Transmission of Signal Nonsmoothness and Transient Improvement in Add-On Servo Control

Tianyu Jiang and Xu Chen

***Abstract*— Plug-in or add-on control is integral for high-performance control in modern precision systems. Despite the capability of greatly enhancing the steady-state performance, add-on compensation can introduce output discontinuity and significant transient response. Motivated by the vast application and the practical importance of add-on control designs, this paper identifies and investigates how general nonsmoothness in signals transmits through linear control systems. We explain the jump of system states in the presence of nonsmooth inputs in add-on servo enhancement, and derive formulas to mathematically characterize the transmission of the nonsmoothness. The results are then applied to devise fast transient responses over the traditional choice of add-on design at the input of the plant. Application examples to a manufacturing control system are conducted, with simulation and experimental results that validate the developed theoretical tools.**

***Index Terms*— Disturbance rejection, nonsmooth inputs, transient control.**

I. Introduction

PLUG-IN or add-on control design is central for servo enhancements in control engineering. In order to provide a storage capacity in the terabyte scale, a modern hard disk drive contains more than 900 000 data tracks within 1 in of the disk. Correspondingly, the width of each track, called track pitch (TP), can easily fall below 30 nm. During read/write operations, servo control must maintain a tracking error that is below 10% TP while strong external disturbances can induce tracking errors that are as large as 70% TP. Such large errors can only be attenuated by adding plug-in control commands. As another example, in high-speed wafer scanning for semiconductor manufacturing, [1] showed that 99.97% of the force commands in the positioning system are contributions of add-on feedforward control.

In feedback algorithms, add-on servo is central for a large class of design schemes that require a baseline feedback controller. Two examples are: disturbance observers [2] and Youla-parameterization-based loop shaping [3], [4]. Either for general low-frequency enhancement [5]–[7], or for the extensions to structured disturbance rejection [8]–[10], disturbance observers usually update the commands at the input side of the plant. Youla parameterization can be parameterized either as an add-on compensation at the plant input side [11], [12], or a combined compensation at the plant input and controller input [13], [14]. In feedforward-related control, adaptive or sensor-based feedforward compensation [15]–[17] can be configured as add-on algorithms either at the plant input or at the reference input (see more details in Section III).

Fundamentally, add-on control brings servo enhancement by introducing new dynamic properties in closed-loop signals. Such a process induces certain degrees of nonsmoothness in the signals. For meeting future demands in high-precision systems, it is essential to understand what types of systems and add-on changes create large transient, and what are the mathematical relationships between the signal nonsmoothness and the induced transient. The importance of such considerations is verified in simulation and experiments in [18] and [19], which compared the transient performance in different feedforward control algorithms. Still, a full theoretical solution to the problem is intrinsically nontrivial, except for simple discontinuities, such as step and ramp signals. Despite the rich literature on designs to achieve the desired steady-state performance, sparse investigations on the transient in add-on compensation are available, and a full understanding of the theoretical add-on transient remains missing. This paper targets to bridge this gap. The focuses are twofold. First, we develop theoretical results about input-to-output discontinuity and reveal its practical importance for the transient performance in control design. Second, new investigations are made to examine the transient characteristics in different add-on control designs. We derive an exact mathematical formula for computing the changes in system outputs when the input and/or its derivatives have discontinuities, and provide computation of the associated transient response. One central result we obtain is that, the common choice of performing add-on control at the input side of the plant yields undesired long transients, if there are delays during turning ON the compensation. Solution of the problem is discussed in detail and verified on a precision motion control platform in semiconductor manufacturing.

The remainder of this paper is organized as follows. Section II describes the wafer scanner hardware on which

Manuscript received January 27, 2016; revised January 22, 2017; accepted February 7, 2017. Manuscript received in final form February 16, 2017. The research is supported in part by the UTC-IASE Breakthrough Award. Recommended by Associate Editor S. Galeani. *(Corresponding author: Xu Chen.)*

The authors are with the Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: tianyu.jiang@uconn.edu; xchen@uconn.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/TCST.2017.2672399

verification of the algorithm is performed. Section III reveals the transient problem in add-on compensation, following which Sections IV and V solve the mathematical problem. Simulation and experimental results are provided in Section VI. Section VII concludes this paper.

Notations and Assumptions: All signals and systems are assumed to be causal and have real-valued coefficients. $\mathcal{L}\{\cdot\}$ and $\mathcal{L}^{-1}\{\cdot\}$ are, respectively, the operators of Laplace and inverse Laplace transforms. For practical purposes, we exclusively consider the first kind or jump discontinuities in signals and their derivatives; and denote $u(t_0^+)$ and $u(t_0^-)$, respectively, as the right- and the left-hand limits of a signal $u(t)$ at t_0. $f^{(i)}(t)$ denotes $d^i f(t)/dt^i$, the generalized ith-order derivative of a function $f(t)$. $f(t)$ is said to have a kth-order discontinuity at t_0 if $y^{(k)}(t_0^+) \neq y^{(k)}(t_0^-)$—in other words, $f(t)$ is of (differentiability) class C^{k-1} but not of class C^k at t_0. $[G_{d\to y}]d$ denotes the time-domain output of the system with respect to the input d. For a closed-loop system consisting of a plant with transfer function P, and a controller (in a negative feedback loop) C, $T \triangleq PC/(1+PC)$ denotes the complementary sensitivity function (the transfer function from the reference to the plant output); $S \triangleq 1/(1+PC)$ is the sensitivity function that defines the dynamics from the output disturbance to the plant output.

Remark 1: We focus on analysis and control of the transient behavior, and assume that the discontinuous change of input properties does not yield system instability (which can be guaranteed by, for instance, a sufficiently long dwell time during switching [20]). For additional information on stability of such switched systems, we refer readers to [21]–[24].

II. Hardware Description and Notations

The developed algorithm in this paper is verified via simulation and experiments on a wafer scanner prototype, a central element for photolithography in the advanced manufacturing of integrated circuits in semiconductor industry. The precision control here synchronizes the motions of a wafer stage and a reticle stage. The motion control allows patterns on integrated circuits to be precisely transformed from a mask on the reticle stage to different locations of the silicon wafer on the wafer stage. A picture of the physical system is provided in [8]. To achieve the nanometer-scale precision requirement, high-performance actuation and measurement tools, including air bearings, epoxy-core linear permanent magnet motors (LPMMs), and laser interferometers, are used. The control commands are executed on a LabVIEW real-time system with field-programmable gate array. Fig. 1 shows the frequency response, from the voltage input of the LPMM to the position of the reticle stage.

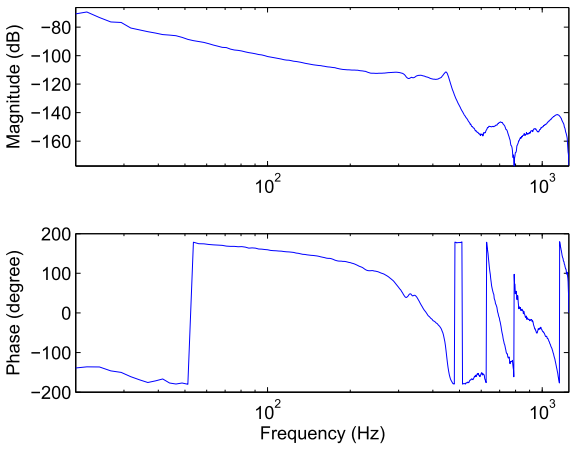

Fig. 1. Frequency response of the reticle stage.

III. Transient in Add-On Compensation

A. Example and Practical Importance

The influence of transient performance is significant in the final achievable control accuracy. Consider an example in Fig. 2. Assuming first that the signals r, $u_{\rm ur}$, and $u_{\rm ue}$ are all zero, we aim at regulating the output y in the presence of the disturbance d. Here, the baseline feedback controller C is best tuned for regular servo performance and system robustness; and $u_{\rm uc}$ is the additional control to compensate d.

If $u_{\rm uc} = -d$, certainly, the disturbance is perfectly rejected. This is ideally the goal for all observer- or feedforward-based disturbance attenuation designs, if injection of control command is at the plant input side. However, in practice:

1) Strong external disturbances may not always present, and $u_{\rm uc}$ is turned ON only when external disturbance reaches the threshold, at which the error tolerance is violated.[1]
2) The control system is usually subjected to different tasks, where different disturbance properties require different add-on designs (indeed, if the add-on scheme is universal for all situations, it should be absorbed as part of the baseline controller).

Hence, for rejecting external disturbances, practically, a switch is used for turning ON or OFF the compensation $u_{\rm uc}$. Consider the case where d is a scaled step signal that occurred at 0.12 s. If the add-on compensation is delayed by 2.4 s (i.e., $u_{\rm uc}$ is added at 2.52 s), even with the "perfect" rejection condition $u_{\rm uc} = -d$, significant transient response can happen as shown in Fig. 3—the experimental results on the wafer scanner system.

Certainly, the above-mentioned example is for demonstration of the problem, and provides only an extreme case where the add-on compensation is turned ON when an integrator in the baseline controller C has already greatly compensated the disturbance, and d is simple enough to be perfectly rejected by simple feedback. These simplifying conditions will be dropped in the remainder of this paper, where the general problem of add-on transient is addressed.

