INTERNATIONAL JOURNAL OF CONTROL, 2016
VOL. 89, NO. 10, 2121–2138
http://dx.doi.org/10.1080/00207179.2016.1149885

Taylor & Francis
Taylor & Francis Group

Fastest recovery after feedback disruption

Zhaoxu Yu[a] and Jacob Hammer[b]

[a]Department of Automation, Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai, P.R. China; [b]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

ABSTRACT

The problem of quickly reducing operating errors during recovery from a feedback disruption is considered. The objective is to design controllers that reduce operating errors as quickly as possible, once feedback has been restored. It is shown that robust optimal feedback controllers that achieve this objective do exist. Furthermore, it is shown that the performance of optimal controllers can be approximated as closely as desired by controllers that generate bang–bang input signals for the controlled system. Controllers that generate bang–bang signals are relatively easy to derive and implement, since bang–bang signals are characterised by a finite list of scalars – their switching times.

ARTICLE HISTORY
Received 15 September 2015
Accepted 31 January 2016

KEYWORDS
Optimal control; feedback disruption; nonlinear control

1. Introduction

The ability of feedback to reduce operating errors in automatic control systems has been widely documented in the scientific literature. No doubt, this ability accounts for the widespread use of feedback in engineering systems as well as for its pervasive manifestations in nature. Yet, disruptions in feedback service cannot always be avoided. Such disruptions may occur as a result of component failures, inauspicious operating conditions, or, in some applications, as part of a pre-planned operating strategy intended to reduce costs and operational burdens. A methodology for reducing operating errors during feedback disruptions has been discussed in Chakraborty and Hammer (2009, 2010). Even so, increased operating errors during periods of feedback disruption are unavoidable. The objective of this paper is to develop controllers that help reduce such errors as quickly as possible, once feedback has been restored.

Potential applications of such controllers abound. For example, a missile that has lost line-of-sight to its target would need to restore low-error target-tracking as quickly as possible, once line-of-sight has been restored. The controllers developed in this paper achieve this objective. Another potential application is in networked control systems, where feedback signals are provided only intermittently, so as to conform to network traffic limitations (see Montestruque & Antsaklis, 2004; Nair, Fagnani, Zampieri, & Evans, 2007; Zhivogyladov & Middleton, 2003, the references cited in these papers, and others). Here, systems naturally develop operating errors during periods of feedback absence; an important task is to reduce these errors as quickly as possible upon feedback re-activation, and this task is accomplished by controllers developed in this paper.

The material discussed in this paper also finds potential applications in biomedicine. Consider, for example, a diabetic patient whose blood glucose concentration has reached dangerously low or dangerously high levels. The deviation of the glucose concentration from its normal level is an error in the operation of the glucose concentration control mechanism, and reducing this error as quickly as possible is important to the health of the patient. The controllers developed in this paper achieve this goal, thus giving rise to optimal treatment protocols to adjust glucose concentration to an acceptable level as quickly as possible.

Another important application can be found in one of the most commonly used control technologies – digital control of continuous-time systems. Here, as is well known, continuous-time systems are controlled by digital computers via a process of periodic sampling: the controlled system's output is sampled periodically, and these samples form the feedback signal of the digital controller. As no feedback is available during the time between samples, the controlled system develops operating errors during the inter-sample period. The controllers developed in this paper facilitate speedy reduction of such errors upon the arrival of the next sample. In this way, employing such controllers can help improve the performance of digitally controlled systems.

An additional class of potential applications can be found in the conduction of economic policy. Here too,

CONTACT Jacob Hammer hammer@mst.ufl.edu

feedback data about economic performance is received at pre-set time intervals, most commonly once every month. Deviations from desired economic policy that develop during such time intervals can be reduced as quickly as possible through the employment of control strategies developed in this paper.

The previous few paragraphs lead us to the following general problem.

Problem 1.1: Design controllers that reduce operating errors as quickly as possible, once feedback is restored after a feedback disruption.

As our discussion so far indicates, Problem 1.1 is relevant to a broad range of applications. Accordingly, there is a need to resolve this problem on a general level. In the past, when encountered by engineers in practice, Problem 1.1 has most commonly been addressed through a variety of specialised techniques that helped settle the issue in specific cases (see, for example, Balakrishnan, Tsourdos, & White, 2012 and the references cited therein). Our aim in this paper is to employ mathematical optimisation techniques to resolve the problem on a general level that is widely applicable. To the best of our knowledge, the general mathematical optimisation problem considered in this paper has not been addressed in the literature before.

In formal terms, the control configuration we consider is depicted in Figure 1, where Σ is the controlled system and C is a controller. As a function of the time t, the input signal of Σ is $u(t)$ and the state of Σ, which also serves as the output, is $x(t)$. The input signal $u(t)$ of Σ is generated by the controller C. Under the present scenario, Σ has been operating for some time in open loop, when feedback is restored momentarily at the time $t = 0$. This momentary closure of the feedback loop provides the controller C with the state sample $x(0)$. Using this sample, C generates an input signal $u(t)$ of Σ, with the objective of driving Σ to reduce as quickly as possible errors that may have accumulated during open-loop operation.

As always, it is important to take into account inaccuracies and uncertainties present in the description of the controlled system Σ. To this end, let Σ_0 be the nominal description of Σ, and, given a real number $\gamma > 0$,

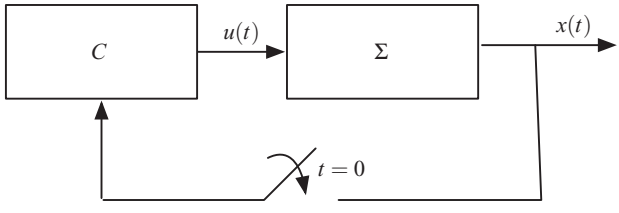

Figure 1. Feedback is restored at $t = 0$.

denote by $\mathcal{F}_\gamma(\Sigma_0)$ the family of all systems whose parameters deviate by no more than γ from their values in Σ_0. Let Σ be a member of $\mathcal{F}_\gamma(\Sigma_0)$, and let $x_0 = x(0)$ be the state of Σ at the time $t = 0$. Clearly, the state $x(t)$ of Σ at the time t depends on x_0 and on the input signal u, so we write

$$x(t) = \Sigma(x_0, u, t). \tag{1.1}$$

More often than not, practical systems impose restrictions on the maximal input signal magnitude they can tolerate. To comply with such restrictions, we assume that members of $\mathcal{F}_\gamma(\Sigma_0)$ permit only input signals with a maximal magnitude of $K > 0$. To state this requirement formally, let R be the set of real numbers, and let R^+ be the set of all non-negative real numbers. For integers n, $m > 0$, let R^n be the set of all real n-dimensional column vectors and let $R^{n \times m}$ be the set of all $n \times m$ matrices of real numbers. Denoting by $|r|$ the absolute value of a real number r, the L^∞-norm of a matrix $v = (v_{ij}) \in R^{n \times m}$ is

$$|v| = \max_{\substack{i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, m}} |v_{ij}|.$$

For a time-dependent matrix $v(t) \in R^{n \times m}$, $t \geq 0$, the L^∞-norm is

$$|v(t)|_\infty := \sup_{t \geq 0} |v(t)|,$$

where $|v(t)|_\infty := \infty$ if the supremum does not exist. The function $v(t)$ is *bounded* if $|v(t)|_\infty < \infty$. In this notation, members of $\mathcal{F}_\gamma(\Sigma_0)$ accept only input signals $u(t) \in R^m$, $t \geq 0$, satisfying

$$|u(t)|_\infty \leq K.$$

After possibly inducing a shift of the state of Σ, we assume that the desired nominal trajectory of Σ is the zero trajectory, namely, the trajectory $x(t) = 0$ for all $t \geq 0$. In these terms, the objective of the controller C is to generate an input signal $u(t)$ that takes every member Σ of the family $\mathcal{F}_\gamma(\Sigma_0)$ from the initial state x_0 to the vicinity of the zero state $x = 0$ as quickly as possible.

Needless to say, due to inaccuracies and uncertainties, it is not possible to bring all members of the family $\mathcal{F}_\gamma(\Sigma_0)$ exactly to the zero state. Instead, we require of the controller to bring all members of the family sufficiently close to the zero state, so the state $x(t)$ of each member satisfies the inequality $x(t)^T x(t) \leq \delta$, where $(\cdot)^T$ denotes the transpose and $\delta > 0$ is a specified real number. It will

be convenient to use the notation

$$|x|_2^2 := x^T x.$$

Definition 1.1: The *δ-vicinity of the origin* is the domain

$$\rho(\delta) := \left\{x \in R^n : |x|_2^2 \leq \delta\right\}.$$

Using the notation (1.1), the minimal time within which a given input signal u can take all members of the family $\mathcal{F}_\gamma(\Sigma_0)$ from the initial state x_0 to the origin's δ-vicinity $\rho(\delta)$ is

$$t_f(x_0, u) := \inf_t \left\{t \geq 0 : \sup_{\Sigma \in \mathcal{F}_\gamma(\Sigma_0)} |\Sigma(x_0, u, t)|_2^2 \leq \delta\right\},$$

where $t_f(x_0, u) := \infty$ if u cannot bring all members of $\mathcal{F}_\gamma(\Sigma_0)$ to $\rho(\delta)$ at the same time. The minimal time within which all members of $\mathcal{F}_\gamma(\Sigma_0)$ can be taken from x_0 to $\rho(\delta)$ by a permissible input signal is then

$$t_f^*(x_0) := \inf_{|u|_\infty \leq K} t_f(x_0, u),$$

where $t_f^*(x_0) := \infty$ if there is no permissible input signal that brings all members of $\mathcal{F}_\gamma(\Sigma_0)$ to $\rho(\delta)$ at the same time.

If $t_f^*(x_0) < \infty$, the question arises whether there is a permissible optimal input signal $u^*(x_0)$ that brings all members of $\mathcal{F}_\gamma(\Sigma_0)$ from x_0 to $\rho(\delta)$ in the minimal time $t_f^*(x_0)$, namely, an input signal $u^*(x_0)$ that satisfies the requirement

$$\left|\Sigma\left(x_0, u^*(x_0), t_f^*(x_0)\right)\right|_2^2 \leq \delta \text{ for all } \Sigma \in \mathcal{F}_\gamma(\Sigma_0).$$

In Section 3, we show that such an optimal signal exists under rather general conditions.

Another important issue is the implementation of the optimal signal $u^*(x_0)$, as such signals may be complicated vector-valued functions of time and hard to compute and implement. An important objective of our discussion is to derive signals that are easy to compute and implement and that provide performance that is as close as desired to optimal performance. In Section 4, we show that an optimal response can be approximated as closely as desired by bang–bang input signals with a finite number of switchings. Such input signals are relatively easy to calculate and implement, since bang–bang signals are basically determined by a finite list of real numbers – their switching times. To summarise, the objectives of this paper can be formulated as follows.

Problem 1.2: Given real numbers δ, $\gamma > 0$ and a family of systems $\mathcal{F}_\gamma(\Sigma_0)$ with initial state x_0, find an optimal input signal $u^*(x_0)$ that takes all members of $\mathcal{F}_\gamma(\Sigma_0)$ to the δ-vicinity of the origin at the minimal time $t_f^*(x_0)$. In addition, find an easy-to-implement input signal that achieves close to optimal performance.

We consider Problem 1.2 in Section 3, where we show that an optimal input signal $u^*(x_0)$ exists for a rather broad class of input-affine nonlinear time-varying systems. As one might imagine, the existence of an optimal solution depends on certain controllability properties of the members of $\mathcal{F}_\gamma(\Sigma_0)$; these controllability properties are discussed in Section 2. Furthermore, in Section 4, we show that an optimal input signal $u^*(x_0)$ can be replaced by a bang–bang signal, without appreciably affecting performance. Recall that a bang–bang signal is a signal whose components switch between their extremal values $-K$ and $+K$. When compared to other classes of signals, bang–bang signals are easier to compute and implement, since a bang–bang signal is basically determined by its switching times.

Our discussion depends on the literature on min–max optimisation, including Kelendzheridze (1961), Pontryagin, Boltyansky, Gamkrelidze, and Mishchenko (1962), Neustadt (1966, 1967), Gamkrelidze (1965), Luenberger (1969), Young (1969), Warga (1972), Chakraborty and Hammer (2009), Chakraborty and Shaikshavali (2009), the references cited in these works, and many others. Yet, to the best of our knowledge, there are no earlier reports in the literature that specifically address the existence, implementation, or approximation of solutions of Problem 1.2.

