_{\rm gc}$ of $K_mG(j\omega)$ and at $s = -5\omega_A$, to minimize their effect in the resonant frequency range. The unity-dc-gain compensator is thus

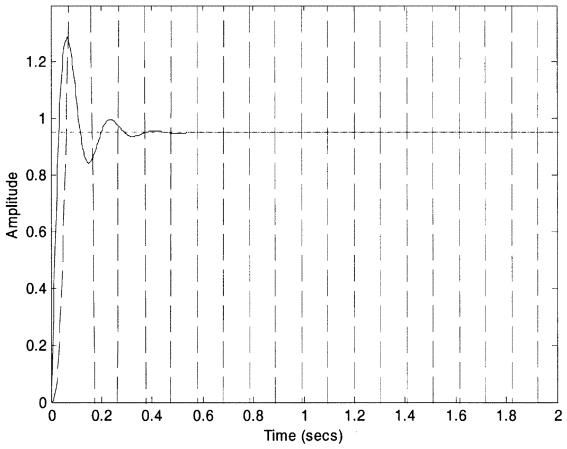
$$G_c(s) = \frac{919.7(s+9.637)(s^2+56.36s+1706.5)}{[(s+449.1)(s+2245.5)(s+15)]}.$$

With a sufficiently high sampling rate, such a controller could be implemented in software or digital hardware in a sampled-data control system.

Using $G_c(s)$ above, the $GM_{\rm dB}$ changed from the unstable -18.1 dB to 36.2 dB, and the PM changed from -41.9° to 43.0° (a far greater change than the requested 30°). In Fig. 6(a), the (unstable) closed-loop pole at A moved to B. Because (11c) is an approximation, $\Delta PM \neq \theta$ [but $\text{sgn}(\Delta PM) = \text{sgn}(\theta)$], and the PRL of $K_mG'(s)$ is not identical to that of $K_mG(s)$, as ideally would be true, though they are quite similar in the region of attained closed-loop poles; as argued above, the agreement is worst near the real axis and far from the resonant region. Also shown (dashed) is the RL of $e^{j\Delta PM}G(s)$, where $\Delta PM = 84.9^\circ$ is the achieved increase in PM. Ideally, for pure UCSPS of ΔPM, the achieved closed-loop pole would be at point C rather than point B, though in this example they are close. However, as seen from its RL in Fig. 6(a), $e^{j\Delta PM}G(s)$ is not conjugate symmetric ($e^{j\Delta PM}$ is a complex gain, as in [8]), and thus is nonphysical and cannot be implemented whereas the corresponding approximate phase-shifter $G_c(s)$ can be implemented.



(a)



(b)

Fig. 6. Conjugate-symmetric phase shifting design example: $G(s) = 3200(s+15)/[(s^2+40s+500)(s^2+4s+68)]$; $K_m = 13.46$. (a) RL and PRL for $K_mG'(s)$ (bold solid), $K_mG(s)$ (solid); RL of $\exp(j\Delta PM) \bullet G(s)$ (dashed). (b) Closed-loop step responses of $K_mG'(s)$ (solid) and $K_mG(s)$ (dashed; unstable).

Fig. 6(b) shows the closed-loop simulation results; the settling time was $t_s = 0.27$ s, and the overshoot was 35.8%; of course the requirement $e_{\rm ss,step} = 0.05$ was achieved exactly. The dashed curve is for proportional-only control for the same value of K_m (unstable result). Also, the first-order lead and lag algorithms in [4] were tried, searching for the best design with $PM_{\rm des}$ ranging from 20° to 80° in 1.2° steps. The results were that the best lag compensator produced $t_s = 6.3$ s and very large-amplitude nonovershooting oscillations on its rise; the best lead compensator produced $t_s = 0.4$ s with 71% overshoot.