B. Ideal-Case Add-On Compensation

Recall Fig. 2. The location of add-on compensation can be at the reference input or the plant input; and the requirement of servo enhancement may come from regulation or tracking

[1] In industrial applications, it is common to run a fault detector to monitor the system performance and switch ON the compensation when the servo performance is degraded to be below a prespecified performance threshold.

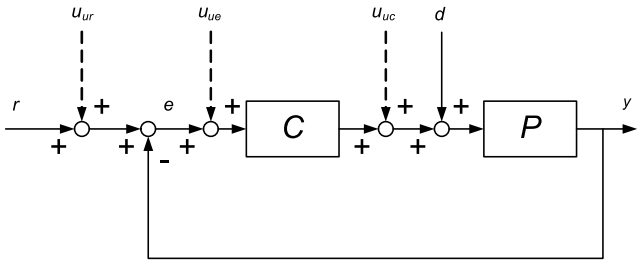

Fig. 2. Add-on control designs in a feedback block diagram.

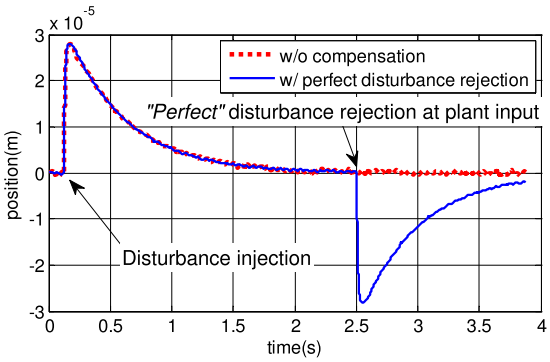

Fig. 3. Demonstration of transient behavior on a wafer scanner.

controls. These additional considerations are now added to form a general block diagram with different configurations of add-on signals. We now formally introduce the signals $u_{\rm ur}$, $u_{\rm ue}$, and $u_{\rm uc}$, which are the added servo-enhancement signals for *updated reference* (UR), *updated error* (UE), and *updated control* (UC), respectively.

Let $G(s)$ be the closed-loop transfer function from the add-on control to the plant output. Assume zero initial conditions at $t = 0$, i.e., $y^{(i)}(0)(= d^i y(0)/dt^i) = 0$, and first focus on attenuating the disturbance d, namely, we aim at achieving

$$[G]u_{\text{add-on}} + [G_{d\to y}]d = 0 \tag{1}$$

where $u_{\text{add}-\text{on}}$ is $u_{\rm uc}$, $u_{\rm ur}$ or $u_{\rm ue}$; $G_{d\to y} = P/(1 + PC)$ is the transfer function from d to y. From Fig. 2, G equals $P/(1 + PC)$ in UC. In UE and UR, the dynamics between $u_{\text{add}-\text{on}}$ and y both equal the complementary sensitivity function, namely, $G = T = PC/(1 + PC)$. Hence, regardless of the design methods, to satisfy (1), the ideal conditions in UC and UR/UE are, respectively, $u_{\rm uc} = -d$ and $[C]u_{ur/ue} = -d$.

C. Transient in Ideal Add-On UC Control

Without loss of generality, suppose the actual disturbance d is as shown in the top subplot in Fig. 4, where at time t_0, the plug-in servo enhancement is turned ON. The ideal-case UC command $u_{\rm uc}$ is the solid line in the second subplot of Fig. 4, which perfectly cancels the disturbance after time t_0.

Let

$$G(s) = \frac{b_n s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} \tag{2}$$

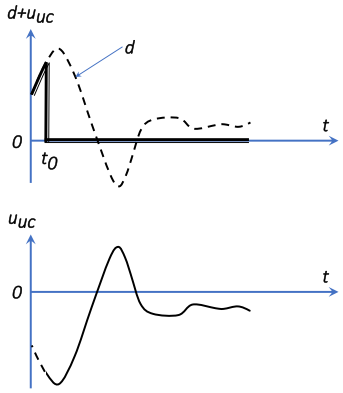

Fig. 4. Input discontinuity in UC.

and consider the response of $G(s)$ to the combined input $u \triangleq d + u_{\rm uc}$. Directly solving the associated ordinary differential equation (ODE) is not feasible as derivatives of u are not well defined at time t_0. We will show how this input discontinuity creates abrupt changes in $y(t)$ and its derivatives. Notice that unlike the discontinuity in step responses, the solution to this input-to-output discontinuity problem is nontrivial.

For a general system, we next derive the exact mathematical result of the transient after $t = t_0$ for u in Fig. 4. Recall (2) and note that $u(t)$ equals zero $\forall t \geqslant t_0$ in Fig. 4. The transient response $y(t)$ in $t \in (t_0, \infty)$ satisfies

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_0 y(t) = 0$$

$$\text{with the initial condition: } \left\{y^{(i)}\left(t_0^+\right)\right\}_{i=0}^{n-1} \tag{3}$$

i.e., the transient is the natural response of the system with the initial condition $\{y^{(i)}(t_0^+)\}_{i=0}^{n-1}$.

Solutions to the ODEs can be obtained using Laplace transforms or direct computation via Calculus. It is, however, central to recognize that the initial condition $y^{(i)}(t_0^+)$ does not equal $y^{(i)}(t_0^-)$, i.e., the actual transient does not simply equal the natural transient response under $y^{(i)}(t_0^-)$—the system states right before the application of add-on compensation (recall the example in Fig. 3). This is due to the input discontinuity of $u^{(i)}(t_0^-)$ jumping to $u^{(i)}(t_0^+)(= 0)$ in Fig. 4. Next, we obtain the formula of $\{y^{(i)}(t_0^+)\}_{i=0}^{n-1}$ based on $u^{(i)}(t_0^-)$ and the dynamics of $G(s)$, and then analyze the resulting transient performance.

IV. Input-to-Output Discontinuity

Theorem 2: Let $u(t)$ and $y(t)$ be the input and the output of a finite-dimensional real-coefficient linear system G, satisfying

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1\dot{y}(t) + a_0 y(t)$$

$$= b_n u^{(n)}(t) + b_{n-1}u^{(n-1)}(t) + \cdots + b_1\dot{u}(t) + b_0 u(t) \tag{4}$$

at time t_0. If $u(t)$ and/or its derivatives have discontinuities: $u(t_0^+) - u(t_0^-) = e_{u,0}$, $\dot{u}(t_0^+) - \dot{u}(t_0^-) = e_{u,1}$, …, $u^{(n)}(t_0^+) - u^{(n)}(t_0^-) = e_{u,n}$, then $y(t)$ and/or its derivatives contain

discontinuities that satisfy

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \dots & a_{n-1} & 1 \end{bmatrix} \begin{bmatrix} y(t_0^+) - y(t_0^-) \\ \dot{y}(t_0^+) - \dot{y}(t_0^-) \\ \vdots \\ y^{(n-1)}(t_0^+) - y^{(n-1)}(t_0^-) \end{bmatrix} = \begin{bmatrix} b_n & 0 & \dots & 0 \\ b_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ b_1 & \dots & b_{n-1} & b_n \end{bmatrix} \begin{bmatrix} e_{u,0} \\ e_{u,1} \\ \vdots \\ e_{u,n-1} \end{bmatrix}. \tag{5}$$

Theorem 2 fully characterizes the input-to-output discontinuity. The matrix on the left-hand side of (5) is nonsingular. Therefore, a unique solution exists for obtaining $\{e_{y,i}\}_{i=0}^{n-1} := \{y^{(i)}(t_0^+) - y^{(i)}(t_0^-)\}_{i=0}^{n-1}$. No knowledge of $u(t)$ is required except at t_0, the instance of discontinuity. More specifically, solutions of $e_{y,i}$ values can be obtained by forward substitution after solving the matrix equality

$$e_{y,0} = b_n e_{u,0} \tag{6a}$$

$$e_{y,1} = b_{n-1} e_{u,0} + b_n e_{u,1} - a_{n-1} e_{y,0} \tag{6b}$$

$$\vdots$$

$$e_{y,n-1} = \sum_{j=0}^{n-1} b_{j+1} e_{u,j} - \sum_{j=0}^{n-2} a_{j+1} e_{y,j}. \tag{6c}$$

Remark 3: Equation (5) provides up to the $(n-1)$th order output discontinuity. If the value of $e_{y,n}$ is of interest, Theorem 2 can be applied to the augmented system $G(s) = (b_n s^{n+1} + b_{n-1} s^n + \cdots + b_1 s^2 + b_0 s)/(s^{n+1} + a_{n-1} s^n + \cdots + a_1 s^2 + a_0 s)$. Similar procedures can provide other higher order discontinuities.

Numerical Verification: Consider the response of a first-order system to a ramp-to-step signal

$$G(s) = \frac{1}{s+a}, \quad u(t) = \begin{cases} \alpha t: & t \in [0, t_0) \\ \alpha t_0: & t \geq t_0. \end{cases} \tag{7}$$

In this example, we have $u(t_0^+) = u(t_0^-)$, $\dot{u}(t_0^+) \neq \dot{u}(t_0^-)$.