The paper is organised as follows. Section 2 introduces the basic framework we use in our discussion. The existence of optimal solutions of Problem 1.2 is discussed in Section 3, while Section 4 shows that optimal performance can be approximated as closely as desired by bang–bang signals. The paper concludes in Section 5 with an example.

2. A formal optimisation framework

2.1 Preliminaries

The discussion in this paper centres on the control of a certain class of input-affine nonlinear systems. In addition to linear systems, this class can be used to model certain types of nonlinear engineering systems, such as flexible joints or special electrical motors (e.g., Modeling, Spong, Hutchinson, & Vidyasagar, 2006). More specifically, our objective is to control a system Σ with $n > 0$

states and $m > 0$ inputs, described by a differential equation of the form

$$\Sigma : \dot{x}(t) = a(t, x(t)) + b(t, x(t))u(t),\ t \geq 0, \quad (2.1)$$

where x: $R^+ \to R^n$: $t \mapsto (x_1(t), x_2(t), \ldots, x_n(t))^T$ is the state of Σ and u: $R^+ \to R^m$: $t \mapsto u(t) = (u_1(t), u_2(t), \ldots, u_m(t))^T$ is the input signal. Here, a: $R^+ \times R^n \to R^n$ and b: $R^+ \times R^n \to R^{n \times m}$ are continuous functions; the initial time is $t = 0$; and the initial state is $x(0) = x_0$. As indicated earlier, the system Σ permits only input signals of magnitude not exceeding $K > 0$, so that $|u|_\infty \leq K$.

To account for the fact that practical systems are subject to uncertainties and inaccuracies, we construe the functions a and b of (2.1) as sums of nominal and error terms

$$\begin{aligned} a(t, x) &= a_0(t, x) + a_\gamma(t, x), \\ b(t, x) &= b_0(t, x) + b_\gamma(t, x), \end{aligned} \quad (2.2)$$

where a_0: $R^+ \times R^n \to R^n$ and b_0: $R^+ \times R^n \to R^{n \times m}$ are given continuous functions describing the nominal system, while a_γ: $R^+ \times R^n \to R^n$ and b_γ: $R^+ \times R^n \to R^{n \times m}$ are unspecified continuous functions describing uncertainties and modelling errors. We assume that the terms in (2.2) satisfy the Lipschitz conditions

$$\begin{aligned} |a_0(t, x_2) - a_0(t, x_1)| &\leq M|x_2 - x_1|, \\ |b_0(t, x_2) - b_0(t, x_1)| &\leq M|x_2 - x_1|, \\ a_0(t, 0) &:= 0, \\ |b_0(t, 0)| &\leq M, \end{aligned} \quad (2.3)$$

$$\begin{aligned} |a_\gamma(t, x_2) - a_\gamma(t, x_1)| &\leq \gamma|x_2 - x_1|, \\ |b_\gamma(t, x_2) - b_\gamma(t, x_1)| &\leq \gamma|x_2 - x_1|, \\ a_\gamma(t, 0) &:= 0, \\ |b_\gamma(t, 0)| &\leq \gamma, \end{aligned} \quad (2.4)$$

for all $t \in R^+$ and all $x_1, x_2 \in R^n$. Here, M and γ are specified real positive numbers, with γ being interpreted as a small number characterising the uncertainty about the model of Σ. The *nominal system* Σ_0 is then given by

$$\begin{aligned} \Sigma_0 : \dot{x}(t) &= a_0(t, x(t)) + b_0(t, x(t))u(t), \\ t \geq 0,\ x(0) &= x_0. \end{aligned} \quad (2.5)$$

Definition 2.1: Let Σ_0 be a nominal system of the form (2.5), where $a_0(t, x)$ and $b_0(t, x)$ are specified continuous functions satisfying (2.3), and let $\gamma > 0$ be a real number. The family $\mathcal{F}_\gamma(\Sigma_0)$ consists of all systems Σ of the form (2.1), where $a(t, x)$ and $b(t, x)$ are given by (2.2), with $a_\gamma(t, x)$ and $b_\gamma(t, x)$ being unspecified continuous functions satisfying (2.4).

We start our discussion of the family $\mathcal{F}_\gamma(\Sigma_0)$ by showing that its members are 'well behaved'. First, some common terminology.

Definition 2.2: A system Σ of the form (2.1) has *no finite escape time* if, for every initial condition x_0, for every bounded input function u, and for every time $t \geq 0$, there is a non-negative real number $N(x_0, u, t) < \infty$ such that $\sup_{\theta \in [0,t]} |\Sigma(x_0, u, \theta)| \leq N(x_0, u, t)$.

In brief, the response of a system with no finite escape time is bounded at all finite times, but it may diverge as $t \to \infty$.

Proposition 2.1: *Members of the family of systems $\mathcal{F}_\gamma(\Sigma_0)$ have no finite escape time.*

Proof: Fix an initial condition $x_0 \in R^n$, and let u: $R^+ \to R^m$ be a bounded input function, say $|u|_\infty \leq K$, where $K \geq 0$ is a real number. Using (2.1), we can write

$$x(t) = x_0 + \int_0^t [a(s, x(s)) + b(s, x(s))u(s)]\, ds.$$

Considering that $a(t, 0) = 0$ by (2.3) and (2.4), we can rewrite

$$\begin{aligned} x(t) = x_0 + \int_0^t \{&[a(s, x(s)) - a(s, 0)] \\ &+ [[b(s, x(s)) - b(s, 0)] + b(s, 0)]\, u(s)\} ds. \end{aligned}$$

Using again (2.3) and (2.4), we obtain

$$\begin{aligned} \sup_{0 \leq \theta \leq t} |x(\theta)| \leq |x_0| &+ \int_0^t (M + \gamma) \sup_{0 \leq \theta \leq t} |x(\theta)| ds \\ &+ \int_0^t (M+\gamma) \left[\left(\sup_{0 \leq \theta \leq t} |x(\theta)| \right) + 1 \right] \left(\sup_{0 \leq \theta \leq t} |u(\theta)| \right) ds. \end{aligned}$$

This yields

$$\begin{aligned} \sup_{0 \leq \theta \leq t} |x(\theta)| \leq |x_0| &+ (M + \gamma) \left(\sup_{0 \leq \theta \leq t} |x(\theta)| \right) t \\ &+ (M + \gamma) \left[\left(\sup_{0 \leq \theta \leq t} |x(\theta)| \right) + 1 \right] Kt, \end{aligned}$$

or

$$\begin{aligned} \sup_{0 \leq \theta \leq t} |x(\theta)| \leq |x_0| &+ (M + \gamma)(1 + K) \left(\sup_{0 \leq \theta \leq t} |x(\theta)| \right) t \\ &+ (M + \gamma) Kt, \end{aligned}$$

so that

$$[1-(M+\gamma)(1+K)t]\left(\sup_{0\le\theta\le t}|x(\theta)|\right) \le |x_0|+(M+\gamma)Kt.$$

Now, choose a value of $t>0$, say $t=\tau>0$, such that $(M+\gamma)(1+K)\tau<1$. Then, we get

$$\sup_{0\le\theta\le\tau}|x(\theta)| \le [|x_0|+(M+\gamma)K\tau]/[1-(M+\gamma)(1+K)\tau]<\infty. \quad (2.6)$$

Next, note that the value of τ depends only on the constants M, γ, and K. Partition the time axis into intervals of length τ, i.e., into the intervals $[0,\tau]$, $[\tau,2\tau]$, … , $[i\tau,(i+1)\tau]$, … , $i=1,2,...$ Then, the value $x(i\tau)$ of $x(t)$ at the end of the interval $[(i-1)\tau,i\tau]$ clearly is the initial value of $x(t)$ for the interval $[i\tau,(i+1)\tau]$. Repeating the arguments that lead to (2.6) for an integer $i\ge1$ yields the inequality

$$\sup_{i\tau\le\theta\le(i+1)\tau}|x(\theta)| \le [|x(i\tau)|+(M+\gamma)K\tau]/[1-(M+\gamma)(1+K)\tau]<\infty,\ i=1,2,...,$$

This inequality shows that, for all integers $i\ge0$, the function $x(t)$ is bounded over the interval $[i\tau,(i+1)\tau]$ whenever $x(i\tau)$ is bounded. Considering that $x(0)$ is bounded, this implies, by induction on the integer i, that $x(t)$ is bounded at all times $t\ge0$. Finally, as the latter is valid for every member Σ of $\mathcal{F}_\gamma(\Sigma_0)$, we conclude that members of $\mathcal{F}_\gamma(\Sigma_0)$ have no finite escape times, and the proposition holds. □

2.2 Controllability considerations

Our discussion in this paper is within the mathematical framework provided by inner product spaces, using the following inner product (see also Chakraborty & Hammer, 2009). For a real number $\alpha>0$ and an integer $m>0$, let $L_2^{\alpha,m}$ be the space of all Lebesgue measurable functions $f,g\colon R^+\to R^m$ with the inner product

$$\langle f,g\rangle := \int_0^\infty e^{-\alpha t}f^T(s)g(s)ds. \quad (2.7)$$

This inner product is well defined for all bounded members f,g of $L_2^{\alpha,m}$.

Recalling that all input signals of members of the family $\mathcal{F}_\gamma(\Sigma_0)$ must be bounded by $K>0$, our main interest is in the class of members of $L_2^{\alpha,m}$ that are bounded by K, namely, in the class of functions

$$U(K) := \left\{u\in L_2^{\alpha,m} : |u|_\infty\le K\right\}. \quad (2.8)$$

This class will serve as the class of input signals for the controlled system Σ of Figure 1.

Throughout our discussion, we assume that the nominal system Σ_0 is controllable in the following sense.

Definition 2.3: Let $K>0$ be a real number. A system Σ_0 of the form (2.5) is *K-controllable* at the initial state x_0 if there is an input signal $u\in U(K)$ that takes Σ_0 from x_0 to the zero state in finite time.

K-controllability of the nominal system Σ_0 entails a somewhat weaker form of controllability of the entire family $\mathcal{F}_\gamma(\Sigma_0)$, as follows.

Proposition 2.2: *Let Σ_0 be a system of the form (2.5), and assume that Σ_0 is K-controllable at the initial state x_0. Then, for every real number $\delta>0$, there is a real number $\gamma>0$ for which the following is true: there is an input signal $u\in U(K)$ and a time $\tau\ge0$ such that $\Sigma^T(x_0,u,\tau)\Sigma(x_0,u,\tau)<\delta$ for all members $\Sigma\in\mathcal{F}_\gamma(\Sigma_0)$.*

Proof: Assume that Σ_0 is K-controllable at the initial state x_0. Then, by Definition 2.3, there is an input function $u\in U(K)$ and a time $\tau\ge0$ such that $\Sigma_0(x_0,u,\tau)=0$. Let Σ be a member of $\mathcal{F}_\gamma(\Sigma_0)$ and denote $x(t):=\Sigma_0(x_0,u,t)$, $x'(t):=\Sigma(x_0,u,t)$, and $\xi(t):=x'(t)-x(t)$. Then, using Definition 2.1 and the facts that $a_\gamma(t,0)=0$ and $|b_\gamma(t,0)|\le\gamma$ by (2.4), we can write

$$\begin{aligned}\dot\xi(t)&=\dot x'(t)-\dot x(t)\\ &=[a_0(t,x'(t))-a_0(t,x(t))]\\ &\quad+[a_\gamma(t,x'(t))-a_\gamma(t,0)]\\ &\quad+[b_0(t,x'(t))-b_0(t,x(t))]u(t)\\ &\quad+\left\{[b_\gamma(t,x'(t))-b_\gamma(t,0)]+b_\gamma(t,0)\right\}u(t).\end{aligned}$$

Then, for any time $t'<t$, we obtain

$$\begin{aligned}\xi(t)=\xi(t')+\int_{t'}^t\Big\{&[a_0(s,x'(s))-a_0(s,x(s))]\\ &+[a_\gamma(s,x'(s))-a_\gamma(s,0)]\\ &+[b_0(s,x'(s))-b_0(s,x(s))]u(s)\\ &+\left\{[b_\gamma(s,x'(s))-b_\gamma(s,0)]+b_\gamma(s,0)\right\}u(s)\Big\}ds.\end{aligned}$$

