H_∞-Based Selective Inversion of Nonminimum-phase Systems for Feedback Controls

Dan Wang and Xu Chen, *Member, IEEE*

***Abstract*— Stably inverting a dynamic system model is fundamental to subsequent servo designs. Current inversion techniques have provided effective model matching for feedforward controls. However, when the inverse models are to be implemented in feedback systems, additional considerations are demanded for assuring causality, robustness, and stability under closed-loop constraints. To bridge the gap between accurate model approximations and robust feedback performances, this paper provides a new treatment of unstable zeros in inverse design. We provide first an intuitive pole-zero-map-based inverse tuning to verify the basic principle of the unstable-zero treatment. From there, for general nonminimum-phase and unstable systems, we propose an optimal inversion algorithm that can attain model accuracy at the frequency regions of interest while constraining noise amplification elsewhere to guarantee system robustness. Along the way, we also provide a modern review of model inversion techniques. The proposed algorithm is validated on motion control systems and complex high-order systems.**

***Index Terms*—H_∞ formulation, model inversion, nonminimum-phase zeros, unstable systems.**

I. Introduction

GIVEN a linear time-invariant system model G, the inversion of G has numerous practical implementations including iterative learning control (ILC) [1]–[3], repetitive control [4], [5], two-degree-of-freedom servo in feedforward control [6], [7], as well as Youla parameterization and disturbance observer in feedback control [8]–[12]. Here, G can be an open-loop plant model or a closed-loop control system. For a minimum-phase system, G^{-1} is stable and ready to be implemented. However, for a system with nonminimum-phase (NMP or unstable) zeros, G^{-1} is unstable and cannot be implemented directly. To find a stable, rational, and causal replacement $\hat{G}^{-1}$ such that $G\hat{G}^{-1}$ approximates 1 is thus a fundamental challenge in inversion-based control designs. Such a challenge is more pronounced in discrete-time systems since 1) integrator-type plant dynamics[1], common in motion control, generate NMP zeros in their zero-order-hold (ZOH) equivalents when the sampling time is sufficiently small; 2) fractional-order delays induce unstable zeros after discretization [13].

Considering the importance and the challenge of model inversion, numerous strategies have been established in modern literature. Based on system representations and scopes of application, we can classify these strategies into two categories: frequency- and time-domain model inversions. The frequency-domain strategies focus on expressing the transfer functions of the stable inverses and hence can be used in both feedback and feedforward controls. Examples in this category include the approximate (e.g., NPZ-ignore, zero-phase-error-tracking control (ZPETC), and zero-magnitude-error-tracking control (ZMETC)) [14]–[17], the ILC-based [18]–[20], and the H_∞-based [21]–[23] model inversions. On the other hand, the time-domain strategies [24]–[27] aim at identifying the optimal control signal that minimizes the error between a given reference and the output. These time-domain algorithms are mainly used as feedforward techniques since a preview of the reference is generally not available in feedback design.

This paper studies the analysis and design of model inversion strategies in the frequency domain. Current strategies in this category aim at achieving effective model matching between $\hat{G}$ and G. Compared with the approximate and the ILC-based model inversions, the H_∞-based model inversion can automatically identify the inverse model without knowing the exact NMP zeros, which particularly benefits systems with complicated pole-zero distributions. However, when the inverses are to be implemented in feedback systems, additional considerations are needed for assuring closed-loop stability and robustness. In pursuit of bridging the gap between accurate model approximations and robust feedback performances, this study builds a new H_∞-based optimal inversion algorithm that advances the field by 1) mitigating control efforts at customized frequencies and thereby enhancing system robustness; 2) reaching high efficiency for complex high-order systems and unstable systems.

Before presenting the main algorithm, we first provide a pole-zero-map-based NMP-zero modulation by replacing high-frequency NMP zeros with stable ones in motion control applications. We verify the feasibility and limit of this intuitive modulation in achieving a stable inverse model and meanwhile capturing the low-frequency system dynamics for high-performance motion control. Then we extend this intuitive modulation to an optimal design of model inversion.

Manuscript received February 15, 2020; accepted April 2, 2020. This work was supported in part by the National Science Foundation (1953155). Recommended by Associate Editor Xianghui Cao. *(Corresponding author: Xu Chen.)*

Citation: D. Wang and X. Chen, "H_∞-based selective inversion of nonminimum-phase systems for feedback controls," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 3, pp. 702–710, May 2020.

The authors are with the Department of Mechanical Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: daw1230@uw.edu; chx@uw.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2020.1003138

[1] When actuators take forces or torques as the input and linear/angular position as the output, integrator-type plant dynamics with a relative degree not less than two show up.

There, replacing the manual adjustment with an automatic and optimal search, we develop a new H_∞-based algorithm that can attain model accuracy at the frequency regions of interest while constraining noise amplification elsewhere to guarantee system robustness. The design goals are achieved by a multi-objective formulation and an all-pass factorization that consider model matching, gain constraints, causality of transfer functions, and factorization of unstable system modes in a unified scheme. The proposed algorithm is validated on motion control systems and complex high-order systems. Moreover, along the path, we unveil previously ignored features of existing inversion strategies by developing a general frequency-domain analysis method, which also gives new insights into comparing the performances of different strategies.

The main contributions of this paper are:

1) conducting an up-to-date review of model inversion strategies and proposing a new frequency-domain analysis method;

2) analyzing the effect of an intuitive NMP-zero modulation and developing a new H_∞-based inversion algorithm;

3) validating the proposed algorithm by presenting detailed case studies with high-fidelity experimental data.

The remainder of this paper is structured as follows. Section II conducts an in-depth review of literature and proposes the new frequency-domain analysis method. Section III elucidates the effect of modulating NMP zeros. The proposed optimal inversion is presented and verified in Section IV. Section V concludes this paper.

II. Review and Comparison of Frequency-Domain Inversion Algorithms—an End-Point Perspective

The frequency-domain inversion algorithms aim at expressing the stable inverse models $F = \hat{G}^{-1}$ in the s- or z-domain (s and z are complex numbers in the Laplace transform and z-transform, respectively). $\hat{G}$ is the minimum-phase system model that approximates G and has a stable inverse. An optimal inverse model is desired for $G\hat{G}^{-1}$ to approximate 1. In this section, we review and compare three typical types of frequency-domain inversion algorithms. In addition, we unveil new features of existing algorithms by developing a general frequency-domain analysis method.

