2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

The Effect of Nonminimum-Phase Zero Locations on the Performance of Feedforward Model-Inverse Control Techniques in Discrete-Time Systems

Jeffrey A. Butterworth, Lucy Y. Pao, and Daniel Y. Abramovitch

Abstract— **Noncollocated sensors and actuators, and/or fast sample rates with plants having high relative degree, can lead to nonminimum-phase (NMP) discrete-time zero dynamics that complicate the control system design. In this paper, we examine three stable approximate model-inverse feedforward control techniques, the *nonmimimum-phase zeros ignore* (NPZ-Ignore), the *zero-phase-error tracking controller* (ZPETC) and the *zero-magnitude-error tracking controller* (ZMETC), which have frequently been used for NMP systems. We analyze how the discrete-time NMP zero locations in the z-plane affect the success of the NPZ-Ignore, ZPETC, and ZMETC model-inverse techniques. We also provide simulation examples using plants based on the system identification of an atomic force microscope and a hard disk drive, showing the tradeoffs in performance relative to NMP zero locations in these different application systems.**

I. INTRODUCTION

Model-inverse control has made several appearances in the literature demonstrating its strength in improving tracking performance, settle time, and other performance metrics; a few examples include [1]–[13]. Specifically, [14] provides a comparison of two typical control architectures that may be used to implement model-inverse control. While our discussion here applies equally to both the closed-loop-injection architecture and the plant-injection architecture described in [14], we will focus this paper on the closed-loop-injection architecture seen in Fig. 1.

When using the closed-loop-injection architecture for model-inverse control, we first design the feedback controller C of Fig. 1 to maximize performance of the stand-alone closed-loop system $H_{CL}(z) = \frac{PC}{1+PC}$. Next, we set F_{CL} equal (or *approximately* equal) to $H_{CL}^{-1}(z)$ for even greater performance gains. The definition of "performance" depends on metrics defined by the designer, but in general the inclusion of F_{CL} using model-inverse control methods will usually improve rise times, settle times, phase lag, tracking performance and more.

This work was supported in part by Agilent Technologies, Inc. and the US National Science Foundation (NSF Grant CMMI-0700877).

J. A. Butterworth is a graduate student of Electrical and Computer Engineering at the University of Colorado at Boulder, Boulder, CO 80309 USA, butterwo@colorado.edu

L. Y. Pao is a professor of Electrical and Computer Engineering at the University of Colorado at Boulder, Boulder, CO 80309 USA, pao@colorado.edu

D. Y. Abramovitch is a senior research engineer in the Nanotechnology Group at Agilent Laboratories, 5301 Stevens Creek Blvd., M/S: 4U-SB, Santa Clara, CA 95051 USA, danny@agilent.com

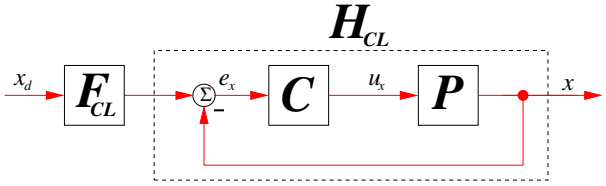


Fig. 1. A block diagram of the closed-loop-injection architecture. When model-inverse control is used, F_{CL} is designed to represent the inverse of the closed-loop system H_{CL} consisting of the feedback controller C and the plant P.

Ideally, the design of F_{CL} would yield a filter *exactly* equal to $H_{CL}^{-1}(z)$. This, for example, would reduce the transfer function from $x_d(t)$ to $x(t)$ (in Fig. 1) to unity, and allow a system's output $x(t)$ to perfectly track a desired trajectory $x_d(t)$. Often the existence of nonminimum-phase (NMP) zeros in the plant force a *stable* approximate inverse to be used in place of the exact inverse.

This paper focuses on discrete-time single-input single-output (SISO) nonmimimum-phase systems. NMP zeros in a discrete-time model can result from sampling a continuous-time system to create the model [15]. Nonminimum-phase zeros may also be a result of sensors and actuators being physically noncollocated. Due to both sources, the appearance of NMP zeros in discrete-time systems can be rather common.

Various stable approximate model-inversion techniques exist, including Tomizuka's popular *zero-phase-error tracking controller* (ZPETC) [3]. A cousin of the ZPETC is the comparatively named *zero-magnitude-error tracking controller* (ZMETC) that has appeared in [2], [5], [6], [9] and [14]. Yet another approximation method is the use of the noncausal series expansion discussed in [7], [16] and [17]. Using a zeroth-order series expansion is effectively the same as choosing to ignore the nonminimum-phase zeros (while accounting for the proper DC gain); approximating the inverse of a system in this way offers a more simplistic approach, but may not be as accurate [2], [18]. In contrast, some have chosen to use the exact unstable inverse and maintain stability of the system by pre-loading initial conditions or using noncausal plant inputs [1], [4], [19].

In general, using an exact unstable inverse method or a higher-order noncausal series expansion method introduces additional complexity when the controllers are implemented. In contrast, the zeroth-order series or nonminimum-phase

zeros ignore (NPZ-Ignore), ZPETC, and ZMETC approximation techniques are natural options for designers of model-inverse controllers. This is largely due to their documented effectiveness (especially in the case of ZPETC and ZMETC) combined with their simplicity of design and implementation [2].