Now, referring to the time $\tau\ge0$ mentioned at the beginning of this proof, it follows by Proposition 2.1 that there is a real number $N>0$ such that $|x(t)|\le N$ and $|x'(t)|\le N$ for all $t\in[0,\tau]$. Consequently, (2.3), (2.4), and the fact that the input signal u is bounded by K imply that, for

all $0 \le t' < t \in [0, \tau]$, we have

$$\begin{aligned}
&\sup_{t' \le \theta \le t} |\xi(\theta)| \le |\xi(t')| \\
&\quad + M \int_{t'}^{t} \left(\sup_{t' \le \theta \le t} |\xi(\theta)| \right) ds \\
&\quad + \gamma \int_{t'}^{t} \left(\sup_{t' \le \theta \le t} |x'(t) - 0| \right) ds \\
&\quad + M \int_{t'}^{t} \left(\sup_{t' \le \theta \le t} |\xi(\theta)| \right) \left(\sup_{0 \le \theta \le t} |u(\theta)| \right) ds \\
&\quad + \gamma \int_{t'}^{t} \left\{ \left(\sup_{t' \le \theta \le t} |x'(t) - 0| \right) + 1 \right\} \left(\sup_{0 \le \theta \le t} |u(\theta)| \right) ds \\
&\le |\xi(t')| + \left(\sup_{t' \le \theta \le t} |\xi(\theta)| \right) M(t - t')(1 + K) \\
&\quad + \gamma (t - t') (N(1 + K) + K).
\end{aligned}$$

Next, choose a real number $\Delta > 0$ for which $M\Delta(1 + K) \le 1/2$ and $p := \tau/\Delta$ is an integer; then take $t = t' + \Delta$. This yields

$$\sup_{t' \le \theta \le t' + \Delta} |\xi(\theta)| \le 2|\xi(t')| + 2\gamma\Delta(N(1 + K) + K). \tag{2.9}$$

Furthermore, partition the interval $[0, \tau]$ into segments of length Δ to obtain the partition $[0, \Delta], [\Delta, 2\Delta], \dots, [(p-1)\Delta, p\Delta]$. Then, using $t' := i\Delta$ for an integer $i \in \{0, 1, \dots, p-1\}$, we can rewrite (2.9) in the form

$$\sup_{i\Delta \le \theta \le (i+1)\Delta} |\xi(\theta)| \le 2|\xi(i\Delta)| + 2\gamma\Delta(N(1 + K) + K),$$
$$i = 0, 1, \dots, p - 1.$$

Iterating this inequality over $i = 0, 1, \dots, p - 1$, and using the fact that $x'(t)$ and $x(t)$ both have the same initial value $x'(0) = x(0) = x_0$, so that $\xi(0) = 0$, we obtain

$$\sup_{0 \le \theta \le \tau} |\xi(\theta)| \le q_{p-1}\gamma\Delta(N(1 + K) + K),$$

where q_p is the integer determined by the recursion $q_{k+1} = 2q_k + 2 = 2(q_k + 1)$, $q_0 = 0$, $k = 0, 1, 2, \dots, p - 1$. As q_{p-1} is clearly finite, the proposition's assertion is valid for $\gamma < \delta/[q_{p-1}\Delta(N(1 + K) + K)]$, and our proof concludes. □

2.3 Formal problem statement

We restate now Problem 1.2 in the following formal form that will serve as the basis of our forthcoming discussion.

Problem 2.1: Let $K > 0$ be a real number, let $\mathcal{F}_\gamma(\Sigma_0)$ be the family of systems of Definition 2.1, let x_0 be the initial state of all members of $\mathcal{F}_\gamma(\Sigma_0)$, and assume that Σ_0 is K-controllable from x_0. Let $U(K)$ of (2.8) be the set of permissible input signals of $\mathcal{F}_\gamma(\Sigma_0)$, and let $\gamma, \delta > 0$ be real numbers that are consistent with Proposition 2.2. For an input signal $u \in U(K)$, denote

$$t_f(x_0, u) := \inf_{t \ge 0} \left\{ \sup_{\Sigma \in \mathcal{F}_\gamma(\Sigma_0)} |\Sigma(x_0, u, t)|_2^2 \le \delta \right\},$$

and set

$$t_f^*(x_0) := \inf_{u \in U(K)} t_f(x_0, u).$$

Then,

(1) find the minimal time $t_f^*(x_0)$.
(2) If $t_f^*(x_0) < \infty$, determine whether there exists an optimal input signal $u^*(x_0) \in U(K)$ for which $t_f^*(x_0) = t_f(x_0, u^*(x_0))$, namely, determine whether there is an input signal that achieves optimal performance.
(3) If an optimal input signal $u^*(x_0)$ exists, find an easy-to-implement signal that approximates optimal performance.

We consider the existence of an optimal input signal $u^*(x_0)$ in Section 3 below, where we show that such an optimal input signal exists under the broad conditions of Problem 2.1. We further show in Section 4 that optimal performance can be approximated as closely as desired by a bang–bang input signal – a piecewise constant input signal whose components switch between the extremal input bounds $-K$ and $+K$. Bang–bang signals are relatively easy to calculate and implement, since they are characterised by a finite string of switching times.

3. Existence of optimal solutions

Using the notation of Problem 2.1, consider the family of system $\mathcal{F}_\gamma(\Sigma_0)$ with the initial state x_0, and assume that γ satisfies the conditions of Proposition 2.2. Then, there is an input function u that takes every member Σ of $\mathcal{F}_\gamma(\Sigma_0)$ from the initial state x_0 to the δ-vicinity of the origin. We show in this section that this implies that Problem 2.1 has an optimal solution $u^*(x_0)$. The proof of the existence of $u^*(x_0)$ relies on two critical facts discussed in this section.

(1) The set $U(K)$ of input functions is compact in an appropriate sense.
(2) The time $t_f(x_0, u)$ is, in an appropriate sense, a continuous function of the input signal u.

Then, the well-known fact that a continuous function attains a minimum over a compact domain implies the existence of the minimal time $t_f^*(x_0)$ of Problem 2.1 as well as the existence of an optimal input function $u^*(x_0)$ that achieves this minimal time. We start with a review of some basic notions.

Definition 3.1: Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$.

(1) A sequence $\{v_n\}_{n=1}^{\infty}$ of members of H *converges weakly* to a member $v \in H$ if $\lim_{n \to \infty} \langle v_n, y \rangle = \langle v, y \rangle$ for every $y \in H$.
(2) A subset W of H is *weakly compact* if every sequence of members of W has a subsequence that converges weakly to a member of W.

The following statement reproduces Lemma 3.2 of Chakraborty and Hammer (2009).

Lemma 3.1: *The set $U(K)$ of (2.8) is weakly compact in the topology of the Hilbert space $L_2^{\alpha,m}$.*

Our discussion depends on the following classical notion.

Definition 3.2: Let S be a subset of a Hilbert space H, and let z be a point of S. A functional $F: S \to R$ is *weakly lower semi-continuous at z* if the following is true for every sequence $\{z_n\}_{n=1}^{\infty} \subseteq S$ that converges weakly to z: whenever $F(z)$ is bounded, there is, for every real number $\varepsilon > 0$, an integer $N > 0$ such that $F(z) - F(z_n) < \varepsilon$ for all $n \geq N$. If F is weakly lower semi-continuous at every point $z \in S$, then F is *weakly lower semi-continuous* over S.

The function F is *weakly continuous* at z if there is, for every real number $\varepsilon > 0$, an integer $N > 0$ such that $|F(z) - F(z_n)| < \varepsilon$ for all $n \geq N$.

We turn now to the first step in proving the existence of an optimal solution of Problem 2.1.

Proposition 3.1: *Under the notation and conditions of Problem 2.1, the function $t(x_0, u)$ is weakly lower semi-continuous as a function of u over $U(K)$.*

In order to prove Proposition 3.1, we need a few auxiliary results.

Lemma 3.2: *Under the notation and conditions of Problem 2.1, let $\{u_i\}_{i=1}^{\infty} \subseteq U(K)$ be a sequence of functions that converges weakly to the function $u \in U(K)$, and let Σ be a member of $\mathcal{F}_\gamma(\Sigma_0)$. Then, the sequence $\{\Sigma(x_0, u_i, t)\}_{i=1}^{\infty}$ converges to $\Sigma(x_0, u, t)$ at every time $t \geq 0$.*

Proof: Let $\{u_i\}_{i=1}^{\infty} \subseteq U(K)$ be a sequence of functions that converges weakly to the function $u \in U(K)$ and denote $x(t, u_i) := \Sigma(x_0, u_i, t)$, $x(t, u) := \Sigma(x_0, u, t)$, and

$$x(t, i) := x(t, u) - x(t, u_i). \tag{3.1}$$

Then, the proof will conclude upon showing that $x(t, i)$ converges to zero. To this end, combining (2.1), (2.2), and the fact that Σ always starts from the same initial state x_0, it follows that $x(0, i) = 0$ for all integers $i \geq 1$ and that

$$\begin{aligned} x(t, i) &= \int_0^t [a(\theta, x(\theta, u)) - a(\theta, x(\theta, u_i))] d\theta \\ &\quad + \int_0^t [b(\theta, x(\theta, u))u(\theta) - b(\theta, x(\theta, u_i))u_i(\theta)] d\theta \\ &= \int_0^t [a(\theta, x(\theta, u)) - a(\theta, x(\theta, u_i))] d\theta \\ &\quad + \int_0^t [b(\theta, x(\theta, u)) - b(\theta, x(\theta, u_i))] u_i(\theta) \\ &\quad + \int_0^t b(\theta, x(\theta, u))[u(\theta) - u_i(\theta)] d\theta. \end{aligned}$$

Using (2.3), (2.4), and the facts that $|u|_\infty \leq K$ and $|u_i|_\infty \leq K$ for all integers $i \geq 1$, we can write

$$\begin{aligned} &\sup_{0 \leq \tau \leq t} |x(\tau, i)| \\ &\leq \int_0^t \sup_{0 \leq \theta \leq t} |a_0(\theta, x(\theta, u)) - a_0(\theta, x(\theta, u_i))| \, d\theta \\ &\quad + \int_0^t \sup_{0 \leq \theta \leq t} \left| a_\gamma(\theta, x(\theta, u)) - a_\gamma(\theta, x(\theta, u_i)) \right| d\theta \\ &\quad + \int_0^t \sup_{0 \leq \theta \leq t} \{ |b_0(\theta, x(\theta, u)) - b_0(\theta, x(\theta, u_i))| \, |u_i(\theta)| \} \, d\theta \\ &\quad + \int_0^t \sup_{0 \leq \theta \leq t} \left\{ \left| b_\gamma(\theta, x(\theta, u)) - b_\gamma(\theta, x(\theta, u_i)) \right| |u(\theta)| \right\} d\theta \\ &\quad + \sup_{0 \leq \tau \leq t} \left| \int_0^\tau b(\theta, x(\theta, u)) \, [u(\theta) - u_i(\theta)] \, d\theta \right| \\ &\leq (M + \gamma) \sup_{0 \leq \tau \leq t} |x(\tau, i)| t + (M + \gamma) \sup_{0 \leq \tau \leq t} |x(\tau, i)| Kt \\ &\quad + \sup_{0 \leq \tau \leq t} \left| \int_0^\tau b(\theta, x(\theta, u))[u(\theta) - u_i(\theta)] d\theta \right|. \end{aligned}$$

Moving terms from the right-hand side to the left-hand side, we can write

$$\begin{aligned} &\{1 - t(M + \gamma)(1 + K)\} \sup_{0 \leq \tau \leq t} |x(\tau, i)| \\ &\quad \leq \sup_{0 \leq \tau \leq t} \left| \int_0^\tau b(\theta, x(\theta, u))[u(\theta) - u_i(\theta)] d\theta \right|. \end{aligned}$$

Now, let $\zeta > 0$ be a value of t for which $\{1 - \zeta[(M + \gamma)(1 + K)\} > 0$, and set $\mu := \{1 - \zeta[(M + \gamma)(1 + K)\}$. Then,

we get

$$\sup_{0\le\tau\le\zeta}|x(\tau,i)| \le \frac{1}{\mu}\left\{\sup_{0\le\tau\le\zeta}\left|\int_0^\tau b(\theta,x(\theta,u))[u(\theta) - u_i(\theta)]d\theta\right|\right\}. \tag{3.2}$$

Furthermore, referring to the inner product (2.7), define the function

$$y_\tau(\theta) := \begin{cases} e^{\alpha\theta}b(\theta,x(\theta,u)) & 0\le\theta\le\tau,\\ 0 & \text{else.}\end{cases}$$

Then, we can write

$$\int_0^\tau b(\theta,x(\theta,u))[u(\theta)-u_i(\theta)]d\theta = \langle u-u_i, y_\tau\rangle.$$

Considering that the sequence $\{u_i\}_{i=1}^\infty$ converges weakly to u, it follows that, for every real number $\beta>0$, there is an integer $N_\tau \ge 0$ such that $|\langle u-u_i, y_\tau\rangle| < \beta$ for all $i \ge N_\tau$. We show next that there is an integer $N\ge 0$ such that $\sup_{0\le\tau\le\zeta}|\langle u-u_i, y_\tau\rangle| < \beta$ for all $i\ge N$.