A. H_∞-Based Model Inversion

1) Algorithm

The model inversion problem for NMP systems has been solved using the H_∞ formulation [21]–[23]. For a continuous-time NMP system $G(s) = (b-s)/(b+s)$ with $b>0$, under a cost function $J = \|W(s)(1-G(s)\hat{G}^{-1}(s))\|_\infty$, where the weighting $W(s) = (k+\xi s)/(k+s)$ is a low-pass filter with $k>0$ and $0 \le \xi < 1$, the optimal inverse of $G(s)$ that minimizes J is a lead filter [22]

$$\hat{G}^{-1}(s) = \frac{k(1-\xi)(b+s)}{(k+b)(k+\xi s)} \tag{1}$$

that has high gains at high frequencies. The frequency response of the optimal $G(s)\hat{G}^{-1}(s)$ is

$$G(j\Omega)\hat{G}^{-1}(j\Omega) = \frac{k(1-\xi)(b-j\Omega)}{(k+b)(k+j\xi\Omega)} \tag{2}$$

where Ω is in rad/s.

2) Frequency-Domain Analysis

To quickly capture the essence of $G\hat{G}^{-1}$, we examine the frequency response of $G\hat{G}^{-1}$ at the two frequency endpoints (0 and ∞ for a continuous-time system or 0 and π in rad for a discrete-time system) and evaluate the characteristics of model matching.

Considering $b=2$, $\xi=0.3$, and different k's, we depict in Fig. 1 the frequency responses of (2). As Ω increases from 0 to ∞, the phase of $G(j\Omega)\hat{G}^{-1}(j\Omega)$ always goes from 0 to $-180°$ (the bottom plot of Fig. 1), and its magnitude goes from $b\left(\frac{1-\xi}{k+b}\right)$(<0 dB) to $\frac{k}{\xi}\left(\frac{1-\xi}{k+b}\right)$, monotonically. Therefore, depending on the values of k and ξ, $G(s)\hat{G}^{-1}(s)$ is a high-pass filter when $\frac{k}{\xi}\left(\frac{1-\xi}{k+b}\right) > b\left(\frac{1-\xi}{k+b}\right)$, i.e., $k > \xi b$ (0.6 in this example), a low-pass filter when $k < \xi b$, and has a constant magnitude when $k = \xi b$ (the top plot of Fig. 1).

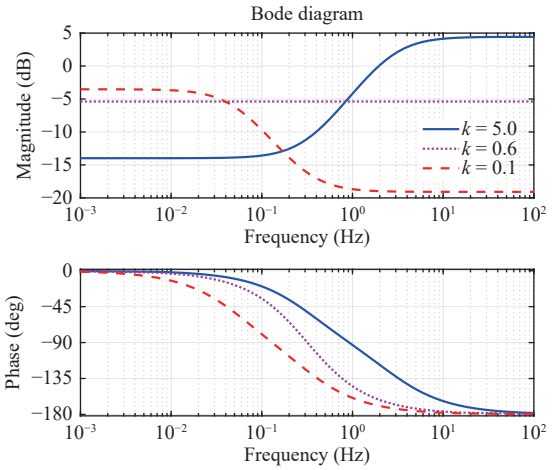

Fig. 1. Frequency responses of $G(j\Omega)\hat{G}^{-1}(j\Omega)$ with $b=2$, $\xi=0.3$, and different values of k.

B. Approximate Model Inversions

1) Algorithms

For discrete-time NMP systems, to obtain the basic structure of the inverse model, approximate model inversions [14]–[17] first factor out the unstable zeros of the system as

$$G(z) = \frac{N(z)}{D(z)} = \frac{N_s(z)N_u(z)}{D(z)} \tag{3}$$

where $N(z)$ and $D(z)$ are coprime polynomials of z, and $N_s(z)$ and $N_u(z)$ contain, respectively, the stable and the unstable zeros. Here, we define $N_u(z)$ as

$$N_u(z) = (z-z_1)(z-z_2)\dots(z-z_n) \tag{4}$$

where $z_1, z_2, \dots, z_n$ are outside the unit circle. Note that

$$N_u(z^{-1}) = (z^{-1}-z_1)(z^{-1}-z_2)\dots(z^{-1}-z_n)$$

has stable zeros. In addition, $N_u(z)N_u(z^{-1})$ is zero-phase.

In the general case, the approximate inverse model of $G(z)$ in (3) has a structure of

TABLE I

$\tilde{N}_u(z)$, $G(z)\hat{G}^{-1}(z)$, AND $Y(z)/R(z)$ IN APPROXIMATE MODEL INVERSIONS. $Y(z)$ AND $R(z)$ ARE TRANSFER FUNCTIONS OF THE OUTPUT AND REFERENCE SIGNALS SHOWN IN FIG. 2.

Method	NPZ-ignore	ZPETC	ZMETC
$\tilde{N}_u(z)$	$N_u(1)$	$\frac{[N_u(1)]^2}{N_u(z^{-1})}$	$N_u(z^{-1})$
$G(z)\hat{G}^{-1}(z)$	$\frac{N_u(z)}{N_u(1)}$	$\frac{N_u(z)N_u(z^{-1})}{[N_u(1)]^2}$	$\frac{N_u(z)}{N_u(z^{-1})}$
$\frac{Y(z)}{R(z)}$	$z^{-m}\frac{N_u(z)}{N_u(1)}$	$z^{-m}\frac{N_u(z)N_u(z^{-1})}{[N_u(1)]^2}$	$z^{-m}\frac{N_u(z)}{N_u(z^{-1})}$
$\frac{Y(e^{j\omega})}{R(e^{j\omega})}$	$e^{-jm\omega}\frac{N_u(e^{j\omega})}{N_u(1)}$	$e^{-jm\omega}\frac{N_u(e^{j\omega})N_u(e^{-j\omega})}{[N_u(1)]^2}$	$e^{-jm\omega}\frac{N_u(e^{j\omega})}{N_u(e^{-j\omega})}$
$\left\lvert\frac{Y(e^{j\pi})}{R(e^{j\pi})}\right\rvert$	$\left\lvert\frac{N_u(-1)}{N_u(1)}\right\rvert$	$\left[\frac{N_u(-1)}{N_u(1)}\right]^2$	1

$$\hat{G}^{-1}(z)=\frac{D(z)}{N_s(z)\tilde{N}_u(z)} \tag{5}$$

where $\tilde{N}_u(z)$ is a design parameter.

Table I summarizes three approximate model inversions with different designs of $\tilde{N}_u(z)$. The NMP zeros ignore method (NPZ-ignore) [16], [17] replaces $N_u(z)$ with $\tilde{N}_u(z)=N_u(1)$ at the cost of magnitude and phase mismatch in $G(z)\hat{G}^{-1}(z)$. The ZPETC [14] assigns instead $\tilde{N}_u(z)=[N_u(1)]^2/N_u(z^{-1})$ and achieves zero-phase error dynamics since $G(z)\hat{G}^{-1}(z)=N_u(z)N_u(z^{-1})/[N_u(1)]^2$ is zero-phase. The ZMETC [16], on the other hand, eliminates all magnitude errors by converting the unstable zeros to their stable reciprocals, namely, $\tilde{N}_u(z)=N_u(z^{-1})$. Note that $N_u(1)$ in NPZ-ignore and $[N_u(1)]^2$ in ZPETC are added to create a unity DC gain of $G(z)\hat{G}^{-1}(z)$.