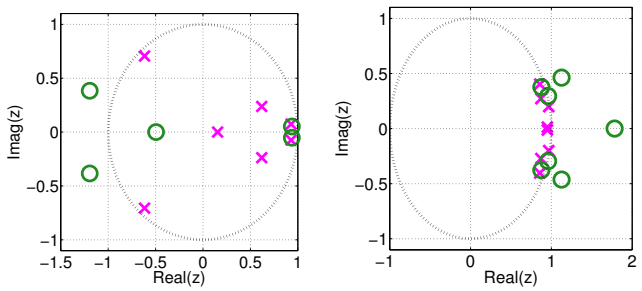
However, the proper choice of the NPZ-Ignore, ZPETC, or ZMETC techniques for maximizing performance objectives depends largely on the system on which it will be applied. In this paper, we discuss how the positioning of a system's nonminimum-phase zeros might indicate to the designer the more effective approximate model-inverse technique when considering NPZ-Ignore, ZPETC, and ZMETC. The performances of these techniques vary considerably as a function of NMP zero location.

This paper is organized as follows. In Section II, we provide further motivation for this paper by presenting the results of simulations using the NPZ-Ignore, ZPETC, or ZMETC model inverse techniques on discrete models of an atomic force microscope and a hard disk drive. The basics of each model-inverse technique are described in Section III. In Section IV, we perform an analysis on a first-order system to show the effect of nonminimum-phase zero location on the three model-inverse methods. An extension to the complex-conjugate, second-order zeros case is presented in Section V. Finally in Section VI, we discuss some additional reasons (beyond zero location) for selecting one of the three model-inverse techniques for a control system.

II. MOTIVATION

The correct choice of the NPZ-Ignore, ZPETC, or ZMETC techniques for maximizing performance objectives depends largely on the system on which it will be applied. As an example, and for motivation of this discussion, we apply the three types of model-inverse control to two closed-loop mechatronic systems in the literature. The first system is a X-Y piezoscanner that is used in the operation of an atomic force microscope (AFM). The plant model for the X direction of this device and a corresponding PID feedback controller are given in [5] which gives us enough information to study a model-inverse application of the closed-loop-injection architecture on the system. The second system is from a hard disk drive (HDD) application. An expression for the complete closed-loop HDD system appears in [2]. The models for both systems were developed through system identification processes and the pole-zero maps for both are shown in Fig. 2. The sample time for the HDD is slower than that of the AFM and as a result the two systems can not be *directly* compared. Regardless, the exact details of each are not critical to our discussion, as we are more interested in the relative location of the nonminimum-phase zeros of each system and how they affect the results of the NPZ-Ignore, ZPETC, or ZMETC techniques.

From Fig. 2, we see that both the HDD and the AFM stage have NMP zeros. The HDD has two unstable zeros in the left-half plane (LHP), while the AFM scanner has three in the right-half plane (RHP). We will see that the



(a) $H_{CL:HDD}(z)$ (b) $H_{CL:AFM}(z)$

Fig. 2. The pole-zero maps of the closed-loop HDD and AFM stage discrete-time systems. The sample time of each is 68μsec and 48μsec, respectively. The order of each is 7 (with a relative degree of 2) and 8 (with a relative degree of 1), respectively.

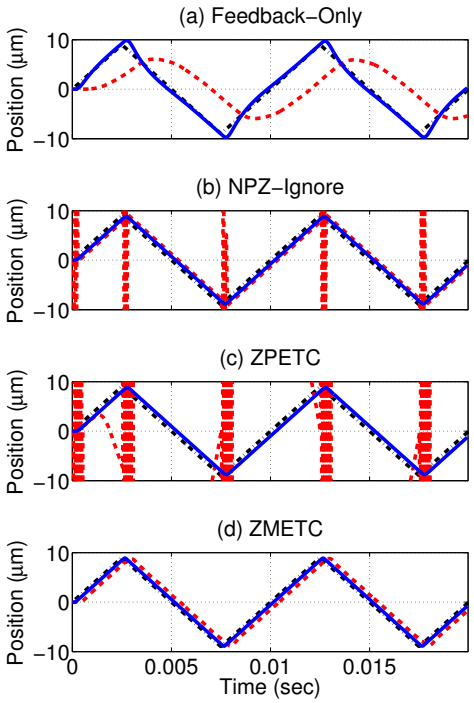


Fig. 3. The simulation results for a 100Hz raster scan input (dash-dot black) for the HDD system (solid blue) and the AFM scanner system (dashed red) for (from top to bottom) feedback-only control and the NPZ-Ignore, ZPETC, and ZMETC combined feedforward/feedback techniques.

difference here is fundamental to the choice for which of the NPZ-Ignore, ZPETC, or ZMETC techniques for model-inverse control should be used. Specifically in Fig. 3, we provide several plots of simulation results for both systems under four different control scenarios. The input x_d (dash-dot black) to both is a 100Hz raster scan (which is rather common to AFM systems [20]). The corresponding outputs x of the HDD and the AFM stage are shown in solid blue and dashed red, respectively. The top plot is a simple feedback-only simulation where $F_{CL} = 1$ for both systems. The three lower plots display simulation results for both systems when using the NPZ-Ignore, ZPETC, or ZMETC techniques for the feedforward controller F_{CL}, respectively.

Fig. 3(a) shows us that given feedback-only control, the

HDD struggles to track the peaks of the 100Hz raster scan while the AFM loses all high frequency features in the scan. In Fig. 3(b), we introduce feedforward model-inverse control in the form of the NPZ-Ignore method. Here we see the HDD (solid blue) tracking the input rather well, but the AFM scanner (dashed red) experiences high-frequency ringing that takes the trajectory out of the frame of the plot at the corners (containing the high-frequency components) of the raster scan. Despite the corners, the rest of the AFM stage scan is tracked rather well. Using the ZPETC method for both systems in Fig. 3(c), we see results similar to the NPZ-Ignore technique. The HDD continues to track well, but in this case, the AFM's trouble at the corners of the raster scan have been amplified considerably and the AFM fails to track the scan at any time. The ZMETC method (in Fig. 3(d)) relieves the AFM stage of its ringing problems and forces a slightly phase-lagged tracking of the raster scan. The HDD continues to track the raster scan despite the fundamental change in the model-inverse procedure. In the next several sections, we discuss why the AFM scanner with RHP nonminimum-phase zeros struggles to track the raster scan using the NPZ-Ignore and ZPETC techniques while the HDD with the LHP nonminimum-phase zeros tracks the scan well regardless of the feedforward method.