Indeed, by contradiction, assume that there is no such integer N. Then, there is a sequence of times $\{\tau_j\}_{j=0}^\infty$, where $0\le\tau_j\le\zeta$ and

$$|\langle u-u_j, y_{\tau_j}\rangle| \ge \beta \tag{3.3}$$

for all $j=0,1,2,\ldots$ Now, as the interval $[0,\zeta]$ is compact, the sequence $\{\tau_j\}_{j=0}^\infty$ must contain a convergent subsequence, say the subsequence $\{\tau_{j_k}\}_{k=0}^\infty$; denote its limit by $\lim_{k\to\infty}\tau_{j_k}=\tau'$. By weak convergence of the sequence $\{u_i\}_{i=1}^\infty$ to u, the subsequence $\{u_{j_k}\}_{k=1}^\infty$ also converges weakly to u. Hence, there is an integer $N'\ge 0$ such that $|\langle u-u_{j_k}, y_{\tau'}\rangle| < \beta/2$ for all $k\ge N'$. Furthermore, according to Proposition 2.1, there is a bound $A>0$ such that $|x(\theta,u)|\le A$ for all $\theta\in[0,\zeta]$. Using this bound together with (2.3) and (2.4), and recalling that $0\le\tau_{j_k},\tau'\le\zeta$ for all integers $k\ge 0$, we obtain

$$\begin{aligned}
&|\langle u-u_{j_k}, y_{\tau'}\rangle - \langle u-u_{j_k}, y_{\tau_{j_k}}\rangle|\\
&\quad= \left|\int_{\tau_{j_k}}^{\tau'} b(\theta,x(\theta,u))[u(\theta)-u_{j_k}(\theta)]d\theta\right|\\
&\quad= \left|\int_{\tau_{j_k}}^{\tau'} [b(\theta,x(\theta,u))-b(\theta,0)]\,[u(\theta)-u_{j_k}(\theta)]d\theta\right|\\
&\qquad+\left|\int_{\tau_{j_k}}^{\tau'} b(\theta,0)[u(\theta)-u_{j_k}(\theta)]d\theta\right|\\
&\quad\le 2(M+\gamma)(A+1)K\,|\tau'-\tau_{j_k}|.
\end{aligned}$$

Now, considering that $\lim_{k\to\infty}\tau_{j_k}=\tau'$, there is an integer $N^*\ge N'$ such that

$$|\tau'-\tau_{j_k}| < \frac{\beta}{4(M+\gamma)(A+1)K}$$

for all $k\ge N^*$. Then,

$$\begin{aligned}
&|\langle u-u_{j_k}, y_{\tau_{j_k}}\rangle|\\
&\quad= |\langle u-u_{j_k}, y_{\tau_{j_k}}\rangle - \langle u-u_{j_k}, y_{\tau'}\rangle + \langle u-u_{j_k}, y_{\tau'}\rangle|\\
&\quad\le |\langle u-u_{j_k}, y_{\tau_{j_k}}\rangle - \langle u-u_{j_k}, y_{\tau'}\rangle| + |\langle u-u_{j_k}, y_{\tau'}\rangle|\\
&\quad< \beta/2+\beta/2=\beta
\end{aligned}$$

for all $k\ge N^*$, in contradiction to (3.3). Hence, for every real number $\beta>0$, there is an integer $N\ge 0$ such that

$$\sup_{0\le\tau\le\zeta}|\langle u-u_i, y_\tau\rangle| < \beta \tag{3.4}$$

for all $i\ge N$.

Next, given any real number $\xi>0$, select a real number β satisfying $0<\beta<\mu\xi$. Then, combining (3.4) with (3.2), it follows that there is an integer $N_\xi\ge 0$ such that

$$\sup_{0\le\tau\le\zeta}|x(\tau,i)| < \xi \tag{3.5}$$

for all $i\ge N_\xi$. This proves that $\lim_{i\to\infty}x(\tau,i)=0$ for all $\tau\in[0,\zeta]$, or, in view of (3.1), that $\lim_{i\to\infty}x(t,u_i)=x(t,u)$ for all $t\in[0,\zeta]$.

Finally, using an argument similar to the one employed in the last part of the proof of Proposition 2.2, this implies that $\lim_{i\to\infty}x(t,u_i)=x(t,u)$ at any finite time $t\ge 0$, and our proof concludes. □

Invoking Definition 3.2, we obtain the following restatement of Lemma 3.2.

Corollary 3.1: *Under the notation and conditions of Problem 2.1, let Σ be a member of $\mathcal{F}_\gamma(\Sigma_0)$. Then, the function $\Sigma(x_0,\nu,t)$ is weakly continuous over $U(K)$ at every time $t\ge 0$.*

In fact, the following somewhat stronger result is a consequence of the proof of Lemma 3.2.

Corollary 3.2: *Under the notation and conditions of Problem 2.1, let Σ be a member of $\mathcal{F}_\gamma(\Sigma_0)$, and let $\tau>0$ be a real number. Then, the function $\Sigma(x_0,\nu,t)$ is uniformly weakly continuous over $U(K)$, in the following sense:*

For every sequence $\{u_i\}_{i=1}^\infty\subseteq U(K)$ that converges weakly to a function $u\in U(K)$, there is, for every real number $\varepsilon>0$, an integer $N(\varepsilon)>0$ such that $\sup_{t\in[0,\tau]}|\Sigma(x_0,u,t)-\Sigma(x_0,u_i,t)|<\varepsilon$ for all integers $i\ge N(\varepsilon)$.

Proof: The corollary follows from (3.5) combined with an argument similar to the one employed in the last part of the proof of Proposition 2.2. □

To continue, we need several facts that are reproduced here from Willard (1970) in a form adapted to the current discussion.

Theorem 3.1:

(1) *A weakly continuous function is weakly lower semi-continuous.*

(2) *Let S and A be topological spaces and assume that, for every member $a \in A$, there is a weakly lower semi-continuous function f_a: $S \to R$. If $\sup_{a\in A} f_a(s)$ exists at each point $s \in S$, then the function $f(s) := \sup_{a\in A} f_a(s)$ is weakly lower semi-continuous on S.*

These general facts allow us to prove the following statement.

Lemma 3.3: *Under the notation and conditions of Problem 2.1, let v be a member of U(K) and define $\psi(t, v) := \sup_{\Sigma\in\mathcal{F}_\gamma(\Sigma_0)} |\Sigma(x_0, v, t)|_2^2$. Then, at every time $t \geq 0$, the function ψ: $U(K) \to R^+$: $v \mapsto \psi(t, v)$ is a weakly lower semi-continuous function over U(K).*

Proof: Indeed, by Corollary 3.1, the function $\Sigma(x_0, v, t)$ is weakly continuous over $U(K)$ at every time $t \geq 0$. Furthermore, as every continuous function of a weakly continuous function is also weakly continuous, it follows that the function $|\Sigma(x_0, v, t)|_2^2 = \Sigma^T(x_0, v, t)\Sigma(x_0, v, t)$ is weakly continuous over $U(K)$ at every time $t \geq 0$. But then, by Theorem 3.1(1), the function $|\Sigma(x_0, v, T)|_2^2$ is also weakly lower semi-continuous on $U(K)$ at every time $t \geq 0$. Finally, recalling that $\psi(t, v) := \sup_{\Sigma\in\mathcal{F}_\gamma(\Sigma_0)} |\Sigma(x_0, v, t)|_2^2$ and invoking Theorem 3.1(2), we conclude that $\psi(t, v)$ is weakly lower semi-continuous on $U(K)$ at every time $t \geq 0$. □

We are ready now to state the proof of Proposition 3.1.

Proof: (of Proposition 3.1): Using the notation and the conditions of Problem 2.1, let $\Sigma \in \mathcal{F}_\gamma(\Sigma_0)$ be a member, and let $u \in U(K)$ be an input signal for Σ. In the notation of Lemma 3.3, we can write

$$t_f(x_0, u) := \inf_t \{t \geq 0 : \psi(t, u(t)) \leq \delta\}.$$

To temporarily simplify our notation, set

$$\theta(u) := \inf_t \{t \geq 0 : \psi(t, u(t)) \leq \delta\}. \tag{2.6}$$

Then, according to Proposition 2.2, there are input signals $u \in U(K)$ for which $\theta(u) < \infty$. Also, by definition, $\theta(u) \geq 0$ for all $u \in U(K)$.

Now, let $u \in U(K)$ be an input signal for which $\theta(u) < \infty$, and consider a sequence of input signals $\{u_i\}_{i=1}^\infty \subseteq U(K)$ that converges weakly to u. Denote $\psi_i(t) := \psi(t, u_i(t))$, $i = 1, 2, \dots$, and $\psi_0(t) := \psi(t, u(t))$. Then, $\theta(u_i) := \inf\{t \geq 0 : \psi_i(t) \leq \delta\}$ and $\theta(u) := \inf\{t \geq 0 : \psi_0(t) \leq \delta\}$. We claim that $\theta(u)$ is a weakly lower semi-continuous function of u over $U(K)$. To prove this claim, it is sufficient to show that the following is true: for every real number $\varepsilon > 0$, there is an integer $N > 0$ such that

$$\theta(u_i) > \theta(u) - \varepsilon \text{ for all } i \geq N. \tag{2.7}$$

To show that the latter is true, select a real number $\varepsilon > 0$. We distinguish between two cases:

Case 3.1. There is an integer $N > 0$ such that $\theta(u_i) \geq \theta(u)$ for all $i \geq N$; then (2.7) is clearly valid for all $i \geq N$.

Case 3.2. Case 3.1 is not valid; then, there is a sequence of integers $j_1, j_2, \dots$ such that $\theta(u_{j_k}) < \theta(u)$ for all integers $k \geq 1$.

We proceed with Case 3.2. Considering that $\theta(u) < \infty$, the inequality $\theta(u_{j_k}) < \theta(u)$ for all integers $k \geq 1$ implies that $\theta(u_{j_k}) < \infty$ for all $k \geq 1$. In addition, combining the fact that $\theta(u) < \infty$ with (2.6) implies that there is a time $\bar{t} \in [\theta(u) - \varepsilon, \theta(u))$ at which $\psi_0(\bar{t}) > \delta$, or

$$\psi_0(\bar{t}) - \delta > 0. \tag{2.8}$$

Furthermore, as $\psi(t, u)$ is weakly lower semi-continuous in u by Lemma 3.3, it follows that, for every real number $\mu > 0$, there is an integer $N > 0$ such that $\psi_0(\bar{t}) - \psi_{j_k}(\bar{t}) < \mu$ for all $k \geq N$. In view of (2.8), we can take $\mu := (\psi_0(\bar{t}) - \delta)/2$; then, we obtain that there is an integer $N > 0$ such that

$$\psi_0(\bar{t}) - \psi_{j_k}(\bar{t}) < (\psi_0(\bar{t}) - \delta)/2 \text{ for all } k \geq N.$$

Rearranging terms, we get

$$\psi_{j_k}(\bar{t}) > (\psi_0(\bar{t}) + \delta)/2 \text{ for all } k \geq N,$$

which, in view of (2.8), implies that $\psi_{j_k}(\bar{t}) > \delta$ for all $k \geq N$, so that $\theta(u_{j_k}) > \bar{t}$. But then, since $\bar{t} \in [\theta(u) - \varepsilon, \theta(u))$, we conclude that $\theta(u_{j_k}) > \theta(u) - \varepsilon$ for all $k \geq N$. Combining this with Case 3.1, it follows that $\theta(u)$ is a weakly lower semi-continuous functional of u. Finally, as $t_f(x_0, u) = \theta(u)$, our proof concludes. □

Summarising our discussion so far, we have seen in Lemma 3.1 that the set $U(K)$ is weakly compact, and we have seen in Proposition 3.1 that $t_f(x_0, u)$ is weakly lower semi-continuous over $U(K)$. This brings us to the point where we can apply the generalised Weierstrass

Theorem (e.g. Zeidler, 1985), which, in the present terminology, states that a weakly lower semi-continuous function attains a minimum in a weakly compact set. Consequently, $t_f(x_0, u)$, as a function of the input signal u, attains a minimum over the set of input signals $U(K)$. This proves that our optimisation problem, Problem 2.1, has a solution: the minimal time $t_f^*(x_0)$ exists, and there is an optimal input signal $u^*(x_0)$ that achieves this minimal time. This proves the following statement, which is the main result of this section.