Furthermore, to make the approximate inverse model $\hat{G}^{-1}(z)$ in (5) realizable and ready to be implemented as a block during feedback/feedforward implementation, a causal inverse model is obtained by multiplying $\hat{G}^{-1}(z)$ with z^{-m}

$$F(z)=z^{-m}\hat{G}^{-1}(z)=z^{-m}\frac{D(z)}{N_s(z)\tilde{N}_u(z)} \tag{6}$$

where

$$m=Order[Denominator\ of\ \hat{G}(z)]-Order[Numerator\ of\ \hat{G}(z)] \tag{7}$$

is the relative degree of $\hat{G}(z)$ and the *Order* function calculates the highest exponent in a transfer function. Next we will prove that m is always larger than 0 in the NPZ-ignore, ZPETC, and ZMETC.

Proof: We can tell from Table I that the relative degree of $\tilde{N}_u(z)$ is 0 in each of the three designs. Thus, from (5), the expression of m in (7) can be reduced to

$$m=Order[D(z)]-Order[N_s(z)]. \tag{8}$$

Also, we have $Order[D(z)]\geq Order[N_s(z)]+Order[N_u(z)]$ from (3) and $Order[N_u(z)]>0$ for NMP systems, yielding $Order[D(z)]>Order[N_s(z)]$, that is, $m>0$ in (8). ■

Here, the result $m>0$ means the delay z^{-m} should always be accounted for to make the inverse model realizable in the feedback/feedforward applications of approximate model inversions. In feedforward applications where a preview of the desire output $y_d(k)$ is available, the delay z^{-m} can be canceled out by letting $r(k)=y_d(k+m)$.

2) Frequency-Domain Analysis

Fig. 2 shows a block diagram to illustrate the goal of the model inversion design, where r, u, and y represent the reference, the input, and the output signals, respectively. Note that subsequently F can be implemented as a block in the feedback/feedforward controller designs, such as the examples in Section IV-C. In Fig. 2, the overall transfer function from the reference signal $r(k)$ to the output signal $y(k)$ is $Y(z)/R(z)=F(z)G(z)=z^{-m}G(z)\hat{G}^{-1}(z)$, which reflects the accuracy of the causal inverse $F(z)$. Table I lists the transfer functions of $Y(z)/R(z)$ in the three approximate model inversions. We take the hard disk drive (HDD) system in Section III as an illustrative example. The transfer function of the system with a sampling frequency ($1/T_s$) of 26.4 kHz is

$$G(z)=z^{-3}\frac{1.447663(z+0.050852)(z+2.494311)}{z^2-1.978354z+0.978808}. \tag{9}$$

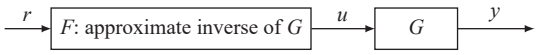

Fig. 2. Block diagram to illustrate the goal of the model inversion design. Note that F can be implemented as a feedback/feedforward controller.

Here, $G(z)$ has one NMP zero at around −2.5, $N_u(z)=z+2.494311$, and m in (8) is 4. $Y(z)/R(z)$ are $z^{-4}(z+2.494311)/3.494311$ for NPZ-ignore, $z^{-4}(z+2.494311)(z^{-1}+2.494311)/3.494311^2$ for ZPETC, and $z^{-4}(z+2.494311)/(z^{-1}+2.494311)$ for ZMETC. Fig. 3 plots the frequency responses of $Y(z)/R(z)$ of the three approximate designs. At low frequencies close to 0, i.e., $z=e^{j\omega}\to 1$, we get the desired result $Y(z)/R(z)\to 1$ for all three methods, and thereby the magnitude and phase responses of $Y(z)/R(z)$ largely overlap with each other (Fig. 3). At the Nyquist frequency π rad (i.e., 13.2 kHz), where $z=e^{j\pi}$, $\left|Y(e^{j\pi})/R(e^{j\pi})\right|$ equals $|N_u(-1)/N_u(1)|$ for NPZ-ignore and equals $[N_u(-1)/N_u(1)]^2$ for ZPETC; that is to say, in log scale, $Y(e^{j\pi})/R(e^{j\pi})$ in ZPETC (−14.72 dB) has twice the magnitude of $Y(e^{j\pi})/R(e^{j\pi})$ in NPZ-ignore (−7.36 dB) (the top plot of Fig. 3). Moreover, in this HDD example, since the NMP zero is a real one at around −2.5 and $m=4$, all three $Y(e^{j\pi})/R(e^{j\pi})$

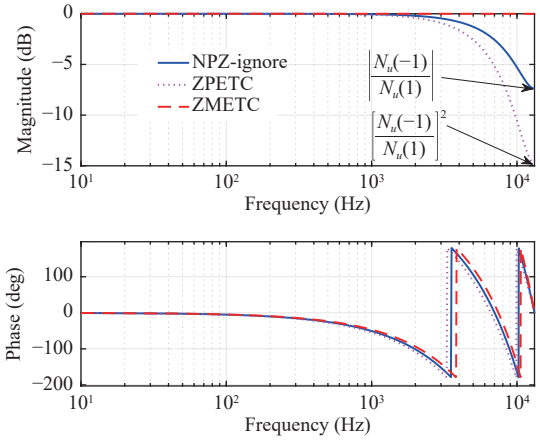

Fig. 3. Frequency responses of $Y(z)/R(z)(=z^{-m}G(z)\hat{G}^{-1}(z))$ (indicating tracking performances) for different approximate model inversions used in the example of the HDD system in (9).

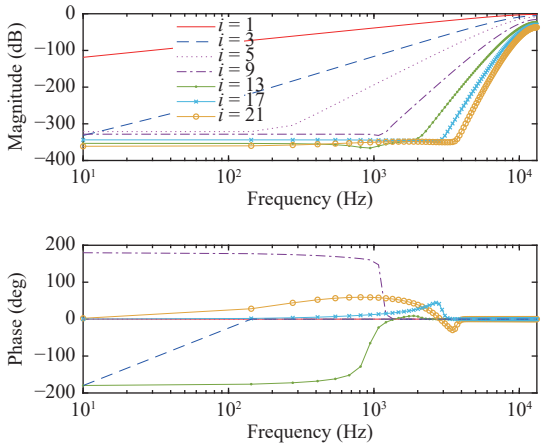

Fig. 4. Frequency responses of $(1-L(z)G(z))^i$ for the example of the HDD system in (9), where $L(z)$ is the learning filter built from ZPETC.

have zero phase at the Nyquist frequency (the bottom plot of Fig. 3).