Convolution or inverse Laplace analysis gives $y(t) = (\alpha/a)t + (\alpha/a^2)e^{-at} - (\alpha/a^2)$, if $t \in [0, t_0)$; $y(t) = (\alpha/a)t_0 + (\alpha/a^2)[e^{-at} - e^{-a(t-t_0)}]$, if $t \geq t_0$; $\dot{y}(t) = (\alpha/a) - (\alpha/a)e^{-at}$, if $t \in [0, t_0)$; $\dot{y}(t) = -(\alpha/a)[e^{-at} - e^{-a(t-t_0)}]$, if $t \geq t_0$; $\ddot{y}(t) = \alpha e^{-at}$, if $t \in [0, t_0)$; and $\ddot{y}(t) = \alpha[e^{-at} - e^{-a(t-t_0)}]$, if $t \geq t_0$. Then

$$\begin{aligned} y(t_0^+) &= y(t_0^-) \\ \dot{y}(t_0^+) &= \dot{y}(t_0^-) \\ \ddot{y}(t_0^+) &= \ddot{y}(t_0^-) + (\dot{u}(t_0^+) - \dot{u}(t_0^-)) \end{aligned} \tag{8}$$

namely, the first-order input discontinuity creates a second-order output discontinuity.

Alternatively, apply Theorem 2 and Remark 3 to the same system. Noticing that $G(s) = 1/(s+a) = s/(s^2+as)$, we have

$$\begin{bmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & a & 1 \end{bmatrix} \begin{bmatrix} e_{y,0} \\ e_{y,1} \\ e_{y,2} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} e_{u,0} \\ e_{u,1} \\ e_{u,2} \end{bmatrix} = \begin{bmatrix} 0 \\ e_{u,0} \\ e_{u,1} \end{bmatrix}.$$

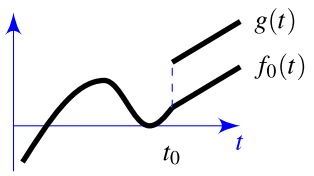

Fig. 5. Decomposition of discontinuity in $g(t)$ at t_0.

Hence $y(t_0^+) - y(t_0^-) = 0$, $\dot{y}(t_0^+) - \dot{y}(t_0^-) = e_{u,0} = 0$, and $\ddot{y}(t_0^+) - \ddot{y}(t_0^-) = e_{u,1} - a e_{u,0} = \dot{u}(t_0^+) - \dot{u}(t_0^-)$. The result matches with that in (8). More important, the computation here removes the necessity to compute the full time-domain solution, which is not only long and complex for high-order systems, but also infeasible for general signals without given time-domain models.

A. Proof and Analysis

We first introduce a representation of discontinuous signals using Dirac delta functions. In the remainder of the texts, we will use $\mu(t)$ to denote the unit step signal, i.e., $\mu(t) = 1$ $\forall t \geq 0$ and $\mu(t) = 0$ $\forall t < 0$; and denote $\delta(t)$ as the Dirac delta function that satisfies $\int_0^t \delta(\tau) d\tau = \mu(t)$ and $\int_0^\infty \delta(\tau - T) g(\tau) d\tau = g(T)$ if $g(t)$ is continuous. As a distribution (a.k.a. generalized function), $\delta(t)$ satisfies $d\mu(t)/dt = \delta(t)$.

Consider a piecewise continuous function $g(t)$ with a first-kind/jump discontinuity at t_0. We can write

$$g(t) = e_{g,0} \mu(t - t_0) + f_0(t) \tag{9}$$

where $f_0(t)$ is continuous at t_0; $e_{g,0} = g(t_0^+) - g(t_0^-)$; and $\mu(t - t_0)$ creates the jump discontinuity at $t = t_0$, as shown in the example in Fig. 5.

Similarly, if $\dot{f}_0(t)$ is furthermore discontinuous at t_0, we have $\dot{f}_0(t) = e_{g,1} \mu(t - t_0) + f_1(t)$, where $f_1(t)$ is continuous at t_0. The derivative of (9) thus must satisfy

$$\dot{g}(t) = e_{g,0} \delta(t - t_0) + e_{g,1} \mu(t - t_0) + f_1(t) \tag{10}$$

where $e_{g,1} \mu(t - t_0)$ gives the first-order discontinuity $\dot{g}(t_0^+) \neq \dot{g}(t_0^-)$.

Further differentiation yields

$$\begin{aligned} \ddot{g}(t) &= e_{g,0} \dot{\delta}(t - t_0) + e_{g,1} \delta(t - t_0) + e_{g,2} \mu(t - t_0) + f_2(t) \\ &\vdots \\ g^{(n)}(t) &= e_{g,0} \delta^{(n-1)}(t - t_0) + \cdots + e_{g,n-1} \delta(t - t_0) \\ &\quad + e_{g,n} \mu(t - t_0) + f_n(t) \end{aligned} \tag{11}$$

where $f_2, \dots, f_n$ are continuous at t_0, and the nonsmoothness of $g(t)$ is characterized by

$$\begin{bmatrix} e_{g,0} \\ e_{g,1} \\ \vdots \\ e_{g,n} \end{bmatrix} = \begin{bmatrix} g(t_0^+) - g(t_0^-) \\ \dot{g}(t_0^+) - \dot{g}(t_0^-) \\ \vdots \\ g^{(n)}(t_0^+) - g^{(n)}(t_0^-) \end{bmatrix}. \tag{12}$$

Equations (9)–(11) can be compactly written as

$$\begin{bmatrix} g(t) \\ g^{(1)}(t) \\ \vdots \\ g^{(n)}(t) \end{bmatrix}^T = \begin{bmatrix} f_0(t) \\ f_1(t) \\ \vdots \\ f_n(t) \end{bmatrix}^T + \begin{bmatrix} \mu(t-t_0) \\ \delta(t-t_0) \\ \vdots \\ \delta^{(n-1)}(t-t_0) \end{bmatrix}^T \times \begin{bmatrix} e_{g,0} & e_{g,1} & \dots & e_{g,n} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & e_{g,1} \\ 0 & \dots & 0 & e_{g,0} \end{bmatrix}. \quad (13)$$

For matrix-vector operations in the form of (13), the following result will appear to be useful:

Fact 4: The following is true:

$$\begin{bmatrix} e_0 & e_1 & \dots & e_n \\ 0 & e_0 & \dots & e_{n-1} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & e_0 \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \\ 1 \end{bmatrix} = \begin{bmatrix} a_0 & \dots & a_{n-1} & 1 \\ \vdots & \ddots & \ddots & 0 \\ a_{n-1} & \ddots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} e_0 \\ \vdots \\ e_{n-1} \\ e_n \end{bmatrix}.$$

We now formally prove Theorem 2.

Proof: Replacing $\{g,\ f_i,\ e_{g,i}\}$ with $\{y,\ f_{y,i},\ e_{y,i}\}$ and $\{u,\ f_{u,i},\ e_{u,i}\}$, respectively, in (13); and applying the resulting equations to $[y(t)\ y^{(1)}(t)\ \dots\ y^{(n)}(t)][a_0\ \dots\ a_{n-1}\ 1]^T = [u(t)\ u^{(1)}(t)\ \dots\ u^{(n)}(t)][b_0\ \dots\ b_{n-1}\ b_n]^T$—the vector form of (4)—give

$$\begin{aligned} &\begin{bmatrix} f_{y,0}(t) \\ f_{y,1}(t) \\ \vdots \\ f_{y,n}(t) \end{bmatrix}^T \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \\ 1 \end{bmatrix} - \begin{bmatrix} f_{u,0}(t) \\ f_{u,1}(t) \\ \vdots \\ f_{u,n}(t) \end{bmatrix}^T \begin{bmatrix} b_0 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\mu(t-t_0)}{\delta(t-t_0)} \\ \vdots \\ \delta^{(n-1)}(t-t_0) \end{bmatrix}^T \left[\begin{array}{c|ccc} e_{u,0} & e_{u,1} & \dots & e_{u,n} \\ \hline 0 & e_{u,0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & e_{u,1} \\ 0 & \dots & 0 & e_{u,0} \end{array}\right] \left[\begin{array}{c} b_0 \\ \hline \vdots \\ b_{n-1} \\ b_n \end{array}\right] \\ &\quad - \begin{bmatrix} \dfrac{\mu(t-t_0)}{\delta(t-t_0)} \\ \vdots \\ \delta^{(n-1)}(t-t_0) \end{bmatrix}^T \left[\begin{array}{c|ccc} e_{y,0} & e_{y,1} & \dots & e_{y,n} \\ \hline 0 & e_{y,0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & e_{y,1} \\ 0 & \dots & 0 & e_{y,0} \end{array}\right] \left[\begin{array}{c} a_0 \\ \hline \vdots \\ a_{n-1} \\ 1 \end{array}\right] \end{aligned} \quad (14)$$

where $f_{y,i}(t)$ and $f_{u,i}(t)$ are continuous at $t = t_0$; $e_{y,i} = y^{(i)}(t_0^+) - y^{(i)}(t_0^-)$; and $e_{u,i} = u^{(i)}(t_0^+) - u^{(i)}(t_0^-)$.