Theorem 3.2: *Under the notation and conditions of Problem 2.1, the following are valid:*

(1) *there is a finite minimal time $t_f^*(x_0)$, and*
(2) *there is an optimal input function $u^*(x_0) \in U(K)$ satisfying $t_f^*(x_0) = t_f(x_0, u^*(x_0))$.*

Theorem 3.2 shows that our optimisation problem has a solution under rather general conditions. Yet, an accurate calculation of the optimal input signal $u^*(x_0)$ may, in general, be rather difficult, and, being a general vector-valued function, $u^*(x_0)$ may also be difficult to implement. In the next section, we show that the response induced by an optimal input signal $u^*(x_0)$ can be approximated as closely as desired by a bang–bang input signal. Bang–bang input signals are relatively easy to calculate and implement, since they are determined by a finite string of real numbers – the switching times.

4. Bang–bang approximation of optimal performance

In this section, we show that the performance induced by an optimal input signal $u^*(x_0)$ of Theorem 3.2 can be approximated as closely as desired by a bang–bang input signal $u^{\pm}(x_0)$. The use of bang–bang input signals substantially simplifies the process of calculating and implementing controllers that achieve close to optimal performance. Throughout this section, we refer to the notation and conditions of Problem 2.1.

In line with standard terminology, a *bang–bang signal* $u^{\pm}(x_0, t)$ is a piecewise constant member of the family of functions $U(K)$ that takes only extremal values. In other words, at any time $t \geq 0$, every component of $u^{\pm}(x_0, t)$ is either $-K$ or $+K$. In this way, a bang–bang signal is characterised by its *switching times* – the times at which its components switch from $-K$ to $+K$ or vice versa. As a result, the calculation of a bang–bang signal basically involves only the calculation of a list of real numbers that represent the switching times. (Of course, one also has to determine at each switching time which way each component switches: from $-K$ to $+K$ or from $+K$ to $-K$.) We will also show that the approximating bang–bang signals have only a finite number of switchings. Needless to say, calculating and implementing a bang–bang signal with a finite list of switching times is significantly simpler than calculating and implementing a more general vector-valued function.

We start our discussion with the following auxiliary result, which includes some of the basic facts needed to prove the existence of a bang–bang input signal that approximates optimal performance.

Lemma 4.1: *Assume the notation and conditions of Problem 2.1. For a member $\Sigma \in \mathcal{F}_\gamma(\Sigma_0)$ with initial state x_0, denote by $x(t) := \Sigma(x_0, u, t)$, $t \geq 0$, the response of Σ to an input signal $u \in U(K)$. Let $b_0(t, x(t))$ and $b_\gamma(t, x(t))$ be as given by (2.1) and (2.2). Then, the following are true.*

(1) *There are real numbers $B_0(x_0) \geq 0$ and $B_\gamma(x_0) \geq 0$ such that $\sup_{0\leq\tau\leq t_f^*(x_0)} |b_0(\tau, x(\tau))| \leq B_0(x_0)$ and $\sup_{0\leq\tau\leq t_f^*(x_0)} |b_\gamma(\tau, x(\tau))| \leq B_\gamma(x_0)$ for all input signals $u \in U(K)$.*
(2) *For every real number $\rho > 0$, there is a real number $\beta(x_0, \rho) > 0$ such that*

$$|b_0(t', x(t')) - b_0(t'', x(t''))| < \rho \text{ and } |b_\gamma(t', x(t')) - b_\gamma(t'', x(t''))| < \rho \tag{4.1}$$

at all times t', $t'' \in [0, t_f^*(x_0)]$ satisfying $|t' - t''| < \beta(x_0, \rho)$, irrespective of what input signal $u \in U(K)$ is used.

Proof: We use the notation and conditions of Problem 2.1, together with the fact that $t_f^*(x_0) < \infty$ by Theorem 3.2. Consider a member $\Sigma \in \mathcal{F}_\gamma(\Sigma_0)$ that starts from the initial state x_0 at the time $t = 0$, and denote by $x(t) := \Sigma(x_0, u, t)$, $t \geq 0$, the response of Σ to an input signal $u \in U(K)$. As $x(t)$ is the solution of the differential equation (2.1) with a bounded input signal $|u|_\infty \leq K$, it is a continuous function of time; consequently, referring to (2.1) and (2.2), the functions $b_0(t, x(t))$ and $b_\gamma(t, x(t))$ are also continuous functions of time.

Furthermore, by Proposition 2.1 (see, in particular, (2.6)), there is a bound $N(x_0) \geq 0$ such that $|x(\tau)| \leq N(x_0)$ at all times $0 \leq \tau \leq t_f^*(x_0)$ for all input signals $u \in U(K)$. By the continuity of the functions $b_0(t, x(t))$ and $b_\gamma(t, x(t))$, this implies that part (1) of the lemma is valid.

Next, to prove part (2) of the lemma, note that the facts mentioned so far imply that the functions $b_0(t, x(t))$ and $b_\gamma(t, x(t))$ are both uniformly continuous over the time

interval $[0, t_f^*(x_0)]$. This means that, for every real number $\rho > 0$, there is a real number $\beta(x_0, \rho, u) > 0$, potentially dependent on the input signal u, such that

$$\begin{aligned}&\left|b_0(t', x(t')) - b_0(t'', x(t''))\right| \\ &\quad < \rho \text{ and } \left|b_\gamma(t', x(t')) - b_\gamma(t'', x(t''))\right| < \rho\end{aligned}$$

at all times t', $t'' \in [0, t_f^*(x_0)]$ satisfying $|t' - t''| < \beta(x_0, \rho, u)$. We claim that $\beta(x_0, \rho, u)$ can be selected to be independent of the input signal u.

To prove this claim, let $u \in U(K)$ be an input signal and denote

$$\begin{aligned}\beta_0(x_0, \rho, u) := \sup\{&t' - t'' : t', t'' \in [0, t_f^*(x_0)] \text{ and } |b_0 \\ &(t', \Sigma(x_0, u, t')) - b_0(t'', \Sigma(x_0, u, t''))| < \rho\},\end{aligned}$$

and set $\beta_0^*(x_0, \rho) := \inf_{u \in U(K)} \beta_0(x_0, \rho, u)$. Similarly, denote

$$\begin{aligned}\beta_\gamma(x_0, \rho, u) := \sup\{&t' - t'' : t', t'' \in \left[0, t_f^*(x_0)\right] \\ &\text{and } |b_\gamma(t', \Sigma(x_0, u, t')) - b_\gamma(t'', \Sigma(x_0, u, t''))| < \rho\},\end{aligned}$$

and set $\beta_\gamma^*(x_0, \rho) := \inf_{u \in U(K)} \beta_\gamma(x_0, \rho, u)$. Now, if $\beta_0^*(x_0, \rho) > 0$ and $\beta_\gamma^*(x_0, \rho) > 0$, then part (2) of the lemma is valid for $\beta(x_0, \rho) := \min\left\{\beta_0^*(x_0, \rho), \beta_\gamma^*(x_0, \rho)\right\}$.

Otherwise, assume first that $\beta_0^*(x_0, \rho) = 0$. Then, there is a sequence of input signals $\{u_i\}_{i=1}^\infty \subseteq U(K)$ such that $\lim_{i \to \infty} \beta_0(x_0, \rho, u_i) = 0$. Considering that $U(K)$ is weakly compact by Lemma 3.1, the sequence $\{u_i\}_{i=1}^\infty$ has a weakly convergent subsequence $\left\{u_{i_k}\right\}_{k=1}^\infty$ that weakly converges to a member $u \in U(K)$. Then, by Lemma 3.2, we have that $\lim_{k \to \infty} \Sigma(x_0, u_{i_k}, t) = \Sigma(x_0, u, t)$ at every time $t \geq 0$. Also, by Corollary 3.2, the function $\Sigma(x_0, \nu, t)$ is a weakly uniformly continuous function of $\nu \in U(K)$ over the interval $[0, t_f^*(x_0)]$. In other words, for every real number $\varepsilon > 0$, there is an integer $N(\varepsilon) > 0$ such that $\sup_{t \in \left[0, t_f^*(x_0)\right]} \left|\Sigma(x_0, u_{i_k}, t) - \Sigma(x_0, u, t)\right| < \varepsilon$ for all $k \geq N(\varepsilon)$. Consequently, as $b_0(t, x)$ is uniformly continuous in x over the domains of current interest, there is an integer $N' > 0$ such that $\left|b_0(t, \Sigma(x_0, u_{i_k}, t)) - b_0(t, \Sigma(x_0, u, t))\right| < \rho/3$ for all $k \geq N'$.

Furthermore, as $\Sigma(x_0, u, t)$ is a continuous function of time, it is uniformly continuous over the compact time interval $[0, t_f^*(x_0)]$. Consequently, there is a real number $\beta > 0$ such that $|b_0(t', \Sigma(x_0, u, t')) - b_0(t'', \Sigma(x_0, u, t''))| < \rho/3$ for all $t', t'' \in [0, t_f^*(x_0)]$ satisfying $|t' - t''| < \beta$. Applying these facts we obtain that, for every integer $k \geq N'$ and for all $|t' - t''| < \beta$, the following is valid:

$$\begin{aligned}&\left|b_0(t', \Sigma(x_0, u_{i_k}, t')) - b_0(t'', \Sigma(x_0, u_{i_k}, t''))\right| \\ &\leq \left|b_0(t', \Sigma(x_0, u_{i_k}, t')) - b_0(t'', \Sigma(x_0, u, t'))\right| \\ &\quad + \left|b_0(t', \Sigma(x_0, u, t')) - b_0(t'', \Sigma(x_0, u, t''))\right| \\ &+ \left|b_0(t'', \Sigma(x_0, u, t'')) - b_0(t'', \Sigma(x_0, u_{i_k}, t''))\right| \\ &\leq \rho/3 + \rho/3 + \rho/3 = \rho,\end{aligned}$$

contradicting the assumption that $\beta_0^*(x_0, \rho) = 0$. As a result, we must have $\beta_0^*(x_0, \rho) > 0$. A similar argument shows that $\beta_\gamma^*(x_0, \rho) > 0$ as well. Thus, part (2) of the lemma is valid for $\beta(x_0, \rho) := \min\left\{\beta_0^*(x_0, \rho), \beta_\gamma^*(x_0, \rho)\right\}$, and our proof concludes. □

The next auxiliary result includes the main facts needed to prove the existence of a bang–bang input signal that approximates optimal performance.

Lemma 4.2: *Assume the notation and conditions of Problem 2.1. Then, for every time $\theta \in [0, t_f^*(x_0))$ and for every real number $\sigma_0 > 0$, there are two real numbers $\eta \in (0, t_f^*(x_0) - \theta]$ and $\mu(\eta) > 0$, and a bang–bang input signal $u^\pm(x_0) \in U(K)$ such that the following are true for every member $\Sigma \in \mathcal{F}_\gamma(\Sigma_0)$:*

(1) *$u^\pm(x_0)$ has a finite number of switchings in the time interval $[\theta, \theta + \eta]$.*

(2) *The difference between the optimal response $x^*(t) := \Sigma(x_0, u^*(x_0), t)$ and the response $x^\pm(t) := \Sigma(x_0, u^\pm(x_0), t)$ to the bang–bang input signal $u^\pm(x_0)$ satisfies*

$$\begin{aligned}\sup_{t \in [\theta, \theta + \eta]} &\left|x^*(t) - x^\pm(t)\right| < \mu(\eta) \left|x^*(\theta) - x^\pm(\theta)\right| \\ &+ \sigma_0.\end{aligned}$$

(3) *The values of η and $\mu(\eta)$ can be selected independently of σ_0, to depend only on the bounds M and γ of (2.3) and (2.4) and on the input signal bound K.*

Proof: Given a time $\theta \in [0, t_f^*(x_0))$ and a real number $\rho > 0$, and using the number $\beta(x_0, \rho) > 0$ of Lemma 4.1, let $\eta \in (0, t_f^*(x_0) - \theta]$ and $\lambda > 0$ be real numbers to be specified later, chosen to satisfy the relationships

$$0 < \lambda < \beta(x_0, \rho) \text{ and } \eta/\lambda \text{ is an integer.} \tag{2.2}$$

Then, define the integer

$$r := \eta/\lambda - 1,$$

and partition the interval $[\theta, \theta+\eta]$ into $r+1$ subintervals of length λ to obtain the partition

$$[\theta, \theta+\eta] = \{[\theta, \theta+\lambda], [\theta+\lambda, \theta+2\lambda], \dots, [\theta+r\lambda, \theta+(r+1)\lambda]\}. \quad (2.3)$$