C. ILC-Based Model Inversion

1) Algorithm

ILC, originally developed for output tracking in repetitive tasks, can be extended to the field of model inversion [18]–[20]. Here, the inverse model $F(z)$ is constructed by designing its impulse response $f(k)$ as the feedforward signal in the following ILC

$$F(z)=\sum_{k=-N/2}^{N/2} f(k)z^{-k}$$
$$f(k)=\lim_{i\to\infty} u_i(k) \tag{10}$$

where $u_i(k)$ is the learned input at the i-th iteration

$$\begin{aligned} u_i(k) &= u_{i-1}(k)+L(z)[r(k)-G(z)u_{i-1}(k)] \\ &= \left[I-(I-L(z)G(z))^i\right]G^{-1}(z)r(k). \end{aligned} \tag{11}$$

Here, the training reference $r(k)$ is designed as the delta impulse $\delta(k)$. The ILC learning filter $L(z)$ is built from the approximate model inversions (Section II-B) such that the stability condition $\|1-L(z)G(z)\|_\infty<1$ is satisfied. With $i\to\infty$, from (10) and (11), $f(k)\to u_\infty(k)\to G^{-1}(z)\delta(k)$, that is, $f(k)$ approximates the impulse response of the unstable $G^{-1}(z)$. Recall that $f(k)$ is the impulse response of $F(z)$. Thus, we obtain $F(z)\approx G^{-1}(z)$.

2) Frequency-Domain Analysis

In the ILC-based model inversion, the transfer function $1-L(z)G(z)$ determines not only the stability condition but also the convergence rate. Fig. 4 shows the frequency responses of $(1-L(z)G(z))^i$, taking again the HDD system in (9) for example. Here, $L(z)$ is built from ZPETC. With increasing iteration number i, the magnitudes of $(1-L(z)G(z))^i$ at low frequencies start to converge to zero. Moreover, a larger i yields a wider low-frequency region with zero magnitude. Therefore, under finite implementation of i, $F(z)$ represents a low-pass approximation of $G^{-1}(z)$ with a tunable bandwidth. One drawback, however, is that system hardware (or a very accurate model G) is needed for iterative experiments to run.

D. Summary of Literature Review and Motivations of This Paper

Table II summarizes the three model inversion strategies. It is noteworthy that these frequency-domain strategies can be implemented in both feedback and feedforward controls. Application of each method certainly depends on the specific problem at hand. Compared with the other two methods, the H_∞-based model inversion can automatically identify the inverse model without knowing the exact NMP zeros, which particularly benefits unstable systems and high-order systems with complicated pole-zero distributions.

For inverse-based feedback control, all the surveyed algorithms have considered accurate model inversion but not robustness against model mismatch that is also crucial for closed-loop performance. In contrast, the algorithm to be proposed in Section IV enhances the system robustness by limiting the magnitude of the inverse model at frequency regions where large model mismatches exist. Before discussing the main algorithm, we provide in Section III some preparatory work on the effect of the NMP zeros.

III. Frequency-Domain Implications of Modulating NMP Zeros

This section studies the influence of modulating the NMP zeros (i.e., shifting the locations of the NMP zeros) on the frequency response of a system. For concreteness, we take the HDD system in [10] as an example, where model inversion underpins servo designs that control precisely the position of the read/write head to provide reliable storage.

The solid line in Fig. 5 shows the frequency response of an experimentally measured HDD system. The nominal model of the motors and actuators in the system is [10]

$$G_c(s)=e^{-10^{-5}s}\frac{3.74488\times10^9}{s^2+565.487s+3.19775\times10^5}. \tag{12}$$

The ZOH equivalent of $G_c(s)$ sampled at 26.4 kHz, namely $G(z)$, is expressed in (9) and has one unstable zero at around -2.5. As plotted in Fig. 5, the frequency response of the NMP

TABLE II
OVERVIEW OF FREQUENCY-DOMAIN INVERSION STRATEGIES: APPROXIMATE [14]–[17], ILC-BASED [18]–[20], AND H_∞-BASED METHODS [21]–[23]. DT AND CT ARE SHORT FOR DISCRETE TIME AND CONTINUOUS TIME, RESPECTIVELY.

Method	DT or CT	Basic structure or design
Approximate	DT	$F(z)=z^{-m}\frac{D(z)}{N_s(z)\bar{N}_u(z)}$
ILC-based	DT	$F(z)=\sum_{k=-N/2}^{N/2} f(k)z^{-k}$, $f(k)=\lim_{i\to\infty} u_i(k)$
H_∞-based	CT/DT	$\min\|\|W(s)(1-G(s)\hat{G}^{-1}(s))\|\|_\infty$
Proposed H_∞-based	DT	$\min_{F(z)\in\mathcal{S}}\left\|\left\| \begin{matrix} W_1(z)(F(z)G(z)-z^{-m}) \\ W_2(z)F(z)G(z) \end{matrix} \right\|\right\|_\infty$ with $F(z)=z^{-m}\hat{G}^{-1}(z)$

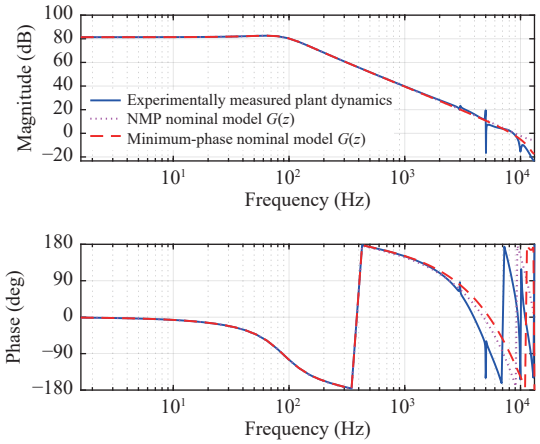

Fig. 5. Frequency responses of actual system dynamics from experiments and nominal system models in the HDD system.

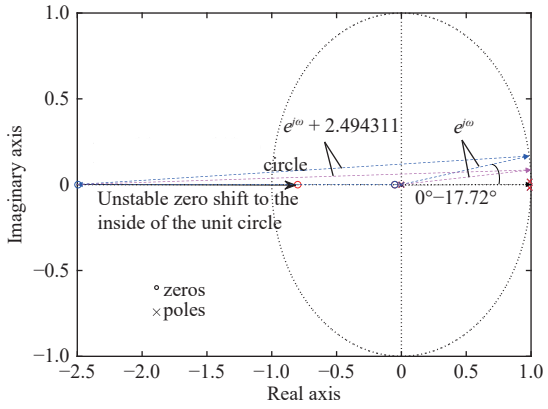

Fig. 6. Illustration of modulating the experimentally identified NMP zero in the HDD system.

$G(z)$ matches well with the actual system dynamics (solid line).