To solve for $e_{y,i}$ values, using Fact 4, we translate (14) to

$$\begin{aligned} &\begin{bmatrix} f_{y,0}(t) \\ f_{y,1}(t) \\ \vdots \\ f_{y,n}(t) \end{bmatrix}^T \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \\ 1 \end{bmatrix} - \begin{bmatrix} f_{u,0}(t) \\ f_{u,1}(t) \\ \vdots \\ f_{u,n}(t) \end{bmatrix}^T \begin{bmatrix} b_0 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\mu(t-t_0)}{\delta(t-t_0)} \\ \vdots \\ \delta^{(n-1)}(t-t_0) \end{bmatrix}^T \left[\begin{array}{ccc|c} b_0 & \dots & b_{n-1} & b_n \\ \hline \vdots & \ddots & \ddots & 0 \\ b_{n-1} & \ddots & \ddots & \vdots \\ b_n & 0 & \dots & 0 \end{array}\right] \begin{bmatrix} e_{u,0} \\ e_{u,1} \\ \vdots \\ e_{u,n} \end{bmatrix} \\ &\quad - \begin{bmatrix} \dfrac{\mu(t-t_0)}{\delta(t-t_0)} \\ \vdots \\ \delta^{(n-1)}(t-t_0) \end{bmatrix}^T \left[\begin{array}{ccc|c} a_0 & \dots & a_{n-1} & 1 \\ \hline \vdots & \ddots & \ddots & 0 \\ a_{n-1} & \ddots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{array}\right] \begin{bmatrix} e_{y,0} \\ e_{y,1} \\ \vdots \\ e_{y,n} \end{bmatrix}. \end{aligned}$$

$\delta(t-t_0)$, $\dot{\delta}(t-t_0)$, $\dots$, $\delta^{(n-1)}(t-t_0)$ are linearly independent, and cannot be expressed as linear combinations of the continuous functions on the left-hand side of the last equality. Hence, their coefficients on the right-hand side must be zero. This corresponds to

$$\begin{aligned} &\begin{bmatrix} a_1 & \dots & a_{n-1} & 1 \\ \vdots & \ddots & \ddots & 0 \\ a_{n-1} & \ddots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} e_{y,0} \\ e_{y,1} \\ \vdots \\ e_{y,n-1} \end{bmatrix} \\ &= \begin{bmatrix} b_1 & \dots & b_{n-1} & b_n \\ \vdots & \ddots & \ddots & 0 \\ b_{n-1} & \ddots & \ddots & \vdots \\ b_n & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} e_{u,0} \\ e_{u,1} \\ \vdots \\ e_{u,n-1} \end{bmatrix}. \end{aligned} \quad (15)$$

Rearranging the rows gives (5). □

Case for Add-On Servo Enhancement: Applying (5) to $u(t)$ in Fig. 4, and noting the input discontinuity of $e_{u,i} = 0 - u^{(i)}(t_0^-), \forall i \geqslant 0$, we have

$$\begin{aligned} &\begin{bmatrix} 1 & 0 & \dots & 0 \\ a_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \dots & a_{n-1} & 1 \end{bmatrix} \begin{bmatrix} e_{y,0} \\ e_{y,1} \\ \vdots \\ e_{y,n-1} \end{bmatrix} \\ &= -\begin{bmatrix} b_n & 0 & \cdots & 0 \\ b_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ b_1 & \cdots & b_{n-1} & b_n \end{bmatrix} \begin{bmatrix} u(t_0^-) \\ \dot{u}(t_0^-) \\ \vdots \\ u^{(n-1)}(t_0^-) \end{bmatrix}. \end{aligned} \quad (16)$$

B. Discussions

Rewriting (5) symbolically as $M_a \mathbf{e}_y = M_b \mathbf{e}_u$ and applying Taylor expansion to M_a^{-1} give $\mathbf{e}_y = M_a^{-1} M_b \mathbf{e}_u = \sum_{k=0}^{\infty} (I - M_a)^k M_b \mathbf{e}_u$. Noting the lower triangular form of M_a, we can further simplify the expression to $\mathbf{e}_y = \sum_{k=0}^{n} (I - M_a)^k M_b \mathbf{e}_u$, as $I - M_a$ is nilpotent and $(I - M_a)^k$ vanishes for $k > n$.

From the results, the generalized output discontinuity is a linear and continuous function of the discontinuities in the input and its derivatives. Bounded input discontinuities generate bounded discontinuities in the output. More specifically, $\|\mathbf{e}_y\|_q \leq \|\sum_{k=0}^{n}(I-M_a)^k M_b\|_{p\to q}\|\mathbf{e}_u\|_p$, where $p, q \in [1,\infty)$; $\|\cdot\|_{p\to q}$ is an induced matrix norm; and $\|\cdot\|_p$ and $\|\cdot\|_q$ are vector norms. The numeric value of the upper bound is problem-dependent. The matrix $\sum_{k=0}^{n}(I-M_a)^k M_b$, however, is always lower triangular and has easy-to-compute matrix norms, as $(I-M_a)^k$ and M_b are both lower triangular (with actually all diagonal entries equal to 1, 0, or b_n). Furthermore, as M_a and M_b do not contain a_0 and b_0, $\|\sum_{k=0}^{n}(I-M_a)^k M_b\|_{p\to q}$ is independent of b_0 and a_0. Discontinuities in $\{y^{(i)}(t)\}_{i=0}^{n-1}$ are, therefore, independent of the DC gain and not related to the magnitude response of the system [see also (6a)–(6c)]. Switched systems/signals can thus generate significant transient responses while providing zero steady-state errors.

1) Influence of High-Order Input Discontinuities: If the input is continuous, i.e., $e_{u,0}=0$, the resulting output is still continuous but not necessarily smooth. For instance, if b_n is large in (6b), large discontinuity in $\dot{y}(t)$ occurs even if there is only a small change of $\dot{u}(t)$. Notice that the result may appear counter to intuitions and perceptions in conventional analysis, which may lead to the assertion that b_1—the scaling coefficient of $\dot{u}(t)$ in (4)—is the dominant factor.

Furthermore, high-order input discontinuities $e_{u,i}$ values only influence high-order derivatives in the output. More specifically, the ith-order output discontinuity $e_{y,i}$ only depends on $\{e_{u,j}: j \leq i\}$, based on the mathematical relations in (6a)–(6c).

2) Influence of the Relative Degree of the System: If $b_n \neq 0$, a discontinuous $u(t)$ will render $y(t)$ and all its derivatives discontinuous. The direct implication is in line with the conventional practice that jump discontinuities are undesired in general switched control. In addition, from (6a), the jump in the output is only linearly dependent on the jump in the input and b_n—the direct gain of the system.

In the case that the system is time-invariant, if the relative degree of the transfer function associated with (4) is r—in other words, b_n, …, b_{n-r+1} all equal zero—then $e_{y,0}$, …, $e_{y,r-1}$ on the left-hand side of (5) must be zero, namely, y, $\dot{y}$, …, $y^{(r-1)}$ are all continuous at t_0, and the input nonsmoothness can only cause discontinuities in the higher order derivatives $y^{(r)}$, …, $y^{(n-1)}$.

3) Case for Nonideal Add-On Control: If the disturbance estimation contains errors or there exists actuation delay t_{ad}, the ideal compensation is no longer feasible, as shown in Fig. 6. Recall Fig. 4. In the case with nonideal add-on control, the condition that the augmented command $u(t) = d(t)+u_{\text{uc}}(t)$ equals zero $\forall t \geqslant t_0$ (i.e., $u^{(i)}(t_0^+)=0, \forall i \geqslant 0$) will not hold. This, however, does not constrain one to compute the transient response due to input discontinuity. To be more specific, one can write the input signal as

$$u(t)=\begin{cases} d(t): & t\in[0,t_0) \\ d(t)+u_{\text{uc}}(t)=\epsilon(t)\neq 0: & t\geqslant t_0 \end{cases} \tag{17}$$

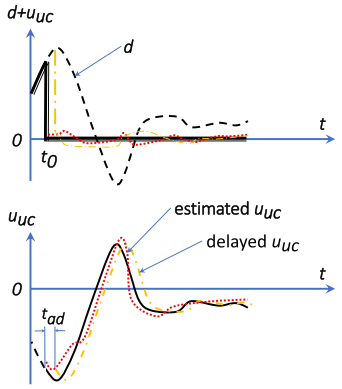

Fig. 6. Nonideal add-on UC control.

where t_0 is the actual implementation time of u_{uc} (i.e., it may contain the actuation delay) and $\epsilon(t)$ is the residual term characterizing the difference between the ideal add-on control command and the actual command. The induced input discontinuities at time t_0 can thus be written as $e_{u,i} = u^{(i)}(t_0^+) - u^{(i)}(t_0^-) = d^{(i)}(t_0^+) + u_{\text{uc}}^{(i)}(t_0^+) - d^{(i)}(t_0^-) = [d^{(i)}(t_0^+) - d^{(i)}(t_0^-)] + u_{\text{uc}}^{(i)}(t_0^+)$. When $d(t)$ is of class C^k at t_0, the term in the square bracket equals zero $\forall 0 \leq i \leq k$, yielding the formula $e_{u,i} = u_{\text{uc}}^{(i)}(t_0^+),\ \forall 0 \leq i \leq k$. Note that although t_0 is not known when there is actuation delay, $u_{\text{uc}}^{(i)}(t_0^+)$ (if exists) is available, since it is the initial condition of the actual add-on signal. Moreover, (5) works without requiring the condition that $u^{(i)}(t_0^+)=0, \forall i \geqslant 0$. Based on the relations in (6a)–(6c), the output discontinuity $\{e_{y,i}: 0 \leq i \leq k\}$ can be derived without the need to know $d(t)$ and t_0.