Using the intervals of this partition, we define a bang–bang signal $u^{\pm}(x_0) = (u_1^{\pm}(x_0,t), u_2^{\pm}(x_0,t), \dots, u_m^{\pm}(x_0,t))^T \in U(K)$ as follows. For each component $u_i^{\pm}(x_0,t)$, we select below a point $\omega_{\ell_i} \in [\theta+\ell\lambda, \theta+(\ell+1)\lambda]$ in each subinterval of the partition (2.3), and define the components of $u^{\pm}(x_0,t)$ over each subinterval $[\theta+\ell\lambda, \theta+(\ell+1)\lambda)$, $\ell = 0,\ 1,\ 2, \dots, r$, by setting

$$u_i^{\pm}(x_0,t) := \begin{cases} K & \text{for } t \in [\theta+\ell\lambda, \omega_{\ell_i}), \text{ and} \\ -K & \text{for } t \in [\omega_{\ell_i}, \theta+(\ell+1)\lambda) \\ & \text{if } \omega_{\ell_i} \neq (\ell+1)\lambda, \end{cases} \quad (2.4)$$

$i = 1,\ 2, \dots,\ m$. To select the point ω_{ℓ_i}, recall that the optimal input signal $u^*(x_0,t) = (u_1^*(x_0,t), u_2^*(x_0,t), \dots, u_m^*(x_0,t))^T$ is a member of $U(K)$, and, as a result, satisfies $|u_i^*(x_0,t)| \leq K$ for all $t \in [0, t_f^*(x_0)]$ and all $i = 1, 2, \dots, m$. This implies that, for each pair of integers $i \in \{1, 2, \dots, m\}$ and $\ell \in \{0, 1, 2, \dots, r\}$, there is a point $\omega_{\ell_i} \in [\theta+\ell\lambda, \theta+(\ell+1)\lambda]$ that satisfies the equality

$$K[2(\omega_{\ell_i} - (\theta+\ell\lambda)) - \lambda] = \int_{\theta+\ell\lambda}^{\theta+(\ell+1)\lambda} u_i^*(s)ds. \quad (2.5)$$

We use this ω_{ℓ_i} in (2.4). Then, the signal $u^{\pm}(x_0)$ of (2.4) is clearly a bang–bang member of $U(K)$, and, in view of (2.5), it has the property

$$\int_{\theta+\ell\lambda}^{\theta+(\ell+1)\lambda} [u_i^*(x_0,s) - u_i^{\pm}(x_0,s)]ds = 0 \quad (2.6)$$

for all $i \in \{1, 2, \dots, m\}$ and all $\ell \in \{0, 1, 2, \dots, r\}$.

Now, consider a member $\Sigma \in \mathcal{F}_\gamma(\Sigma_0)$ that starts from the initial state x_0 at the time $t = 0$, and compare the responses of Σ obtained from the two input signals $u^*(x_0)$ and $u^{\pm}(x_0)$. According to our notation, $x^*(t) = \Sigma(x_0, u^*(x_0), t)$ is the response of Σ to $u^*(x_0)$, and $x^{\pm}(t) = \Sigma(x_0, u^{\pm}(x_0), t)$ is the response of Σ to $u^{\pm}(x_0)$. At a time $\theta \in [0, t_f^*(x_0))$, the respective states of Σ are $x^*(\theta)$ and $x^{\pm}(\theta)$. Using (2.1) and (2.2) and setting

$$\xi(t) := x^*(t) - x^{\pm}(t), \quad (2.7)$$

we can write the following for a time $t \in [\theta, t_f^*]$.

$$\begin{aligned}\xi(t) =& \xi(\theta) + \int_\theta^t \Big[\big(a_0(s,x^*(s)) + a_\gamma(s,x^*(s))\big) \\ &- \big(a_0(s,x^{\pm}(s)) + a_\gamma(s,x^{\pm}(s))\big) \\ &+ \big(b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big)u^*(x_0,s) \\ &- \big(b_0(s,x^{\pm}(s)) + b_\gamma(s,x^{\pm}(s))\big)u^{\pm}(x_0,s)\Big]ds,\end{aligned}$$

so that

$$\begin{aligned}\sup_{t\in[\theta,\theta+\eta]} |\xi(t)| \leq\ & |\xi(\theta)| \\ &+ \sup_{t\in[\theta,\theta+\eta]} \Big| \int_\theta^t \Big[\big(a_0(s,x^*(s)) + a_\gamma(s,x^*(s))\big) \\ &- \big(a_0(s,x^{\pm}(s)) + a_\gamma(s,x^{\pm}(s))\big) \\ &+ \big(b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big)u^*(x_0,s) \\ &- \big(b_0(s,x^{\pm}(s)) + b_\gamma(s,x^{\pm}(s))\big)u^{\pm}(x_0,s)\Big]ds\Big|.\end{aligned}$$

Then, we get

$$\begin{aligned}&\sup_{t\in[\theta,\theta+\eta]} |\xi(t)| \leq |\xi(\theta)| \\ &+ \sup_{t\in[\theta,\theta+\eta]} \Big\{ \Big| \int_\theta^t \Big[\big(a_0(s,x^*(s)) + a_\gamma(s,x^*(s))\big) \\ &- \big(a_0(s,x^{\pm}(s)) + a_\gamma(s,x^{\pm}(s))\big)\Big]ds\Big| \\ &+ \Big| \int_\theta^t \big[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big]u^*(x_0,s) \\ &- \big[b_0(s,x^{\pm}(s)) + b_\gamma(s,x^{\pm}(s))\big]u^{\pm}(x_0,s)ds\Big|\Big\} \\ &\leq |\xi(\theta)| + \sup_{t\in[\theta,\theta+\eta]} \Big\{ \int_\theta^t \Big| \big(a_0(s,x^*(s)) + a_\gamma(s,x^*(s))\big) \\ &- \big(a_0(s,x^{\pm}(s)) + a_\gamma(s,x^{\pm}(s))\big)\Big|ds \\ &+ \Big| \int_\theta^t \Big[\big(b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big)u^*(x_0,s) \\ &- \big(b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big)u^{\pm}(x_0,s)\Big]ds\Big| \\ &+ \Big| \int_\theta^t \Big[\big(b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big)u^{\pm}(x_0,s) \\ &- \big(b_0(s,x^{\pm}(s)) + b_\gamma(s,x^{\pm}(s))\big)u^{\pm}(x_0,s)\Big]ds\Big|\Big\} \\ &\leq |\xi(\theta)| + \sup_{t\in[\theta,\theta+\eta]} \int_\theta^t (M+\gamma)\,|x^*(s) - x^{\pm}(s)|\,ds\end{aligned}$$

$$
\begin{aligned}
&+ \sup_{t\in[\theta,\theta+\eta]} \left| \int_\theta^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right] \right.\\
&\left. \times \left[u^*(x_0,s) - u^\pm(x_0,s)\right] ds \right|\\
&+ \sup_{t\in[\theta,\theta+\eta]} \int_\theta^t \left|b_0(s,x^*(s)) - b_0(s,x^\pm(s))\right| \left|u^\pm(x_0,s)\right| ds\\
&+ \sup_{t\in[\theta,\theta+\eta]} \int_\theta^t \left|b_\gamma(s,x^*(s)) - b_\gamma(s,x^\pm(s))\right| \left|u^\pm(x_0,s)\right| ds\\
\le\ & |\xi(\theta)| + (M+\gamma)\int_\theta^{\theta+\eta} \sup_{s\in[\theta,\theta+\eta]} \left|x^*(s) - x^\pm(s)\right| ds\\
&+ \sup_{t\in[\theta,\theta+\eta]} \left| \int_\theta^t \left[b_0(s,x^*(s)) \right.\right.\\
&\left.\left. + b_\gamma(s,x^*(s))\right] \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds \right|\\
&+ \int_\theta^{\theta+\eta} \sup_{s\in[\theta,\theta+\eta]} \left|b_0(s,x^*(s)) \right.\\
&\left. - b_0(s,x^\pm(s))\right| \sup_{s\in[\theta,\theta+\eta]} \left|u^\pm(x_0,s)\right| ds\\
&+ \int_\theta^{\theta+\eta} \sup_{s\in[\theta,\theta+\eta]} \left|b_\gamma(s,x^*(s)) \right.\\
&\left. - b_\gamma(s,x^\pm(s))\right| \sup_{s\in[\theta,\theta+\eta]} \left|u^\pm(x_0,s)\right| ds.
\end{aligned}
$$

Using (2.3) and (2.4), recalling that $|u^*(x_0)|_\infty \le K$ and $|u^\pm(x_0)|_\infty \le K$, and using (2.7), we can write

$$
\begin{aligned}
&\sup_{t\in[\theta,\theta+\eta]} |\xi(t)| \le |\xi(\theta)| + (M+\gamma)\left(\sup_{s\in[\theta,\theta+\eta]} |\xi(s)|\right)\eta\\
&+ \sup_{t\in[\theta,\theta+\eta]} \left| \int_\theta^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right] \right.\\
&\left. \times \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds \right|\\
&+ M\int_\theta^{\theta+\eta} \sup_{s\in[\theta,\theta+\eta]} |\xi(s)|\, K ds\\
&+ \gamma\int_\theta^{\theta+\eta} \sup_{s\in[\theta,\theta+\eta]} |\xi(s)|\, K ds.
\end{aligned}
$$

From this, we get

$$
\begin{aligned}
&\sup_{t\in[\theta,\theta+\eta]} |\xi(t)| \le |\xi(\theta)|\\
&\quad + (M+\gamma)(1+K)\eta\left(\sup_{t\in[\theta,\theta+\eta]} |\xi(t)|\right)\\
&\quad + \sup_{t\in[\theta,\theta+\eta]} \left| \int_\theta^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right] \right.\\
&\quad \left. \times \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds \right|.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&(1-(M+\gamma)(1+K)\eta) \sup_{t\in[\theta,\theta+\eta]} |\xi(t)| \le |\xi(\theta)|\\
&\quad + \sup_{t\in[\theta,\theta+\eta]} \left| \int_\theta^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right] \right.\\
&\quad \left. \times \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds \right|.
\end{aligned}
$$

Now, choose a value of $\eta \in (0, t_f^*(x_0) - \theta]$ such that $(M+\gamma)(1+K)\eta < 1$ and denote

$$
\mu(\eta) := \frac{1}{1-(M+\gamma)(1+K)\eta}. \tag{2.8}
$$

Note that, for this choice of η and $\mu(\eta)$, statement (3) of the lemma is valid. Then,

$$
\begin{aligned}
&\sup_{t\in[\theta,\theta+\eta]} |\xi(t)| \le \mu(\eta)\,|\xi(\theta)| + \mu(\eta)\\
&\quad \times \sup_{t\in[\theta,\theta+\eta]} \left| \int_\theta^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right] \right.\\
&\quad \left. \times \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds \right|.
\end{aligned} \tag{2.9}
$$

We examine now the supremum that appears on the right-hand side of (2.9), employing the partition (2.3) in combination with (2.4)–(2.6). For a time $t \in (\theta, \theta+\eta]$, let $q(t) \in \{0,1,2,\dots,r\}$ be the integer for which $t \in [q(t)\lambda, (q(t)+1)\lambda]$. Then, we can write

$$
\begin{aligned}
&\sup_{t\in[\theta,\theta+\eta]} \left| \int_\theta^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right] \right.\\
&\quad \left. \times \left(u^*(s) - u^\pm(x_0,s)\right) ds \right|\\
&= \sup_{t\in[\theta,\theta+\eta]} \left| \sum_{i=0}^{q(t)-1} \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right] \right.\\
&\quad \times \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds\\
&\quad + \int_{\theta+q(t)\lambda}^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right]\\
&\quad \left. \times \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds \right|\\
&= \sup_{t\in[\theta,\theta+\eta]} \left| \sum_{i=0}^{q(t)-1} \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \left\{b_0(\theta+i\lambda, x^*(\theta+i\lambda)) \right.\right.\\
&\quad - b_0(\theta+i\lambda, x^*(\theta+i\lambda)) + b_\gamma(\theta+i\lambda, x^*(\theta+i\lambda))\\
&\quad - b_\gamma(\theta+i\lambda, x^*(\theta+i\lambda)) + b_0(s,x^*(s))\\
&\quad \left.\left. + b_\gamma(s,x^*(s))\right\} \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds\right.\\
&\quad + \int_{\theta+q(t)\lambda}^t \left[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\right]\\
&\quad \left. \times \left(u^*(x_0,s) - u^\pm(x_0,s)\right) ds \right|
\end{aligned}
$$

$$
\begin{aligned}
&\le \sup_{t\in[\theta,\theta+\eta]} \Bigg| \sum_{i=0}^{q(t)-1} b_0(\theta+i\lambda, x^*(\theta+i\lambda)) \\
&\quad \times \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg| \\
&\quad + \sup_{t\in[\theta,\theta+\eta]} \Bigg| \sum_{i=0}^{q(t)-1} b_\gamma(\theta+i\lambda, x^*(\theta+i\lambda)) \\
&\quad \times \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg| \\
&\quad + \sup_{t\in[\theta,\theta+\eta]} \Bigg| \sum_{i=0}^{q(t)-1} \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \big[b_0(s,x^*(s)) \\
&\quad - b_0(\theta+i\lambda, x^*(\theta+i\lambda))\big]\big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg| \\
&\quad + \sup_{t\in[\theta,\theta+\eta]} \Bigg| \sum_{i=0}^{q(t)-1} \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \big[b_\gamma(s,x^*(s)) \\
&\quad - b_\gamma(\theta+i\lambda, x^*(\theta+i\lambda))\big]\big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg| \\
&\quad + \sup_{t\in[\theta,\theta+\eta]} \Bigg| \int_{\theta+q(t)\lambda}^{t} \big[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big] \\
&\quad \times \big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg|. \qquad (2.10)
\end{aligned}
$$