We investigate next the frequency-domain implications of the NMP-zero locations by analyzing $N_u(e^{j\omega})=e^{j\omega}+2.494311$ in (9). Consider the rule of thumb that the closed-loop bandwidth B_p is around 10% of the Nyquist frequency ($1/2T_s$ Hz) or $\omega_p=2\pi B_p T_s\approx 2\pi\frac{0.1}{2T_s}T_s=18°$; in this example, $B_p=1300$ Hz, and $\omega_p=2\pi\times 1300/26\,400=17.72°$. In other words, ω sweeps only a small arc on the unit circle from 0 to 17.72° in the main performance region, yielding mild changes to the vector $e^{j\omega}+2.494311$, as shown in Fig. 6. Therefore, when shifting the NMP zero to a stable one, e.g., at -0.8 (Fig. 6), we can get a minimum-phase nominal model $\hat{G}_0(z)$ that has a stable inverse and largely maintains frequency response of the system in desired low-frequency regions

$$\hat{G}_0(z)=z^{-3}\frac{1.447663(z+0.050852)(z+0.8)}{z^2-1.978354z+0.978808}.$$

Normalizing $\hat{G}_0(z)$ to retain the DC gain of $G(z)$ in (9), we get

$$\hat{G}(z)=z^{-3}\frac{(z+0.050852)(z+0.8)}{0.355831z^2-0.703959z+0.348290}. \tag{13}$$

As shown in Fig. 5, $\hat{G}(e^{j\omega})$ (dashed line) matches well with the NMP $G(e^{j\omega})$ (dotted line) and the actual system dynamics (solid line) below 3000 Hz. This frequency is large enough for most servo-enhancement schemes in single-stage HDDs.

In summary, a stable inverse is readily achievable through modulating the NMP zeros as long as the NMP zeros do not occur in the desired low-frequency regions. This result justifies the basic idea of the H_∞-based optimal inversion, where the manual modulation is upgraded to an automatic and optimal search, as shall be proposed next.

IV. PROPOSED H_∞-BASED OPTIMAL INVERSION

Based on the frequency-domain analysis in Section III, this section develops an H_∞-based optimal inversion. The design principle is to automatically search for the optimal inverse model to selectively fit different frequency regions. At frequencies where no NMP zeros exist and no large model uncertainties occur, we impose an accurate model matching between the minimum-phase model $\hat{G}(z)$ and the original NMP model; at other frequencies, we limit the magnitude response of the inverse model to increase the system robustness. We explore the design procedures, case studies, and frequency-domain analyses of the proposed algorithm, first for NMP systems and then for unstable systems.

A. H_∞-Based Optimal Inversion for NMP Systems

1) Algorithm

Let $\mathcal{S}$ denote the set of stable, proper, and rational discrete-time transfer functions. We search among $\mathcal{S}$ to find the optimal inverse model $F(z)=z^{-m}\hat{G}^{-1}(z)$ that satisfies

a) $F(z)$ is realizable/proper: This relates to the z^{-m} term in $F(z)$. To minimize the delays, m can be tuned and usually equals the relative degree of $G(z)$.

b) Model matching: $\min\|W_1(z)(F(z)G(z)-z^{-m})\|_\infty$. Namely, we minimize the maximum magnitude of the model mismatch

$F(z)G(z)-z^{-m}$ weighted by $W_1(z)$. The weighting $W_1(z)$ determines the frequency regions for accurate model matching. If $G^{-1}(z)$ is stable, the direct solution is $F(z)=z^{-m}G^{-1}(z)$.

c) Gain constraint: $\min\|W_2(z)F(z)G(z)\|_\infty$. Here, the magnitude of $F(z)G(z)$ is scaled by the weighting $W_2(z)$. For instance, $W_2(z)$ can be a high-pass filter to constrain noise amplification at high frequencies. The solution for this condition alone is $F(z)=0$, that is, $F(z)$ does not amplify any input signals.

Integrating the above three goals yields the multi-objective optimization principle

$$\min_{F(z)\in\mathcal{S}}\left\|\begin{bmatrix} W_1(z)(F(z)G(z)-z^{-m}) \\ W_2(z)F(z)G(z)\end{bmatrix}\right\|_\infty. \tag{14}$$

The optimal inverse model $F(z)$ given by (14) preserves accurate model information in the frequency regions specified by $W_1(z)$ and, on the other hand, penalizes excessive high gains of $F(z)$ at frequencies determined by $W_2(z)$. Typically, $W_1(z)$ is a low-pass filter, and $W_2(z)$ is a high-pass one, as shown in the example of Fig. 7. For one system model, the weightings can be flexibly designed, yielding different inverse models $F(z)$.

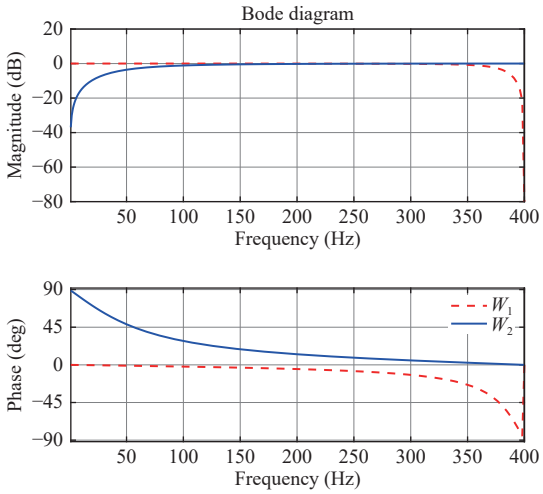

Fig. 7. Frequency responses of the weightings W_1 and W_2 in the active suspension system.

The optimization principle in (14) can be solved within the framework of H_∞ controls. $F(z)$ can be solved by the *hinfsyn* function in the robust control toolbox of MATLAB and tuned for the target performance by changing the input arguments *gamTry* and *gamRange* of the function. Fig. 8 shows the block diagram realization of (14). Here, the *hinfsyn* function minimizes the two error signals e_1 and e_2. The solution of $F(z)$ exists as long as $G(z)$, $W_1(z)$, and $W_2(z)$ are stable. After (14) is solved, a lower-order $F(z)$ can be reached by applying standard model-reduction techniques, if needed.

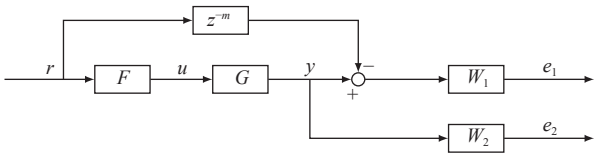

Fig. 8. Block diagram for the H_∞-based optimal inverse design.

Remark: When the system model is subjected to perturbations, we can use a multiplicative uncertainty model to lump the various dynamic uncertainties.

$$G_p(z)=G(z)(1+W_I(z)\Delta_I(z)) \tag{15}$$

where $\|\Delta_I\|_\infty\le 1$ [28]. Fig. 9 shows the block diagram of the proposed H_∞-based optimal inverse with uncertainties taken into consideration. The problem now is to find a stabilizing inverse model $F(z)$ such that the H_∞ norm of the transfer function between r and $[e_1, e_2]^T$ is less than 1 for all Δ_I, that is

$$\min_{F(z)\in\mathcal{S}}\left\|\begin{bmatrix} W_1(z)(F(z)G_p(z)-z^{-m}) \\ W_2(z)F(z)G_p(z)\end{bmatrix}\right\|_\infty \tag{16}$$

which is no longer a standard H_∞ optimization but a robust performance problem. The μ-synthesis and *DK*-iteration procedures can be utilized to solve the problem [28], [23].