III. THREE STABLE APPROXIMATE MODEL-INVERSION TECHNIQUES

Defining the closed-loop system in Fig. 1 as $H_{CL}(z) = \frac{PC}{1+PC}$ and writing out the transfer function from $x_d(t)$ to $x(t)$, we arrive at (1).

$$\frac{X(z)}{X_d(z)} = H_{CL}(z)F_{CL}(z). \tag{1}$$

Assuming $H_{CL}(z)$ is exactly proper and all zeros are minimum phase, we could define $F_{CL}(z) = H_{CL}^{-1}(z)$, and achieve perfect tracking of an input signal. However, many systems are not exactly proper and have nonmimimum-phase zeros that require a stable approximate model inversion technique. The design of the three techniques of focus begin with the same basic structure. First, we write the dynamics of the system as in (2), partitioning $B(z)$ into the polynomial $B_s(z)$ containing the stable (invertible) zeros and the polynomial $B_u(z)$ containing the unstable (noninvertible) zeros:

$$H_{CL} = \frac{B(z)}{A(z)} = \frac{B_s(z)B_u(z)}{A(z)}. \tag{2}$$

The polynomial $A(z)$ contains all the poles of the closed-loop system model, and $B_u(z)$ can be written in the form of the n^{th}-order polynomial

$$B_u(z) = b_{un}z^n + b_{u(n-1)}z^{n-1} + \cdots + b_{u0}, \tag{3}$$

where n is the number of NMP zeros. We can then write the basic structure for NPZ-Ignore, ZPETC, and ZMETC as:

$$F_{CL}(z) = \tilde{H}_{CL}^{-1}(z) = \frac{z^{-q}A(z)}{B_s(z)B_u^*(z)}, \tag{4}$$

TABLE I
APPROXIMATE INVERSES & OVERALL TRANSFER FUNCTIONS FOR EACH APPROXIMATE INVERSE METHOD

METHOD	$F_{CL}(z) = \tilde{H}_{CL}^{-1}(z)$	$\frac{X(z)}{X_d(z)}$
NPZ-Ignore	$\frac{z^{-q}A(z)}{B_s(z)B_u(1)}$	$\frac{z^{-q}B_u(z)}{B_u(1)}$
ZPETC	$\frac{z^{-q}A(z)B_u^f(z)}{B_s(z)(B_u(1))^2}$	$\frac{z^{-q}B_u(z)B_u^f(z)}{(B_u(1))^2}$
ZMETC	$\frac{z^{-q}A(z)}{B_s(z)B_u^f(z)}$	$\frac{z^{-q}B_u(z)}{B_u^f(z)}$

where the $\sim$ indicates an *approximate* inverse of the system. Since the number of roots of the polynomial $A(z)$ do not always equal the number of roots of the polynomial $(B_s(z)B_u^*(z))$, q units of delay in (4) are required to ensure a causal implementation of $F_{CL}(z)$. $B_u^*(z)$ is defined depending on the type of stable model-inverse technique to be used. The precise choice of $B_u^*(z)$ will be described in detail in the following subsections.

A. The NPZ-Ignore Technique

When using the NPZ-Ignore technique, the designer ignores any nonminimum-phase zeros in the system model and makes the proper adjustments for how this might affect the DC gain of the overall system. In general, this is the least precise of all the approximate model-inverse techniques discussed here as there is no accounting for an entire portion of the system dynamics. Nevertheless, this is the simplest of the three and may have benefits if there are limited computational resources for implementing the controller.

For a NPZ-Ignore design, $B_u^*(z)$ in (4) reduces to

$$B_{u:Ign}^*(z) = B_u(z)\big|_{z=1} = B_u(1). \tag{5}$$

Here we are ignoring the nonminimum-phase zero dynamics and reducing $B_u^*(z)$ to a scalar that compensates for losses in the DC gain associated with not including the dynamics of $B_u(z)$. The resulting NPZ-Ignore design for $\tilde{H}_{CL:Ign}^{-1}(z)$ appears in Table I along with the overall transfer function that corresponds to (1). For $F_{CL}(z)$ to be causal, the delay q must be equal to the order of $A(z)$ minus the order of $B_s(z)$, $(q = O(A(z)) - O(B_s(z)))$.

B. The ZPETC Technique

The structure of the ZPETC technique is very similar to the NPZ-Ignore method, but the ZPETC is higher order because it retains the dynamics of the nonminimum-phase zeros. When using the ZPETC method, $B_u^*(z)$ in (4) becomes:

$$B_{u:ZP}^*(z) = \frac{\left(B_u(z)\big|_{z=1}\right)^2}{B_u^f(z)} = \frac{(B_u(1))^2}{B_u^f(z)} \tag{6}$$

where $B_u^f(z)$ is defined by

$$B_u^f(z) = b_{u0}z^n + b_{u1}z^{n-1} + \cdots + b_{un}. \tag{7}$$

Note that the difference between (3) and (7) is the "flipping" of the coefficients. It is this action which converts the unstable $B_u(z)$ into the stable $B_u^f(z)$.