Invoking (2.6) leads to

$$
\begin{aligned}
&\le \sum_{i=0}^{q(t)-1} \sup_{t\in[\theta,\theta+\eta]} \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \big|b_0(s,x^*(s)) \\
&\quad - b_0(\theta+i\lambda, x^*(\theta+i\lambda))\big|\,\big|u^*(x_0,s)-u^{\pm}(x_0,s)\big|\,ds \\
&\quad + \sum_{i=0}^{q(t)-1} \sup_{t\in[\theta,\theta+\eta]} \int_{\theta+i\lambda}^{\theta+(i+1)\lambda} \big|b_\gamma(s,x^*(s)) \\
&\quad - b_\gamma(\theta+i\lambda, x^*(\theta+i\lambda))\big|\,\big|u^*(x_0,s)-u^{\pm}(x_0,s)\big|\,ds \\
&\quad + \sup_{t\in[\theta,\theta+\eta]} \Bigg| \int_{\theta+q(t)\lambda}^{t} \big[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big] \\
&\quad \times \big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg|.
\end{aligned}
$$

Recalling Lemma 4.1, (4.1) and (2.2), we get for (2.10) the following bound:

$$
\begin{aligned}
&\sup_{t\in[\theta,\theta+\eta]} \Bigg| \int_{\theta}^{t} \big[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big] \\
&\quad \times \big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg| \\
&\le q(t)\lambda\rho 2K + q(t)\lambda\rho 2K + \big[B_0(x_0)+B_\gamma(x_0)\big]2K\lambda.
\end{aligned}
$$

Noting that $0 \le q(t)\lambda \le \eta$, we obtain

$$
\begin{aligned}
&\sup_{t\in[\theta,\theta+\eta]} \Bigg| \int_{\theta}^{t} \big[b_0(s,x^*(s)) + b_\gamma(s,x^*(s))\big] \\
&\quad \times \big(u^*(x_0,s)-u^{\pm}(x_0,s)\big)\,ds \Bigg| \\
&\le 4K\rho\eta + 2K\big[B_0(x_0)+B_\gamma(x_0)\big]\lambda. \qquad (2.11)
\end{aligned}
$$

Now, referring to the number σ_0 of the lemma statement and to $\mu(\eta)$ of (2.8), and recalling that $\eta > 0$, choose the number ρ in (4.1) to satisfy

$$0 < \rho < \frac{\sigma_0}{8\mu(\eta)K\eta}.$$

Then, choose $\lambda > 0$ such that η/λ is an integer and

$$0 < \lambda < \min\left\{\beta(x_0,\rho), \frac{\sigma_0}{4\mu(\eta)K\big[B_0(x_0)+B_\gamma(x_0)\big]}\right\},$$

which is clearly consistent with (2.2). Substituting these values into (2.9) and using (2.11), we obtain that

$$\sup_{t\in[\theta,\theta+\eta]} |\xi(t)| < \mu(\eta)\,|\xi(\theta)| + \sigma_0,$$

and our proof concludes. □

We can state now the main result of this section, which shows that, without significantly compromising performance, the optimal input signal $u^*(x_0)$ can be replaced by a bang–bang input signal.

Theorem 4.1: *Assume the notation and conditions of Problem 2.1. Then, for every real number $\sigma > 0$, there is a bang–bang input signal $v^{\pm}(x_0, t) \in U(K)$ for which the following are true:*

(1) *$v^{\pm}(x_0, t)$ has a finite number of switchings over the interval $[0, t_f^*(x_0)]$.*

(2) *For every member $\Sigma \in \mathcal{F}_\gamma(\Sigma_0)$, the discrepancy between the response $x^*(t) = \Sigma(x_0, u^*(x_0), t)$ to an optimal input signal $u^*(x_0)$ and the response $x_v^{\pm}(t) = \Sigma(x_0, v^{\pm}(x_0), t)$ to the bang–bang input signal $v^{\pm}(x_0)$ satisfies the inequality*

$$\sup_{t\in[0,t_f^*(x_0)]} \big|x^*(t) - x_v^{\pm}(t)\big| < \sigma.$$

Proof: Let $\Sigma \in \mathcal{F}_\gamma(\Sigma_0)$ be a member starting from the initial state x_0 at the time $t = 0$, let $\sigma_0 > 0$ be a real number to be specified later, and let $x^*(t) = \Sigma(x_0, u^*(x_0), t)$ be the response of Σ to an optimal input signal $u^*(x_0)$. Then, by Lemma 4.2 with $\theta = 0$, there is a real number $\eta \in \left(0, t_f^*(x_0)\right]$ and a bang–bang input signal $u^{\pm,1}(x_0, t)$ with a finite number of switchings over the time interval

$t \in [0, \eta]$, for which the following is true: the response $x^{\pm,1}(t) := \Sigma(x_0, u^{\pm,1}(x_0), t)$ of Σ to $u^{\pm,1}(x_0, t)$, starting at $t = 0$ from the initial state x_0, satisfies $\sup_{t\in[0,\eta]} |x^*(t) - x^{\pm,1}(t)| < \mu(\eta)|x^*(0) - x^{\pm,1}(0)| + \sigma_0$. Taking into account the fact that $x^*(0) = x^{\pm,1}(0) = x_0$, this yields

$$\sup_{t\in[0,\eta]} |x^*(t) - x^{\pm,1}(t)| < \sigma_0. \tag{2.12}$$

(The construction of $u^{\pm,1}(x_0, t)$ is described in (2.4).)

Now, starting the construction of the input signal $v^{\pm}(x_0, t)$ of the theorem, set

$$v^{\pm}(x_0, t) := u^{\pm,1}(x_0, t) \text{ for all } t \in [0, \eta],$$

and denote by $x_v^{\pm}(t) := \Sigma(x_0, v^{\pm}(x_0), t)$, $t \in [0, \eta]$, the response of Σ to the input signal $v^{\pm}(x_0)$ starting at $t = 0$ from the initial state x_0. Clearly then, $x_v^{\pm}(t) = x^{\pm,1}(t)$, $t \in [0, \eta]$, so that, by (2.12), we have

$$\sup_{t\in[0,\eta]} |x^*(t) - x_v^{\pm}(t)| < \sigma_0. \tag{2.13}$$

Recalling from Lemma 4.2(3) that the numbers η and $\mu(\eta)$ depend only on M, γ, and K and not on the value of σ_0, partition the time interval $[0, t_f^*(x_0)]$ into subintervals of length η to obtain the partition

$$[0, t_f^*(x_0)] = \big\{[0, \eta], [\eta, 2\eta], \ldots, \times [(k-1)\eta, k\eta], [k\eta, t_f^*(x_0)]\big\},$$

where k is the largest integer satisfying $k < t_f^*(x_0)/\eta$. Now, we extend the input signal $v^{\pm}(x_0)$ to the interval $(\eta, 2\eta]$, by applying Lemma 4.2 with $\theta = \eta$. Let $u^{\pm,2}(x_0, t)$ be a bang–bang input signal that satisfies the conditions of Lemma 4.2 over the time interval $[\eta, 2\eta]$; then, $u^{\pm,2}(x_0, t)$ has a finite number of switchings in this interval. Extend the signal $v^{\pm}(x_0)$ to the interval $(\eta, 2\eta]$ by setting

$$v^{\pm}(x_0, t) := u^{\pm,2}(x_0, t),\ t \in (\eta, 2\eta];$$

denote by $x_V^{\pm}(t) := \Sigma(x_0, v^{\pm}(x_0), t)$, $t \in [0, 2\eta]$, the response of Σ to $v^{\pm}(x_0)$ starting from the initial state x_0. Then, according to Lemma 4.2 with $\theta = \eta$, we have $\sup_{t\in[\eta,2\eta]} |x^*(t) - x^{\pm}(t)| < \mu(\eta)|x^*(\eta) - x_v^{\pm}(\eta)| + \sigma_0$. In view of (2.13), we get

$$\sup_{t\in[\eta,2\eta]} |x^*(t) - x^{\pm}(t)| < \mu(\eta)\sigma_0 + \sigma_0. \tag{2.14}$$

Continuing in a similar manner, the next step is to extend $v^{\pm}(x_0)$ to the interval $(2\eta, 3\eta]$. This is accomplished by using Lemma 4.2 with $\theta = 2\eta$ to build a bang–bang input signal $u^{\pm,3}(x_0, t)$ with a finite number of switchings on $[2\eta, 3\eta]$, and setting $v^{\pm}(x_0, t) := u^{\pm,3}(x_0, t)$, $t \in (2\eta, 3\eta]$. In view of (2.14) and Lemma 4.2 with $\theta = 2\eta$, this leads to the inequality

$$\sup_{t\in[2\eta,3\eta]} |x^*(t) - x^{\pm}(t)| < \mu(\eta)[\mu(\eta)\sigma_0 + \sigma_0] + \sigma_0.$$

More generally, for an integer $i \in \{1, \ldots, k\}$, we use Lemma 4.2 with $\theta = i\eta$ to build a bang–bang input signal $u^{\pm,(i+1)}(x_0, t)$ over the interval $[i\eta, \min\{(i+1)\eta, t_f^*(x_0)\})$ having a finite number of switchings over this interval, and set $v^{\pm}(x_0, t) := u^{\pm,(i+1)}(x_0, t)$, $t \in (i\eta, \min\{(i+1)\eta, t_f^*(x_0)\}$. Using Lemma 4.2 with $\theta = i\eta$, we get the inequality

$$\sup_{t\in\left[i\eta,\min\left\{(i+1)\eta,t_f^*(x_0)\right\}\right]} |x^*(t) - x^{\pm}(t)| < \chi_i,$$

where

$$\chi_i := \mu(\eta)|x^*(i\eta) - x^{\pm,i}(i\eta)| + \sigma_0 = \mu(\eta)\chi_{i-1} + \sigma_0.$$

Thus, χ_i is determined by the recursion

$$\chi_{i+1} = \mu(\eta)\chi_i + \sigma_0,\ i = 0, 1, 2, \ldots,$$
$$\chi_0 = 0.$$

Referring to the number σ of the theorem's statement, a slight reflection shows that σ_0 can be selected so that $\chi_i < \sigma$ for all $i \in \{0, 1, \ldots, k\}$. Then, with such value of σ_0, the bang–bang input signal

$$v^{\pm}(x_0, t) = \begin{cases} u^{\pm,1}(x_0, t) & \text{for } t \in [0, \eta], \\ u^{\pm,2}(x_0, t) & \text{for } t \in (\eta, 2\eta], \\ \vdots & \\ u^{\pm,k}(x_0, t) & \text{for } t \in ((k-1)\eta, k\eta], \\ u^{\pm,(k+1)}(x_0, t) & \text{for } t \in (k\eta, t_f^*(x_0)], \end{cases}$$

satisfies the requirements of the theorem. As $v^{\pm}(x_0)$ is built from a finite number of bang–bang segments, and each of these segments has a finite number of switchings, we conclude that $v^{\pm}(x_0)$ also has a finite number of switchings over the interval $[0, t_f^*(x_0)]$. Furthermore, with the above selection of σ_0, we get

$$\sup_{t\in\left[0,t_f^*(x_0)\right]} |x^*(t) - x_v^{\pm}(t)| < \sigma.$$

This concludes our proof. □

Theorem 4.1 shows that bang–bang input signals can drive a system to a performance that is as close as desired