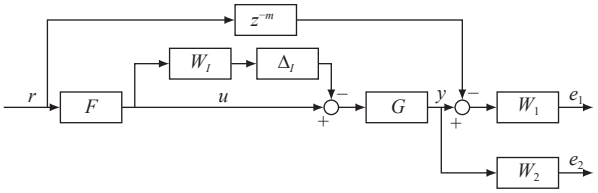

Fig. 9. Block diagram for the H_∞-based optimal inverse design considering uncertainty.

2) Case Study With Frequency-Domain Analysis

This case study shows efficiency of the proposed algorithm for high-order NMP systems with complicated pole-zero distributions. We take for example the active suspension system in [29] that serves as a benchmark on adaptive regulation. The control goal there is to attenuate the vibrations transmitted to the base frame, and model inversion is critical for the best results achieved in the benchmark [30]. Although the system is open-loop stable, the existence of the NMP zeros challenges model inversion in general feedback and feedforward control.

Via standard system identification methods, the system model $G(z)$ is experimentally identified with a sampling rate of 800 Hz and has an order of 22. As shown in the pole-zero plot in Fig. 10, four NMP zeros show up in $G(z)$. Furthermore, with the two weighting functions designed as in Fig. 7, we solve the optimization principle in (14) and obtain the optimal inverse $F(z)$. After that, we reduce the order of $F(z)$ to 23 by applying the model-reduction function *reduce* in MATLAB. The pole-zero plot of the 23rd-order $F(z)$ is also shown in Fig. 10. Then the minimum-phase system model is secured by $\hat{G}(z)=z^{-m}F^{-1}(z)$ $(m=2)$.

As shown in Fig. 11, $\hat{G}(z)$ obtained from the proposed H_∞-based optimal inversion (red dashed line) matches well with the identified NMP $G(z)$ (blue solid line). Moreover, at high frequencies near the Nyquist frequency, $\hat{G}(z)$ from the proposed method (red dashed line) has higher magnitudes than that from the existing H_∞-based method without the gain-

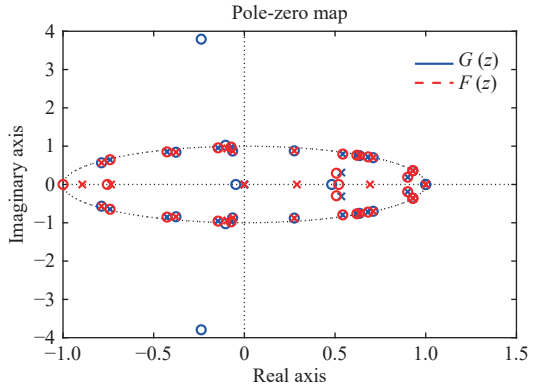

Fig. 10. Pole-zero plot of the experimentally identified system model and its minimum-phase approximation of the active suspension system.

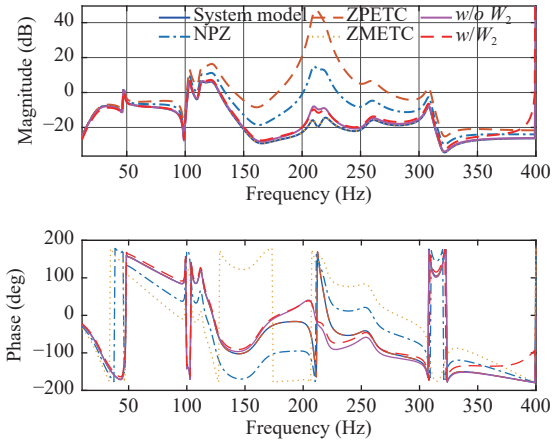

Fig. 11. Frequency responses of the experimentally identified system model $G(z)$ and its minimum-phase approximations of the active suspension system. Models obtained from ZMETC and ZPETC, respectively, have the same magnitude and phase responses as the system model. Proposed H_∞-based optimal inversion: red dashed line. Previous H_∞-based method without gain constraint: magenta solid line.

constraint condition (magenta solid line). That is to say, the second weighting W_2 has served to limit the magnitudes of the inverse model $F(z)$, as it was designed to. Fig. 11 also brings the approximate methods (Section II-B) into comparison. The minimum-phase model $\hat{G}(z)$ from ZMETC has the same magnitude response as the system model $G(z)$ but has large phase errors, whereas ZPETC yields a $\hat{G}(z)$ with no phase error but large magnitude mismatch. $\hat{G}(z)$ obtained from NPZ-ignore has large errors in both magnitude and phase. The proposed H_∞-based optimal inversion outperforms the other methods by not only striking a balance between magnitude and phase matches but also mitigating control efforts (i.e., magnitudes of $F(z)$) at high frequencies for system robustness.

B. H_∞-Based Optimal Inversion for Unstable Systems

1) Algorithm

For unstable $G(z)$, Fig. 8 and (14) are ill conditioned, and the MATLAB function *hinfsyn* returns an empty solution of $F(z)$. The first intuition for applying the H_∞-based optimal inversion is perhaps to ignore the unstable poles of $G(z)$ and take the remaining part as a fictitious system model. However, ignoring the unstable poles alters the relative degree of the system and may generate a non-causal system. Furthermore, numerical issues may arise after changing the magnitudes of the system. To overcome these difficulties, this section introduces an approach by using an all-pass factorization.

We first factor out the unstable poles of $G(z)$.

$$G(z) = z^{-m}G_0(z)\prod_i \frac{1}{z+p_i} \tag{17}$$

where $|p_i| > 1$ and $G_0(z)$ contains all the zeros and stable poles of $G(z)$.

Performing the all-pass factorization gives

$$G(z) = G_s(z)\prod_i \frac{\bar{p}_i z+1}{z+p_i} \tag{18}$$

$$G_s(z) = z^{-m}G_0(z)\prod_i \frac{1}{\bar{p}_i z+1} \tag{19}$$

where $\bar{p}_i$ is the complex conjugate of p_i. Here, the unstable poles in $G(z)$ are replaced by their reciprocals in $G_s(z)$. The product term $\prod_i(\bar{p}_i z+1)/(z+p_i)$ in (18) has unity magnitude, that is, the stable $G_s(z)$ has the same magnitude response as the unstable $G(z)$. Then we can substitute $G(z)$ with $G_s(z)$ when implementing the procedure proposed in Section IV-A.