The ZPETC design for $\tilde{H}^{-1}_{CL:ZP}(z)$ appears in Table I along with the corresponding overall transfer function. In order for the implementation of $F_{CL}(z)$ to be causal, the delay $q = O(A(z)B^f_u(z)) - O(B_s(z))$.

The reader should note that for both NPZ-Ignore and ZPETC, (1) is a finite-impulse-response (FIR) filter which can be advantageous for various applications [2], [7].

C. The ZMETC Technique

In contrast to the ZPETC method that converts NMP zeros of the model into stable *zeros* of the approximate inverse, the ZMETC technique transforms the nonminimum-phase zeros of the model into stable *poles* of the approximate inverse. When using the ZMETC method, $B^*_u(z)$ in (4) becomes:

$$B^*_{u:ZM}(z) = B^f_u(z), \tag{8}$$

where $B^f_u(z)$ is defined in (7).

The expression for $\tilde{H}^{-1}_{CL:ZM}(z)$ appears in Table I along with the overall transfer function that corresponds to (1). In this case, $q = O(A(z)) - O(B_s(z)B^f_u(z))$.

No compensation for changes in DC gain are required for the ZMETC method because the DC gain of (1) remains at unity as per design. Unlike the NPZ-Ignore and ZPETC methods, the ZMETC technique leads to an overall transfer function that is an infinite-impulse-response (IIR) filter.

IV. ANALYSIS OF A FIRST-ORDER EXAMPLE

In this section, we will break down a simple first-order example for each model-inverse technique. To start, we assume a sampling time $T_s = 1.0$ seconds and that the parameter representing the position of the zero $a \in (-\infty, -1) \cup (1, \infty)$ is real and outside the unit circle in the system

$$H_{CL} = \frac{(z-a)}{(z-p)}. \tag{9}$$

In this case, $A(z)$, $B_s(z)$, $B_u(z)$, and $B^f_u(z)$ are defined as:

$$\begin{aligned} A(z) &= (z-p),\ B_s(z) = 1, \\ B_u(z) &= (z-a),\ \text{and } B^f_u(z) = (-az+1). \end{aligned} \tag{10}$$

Using each of the three techniques described in Section III, we can define $\tilde{H}^{-1}_{CL}(z)$ for each as shown in Table II. Looking at the overall transfer functions in Table II, we see that the pole p plays no role as it has been canceled out. Recall, that as per the definition of $a \in (-\infty, -1) \cup (1, \infty)$, the equations in Table II are *not defined* when $-1 < a < 1$. This is because when $-1 < a < 1$, exact inverses can be used.

If we assume $p = 0.5$ and $a = \mp 1.1$ (to achieve LHP and RHP nonminimum-phase zeros, respectively), we can plot pole-zero maps for each model and model-inverse method (Fig. 4). As expected from Table II, we see that in all cases the transfer function $X(z)/X_d(z)$ retains the NMP zero in the dynamics. As a result, the NMP zero still influences the system's dynamics near the frequency specified by its location. Specifically, the LHP real zero in $X(z)/X_d(z)|_{Ign}$ appears at an extremely high frequency. In contrast, the RHP real zero in $X(z)/X_d(z)|_{Ign}$ appears at a frequency two orders of magnitude lower than that of its LHP counterpart. As a result, the RHP nonminimum-phase zero plays a much larger role in the dynamics of the system than the LHP zero, specifically in the form of amplification of high-frequency components of input signals (see Fig. 5).

TABLE II
APPROXIMATE INVERSES & OVERALL TRANSFER FUNCTIONS FOR THE 1st-ORDER EXAMPLE FOR EACH APPROXIMATE INVERSE METHOD

METHOD	$F_{CL}(z) = \tilde{H}^{-1}_{CL}(z)$	$\frac{X(z)}{X_d(z)}$
NPZ-Ignore	$\frac{z^{-1}(z-p)}{(1-a)}$	$\frac{(z-a)}{(1-a)z}$
ZPETC	$\frac{z^{-2}(z-p)(-az+1)}{(1-a)^2}$	$\frac{(z-a)(-az+1)}{(1-a)^2z^2}$
ZMETC	$\frac{(z-p)}{(-az+1)}$	$\frac{(z-a)}{(-az+1)}$

In Figs. 4 and 5, we see the problem with a RHP nonminimum-phase zero becomes worse when using the ZPETC method. Here the low-frequency RHP nonminimum-phase zero gets reflected into the unit circle as an additional zero. This compounds the problems of the NPZ-Ignore technique as another zero (albeit stable) gets placed at the exact same frequency of the RHP nonminimum-phase zero. This additional zero counteracts the phase drop associated with the RHP nonminimum-phase zero, but in the magnitude sense we get even more amplification of high-frequency components of input signals. In contrast, the ZPETC method suffers no ill effects of the LHP nonminimum-phase zero as it and the additional stable zero occur at such extremely high frequencies that their influence on dynamics is limited to attenuating the extremely high frequency region.

In contrast, we see in Figs. 4(e) and 4(f) that the ZMETC method reflects NMP zeros into the unit circle as poles at the same frequency. This is how this method is able to attain "*zero-magnitude* error" at all frequencies. The increasing magnitude effects of any nonminimum-phase zero is canceled by a stable pole. One should note that in Fig. 5, the ZPETC method for both LHP and RHP systems does not appear to have "*zero-phase* error" at all frequencies. This is only due to the delay added to F_{CL} to make it causal. If no delay were required (*e.g.*, $q = 0$), the ZPETC curves in Fig. 5 would remain at 0 degrees for all frequencies.