As an example, recall the system and the input in (7). If we regard $u(t)=\alpha t,\ t\geq 0$ and $u(t)=\alpha t_0,\ t\geq t_0$, respectively, as the disturbance $d(t)$ and residual term $\epsilon(t)$ in (17), then the corresponding add-on signal (which we design) is $u_{\text{uc}}(t) = \epsilon(t) - d(t) = \alpha t_0 - \alpha t,\ t \geq t_0$. One observes that $u_{\text{uc}}^{(i)}(t_0^+)$ is exactly the ith-order input discontinuity. Similar to the application of Theorem 2 and Remark 3 to (7), this information of add-on control $u_{\text{uc}}(t)$ is adequate and sufficient to identify the induced output discontinuities.

V. Time-Domain Response and Transient Speed

Based on analysis in Section IV, motion control systems that are powered by motors with double integrator (inertia system) or $1/(ms^2+bs)$ types of nominal dynamics always generate continuous outputs; systems with very fast input–output dynamics, such as piezoelectric actuators (whose nominal dynamics can be commonly modeled as a constant-gain system), are sensitive to nonsmoothness in the input. On the other hand, many servo controllers can be considerably sensitive to input discontinuities. For instance, consider the causal implementation of a PID controller $C_{\text{PID}}(s) = k_p + k_i/s + k_d s/(\epsilon s+1) = [(k_p\epsilon+k_d)s^2+(k_p+k_i\epsilon)s+k_i]/[s(\epsilon s+1)]$, where $\epsilon \ll 1$ is a small positive number. The direct gain of the controller $k_p + k_d/\epsilon$ can be enormous in practice. Let e and u be the input and output of $C_{\text{PID}}(s)$. Then, (5) gives $u(t_0^+)-u(t_0^-) = (k_p+(k_d/\epsilon))[e(t_0^+)-e(t_0^-)]$, $\dot{u}(t_0^+)-\dot{u}(t_0^-) = (k_i-(k_d/\epsilon^2))[e(t_0^+)-e(t_0^-)] + (k_p+(k_d/\epsilon))[\dot{e}(t_0^+)-\dot{e}(t_0^-)]$.

Therefore, both the controller output and its derivatives will exhibit significant jump discontinuities when the input has only small discontinuities. A lead-lag controller can reduce the amplification of the input discontinuities, but still has a direct-gain component. The induced large magnitude in the control command can easily trigger the input saturation and the resulting nonlinear effects in the feedback loop (and even instability and hardware damage). Moreover, omitted in conventional transient analysis, abnormal faults triggered by such discontinuity amplification are much more difficult to detect and analyze (see also the second example in Section VI).

The analysis in the preceding paragraphs focused on characterizing the output at the switching instance. Together with the quantitative analysis, the following qualitative measures can be made about intrinsic properties of the system and signals that will induce small/smooth transient: large relative degree in the transfer function (LTI case), small coefficients for high-order derivatives, and small low-order discontinuities in the input signal. To further reveal characteristics of the transient after switching, we next derive the exact time-domain response due to the input nonsmoothness.

Assume that the system coefficients are constant in (4). The transient response after t_0 is the solution to the ODE

$$y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_0 y(t) = 0 \tag{18}$$

with the initial condition $\{y^{(i)}(t_0^+)(= y^{(i)}(t_0^-) + e_{y,i})\}_{i=0}^{n-1}$.

Proposition 5: Let the same assumptions in Theorem 2 hold. Let $G(s) = B(s)/A(s)$ be stable with $A(s) = s^n + a_{n-1}s^{n-1} + \cdots + a_0$ and $B(s) = b_n s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0$. The transient response due to $\{e_{u,i}\}_{i=0}^{n}$—the input nonsmoothness at t_0—has the Laplace transform

$$Y(s, e_i) = \frac{e^{-t_0 s}\begin{bmatrix} b_n s^{n-1} + \cdots + b_2 s + b_1 \\ b_n s^{n-2} + \cdots + b_3 s + b_2 \\ \vdots \\ b_n s + b_{n-1} \\ b_n \end{bmatrix}^T \begin{bmatrix} e_{u,0} \\ e_{u,1} \\ \vdots \\ e_{u,n-2} \\ e_{u,n-1} \end{bmatrix}}{A(s)}. \tag{19}$$

The full transient response after t_0 is

$$\begin{aligned} Y(s) &= Y(s, e_i) \\ &\quad + \frac{e^{-t_0 s}}{A(s)}\begin{bmatrix} s^{n-1} + a_{n-1}s^{n-2} + \cdots + a_2 s + a_1 \\ s^{n-2} + a_{n-1}s^{n-3} + \cdots + a_2 \\ \vdots \\ s + a_{n-1} \\ 1 \end{bmatrix}^T \\ &\quad \times \begin{bmatrix} y(t_0^-) \\ \dot{y}(t_0^-) \\ \vdots \\ y^{(n-2)}(t_0^-) \\ y^{(n-1)}(t_0^-) \end{bmatrix}. \end{aligned} \tag{20}$$

Proof: Consider t_0^+ as the initial time. The Laplace-domain quantities $\mathcal{L}\{y(t)\} = Y(s)$, $\mathcal{L}\{\dot{y}(t)\} = sY(s) - y(t_0^+)$, and $\mathcal{L}\{y^{(n)}(t)\} = s^n Y(s) - s^{n-1}y(t_0^+) - s^{n-2}\dot{y}(t_0^+) - \cdots - y^{(n-1)}(t_0^+)$ give

$$\begin{aligned} &\mathcal{L}[y(t), \dot{y}(t), \ldots, y^{(n)}(t)]^T \\ &\quad = [1, s, \ldots, s^n]^T Y(s) \\ &\qquad - \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ 1 & 0 & & \vdots \\ s & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ s^{n-1} & \cdots & s & 1 \end{bmatrix} \begin{bmatrix} y(t_0^+) \\ \dot{y}(t_0^+) \\ \vdots \\ y^{(n-1)}(t_0^+) \end{bmatrix}. \end{aligned} \tag{21}$$

Writing the Laplace transform of (18) as $[a_0, a_1, \ldots, a_{n-1}, 1]\mathcal{L}[y(t), \dot{y}(t), \ldots, y^{(n-1)}(t), y^{(n)}(t)]^T = 0$ and using (21), we can solve for $Y(s)$. The solution is

$$\begin{aligned} &Y(s) \\ &= \frac{e^{-t_0 s}\begin{bmatrix} s^{n-1} \\ s^{n-2} \\ \vdots \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n-1} & 1 \end{bmatrix} \begin{bmatrix} y(t_0^+) \\ \dot{y}(t_0^+) \\ \vdots \\ y^{(n-1)}(t_0^+) \end{bmatrix}}{A(s)}. \end{aligned} \tag{22}$$

Substituting in $y^{(i)}(t_0^+) = y^{(i)}(t_0^-) + e_{y,i}$ yields the decomposition $Y(s) = Y(s, t_0^-) + Y(s, e_i)$, where $Y(s, t_0^-)$ and $Y(s, e_i)$—obtained by replacing $y(t_0^+)$ in (22) with $y(t_0^-)$ and e_i, respectively—are the Laplace transforms of the natural transient and the transient due to nonsmoothness in the input. Writing out $Y(s, t_0^-)$ explicitly gives (20). To obtain the specific form of $Y(s, e_i)$ in (19), using (5) and (22) gives

$$\begin{aligned} &Y(s, e_i) \\ &= \frac{e^{-t_0 s}\begin{bmatrix} s^{n-1} \\ s^{n-2} \\ \vdots \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n-1} & 1 \end{bmatrix} \begin{bmatrix} e_{y,0} \\ e_{y,1} \\ \vdots \\ e_{y,n-1} \end{bmatrix}}{A(s)} \\ &= \frac{e^{-t_0 s}\begin{bmatrix} s^{n-1} \\ s^{n-2} \\ \vdots \\ 1 \end{bmatrix}^T \begin{bmatrix} b_n & 0 & \cdots & 0 \\ b_{n-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ b_1 & \cdots & b_{n-1} & b_n \end{bmatrix} \begin{bmatrix} e_{u,0} \\ e_{u,1} \\ \vdots \\ e_{u,n-1} \end{bmatrix}}{A(s)} \end{aligned} \tag{23}$$

which, after simplifications, is equivalent to (19). □

Notice that Proposition 5 quantifies the postswitching response using only properties of the preswitching system and signal. The prophetic result can provide guidance on the switching instance as well as conditions for small/smooth postswitching transient response. In (19) and (20), the weighting of the input discontinuities is characterized by the norms of the scaling transfer functions. Take (19) for instance. The transient $Y(s, e_i)$ is a linear combination of the delayed

impulse responses of

$$G_{B,n}(s) := \frac{b_n}{A(s)}, \quad G_{B,n-1}(s) := \frac{b_n s + b_{n-1}}{A(s)}, \ldots . \quad (24)$$

The overall transient speed depends on the poles and zeros in the individual modes. The 2 norm of the impulse response of $G_{B,i}$ equals the $\mathcal{H}_2$ norm of the stable transfer function $G_{B,i}(s)$. Thus, the 2 norm of the overall time-domain response $y_{e_i}(t) = \mathcal{L}^{-1}\{Y(s,e_i)\}$ is upper bounded by $||y_{e_i}(t)||_2 \le ||G_{B,n}(s)||_2|e_{u,n-1}| + ||G_{B,n-1}(s)||_2|e_{u,n-2}| + \ldots ||G_{B,1}(s)||_2|e_{u,0}|$. Computing the $\mathcal{H}_2$ norms of the $G_{B,i}(s)$ values provides a quantitative understanding of the significant terms in the discontinuities of the input and its derivatives; and hence can guide designers about the selection of the switching instance.