As a second example, suppose $G(s) = (s+5)/[s(s+9)(s^2+4s+20)]$, which has two real poles and is type 1 ($e_{\rm ss,step} = 0$ if stable), and $K_m = 3.6$ for $e_{\rm ss,ramp} = 0.2$. The PM changed from -6.6° to $+52^\circ$ (an increase of 58.6°) when a phase shift of $+70^\circ$ was requested, and again the current method outperformed the best lead or lag result using the PM-based algorithms in [4]. Again, the PRLs before and after phase shifting are similar in the resonance (design) region. Results, including

PRL similarity and achievable phase shifts, vary for different systems (better or worse than presented here), but this technique often provides a simple way to achieve large increases in PM—often much more than a first-order compensator could provide. However, the main purpose of these examples and especially of Fig. 6(a) is to illustrate further the ideas discussed in this paper concerning PRL and PM.

IX. Conclusion

The visualizations and explanations in this paper help the student comprehend PM. Fig. 1 clarifies the relation between PM, RL, and PRL; Figs. 2 and 3 clarify the relation between RL/PRL and Bode plots for evaluating both PM and GM; Fig. 4 clarifies the evaluation of PM in the case of resonances that result in multiple gain-crossover frequencies; Fig. 5 cautions about the use of simplistic PM estimates in practical systems; Fig. 6 introduces the UCSPS concept and its application in compensator design for improvement of PM. Within the introductory level appropriate for undergraduates, the ideas and graphics in this paper stimulate student interest and may encourage independent thought about and study of the principles of control theory.

Appendix

In [1], the phase-root locus was defined as the set of contours of s for which $|K_mG(s)|$ is unity. The PRL plot was generated in [1] by evaluating $K_mG(s)$ on a dense grid and selecting the unity-magnitude contour from a call to MATLAB's "contour" routine. In this appendix a direct, more efficient and memory-saving approach is given to calculating PRL that avoids such array calculations.

One begins by setting the magnitude of $K_mG(s) = K_mB(s)/A(s)$ to unity [enforcing the PRL magnitude condition (2)] and recalling that the magnitude squared is equal to a quantity times its complex conjugate:

$$K_m^2\,\frac{B(s)B^*(s)}{A(s)A^*(s)} = 1. \tag{A1}$$

For each open-loop pole $s_{p,k}$ ($k \in [1, N]$), the associated PRL loop can be "drawn" by letting $s = s_{p,k} + re^{j\phi}$, where $0 \le \phi < 2\pi$. There is one such loop for each pole, except that these loops will overlap when K_m is sufficiently large (but there is no harm in drawing a contour twice). Write

$$A(s) = \prod_{i=1}^{N}(s - s_{p,i}),\ B(s) = \prod_{i=1}^{M}(s - s_{z,i}) \tag{A2}$$

in which $s_{z,i}$ are the open-loop zeros. Substituting $s = s_{p,k} + re^{j\phi}$ and (A2) into (A1) and crossmultiplying

$$\begin{aligned} K_m^2 \prod_{i=1}^{M}&(re^{j\phi} + s_{p,k} - s_{z,i})(re^{-j\phi} + s_{p,k}^* - s_{z,i}^*) \\ &= \prod_{i=1}^{N}(re^{j\phi} + s_{p,k} - s_{p,i})(re^{-j\phi} + s_{p,k}^* - s_{p,i}^*). \end{aligned} \tag{A3}$$

By holding ϕ fixed at a particular value, (A3) may be turned into a polynomial for r:

$$\begin{aligned} K_m^2 \prod_{i=1}^{M}&(r^2 + r\cdot 2\mathrm{Re}\{e^{-j\phi}[s_{p,k} - s_{z,i}]\} \\ &+ |s_{p,k}|^2 + |s_{z,i}|^2 - 2\mathrm{Re}\{s_{p,k}s_{z,i}^*\}) \\ =\prod_{i=1}^{N}&(r^2 + r\cdot 2\mathrm{Re}\{e^{-j\phi}[s_{p,k} - s_{p,i}]\} \\ &+ |s_{p,k}|^2 + |s_{p,i}|^2 - 2\mathrm{Re}\{s_{p,k}s_{p,i}^*\}). \end{aligned} \tag{A4}$$

In MATLAB, the expansion of (A4) into a polynomial in r is straightforward by using "conv" and subtracting polynomial coefficient arrays. The fixed value of ϕ and the roots r that are positive and real are selected, and $s = s_{p,k} + re^{j\phi}$ are displayed. Then the next value of ϕ on the range [0, 2π) is selected, and the process above is repeated. Finally, the entire process just described is repeated for all open-loop poles $s_{p,k}$.