The source of these high-frequency magnitude issues becomes clear if we look at the real and imaginary parts of $X(e^{j\omega})/X_d(e^{j\omega})$ for each model-inverse method. Defining

$$\Gamma = \Re\left[\frac{X(e^{j\omega})}{X_d(e^{j\omega})}\right] \text{ and } \Phi = \Im\left[\frac{X(e^{j\omega})}{X_d(e^{j\omega})}\right]$$

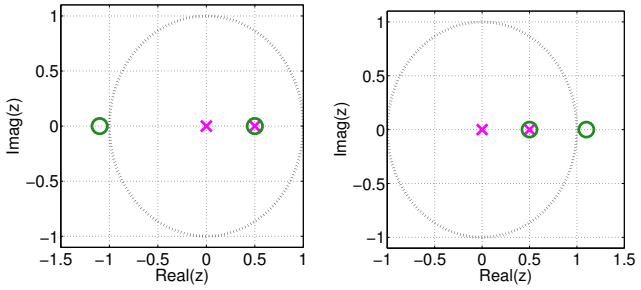
for each model-inverse method, and utilizing some trigonometric identities, we can write Γ and Φ for each method (see Table III). Looking at the magnitude for ZMETC in Table III, it again becomes clear why this method is called the *zero-magnitude-error* tracking controller as its magnitude remains at unity for all frequencies regardless of the value of a. In contrast, we see a very different behavior in the NPZ-Ignore and ZPETC methods when the parameter a is varied.

Specifically, we note that as ω approaches the Nyquist

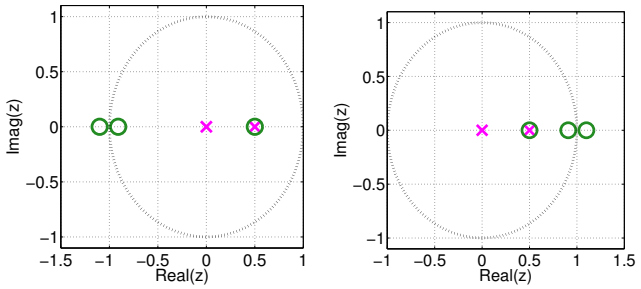
TABLE III

SUMMARY OF $X(e^{j\omega})/X_d(e^{j\omega})$ FOR THE FIRST-ORDER EXAMPLE FOR EACH APPROXIMATE INVERSE METHOD

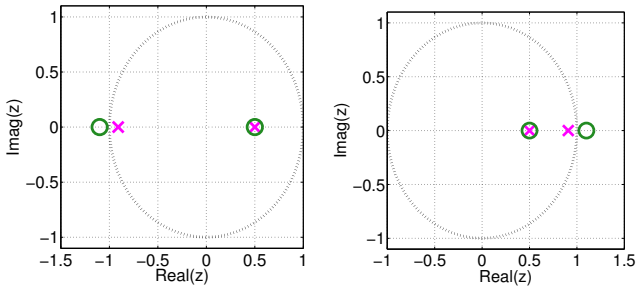
METHOD	$\Re\left[X(e^{j\omega})/X_d(e^{j\omega})\right]$	$\Im\left[X(e^{j\omega})/X_d(e^{j\omega})\right]$	MAGNITUDE $\left(\sqrt{\Gamma^2+\Phi^2}\right)$
NPZ-Ignore	$\Gamma_{Ign}=\frac{1-a\cos(\omega)}{1-a}$	$\Phi_{Ign}=\frac{a\sin(\omega)}{1-a}$	$\sqrt{\frac{1+a^2-2a\cos(\omega)}{(1-a)^2}}$
ZPETC	$\Gamma_{ZP}=\frac{\cos(\omega)}{(1-a)^2}(1+a^2-2a\cos(\omega))$	$\Phi_{ZP}=\frac{-\sin(\omega)}{(1-a)^2}(1+a^2-2a\cos(\omega))$	$\frac{1+a^2-2a\cos(\omega)}{(1-a)^2}$
ZMETC	$\Gamma_{ZM}=\frac{(1+a^2)\cos(\omega)-2a}{1+a^2-2a\cos(\omega)}$	$\Phi_{ZM}=\frac{(1-a^2)\sin(\omega)}{1+a^2-2a\cos(\omega)}$	1



(a) LHP Zero, $X(z)/X_d(z)|_{Ign}$ (b) RHP Zero, $X(z)/X_d(z)|_{Ign}$



(c) LHP Zero, $X(z)/X_d(z)|_{ZP}$ (d) RHP Zero, $X(z)/X_d(z)|_{ZP}$



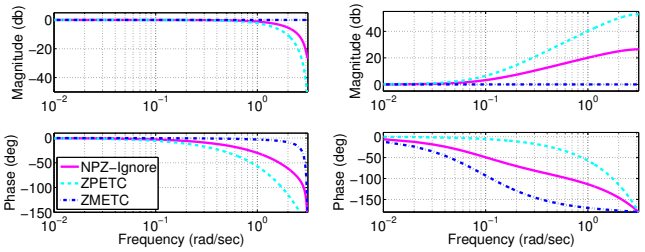
(e) LHP Zero, $X(z)/X_d(z)|_{ZM}$ (f) RHP Zero, $X(z)/X_d(z)|_{ZM}$

Fig. 4. The pole-zero maps of $X(z)/X_d(z)$ for the simple first-order example for each model-inverse method. The three plots in the left-hand column represent the systems with LHP nonminimum-phase zeros whereas the right column of plots are the contrasting systems with RHP zeros. Here, we assume $p=0.5$ and $a=\mp 1.1$.