To examine the transient speed in add-on control, first consider the *natural response* $Y(s,t_0^-)$, which is (22) with t_0^+ replaced by t_0^-, namely

$$Y\left(s,t_0^-\right) = e^{-t_0 s}\left\{G_0(s)y\left(t_0^-\right) + G_1(s)\dot{y}\left(t_0^-\right) + \ldots\right\} \quad (25)$$

where

$$G_0(s) = \frac{s^{n-1} + a_{n-1}s^{n-2} + \cdots + a_1}{A(s)} = \frac{1}{s} - \frac{a_0}{A(s)}\frac{1}{s} \quad (26)$$

$$G_1(s) = \frac{s^{n-2} + a_{n-1}s^{n-3} + \cdots + a_2}{A(s)} = \frac{1}{s}G_0(s) - \frac{a_1}{A(s)}\frac{1}{s} \quad (27)$$

$$\vdots$$

$$G_i(s) = \frac{1}{s}G_{i-1}(s) - \frac{a_i}{A(s)}\frac{1}{s} \quad \forall i \in \{1,2,\ldots,n-1\}. \quad (28)$$

In the time domain, from Final Value Theorem, all elements $\mathcal{L}^{-1}\{G_i(s)\}$ $\forall i$ in (25) have zero steady-state values. Hence, the transient indeed eventually converges to zero. From (26), $\mathcal{L}^{-1}\{G_0(s)\}$ is the difference between a unit step and the step response of $a_0/A(s)$, whose transient duration depends on the poles from $A(s) = 0$. $\mathcal{L}^{-1}\{G_1(s)\}$ from (27) is the difference between $\mathcal{L}^{-1}\{(1/s)G_0(s)\}$—the integral of $\mathcal{L}^{-1}\{G_0(s)\}$ (a signal with zero steady-state value)—and the step response of $a_1/A(s)$. Due to the integral effect, the transient speed of $\mathcal{L}^{-1}[G_1(s)]$ is slower than $\mathcal{L}^{-1}[G_0(s)]$. Analogous results hold for the general case (28).

For the *response due to input discontinuities*, with $e_{u,i} = 0 - u^{(i)}(t_0^-)$, similar construction gives that $Y(s,e_i)$ in (23) is

$$Y(s,e_i) = -e^{-t_0 s}\left\{G_{B,0}(s)u\left(t_0^-\right) + G_{B,1}\dot{u}\left(t_0^-\right) + \ldots\right\} \quad (29)$$

$$G_{B,0}(s) = \frac{b_n s^{n-1} + \cdots + b_2 s + b_1}{A(s)} = G(s)\frac{1}{s} - \frac{b_0}{A(s)}\frac{1}{s} \quad (30)$$

$$G_{B,i}(s) = \frac{1}{s}G_{B,i-1}(s) - \frac{b_i}{A(s)}\frac{1}{s} \quad \forall i \in \{1,2,\ldots,n-1\}.$$

Thus, $\mathcal{L}^{-1}\{G_{B,0}(s)\}$ is the transient difference between the step responses of $G(s)$ and $b_0/A(s)$. In (26) and (30), the scaling a_0 and b_0 changes only the relative magnitude of the response. For a fast overall transient response, (29) needs to match the transient speed of (25) as $Y(s) = Y(s,t_0^-) + Y(s,e_i)$. If the step response of $G(s)$ is slow, namely, $\mathcal{L}^{-1}\{G(s)/s\}$ in (30) is slow compared with $\mathcal{L}^{-1}\{1/s\}$ in (26), then the transient of $\mathcal{L}^{-1}\{G_{B,0}(s)\}$ will be slower than that of $\mathcal{L}^{-1}\{G_0(s)\}$.

Actually, to have the same speed of response, the ideal case can be seen to be that $\boldsymbol{G(s) = 1}$, i.e., the add-on compensation is directly applied on the output y (which is, of course, not feasible in practice).

Recall in Fig. 2, that $G(s)$—the dynamics between the add-on control command and the plant output—is $P/(1+PC)$ in UC and $PC/(1+PC)$ in UR/UE, respectively. Among the closed-loop transfer functions in a general feedback block diagram in Fig. 2, $\boldsymbol{PC/(1+PC)}$ has the dynamic response that is closest to $\boldsymbol{G(s) = 1}$, and hence will provide the fastest transient response from the perspective of add-on injection.

On the other hand, from the pole-zero point of view, UR also has faster transient. As $G(s) = P/(1+PC)$ in UC and $G(s) = PC/(1+PC)$ in UR, the zeros of $G(s)$ in UC and UR contain, respectively, the poles of C and the zeros of C. The feedback controller C, if feasible, is preferred to be designed to have stable zeros, as open-loop unstable zeros slow down the transient and will yield various fundamental limitations in the steady-state performance of a feedback system (see [25]–[27]). For the poles of C, marginally stable poles or poles close to the imaginary axis are often needed for high-gain feedback at low frequencies (consider the case of PID control).

If again the add-on command is not ideal or there exists actuation delay, similar to (17), we can decompose the actual add-on command into two parts: $u_{add-on} = u_{ideal} + \epsilon$. In this case, (18) becomes $y^{(n)}(t) + a_{n-1}y^{(n-1)}(t) + \cdots + a_1\dot{y}(t) + a_0 y(t) = b_n\epsilon^{(n)}(t) + b_{n-1}\epsilon^{(n-1)}(t) + \cdots + b_1\dot{\epsilon}(t) + b_0\epsilon(t)$. The transient response can now be decomposed into three parts: $Y(s) = Y(s,t_0^-) + Y(s,e_i) + Y(s,\epsilon)$, where the term $Y(s,\epsilon)$ is the frequency domain response of the residual disturbance $\epsilon(t)$. Assume that the relative magnitude of the residual disturbance $\epsilon(t)$ over u_{ideal} is small and on the same scale for both UC and UR. The conclusion that UR provides faster transient response than UC is then still valid.

VI. Simulation and Experiments

In this section, the theory is verified on the wafer scanner system. The plant has a nominal model $P(s) = 1/(0.2556s^2 + 0.279s)$ and is stabilized (in negative feedback) by a PID controller $C(s) = 10\,000(1 + 2/s + 0.012s)$. The noncausal differentiation action in the PID controller is approximated by $s/(\epsilon s + 1)$ with $\epsilon = 1/9000$.

Evaluating the transient properties, such as the impulse and the step responses, reveals that $PC/(1+PC)$ provides much faster transient response with respect to rapid changing input signals. For actual disturbance rejection, abrupt step disturbances are first injected to the plant similar to that in Fig. 3, at around 0.12 s. This time, the add-on compensation is turned ON much faster than the case in Fig. 3. Fig. 7 shows the effect of add-on compensation in simulation. As the baseline controller already contains an integrator, the step disturbance is asymptotically rejected in the first subplot. For verifying the transient performance, add-on compensation is turned ON at $t_0 = 0.5$ s and $t_0 = 0.25$ s, using the UC and UR configurations in Section III-B.

Both add-on compensation schemes affect the system in the direction of reducing the error. Comparing the

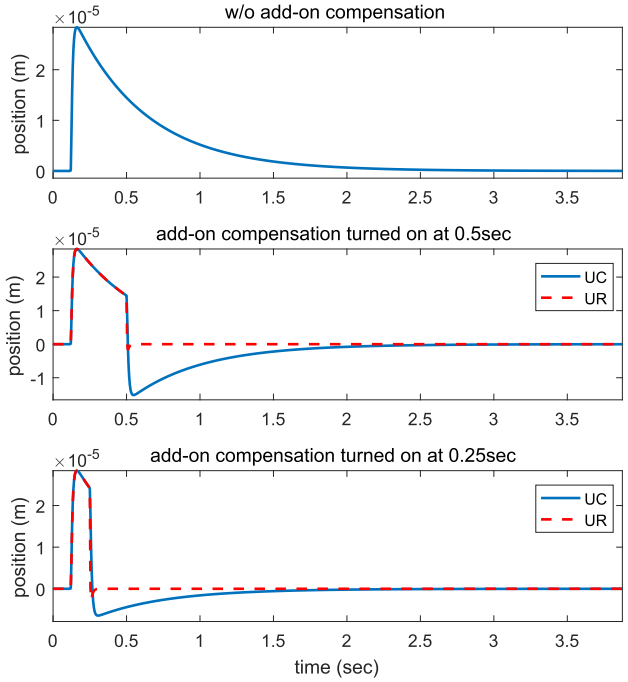

Fig. 7. Effects of add-on compensation in simulation.