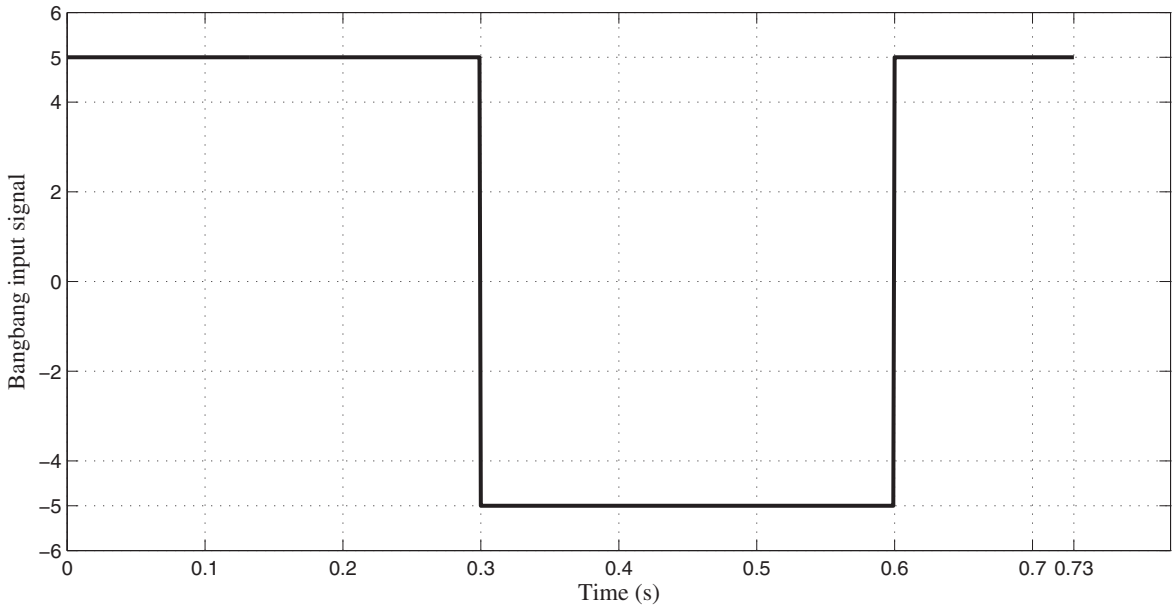

Figure 2. A bang–bang input signal $v^{\pm}(x_0)$ that approximates optimal performance.

to optimal performance. This is an important conclusion, since it is much easier to calculate and implement a bang–bang signal that approximates optimal performance, than to calculate and implement an optimal input signal. Indeed, one can find an appropriate bang–bang input signal that approximates optimal performance through a trial-and-error optimisation process over the class of bang–bang input signals.

5. Example

Consider the family of systems $\mathcal{F}$ described by the differential equations

$$\mathcal{F}: \begin{array}{l} \dot{x}_1(t) = -c(1 + 0.5\cos(t))x_1(t) + (1-t)u(t), \\ \dot{x}_2(t) = d(1 - 0.5\sin(t))x_2(t) + (1-t)u(t), \end{array} \quad t \geq 0,$$

where c and d are real numbers describing parameter uncertainties subject to the inequalities

$$1.4 \leq c \leq 1.6,$$
$$0.9 \leq d \leq 1.1.$$

The input signal $u(t)$ is subject to the bound

$$|u(t)|_\infty \leq 5.$$

The state vector at the time t is $x(t) = (x_1(t), x_2(t))^T$, and all members of the family $\mathcal{F}$ start at $t = 0$ from the initial state $x_0 = (3.5, -1)^T$. The control objective is to find an optimal input signal $u^*(x_0)$ that brings the state $x(t)$ of every member of the family $\mathcal{F}$ as quickly as possible from x_0 to $\rho(1.25)$, namely, to the 1.25 −vicinity of the origin.

Numerical optimisation shows that the shortest time to achieve this control objective is $t_f^*(x_0) = 0.73$. As Figure 3 depicts, a similar time can be achieved by the bang–bang input signal $v^{\pm}(x_0)$ of Figure 2; note that $v^{\pm}(x_0)$ has two switchings in the time interval $[0, t_f^*(x_0)] = [0, 0.73]$.

The optimisation process used to derive the results of this example is a relatively simple process that consists of numerical optimisation performed over the switching times of the bang–bang input signal $v^{\pm}(x_0)$. In somewhat simplified terms, this process proceeds as follows.

Denote by κ the number of switching times of a bang–bang input signal $v^{\pm}(x_0)$, and note that a bang–bang signal is determined by its switching times and by the signs of its components following each switching time (in this example, the input signal has only one component). As before, denote by $x(t)$ the response of members of the family $\mathcal{F}$. Using physical, engineering, or general mathematical considerations, estimate a time ϑ within which the specification $x^T(t)x(t) \leq 1.25$ can be met for all members of $\mathcal{F}$.

(1) Set $\kappa = 0$.
(2) Use a numerical process to vary the placement of the κ switching times, as well as the sign of the signal after each switching time, until the response $x(t)$ of each member of the family $\mathcal{F}$ satisfies one of the following:

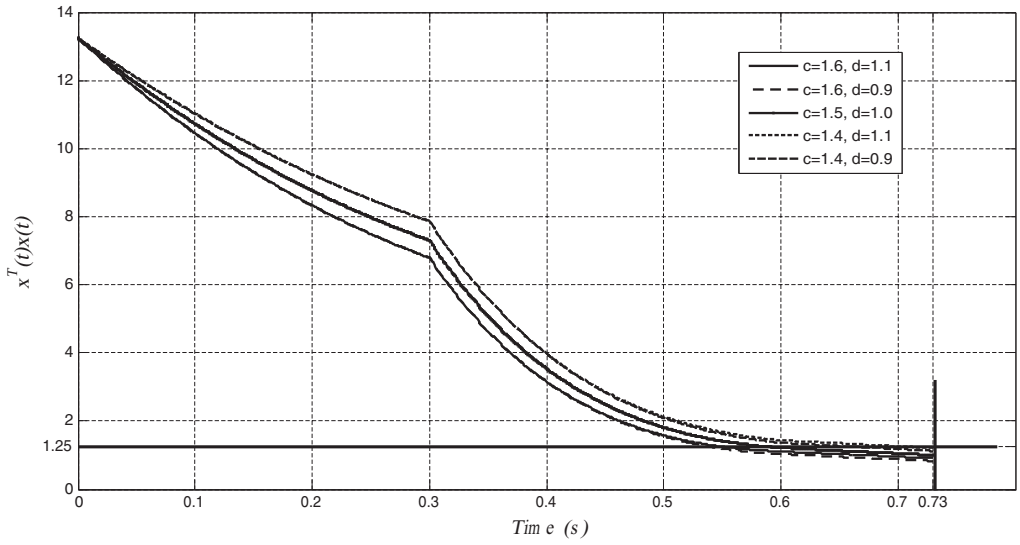

Figure 3. $x^T(t)x(t)$ for the family $\mathcal{F}$ with the bang–bang input signal $v^{\pm}(x_0)$ of Figure 2.

(a) There is a time $t(\kappa) \in [0, \vartheta]$ at which the specification $x^T(t(\kappa))x(t(\kappa)) \leq 1.25$ is met by all members of the family $\mathcal{F}$; then, continue to (3).

(b) No such time $t(\kappa)$ can be found; then, replace κ by $\kappa + 1$ and repeat from (2).

(3) Use a numerical optimisation process to vary the placement of the κ switching times, as well as the sign of the signal after each switching time, to find the minimal value of the time $t(\kappa)$ for κ switching times. Denote this minimal value by $T(\kappa)$ and denote by $v^{\pm}(\kappa, x_0)$ the corresponding bang–bang input signal.

(4) Terminate the process if $\kappa \geq 1$ and $|T(\kappa) - T(\kappa - 1)| \leq e$, where e is a specified indicator of acceptable deviation from optimality.

(5) Replace κ by $\kappa + 1$ and repeat from (3).

Upon termination of this iterative process, the time $T(\kappa - 1)$ approximates the minimal time $t_f^*(x_0)$, and the bang–bang input signal $v^{\pm}(x_0) := v^{\pm}(\kappa - 1, x_0)$ approximates optimal performance. Considering that the latter is a bang–bang signal, it is relatively easy to implement.

If necessary, more elaborate numerical optimisation techniques can be employed to derive the switching times and signs of a bang–bang input signal that achieves a close approximation of optimal performance.

6. Conclusion

In this paper, we investigated the problem of reducing operating errors as quickly as possible during recovery from a feedback disruption. The objective was to develop controllers which, upon restoration of the feedback signal, reduce in minimal time operating errors that have accumulated during the time feedback signals were absent. The main results of the paper are twofold:

(1) robust optimal controllers that reduce operating errors in minimal time upon feedback recovery do exist under rather broad conditions; and

(2) the performance of such optimal controllers can be approximated as closely as desired by controllers that generate bang–bang input signals for the controlled system.

The fact that optimal performance can be closely approximated by controllers that generate bang–bang signals is a significant advantage. Indeed, controllers that generate bang–bang signals are relatively easy to design and implement, since a bang–bang signal is basically determined by a finite string of scalars – the switching times.

The results of this paper have potential applications in a number of control engineering specialties. One important such specialty is the design of digital controllers for continuous-time systems. Here, controllers interact with the controlled system through a process of periodic sampling of the controlled system's output signal. It goes without saying that a feedback disruption occurs between every two samples. During this inter-sample time, the controlled system operates without feedback and develops operating errors; reducing these errors as quickly as possible upon arrival of the next output signal sample will improve system performance. The latter can be achieved by controllers developed in this paper.

Finally, many directions of future research are open in subjects related to the topics discussed in this paper.

One such direction would be to generalise the results of this paper to nonlinear systems that are not necessarily input affine. Another direction of future research would be to combine the results of this paper with the results of Chakraborty and Hammer (2009, 2010) to create a unified controller that keeps operating errors low during feedback disruptions and then reduces these errors as quickly as possible, once feedback has been restored.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work of Zhaoxu Yu was supported in part by the Natural Science Foundation of the P. R. China [grant number 61304071] and by the Fundamental Research Funds for the Central Universities.

References

Balakrishnan, S., Tsourdos, A., & White, B. (Eds.). (2012). *Advances in missile guidance, control, and estimation* (1st ed.). Boca Raton, FL: CRC Press.

Chakraborty, D., & Hammer, J. (2009). Optimal control during feedback failure. *International Journal of Control, 82*(8), 1448–1468.

Chakraborty, D., & Hammer, J. (2010). Robust optimal control: Low-error operation for the longest time. *International Journal of Control, 83*(4), 731–740.

Chakraborty, D., & Shaikshavali, C. (2009). An approximate solution to the norm optimal control problem. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (pp. 4490–4495). San Antonio, TX.

Gamkrelidze, R. (1965). On some extremal problems in the theory of differential equations with applications to the theory of optimal control. *SIAM Journal on Control, 3*, 106–128.

Kelendzheridze, D. (1961). On the theory of optimal pursuit. *Soviet Mathematics Doklady, 2*, 654–656.

Luenberger, D.G. (1969). *Optimization by vector space methods*. New York, NY: Wiley.

Modeling, T.R., Spong, C.M.W., Hutchinson, S., & Vidyasagar, M. (2006). *Robot modeling and control*. New York, NY: Wiley.

Montestruque, L., & Antsaklis, P. (2004). Stability of model-based networked control systems with time-varying transmission times. *IEEE Transactions on Automatic Control, 49*(9), 1562–1572.

Nair, G., Fagnani, F., Zampieri, S., & Evans, R.J. (2007). Feedback control under data rate constraints: An overview. *Proceedings of the IEEE, 95*(1), 108–137.

Neustadt, L. (1966). An abstract variational theory with applications to a broad class of optimization problems I, general theory. *SIAM Journal on Control, 4*, 505–527.

Neustadt, L. (1967). An abstract variational theory with applications to a broad class of optimization problems II, applications. *SIAM Journal on Control, 5*, 90–137.

Pontryagin, L., Boltyansky, V., Gamkrelidze, R., & Mishchenko, E. (1962). *The mathematical theory of optimal processes*. New York, NY: Wiley.

Warga, J. (1972). *Optimal control of differential and functional equations*. New York, NY: Academic Press.

Willard, S. (1970). *General topology*. Reading, MA: Addison-Wesley.

Young, L. (1969). *Lectures on the calculus of variations and optimal control theory*. Philadelphia, PA: W. B. Saunders.

Zeidler, E. (1985). *Nonlinear functional analysis and its applications III*. New York, NY: Springer-Verlag.

Zhivogyladov, P., & Middleton, R. (2003). Networked control design for linear systems. *Automatica, 39*, 743–750.