frequency (π radians/second in the case of $T_s=1$ second), the imaginary term Φ_{Ign} goes to zero regardless of the value of a. That leaves the real term Γ_{Ign} to dominate the magnitude of $X(z)/X_d(z)|_{Ign}$ and

$$\left\|\lim_{\omega\to\pi}\Gamma_{Ign}(\omega)\right\|=\left\|\frac{1+a}{1-a}\right\|. \tag{11}$$

When $-\infty<a<-1$, we see that (11) is less than one and as a result we expect a "rolling off" of the magnitude near the



(a) LHP Zero, $a=-1.1$ (b) RHP Zero, $a=1.1$

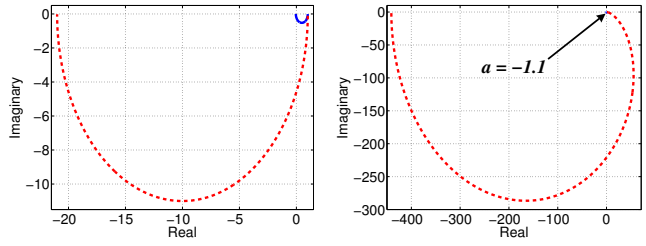
Fig. 5. The Bode plots of the transfer function $X(z)/X_d(z)$ of the first-order example for each model-inverse method with a LHP and RHP nonminimum-phase zero ($a=-1.1$ and $a=1.1$, respectively).

Nyquist frequency. In contrast, when $1<a<\infty$, (11) achieves a value greater than one that suggests amplification of any high-frequency components of the input signal. A similar analysis on the ZPETC method shows that as ω approaches π radians/second

$$\left\|\lim_{\omega\to\pi}\Gamma_{ZP}(\omega)\right\|=\left\|\lim_{\omega\to\pi}\Gamma_{Ign}(\omega)\right\|^2=\left\|\frac{1+a}{1-a}\right\|^2, \tag{12}$$

which indicates the squaring effects of a RHP nonminimum-phase zero when compared to the NPZ-Ignore method (which also can be seen in the magnitude column of Table III). We can see this behavior graphically if we look at the positive-frequency portion of the Nyquist plots of both $X(z)/X_d(z)|_{Ign}$ and $X(z)/X_d(z)|_{ZP}$ for $a=\mp 1.1$ in Fig. 6. Specifically, in Fig. 6(a) we see how the line representing the system with a RHP zero (dashed red) leaves the area near zero and converges to a real value of -21 as ω converges on π radians/second. In contrast, the line representing the system with a LHP zero (solid blue) never leaves a ball of radius one centered on the origin. When we turn to the ZPETC method in Fig. 6(b), we see that the dashed red line achieves such a large value, that the solid blue line (which always has magnitudes less than or equal to one) is so small and not visible on this scale.

Fig. 7 is a plot of the limit of the magnitude as ω approaches the Nyquist frequency of $X(z)/X_d(z)$ in (9) for each control method as a is varied. Here the region where a value of a indicating a minimum-phase zero in (9) is marked in shaded yellow. This region is not valid for comparison of the three approximate model-inversion techniques as the exact inverse can be used. Again, we see that the ZMETC method achieves unity gain for $X(z)/X_d(z)|_{ZM}$ as a is varied. We also note that for the NPZ-Ignore and ZPETC



(a) NPZ-Ignore (b) ZPETC

Fig. 6. The positive-frequency portion of the Nyquist plots of both $X(z)/X_d(z)|_{Ign}$ and $X(z)/X_d(z)|_{ZP}$ for $a = \mp 1.1$ in (9). Solid blue represents $a = -1.1$, while dashed red represents $a = 1.1$. In (b), the solid blue line ($a = -1.1$) is not visible due to the large scale of the dashed red line ($a = 1.1$). Note that the axis scales of (a) and (b) are not the same.

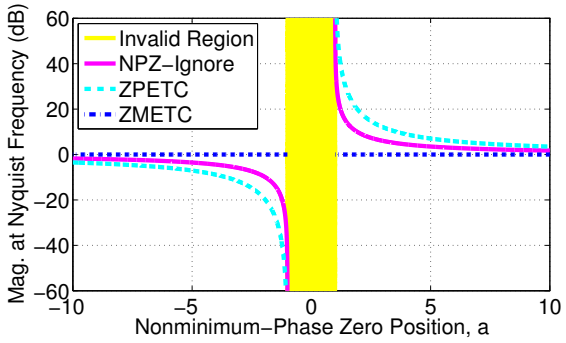


Fig. 7. The limit of the magnitude of $X(z)/X_d(z)$ as ω approaches the Nyquist frequency for the first-order example for each control method as a is varied. The shaded yellow region indicates the area where the zero would become minimum phase and is not a valid area for comparison.

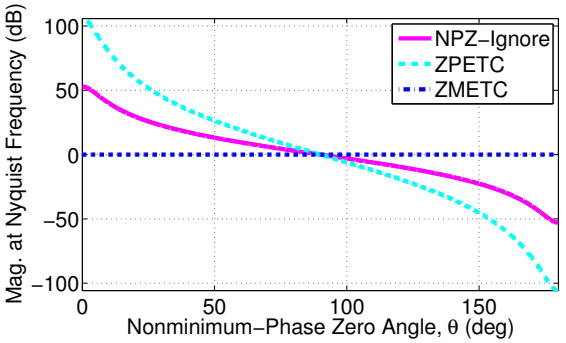


Fig. 8. The limit of the magnitude of $X(z)/X_d(z)$ as ω approaches the Nyquist frequency for the complex-conjugate zeros example for each control method as θ is varied and r_z is a constant 1.1. θ is measured from the positive-real axis and as a result, the left-hand portion of this plot represents RHP complex-conjugate zeros.