For unstable systems, the design steps of the H_∞-based optimal model inversion are modified as

a) Write the pole-zero representation of $G(z)$, determine the relative degree m of $G(z)$, and then factor out the unstable poles as in (17);

b) Perform the all-pass factorization by transforming $G(z)$ in (17) to $G_s(z)$ in (19);

c) Substitute $G_s(z)$ into (14), and solve (14) to find $F_s(z) = z^{-m}\hat{G}_s^{-1}(z)$;

d) Take into account the effect of the unstable poles in (18) by $F(z) = F_s(z)\prod_i(z+p_i)/(\bar{p}_i z+1)$. The minimum-phase system model is then $\hat{G}(z) = z^{-m}F^{-1}(z)$.

2) Case Study With Frequency-Domain Analysis

In this case study, we show how to implement the H_∞-based optimal inversion in unstable systems.

Consider a discrete-time transfer function

$$G(z) = \frac{z^{-1}(z+1.5)}{z-1.2} \tag{20}$$

with a relative degree of $m = 1$ and a sampling rate of 26.4 kHz. $G(z)$ contains an unstable pole 1.2 at low frequency and an unstable zero –1.5 at high frequency. Following the aforementioned design steps for unstable systems, we first get

$$G_s(z) = \frac{z^{-1}(z+1.5)}{(1-1.2z)}. \tag{21}$$

Substituting the stable $G_s(z)$ into the *hinfsyn* function yields a nonempty solution of $F_s(z)$ that satisfies the optimization principle in (14): $F_s(z) = z^{-m}\hat{G}_s^{-1}(z)$. Here, we design the weighting functions as

$$W_1(z) = \frac{0.5138z+0.5137}{z+0.0264} \quad W_2(z) = \frac{z-0.6423}{z-0.2846}$$

using the MATLAB function *makeweight*. The obtained $F_s(z)$ is further normalized to have the same magnitude as the unstable $G_s^{-1}(z)$ at 800 Hz. The inverse filter is thus given by $F(z) = F_s(z)(z-1.2)/(1-1.2z)$. Using *minreal* in MATLAB, we reduce the order of the inverse filter $F(z)$ from 6 to 3 and obtain

$$F(z) = \frac{0.7439z^3 - 1.086z^2 + 0.227z + 0.006236}{z^3 + 0.5056z^2 - 0.1335z - 0.003618}.$$

The minimum-phase system model is thereby $\hat{G}(z) = z^{-1}F^{-1}(z)$. As shown in Fig. 12, $\hat{G}(z)$ (dashed line) matches well with $G(z)$ (solid line) particularly at frequencies below 5000 Hz, which is large enough for general feedback designs. Besides, compared with the existing H_∞-based method (dotted line), near the Nyquist frequency, the high gain of $\hat{G}(z)$ from the proposed method (dashed line) indicates a small magnitude of $F(z)$, which matches with the gain-constraint design criterion in Section IV-A.

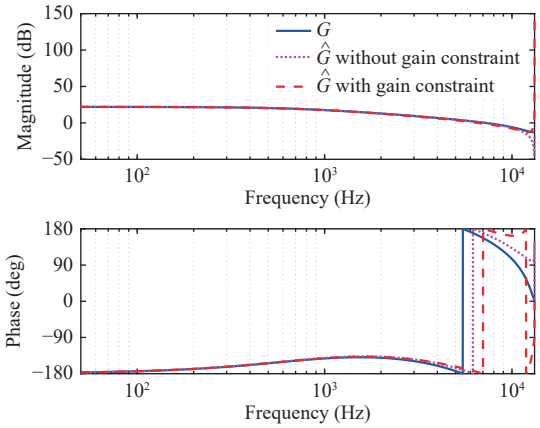

Fig. 12. Frequency responses of the system model $G(z) = z^{-1}(z+1.5)/(z-1.2)$ and its minimum-phase approximations. Proposed H_∞-based approach: red dashed line. Previous H_∞-based method without gain constraint: magenta dotted line.

C. Feedback Applications of the Proposed Algorithms

Model inversion is fundamental to subsequent servo designs, such as Youla-Kucera parameterization and adaptive disturbance observers [8]–[12]. This section provides application examples that experimentally verify the preliminary NMP-zero modulation (Section III) and the H_∞-based optimal inversion (Section IV).

In laser-based additive manufacturing, a galvo scanner system applies mirrors to reflect input laser beams to follow predefined trajectories. In [12], the authors first identify experimentally the NMP system model. After that, the minimum-phase model is obtained by moving the unstable zero from –4.419 to –0.6. Based on the minimum-phase model, [12] builds an outer-loop inverse-based Youla-Kucera parameterization scheme to reject single-frequency narrow-band disturbances.

Reference [10] studies the track-following problem in a single-stage HDD system. The system model in (9) has one NMP zero, which is shifted inside the unit circle to make the inverse model strictly stable, as shown in Fig. 6. Then with the stable inverse model, [10] designs an adaptive disturbance observer based on the internal model principle to reject multiple narrow-band disturbances.

In the active suspension benchmark discussed in [30], the minimum-phase model (red dashed line in Fig. 11) is obtained by applying the proposed H_∞-based optimal inversion. The model is then used to build an adaptive disturbance observer with an infinite impulse response structure to reject unknown or time-varying narrow-band vibrations.

V. Conclusion

In this paper, we discussed new frequency-domain analysis and design approaches to invert a nonminimum-phase (NMP) linear time-invariant system, with a focus on robustness and needed design constraints in feedback implementations. We reveal that among existing model inversion techniques, the H_∞-based method stands out by automatically identifying the inverse model without knowing the exact NMP zeros. Furthermore, we illustrated that modulating the location of the NMP zero only changes the system response at selective frequency regions. Leveraging this fact, for general NMP systems, we propose a discrete-time H_∞-based optimal inversion to automatically design the inverse model for selective frequency regions defined by two weighting functions. Verifications in complex high-order systems and unstable systems show the strengths of the proposed algorithm.

References

[1] D. Bristow, M. Tharayil, and A. G. Alleyne, "A survey of iterative learning control," *IEEE Control Systems Magazine*, vol. 26, no. 3, pp. 96–114, 2006.

[2] D. Shen, "Iterative learning control with incomplete information: a survey," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 5, pp. 885–901, 2018.

[3] R. de Rozario and T. Oomen, "Data-driven iterative inversion-based control: achieving robustness through nonlinear learning," *Automatica*, vol. 107, pp. 342–352, 2019.

[4] D. Wang and X. Chen, "A multirate fractional-order repetitive control for laser-based additive manufacturing," *Control Engineering Practice*, vol. 77, pp. 41–51, 2018.

[5] S. Zhu, X. Wang, and H. Liu, "Observer-based iterative and repetitive learning control for a class of nonlinear systems," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 5, pp. 990–998, 2018.

[6] Y. Li and M. Tomizuka, "Two-degree-of-freedom control with robust feedback control for hard disk servo systems," *IEEE/ASME Trans. Mechatronics*, vol. 4, no. 1, pp. 17–24, 1999.

[7] C. Wang, M. Zheng, Z. Wang, and M. Tomizuka, "Robust two-degreeof-freedom iterative learning control for flexibility compensation of industrial robot manipulators," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2016, pp. 2381–2386.