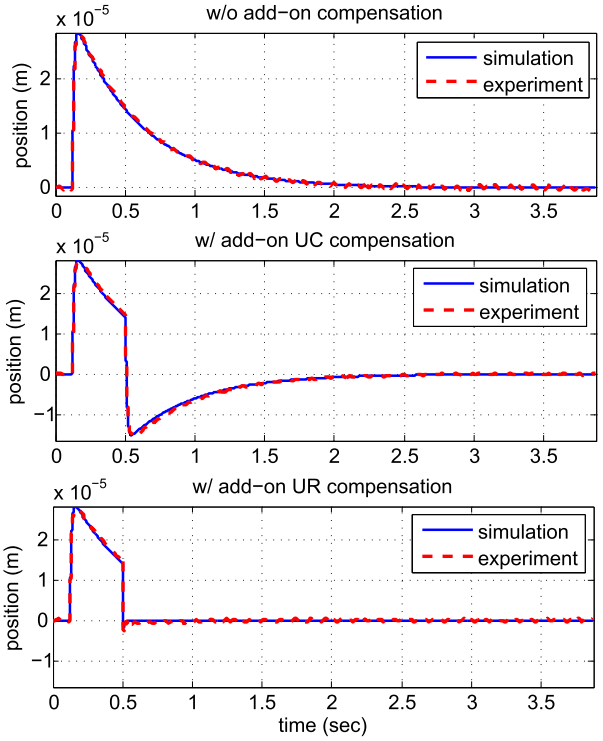

Fig. 8. Simulation and experimental comparison of transients in UC and UR compensation.

two bottom subplots in Fig. 7, it can be seen that the earlier the compensation is turned ON, the more significant the rejection effect is. For UC compensation, however, there is a large undershoot due to the system dynamics, yielding a much slower and worse transient compared with UR. Experiments on the physical system also strongly verify the results as shown in Fig. 8, where it is observed that the simulation and experimental results almost overlap with each other, and UR provides almost zero transient response in the actual system. Notice that compared with the results in Fig. 3, although the transient of UC is better in Fig. 8 (thanks to an earlier application of the add-on enhancement), the UC transient is still significantly long compared with that of UR. The adverse effect of UC transient always exists as long as there are delays in turning ON the add-on control.

Furthermore, applying the obtained results (23), we can get the transient due to input discontinuity in UR

$$\mathcal{L}^{-1}\left\{\frac{e^{-t_0 s}[(120s+10\,000)e_{u,0}+120e_{u,1}]}{0.2556s^3+120.3s^2+10\,000s+20\,000}\right\} \tag{31}$$

and in UC

$$\mathcal{L}^{-1}\left\{\frac{e^{-t_0 s}\tilde{e}_{u,0}}{0.2556s^3+120.3s^2+10\,000s+20\,000}\right\} \tag{32}$$

where $e_{u,0}$, $e_{u,1}$, and $\tilde{e}_{u,0}$ are the input discontinuities at time 0.5 s. Fig. 9 provides the computed transient response due to input discontinuity. The red dashed lines provide a zoomed-in view of the experimental data in Fig. 8. The blue solid lines are directly computed using only the nominal system model and input derivatives at time 0.5 s. It is clear that except for some effects from noises in actual hardware (there is a permanent 18-Hz force ripple in the motor), the calculated responses closely match those in the actual system.

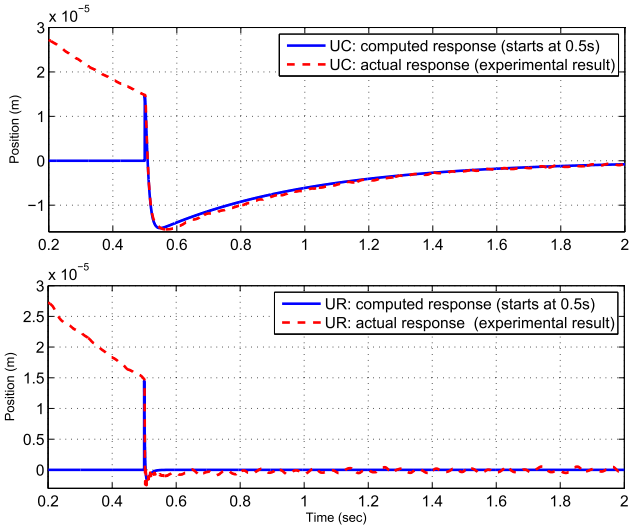

Fig. 9. Verification of transient computation algorithm in UC and UR. Red dashed lines are from actual experiments. Blue solid lines are computed from the developed transient computation equations (the equations compute only the add-on transient response, which starts at 0.5 s).

Of course, as discussed in the theoretical derivations, the transient problem exists not just for the case of step disturbances but for any add-on design with input discontinuities. Fig. 10 reveals the add-on transient in compensating frequency-dependent disturbances (a 500-Hz vibration). A similar superior performance of UR add-on design is observed. In this example, it is no longer possible to straightforwardly tell the direction of the adverse transient in UC, as the high-frequency input and its derivatives change very

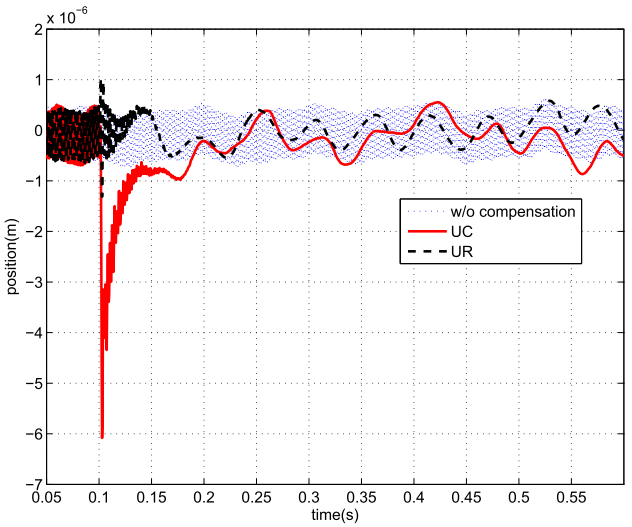

Fig. 10. Experimental comparison of add-on vibration compensations: compensation turned ON at 0.1 s, to attenuate a 500-Hz external vibration (the residual errors are from an internal 18-Hz motor force ripple).

rapidly with respect to time; and the obtained conclusions in the paper are increasingly important for avoiding large servo errors during controller implementation.

As a second example, we apply the developed tools to analyze a switched control scheme. Let $d = 0$ in Fig. 2. Consider the case of tracking a reference r as shown in the top plot of Fig. 11(a), which consists of a 10-Hz periodic signal and a 100-Hz signal that starts at around 0.6 s. r is designed to contain no discontinuities itself. To track the more aggressive 100-Hz reference signal, the feedback controller C switches to a more aggressive mode $C_2 = 40\,000 \times (1 + 3/s + 0.02\ \text{s}/(18\,000\ \text{s} + 1))$ at around 0.75 s, resulting in the improved tracking in Fig. 11(a). However, a detailed look at the control output indicates a significant increase of $|u(t)|$ as shown in Fig. 11(b). As the saturation limits of the control input are -10 and 10 V, such high-amplitude control inputs are extremely dangerous for application in practice, despite that the tracking error appears to be well controlled in simulation. Applying Theorem 2 to analyze the overlooked danger, one can find that due to the jump in the input to C_2, a significant discontinuity occurs in the output of C_2: $u(t_0^+) - u(t_0^-) = -991.2$ V; $\dot{u}(t_0^+) - \dot{u}(t_0^-) = 1.76255 \times 10^7$ V/s. The calculated -991.2 V jump in the control command can be seen to match well with the actual signal in Fig. 11(b). Furthermore, applying Proposition 5 gives the star-marked solid line in Fig. 11(c), which shows that the transient induced from the discontinuity in C_2 indeed is the main contributor of the abruptness in the overall control command.