methods, we see peaking magnitudes as a approaches 1 from the right; this will serve to amplify any high-frequency components and noise in the input signal. Additionally, we see the "rolling-off" of magnitude as a approaches -1 from the left. Depending on the system, this rolling off may actually be a desired design feature of these methods because it may aid in attenuation of high-frequency noise in the system. In contrast, the ZMETC will continue to allow this high-frequency noise to pass through the system. We should note that the "rolling-off" of magnitude associated with LHP nonminimum-phase zeros in the NPZ-Ignore and ZPETC methods does not generally affect tracking magnitudes as tracking at such high-frequencies is often beyond what the physical system is capable of doing. Clearly, the effect of these NMP zeros (regardless of their existence in the LHP or the RHP) becomes trivial as $\|a\|$ becomes large.

V. EXTENSION TO COMPLEX ZEROS

Extending to an example with complex-conjugate NMP zeros, we see the same trends of the first-order system . If we define $r_z > 1$ as the distance from the origin to the complex zeros and θ as the angle (measured from the positive-real axis in radians) defining the direction of r_z, we can write

$$\begin{aligned} H_{CL} &= \frac{(z - r_z e^{j\theta})(z - r_z e^{-j\theta})}{\Psi(z)} \\ &= \frac{(z^2 - 2r_z\cos(\theta)z + r_z^2)}{\Psi(z)}, \end{aligned} \tag{13}$$

where $\Psi(z)$ represents a polynomial of stable poles of order two or more. As we discovered previously, the exact poles in the system are not of interest (assuming they are stable) as they will be canceled in any of the three model-inverse methods. The angle θ defines the LHP or RHP location of the complex zeros.

Performing an analysis similar to that for the first-order system in Section IV, we can show that the conclusions in Section IV extend to the case of complex-conjugate zeros by considering when ω approaches the Nyquist frequency similar to (11) and (12). We find that the imaginary terms Φ go to zero and the real terms dominate the magnitude:

$$\begin{aligned} \left\| \lim_{\omega\to\pi} \Gamma_{Ign}(\omega) \right\| &= \left\| \frac{(1 + r_z^2 + 2r_z\cos(\theta))}{(1 + r_z^2 - 2r_z\cos(\theta))} \right\|, \\ \left\| \lim_{\omega\to\pi} \Gamma_{ZP}(\omega) \right\| &= \left\| \lim_{\omega\to\pi} \Gamma_{Ign}(\omega) \right\|^2, \text{ and} \\ \left\| \lim_{\omega\to\pi} \Gamma_{ZM}(\omega) \right\| &= 1. \end{aligned} \tag{14}$$

These limits are only valid when $r_z > 1$ since exact inversion can be used when $r_z < 1$.

Fig. 8 shows the limit of the magnitude of $X(z)/X_d(z)$ as ω approaches the Nyquist frequency for constant $r_z = 1.1$ and varying θ. Again, we see that the ZMETC method achieves unity gain for $X(z)/X_d(z)|_{ZM}$ as θ is varied. We also note that for the NPZ-Ignore and ZPETC methods, the effect of NMP zeros becomes trivial as the zeros move close to the imaginary axis. Similar to Fig. 7, in Fig. 8 we also see the "rolling off" effect of LHP zeros for the NPZ-Ignore and ZPETC methods, and the amplification effect in the RHP.

VI. DISCUSSION AND CONCLUSIONS

Given the analysis in this paper, we have shown that:

(i) For systems with RHP NMP zeros near the unit circle, ZMETC is the only viable choice of the three feedforward methods discussed here.

(ii) When NMP zeros are far from the unit circle (in either the RHP or LHP), their affect on the feedforward methods is reduced.

(iii) If the NMP zeros are in the LHP and near the unit circle, the selection of a feedforward method should be based on the desired control objectives.

Given our analysis, an inexperienced designer of approximate model-inverse controllers might make the conclusion that he/she will *always* use the ZMETC method as it guarantees unity gain of $H_{CL}(z)\tilde{H}_{CL}^{-1}(z)$ for all ω frequencies and any LHP or RHP location of the nonminimum-phase zeros. For NMP zeros in the LHP or far outside the unit circle, other considerations may be important.

When NMP zeros are far outside the unit circle (case (ii)), one should consider computational requirements and phase-lag issues. For any implementation that has limited computational resources, the NPZ-Ignore method is the simplest and may be the best option. In terms of phase-lag issues, depending on the system model, a different amount of delay (z^{-q}) may be needed to make any of the three feedforward filters causal. This consideration could be crucial for any time critical applications [2], [8]. In contrast, some AFM scan applications are not phase critical [14], so phase delays may not be problematic.

When NMP zeros are in the LHP and near the unit circle (case (iii)), once should consider computational requirements, phase lag issues, *and* high-frequency noise concerns. As mentioned in Section IV, ZMETC propagates input noise through the system. For case (iii), the ZPETC or NPZ-Ignore methods might be a better option as they tend to "roll-off" these high-frequency regions providing attenuation to any high-frequency noise.

Fig. 3 demonstrates that of the three model-inverse techniques, any one can be applied to the HDD application with some level of success. Here, the designer can make a choice of the three based on their ultimate control goals. In general, ZMETC would not be an ideal choice for the HDD if the ZPETC provides sufficient results as the noise attenuation of the ZPETC method would likely be desired on an actual implementation. An HDD application might also have limited computational resources, so the NPZ-Ignore technique might be the best option under those circumstances. Unfortunately, the RHP zeros of the AFM limit it to only the ZMETC method, and the performance of the control system could be limited by high-frequency disturbances.