The resulting PRL plot, obtained with a minimum of computation, is accurate and has been checked against the original procedure in [1]. The only requirement is that the system order N be small enough so that root-finding is accurate; MATLAB's routine "roots" has been found accurate up to around order 35 or so, which means for N up to about 17—clearly usually sufficient. In the event of very high order, the original procedure in [1] will always work.

One additional detail is the manner in which the contours are drawn if, as in MATLAB, contours (lines between points) rather than individual points are desired to be displayed. For the kth open-loop pole $s_{p,k}$, the minimum positive real value of r for $\phi = 0$ is selected, and the resulting $s = s_{p,k} + re^{j\phi} = s_{p,k} + r$ is called the "previous" plotted point. Of the possibly numerous positive real values of r on the next value of ϕ, the one closest to the previous r value is selected (to make continuous contours, and to avoid plotting PRL loops for other open-loop poles). Technically, for just plotting points, any one open-loop pole could be used to generate the entire set of PRL loops; but for open-loop poles far from the one selected about which ϕ is defined, the resolution is naturally reduced. Moreover, separate loops are desired for the individual open-loop poles, when appropriate. Now a line is drawn from the "previous" value of s to the current one, the "previous" value of s is set to the current value of s, and the process repeats for all ϕ and then for all $s_{p,k}$. This technique for computing PRL was used for some of the plots in this paper, but identical results are obtained using the array approach in [1].

References

[1] T. J. Cavicchi, "Phase-root locus and relative stability," *IEEE Contr. Syst. Mag.*, vol. 16, pp. 69–76, Aug. 1996.

[2] ——, (2000) Supplementary Materials for "Phase Margin Revisited: Phase-Root Locus, Bode Plots, and Phase Shifters". [Online]. Available: http://members.tripod.com/tjc2000

[3] ——, "Minimum return difference as a compensator design tool," *IEEE Trans. Education*, vol. 44, pp. 120–128, May 2001.

[4] N. S. Nise, *Control Systems Engineering*, 3rd ed. New York: Wiley, 2000.

[5] K. Ogata, *Modern Control Engineering*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1997.

[6] R. C. Dorf and R. H. Bishop, *Modern Control Systems*, 9th ed. Upper Saddle River, NJ: Prentice-Hall, 2001.
[7] C. L. Phillips and R. D. Harbor, *Feedback Control Systems*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2000.
[8] M. L. Nagurka and T. R. Kurfess, "Gain and phase margins of SISO systems from modified root locus plots," *IEEE Contr. Syst. Mag.*, vol. 12, pp. 123–127, June 1992.
[9] W. Bolton, *Control Engineering*, 2nd ed. New York: Addison-Wesley, 1998.
[10] M. Driels, *Linear Control Systems Engineering*. New York: McGraw-Hill, 1996.
[11] B. C. Kuo, *Automatic Control Systems*, 7th ed. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[12] B. Van der Pol and H. Bremmer, *Operational Calculus Based on the Two-Sided Laplace Integral*. Cambridge, U.K.: Cambridge Univ. Press, 1959, p. 114.

Thomas J. Cavicchi received the B.S. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1982, and the M.S. and Ph.D. degrees in electrical engineering from University of Illinois, Urbana, in 1984 and 1988, respectively.

He is a Professor of electrical engineering at Grove City College, Grove City, PA, where he teaches courses on digital and analog control systems and digital communication systems; he also has recently taught courses on signals and systems and electrical engineering for nonelectrical engineering majors. He is the author of *Digital Signal Processing* (New York: Wiley, 2000) and *Fundamentals of Electrical Engineering: Principles and Applications* (Englewood Cliffs, NJ: Prentice-Hall, 1993). He also coteaches the senior EE labs including the capstone senior design projects, has taught graduate classes on digital signal processing and digital spectral analysis, and has conducted research on DSP and ultrasonic diffraction scattering for medical imaging.

Dr. Cavicchi is a member of Sigma Xi.