With the prediction in Fig. 11(c), one can turn ON the input to C_2 first and slightly delay the engagement of the output of C_2, to avoid injecting the high-amplitude signals in the closed loop. For instance, a 20-step delay in turning ON the output of C_2 gives the servo results in Fig. 12, where in the top plot, the control command is seen to be maintained well under the saturation limits (actually no visual discontinuity or overshoot is observable from the new control command); and in the bottom plot, the error remains to be controlled with a slight 0.05 s longer transient compared with Fig. 11(a).[2]

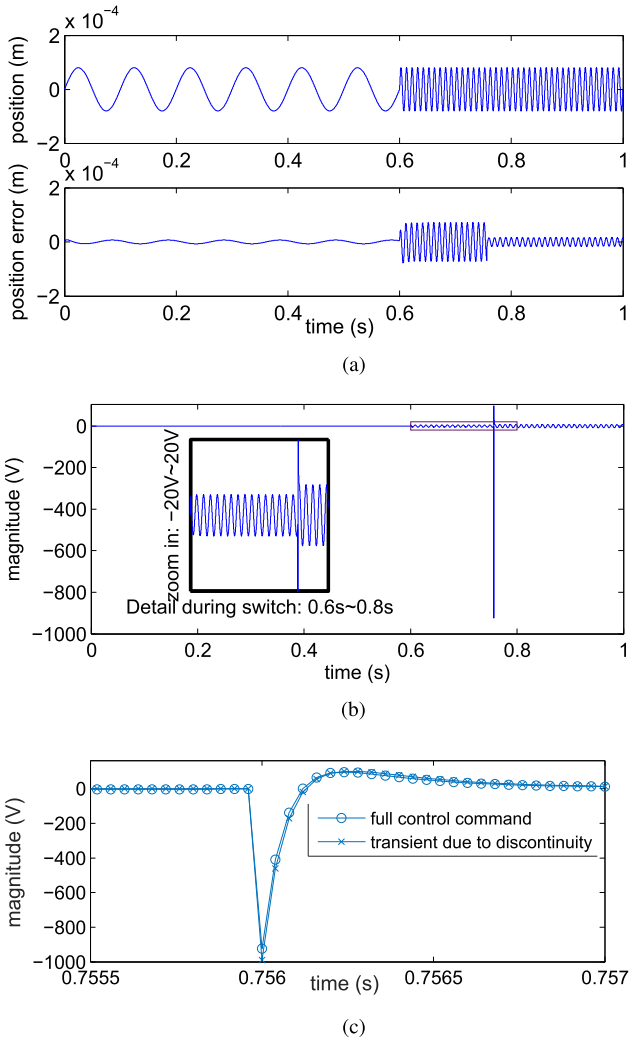

Fig. 11. Closed-loop signals with direct controller switching. (a) Reference and tracking error. (b) Corresponding control input. (c) Decomposition of control command: the transient due to discontinuity dominates in the postswitching transient control command.

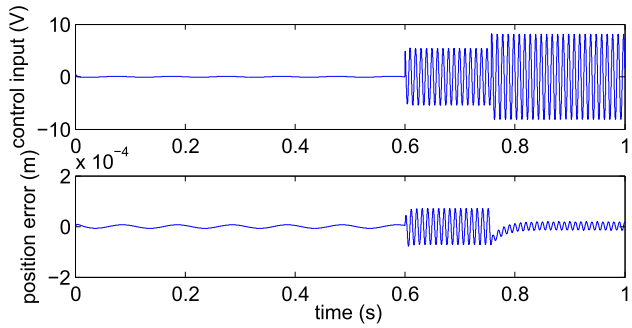

Fig. 12. Closed-loop signals with smoothened switching.

[2]Certainly, the transient can be further controlled using advanced switching mechanism. This paper focuses on providing the fundamental root causes and mathematical analysis tools.

VII. Conclusion

This paper addresses the general input-to-output discontinuity problem and applies the results to the transient improvement of add-on control designs. Simulation and experimental results are provided to show the validity of the theoretical analysis. Essentially, undesired transients occur as long as there are input discontinuities acting upon a dynamic system with poor transient properties. The problem is not only important for add-on compensation schemes, but also for other applications, such as the switching between multiple controllers.

References

[1] M. F. Heertjes, D. Hennekens, and M. Steinbuch, "MIMO feed-forward design in wafer scanners using a gradient approximation-based algorithm," *Control Eng. Pract.*, vol. 18, no. 5, pp. 495–506, 2010.

[2] K. Ohnishi, "Robust motion control by disturbance observer," *J. Robot. Soc. Jpn.*, vol. 11, no. 4, pp. 486–493, 1993.

[3] D. C. Youla, J. Bongiorno, Jr., and H. Jabr, "Modern Wiener–Hopf design of optimal controllers part I: The single-input-output case," *IEEE Trans. Autom. Control*, vol. 21, no. 1, pp. 3–13, Feb. 1976.

[4] V. Kučera, "Stability of discrete linear feedback systems," in *Proc. 6th IFAC World Congr.*, vol. 1. 1975, paper 44.1.

[5] C. J. Kempf and S. Kobayashi, "Disturbance observer and feedforward design for a high-speed direct-drive positioning table," *IEEE Trans. Control Syst. Technol.*, vol. 7, no. 5, pp. 513–526, Sep. 1999.

[6] K. S. Eom, I. H. Suh, and W. K. Chung, "Disturbance observer based path tracking control of robot manipulator considering torque saturation," *Mechatronics*, vol. 11, no. 3, pp. 325–343, 2001.

[7] K. Yang, Y. Choi, and W. K. Chung, "On the tracking performance improvement of optical disk drive servo systems using error-based disturbance observer," *IEEE Trans. Ind. Electron.*, vol. 52, no. 1, pp. 270–279, Feb. 2005.

[8] X. Chen and M. Tomizuka, "New repetitive control with improved steady-state performance and accelerated transient," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 2, pp. 664–675, Mar. 2014.

[9] 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 Syst. Technol.*, vol. 20, no. 2, pp. 408–415, Mar. 2012.

[10] X. Chen and M. Tomizuka, "Optimal decoupled disturbance observers for dual-input single-output systems," *ASME J. Dyn. Syst., Meas., Control*, vol. 136, no. 5, p. 051018, 2014.

[11] I. D. Landau, A. C. Silva, T.-B. Airimitoaie, G. Buche, and M. Noe, "Benchmark on adaptive regulation—Rejection of unknown/time-varying multiple narrow band disturbances," *Eur. J. Control*, vol. 19, no. 4, pp. 237–252, 2013.

[12] X. Chen and M. Tomizuka, "Control methodologies for precision positioning systems," in *Proc. Amer. Control Conf.*, Jun. 2013, pp. 3710–3717.

[13] R. de Callafon and C. E. Kinney, "Robust estimation and adaptive controller tuning for variance minimization in servo systems," *J. Adv. Mech. Design, Syst., Manuf.*, vol. 4, no. 1, pp. 130–142, 2010.

[14] B. D. O. Anderson, "From Youla–Kucera to identification, adaptive and nonlinear control," *Automatica*, vol. 34, no. 12, pp. 1485–1506, 1998.

[15] M. Bodson, A. Sacks, and P. Khosla, "Harmonic generation in adaptive feedforward cancellation schemes," *IEEE Trans. Autom. Control*, vol. 39, no. 9, pp. 1939–1944, Sep. 1994.

[16] J. Zhang, R. Chen, G. Guo, and T.-S. Low, "Modified adaptive feedforward runout compensation for dual-stage servo system," *IEEE Trans. Magn.*, vol. 36, no. 5, pp. 3581–3584, Sep. 2000.

[17] B. Widrow and E. Walach, *Adaptive Inverse Control: A Signal Processing Approach, Reissue Edition*. Hoboken, NJ, USA: Wiley, Nov. 2007.

[18] B. P. Rigney, L. Y. Pao, and D. A. Lawrence, "Settle time performance comparisons of stable approximate model inversion techniques," in *Proc. Amer. Control Conf.*, Jun. 2006, pp. 600–605.

[19] B. P. Rigney, L. Y. Pao, and D. A. Lawrence, "Nonminimum phase dynamic inversion for settle time applications," *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 5, pp. 989–1005, Sep. 2009.

[20] M. Cao and A. S. Morse, "Dwell-time switching," *Syst. Control Lett.*, vol. 59, no. 1, pp. 57–65, 2010.

[21] H. Lin and P. J. Antsaklis, "Stability and stabilizability of switched linear systems: A survey of recent results," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 308–322, Feb. 2009.

[22] P. Bolzern, P. Colaneri, and G. De Nicolao, "Almost sure stability of Markov jump linear systems with deterministic switching," *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 209–214, Jan. 2013.

[23] D. Liberzon and A. S. Morse, "Basic problems in stability and design of switched systems," *IEEE Control Syst.*, vol. 19, no. 5, pp. 59–70, Oct. 1999.

[24] Z. Sun and S. S. Ge, "Analysis and synthesis of switched linear control systems," *Automatica*, vol. 41, no. 2, pp. 181–195, 2005.

[25] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory*. New York, NY, USA: Macmillan, 1992, vol. 134.

[26] G. Stein, "Respect the unstable," *IEEE Control Syst.*, vol. 23, no. 4, pp. 12–25, Aug. 2003.

[27] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, 2nd ed. Chichester, U.K.: Wiley, 2005.

Xu Chen received the bachelor's degree from Tsinghua University, Beijing, China, in 2008, and the M.S. and Ph.D. degrees in mechanical engineering from the University of California at Berkeley (UC Berkeley), Berkeley, CA, USA, in 2010 and 2013, respectively.

From 2013 to 2014, he was a Lecturer with UC Berkeley. In 2014, he joined as a Faculty Member of the Department of Mechanical Engineering, University of Connecticut, Storrs, CT, USA. His current research interests include digital control, learning control, adaptive and optimal controls, with applications to advanced manufacturing, mechatronics, and precision engineering.

Dr. Chen was a recipient of the Young Investigator Award in 2014 ISCIE/ASME International Symposium on Flexible Automation and the 2012 Chinese Government Award for Outstanding Students Abroad.

Tianyu Jiang received the B.S. degree from the University of Science and Technology of China, Hefei, China in 2014. He is currently pursuing the Ph.D. degree with the Department of Mechanical Engineering, University of Connecticut, Storrs, CT, USA.

His current research interests include the theory and practice of dynamic systems and controls with application to additive and advanced manufacturing.