We should note that in [17], the authors discuss the effectiveness of their higher-order noncausal-series approximation for model-inverse control on a system with RHP nonminimum-phase zeros. This offers another model-inverse control method for the AFM stage system, but as we mentioned previously will introduce additional complexity into an implementation in the form of a very high-order filter. Further, in the same paper, the authors preprocess the desired trajectory signal ($x_d(t)$) with a low-pass filter in order to allow the use of the ZPETC method with a system containing RHP nonminimum-phase zeros. This may also provide an alternative to the designer, but the benefits to this approach may be detrimental to performance in the case of an AFM scanning stage in which tracking of the high-frequency components of the raster scan is equally important as the low-frequency components.

In summary, this paper has provided a better understanding of the effect of NMP zero locations on the NMP-Ignore, ZPETC, and ZMETC feedforward techniques.

REFERENCES

[1] S. Devasia, D. Chen, and B. Paden, "Nonlinear inversion-based output tracking," *IEEE Trans. Auto. Ctrl.*, vol. 41, no. 7, pp. 930–942, July 1996.

[2] B. P. Rigney, L. Y. Pao, and D. A. Lawrence, "Nonminimum phase dynamic inversion for settle time applications," *Provisionally accepted for publication in: IEEE Trans. Ctrl. Sys. Tech.*, 2008, pre-print available here: http://ece.colorado.edu/∼pao/journals.html.

[3] M. Tomizuka, "Zero phase error tracking algorithm for digital control," *ASME J. Dyn. Sys., Meas., & Ctrl.*, vol. 109, pp. 65–68, March 1987.

[4] Q. Zou and S. Devasia, "Preview-based optimal inversion for output tracking: Application to scanning tunneling microscopy," *IEEE Trans. Ctrl. Sys. Tech.*, vol. 12, no. 3, pp. 375–386, May 2004.

[5] L. Y. Pao, J. A. Butterworth, and D. Y. Abramovitch, "Combined feedforward/feedback control of atomic force microscopes," in *Proc. Amer. Ctrl. Conf.*, July 2007, pp. 3509–3515.

[6] B. Potsaid and J. T. Wen, "High performance motion tracking control," in *Proc. IEEE Int. Conf. Ctrl. Apps.*, Sept. 2004, pp. 718–723.

[7] B. P. Rigney, L. Y. Pao, and D. A. Lawrence, "Model inversion architectures for settle time applications with uncertainty," in *Proc. IEEE Conf. Dec. & Ctrl.*, Dec. 2006, pp. 6518–6524.

[8] ——, "Settle-time performance comparisons of approximate inversion techniques for LTI nonminimum phase systems," in *Proc. Amer. Ctrl. Conf.*, June 2006, pp. 600–605.

[9] J. T. Wen and B. Potsaid, "An experimental study of a high performance motion control system," in *Proc. Amer. Ctrl. Conf.*, June 2004, pp. 5158–5163.

[10] D. Croft and S. Devasia, "Vibration compensation for high speed scanning tunneling microscopy," *Rev. Sci. Instr.*, vol. 70, no. 12, pp. 4600–4605, Dec. 1999.

[11] D. Croft, G. Shedd, and S. Devasia, "Creep, hysteresis, and vibration compensation for piezoactuators: Atomic force microscopy application," *ASME J. Dyn. Sys., Meas., & Ctrl.*, vol. 123, pp. 35–43, March 2001.

[12] S. Tien, Q. Zou, and S. Devasia, "Iterative control of dynamics-coupling-caused errors in piezoscanners during high-speed AFM operation," *IEEE Trans. Ctrl. Sys. Tech.*, vol. 13, no. 6, pp. 921–931, Nov. 2005.

[13] Y. Wu and Q. Zou, "Iterative control approach to compensate for the hysteresis and the vibrational dynamics effects of piezo actuators," in *Proc. Amer. Ctrl. Conf.*, June 2006, pp. 424–429.

[14] J. A. Butterworth, L. Y. Pao, and D. Y. Abramovitch, "Architectures for tracking control in atomic force microscopes," in *Proc. IFAC World Cong.*, July 2008.

[15] K. J. Åström, P. Hagaander, and J. Sternby, "Zeros of sampled systems," *Automatica*, vol. 20, no. 1, pp. 31–38, 1984.

[16] E. Gross, M. Tomizuka, and W. Messner, "Cancellation of discrete time unstable zeros by feedforward control," *ASME J. Dyn. Sys., Meas., & Ctrl.*, vol. 116, pp. 33–38, March 1994.

[17] E. Gross and M. Tomizuka, "Experimental beam tip tracking control with a truncated series approximation to uncancelable dynamics," *IEEE Trans. Ctrl. Sys. Tech.*, vol. 2, no. 4, pp. 382–391, Dec. 1994.

[18] B. Haack and M. Tomizuka, "The effect of adding zeros to feedforward controllers," *ASME J. Dyn. Sys., Meas., & Ctrl.*, vol. 113, no. 4, pp. 6–10, March 1991.

[19] L. R. Hunt, G. Meyer, and R. Su, "Noncausal inverses for linear systems," *IEEE Trans. Auto. Ctrl.*, vol. 41, no. 4, pp. 608–611, April 1996.

[20] D. Y. Abramovitch, S. B. Andersson, L. Y. Pao, and G. Schitter, "A tutorial on the mechanisms, dynamics, and control of atomic force microscopes," in *Proc. Amer. Ctrl. Conf.*, July 2007, pp. 3488–3502.