[8] T. Jiang, H. Xiao, J. Tang, L. Sun, and X. Chen, "Local loop shaping for rejecting band-limited disturbances in nonminimum-phase systems with application to laser beam steering for additive manufacturing," *IEEE Trans. Control Systems Technology*, pp. 1–14, 2019.

[9] K. Ohnishi, "Robust motion control by disturbance observer," *J. the Robotics Society of Japan*, vol. 11, no. 4, pp. 486–493, 1993.

[10] X. Chen and M. Tomizuka, "A minimum parameter adaptive approach for rejecting multiple narrow-band disturbances with application to hard disk drives," *IEEE Trans. Control Systems Technology*, vol. 20, no. 2, pp. 408–415, 2011.

[11] A. Apte, U. Thakar, and V. Joshi, "Disturbance observer based speed control of PMSM using fractional order PI controller," *IEEE/CAA J.*

Autom. Sinica, vol. 6, no. 1, pp. 316–326, 2019.

[12] D. Wang and X. Chen, "A tutorial on loop-shaping control methodologies for precision positioning systems," *Advances in Mechanical Engineering*, vol. 9, no. 12, pp. 1–12, 2017.

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

[14] M. Tomizuka, "Zero phase error tracking algorithm for digital control," *ASME J. Dynamic Systems*, *Measurement*, *and Control*, vol. 109, no. 1, pp. 65–68, 1987.

[15] L. Dai, X. Li, Y. Zhu, and M. Zhang, "Quantitative analysis on tracking error under different control architectures and feedforward methods," in *Proc. IEEE American Control Conf.*, 2019, pp. 5680–5686.

[16] J. A. Butterworth, L. Y. Pao, and D. Y. Abramovitch, "Analysis and comparison of three discrete-time feedforward model-inverse control techniques for nonminimum-phase systems," *Mechatronics*, vol. 22, no. 5, pp. 577–587, 2012.

[17] J. A. Butterworth, L. Y. Pao, and D. Y. Abramovitch, "The effect of nonminimum-phase zero locations on the performance of feedforward model-inverse control techniques in discrete-time systems," in *Proc. IEEE American Control Conf.*, 2008, pp. 2696–2702.

[18] S. Devasia, "Iterative machine learning for output tracking," *IEEE Trans. Control Systems Technology*, vol. 27, no. 2, pp. 516–526, 2017.

[19] K.-S. Kim and Q. Zou, "A modeling-free inversion-based iterative feedforward control for precision output tracking of linear timeinvariant systems," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 6, pp. 1767–1777, 2013.

[20] C.-W. Chen and T.-C. Tsao, "Data-based feedforward controller reconstruction from iterative learning control algorithm," in *Proc. IEEE Int. Conf. Advanced Intelligent Mechatronics*, 2016, pp. 683–688.

[21] M. Zheng, F. Zhang, and X. Liang, "A systematic design framework for iterative learning control with current feedback," *IFAC J. Systems and Control*, vol. 5, pp. 1–10, 2018.

[22] B. Francis and G. Zames, "On H_∞-optimal sensitivity theory for SISO feedback systems," *IEEE Trans. Automatic Control*, vol. 29, no. 1, pp. 9–16, 1984.

[23] M. Zheng, C. Wang, L. Sun, and M. Tomizuka, "Design of arbitraryorder robust iterative learning control based on robust control theory," *Mechatronics*, vol. 47, pp. 67–76, 2017.

[24] R. de Rozario, A. J. Fleming, and T. Oomen, "Finite-time learning control using frequency response data with application to a nanopositioning stage," *IEEE/ASME Trans. Mechatronics*, vol. 24, no. 5, pp. 2085–2096, 2019.

[25] J. Dewey, K. Leang, and S. Devasia, "Experimental and theoretical results in output-trajectory redesign for flexible structures," *J. Dynamic Systems*, *Measurement*, *and Control*, vol. 120, no. 4, pp. 456–461, 1998.

[26] K. S. Ramani, M. Duan, C. E. Okwudire, and A. Galip Ulsoy, "Tracking control of linear time-invariant nonminimum phase systems using filtered basis functions," *J. Dynamic Systems*, *Measurement*, *and Control*, vol. 139, no. 1, 2017. DOI: 10.1115/1.4034367.

[27] J. van Zundert and T. Oomen, "On inversion-based approaches for feedforward and ILC," *Mechatronics*, vol. 50, pp. 282–291, 2018.

[28] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control*: *Analysis and Design*. Wiley New York, vol. 2, 2007.

[29] I. D. Landau, A. C. Silva, T.-B. Airimitoaie, G. Buche, and M. Noe, "An active vibration control system as a benchmark on adaptive regulation," in *Proc. IEEE 2013 European Control Conf.*, 2013, pp. 2873–2878.

[30] X. Chen and M. Tomizuka, "Selective model inversion and adaptive disturbance observer for time-varying vibration rejection on an activesuspension benchmark," *European J. Control*, vol. 19, no. 4, pp. 300–312, 2013.

Dan Wang received the B.S. degree in material science and engineering from Shandong University, in 2012 and the M.S. degree in mechanical engineering from Tsinghua University, in 2015. She is currently working toward the Ph.D. degree at the University of Washington, USA. She received the Best Paper Award in the *2018 International Symposium on Flexible Automation*. Her main research interests include control theory, modeling, signal processing, robotics, artificial intelligence, FEM, and AM.

Xu Chen (M'09) is an Assistant Professor in the Department of Mechanical Engineering at the University of Washington, USA. He received the M.S. and Ph.D. degrees in mechanical engineering from the University of California, Berkeley, in 2010 and 2013, respectively, and the bachelor degree with honors from Tsinghua University in 2008. His research interests include dynamic systems and controls, advanced manufacturing, robotics, and intelligent mechatronics. His work – funded by NSF, DOE, DOD, state, and industries – has led to 4 Best Paper Awards, patented and massively deployed servo algorithms in the information storage industry, top-ranked adaptive control methods in international benchmark evaluations, and the graduation of 2 University Scholars. Dr. Chen is a Recipient of the National Science Foundation CAREER Award, the Young Investigator Award from ISCIE/*ASME International Symposium on Flexible Automation*, and the inaugural UTC Institute for Advanced Systems Engineering Breakthrough Award in 2016. He is Publicity and Local Arrangements Chairs of the 2020 and the *2023 IEEE/ASME International Conferences on Advanced Intelligent Mechatronics*, and Exhibits Chair of the *2021 IEEE American Control Conference*. He served the ASME Dynamic Systems and Control (DSC) Division in roles including Chair of the Vibration Technical Committee, News Editor of the *DSC Magazine*, Editor of the *DSC Newsletter*, and Student and Young Members Chair of the 2016 and the *2020 ASME DSC Conferences*. He is a member of IEEE, ASME, SME, and SIAM.