

DISCRETE-TIME REFERENCE GOVERNORS AND THE NONLINEAR CONTROL OF SYSTEMS WITH STATE AND CONTROL CONSTRAINTS

ELMER G. GILBERT AND ILYA KOLMANOVSKY

Department of Aerospace Engineering, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

AND

KOK TIN TAN

Defence Science Organisation, Republic of Singapore, Singapore 0511, Republic of Singapore

SUMMARY

Discrete-time, linear control systems with specified pointwise-in-time constraints, such as those imposed by actuator saturation, are considered. The constraints are enforced by the addition of a nonlinear 'reference governor' that attenuates, when necessary, the input commands. Because the constraints are satisfied, the control system remains linear and undesirable response effects such as instability due to saturation are avoided. The nonlinear action of the reference governor is defined in terms of a finitely determined maximal output admissible set and can be implemented on-line for systems of moderately high order. The main result is global in nature: if the input command converges to a statically admissible input and the initial state of the system belongs to the maximal output admissible set, the eventual action of the reference governor is a unit delay.

KEY WORDS nonlinear control; reference governor; control saturation; state constraints

1. INTRODUCTION

Linear models play a vital role in control system design. Powerful design techniques are available for synthesizing controllers which achieve a variety of performance and robustness objectives. Treatment of systems with significant nonlinearities is in a less satisfactory state. Special assumptions concerning the nature of the nonlinearities and the design objectives are needed. This paper addresses the implementation of nonlinear controllers for systems in which the equations of motion are linear but there are pointwise-in-time constraints on control and/or state variables. These assumptions are quite realistic in many practical applications. Certainly, they apply to systems in which the principal nonlinearity is actuator saturation. Our objective is to obtain controllers which preserve desirable small-input behaviour of linear-system designs, respond well to large inputs, have simple structure and, for systems of significant order, can be implemented digitally at reasonably high sampling rates.

Consider the example system shown in Figure 1. The purpose of feedback control is good tracking performance, i.e., a small closed-loop error, $z(t)$, for a large class of reference

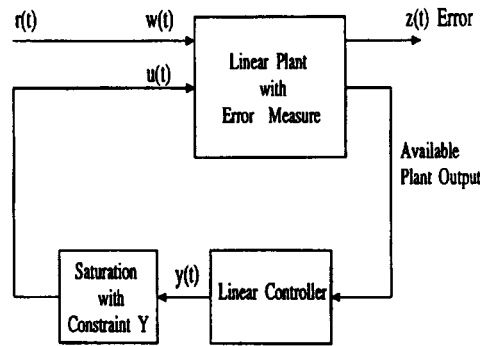


Figure 1. Closed-loop system with controller saturation

commands, $w(t) = r(t)$. The output of the controller, $y(t)$, is subject to saturation: $u(t) = \text{sat } y(t)$, where $\text{sat } y = y$ if and only if $y \in Y$. Typically, the origin is contained in the interior of Y . Thus, when the reference commands are small, saturation does not occur. Suppose, for this small-input situation, an effective linear controller has been designed. What happens when the reference commands are large enough to cause saturation? Often the effect is unacceptable: $z(t)$ may diverge or become disproportionately large and persist for a long time. A familiar, simple example is integrator windup. Contemporary design methods, which deal effectively with poorly conditioned or unstable multivariable plants, expand the potential for response pathology.²³

In engineering practice, undesirable effects of saturation are often attacked by a combination of compromises such as: relaxation of performance objectives in the linear system design, restrictions on the class of inputs, use of more powerful actuators. A less common strategy is to acknowledge the inherently nonlinear character of the problem and introduce nonlinear control. Many approaches have been proposed. They include, together with samplings of the literature: stable feedback regulation^{7,13,17,22,24,25,27} moving-horizon optimal control,¹⁸⁻²⁰ anti-windup schemes^{1,2,6,8,28} model predictive control^{9,12,21} and the exploitation of positively invariant sets.^{3-5,10,11,14} Most of these approaches do not apply to the tracking problem of Figure 1. Those that do have weaknesses: lack of a general underlying theory, need for extensive on-line computations, and inability to incorporate realistic performance and robustness objectives.

One promising approach is the continuous-time error governor of Kapasouris, Athans and Stein.¹⁵ In the context of Figure 1, it involves the introduction of a variable scalar gain, at the input of the controller. The gain, $\kappa(t) \in [0, 1]$, is determined for each t so that $y(\tau) \in Y$ for all $\tau \geq t$. The determination is based on the maximal output admissible set¹⁰ of the controller, i.e., the set of all initial controller states for which the unforced controller satisfies $y(t) \in Y$, $t \geq 0$. With small reference commands $\kappa(t) \equiv 1$; thus the small-input performance of the linear design is preserved. For large commands there are values of t for which $\kappa(t) < 1$; however, $y(t) \in Y$, $t \geq 0$, and saturation is avoided. Simulation studies demonstrate much improved large-input response, particularly in multivariable systems where saturation causes changes in the direction of controller output.

The error governor has several limitations: the computational times for determining $\kappa(t)$ are excessively large, the plant must be asymptotically stable and, from a theoretical point of view, stability and the response results are only local in character. The first difficulty can be overcome by using discrete-time implementations of the error governor.¹⁰ Tan²⁶ has suggested a further modification which, with certain restrictions, eliminates the need for a stable plant. In addition,

he shows that for $r(t) \equiv 0$ his modified error-governor system is asymptotically stable in the large.

In this paper we examine the action of discrete-time reference governors. Figure 2 displays the basic arrangement, reduced to its simplest form. It consists of a controlled process, a pointwise-in-time constraint, $y(t) \in Y$, and the reference governor. The purpose of the governor is to attenuate the input command, but only when necessary, so that the constraint is always satisfied. When the controlled process is the closed-loop system of Figure 1, the scheme guarantees that saturation does not take place. Thus, the linear model of the closed-loop system and its performance characteristics remain valid. The nonlinear behaviour of the overall system is determined entirely by the reference governor, whose response characteristics are relatively simple. A key advantage of the approach is that it can be applied to feedback systems which have already been designed for optimum, linear-system performance. Since the reference governor is added after the linear design is complete, compromises need not be made in the linear design process.

Results and developments of the paper are organized as follows. Section 2 formulates the control problem and proposes two discrete-time reference governors, one static and one dynamic. While the static governor is simple and intuitively obvious, it has a somewhat surprising defect: for constant reference commands the output of the reference governor may fail to converge. The dynamic governor, which is the natural discrete-time extension of the continuous-time reference governor of Kerasi, Athans and Stein,¹⁶ is the principal concern of this paper. Its implementation, which is also based on a maximal output admissible set, is discussed in Section 3. Under appropriate conditions, it is shown that the required nonlinearity can be computed in an efficient and systematic fashion. Section 4 presents results on system response. Unlike others,^{10,15,16} they are global in nature. For instance, suppose $r(t) \rightarrow r_0$, where r_0 is any statically admissible input for the controlled process. Then, $w(t) \rightarrow r_0$. Proofs of the theorems in Section 4 are given in Section 5. Generalizations of the dynamic governor are considered in Section 6. In Section 7 the reference governor is applied to the feedback control of a helicopter which is unstable and has bounds on both of its two inputs. The helicopter with its controller is order 10, so the problem is nontrivial. Large-input responses are good and the instability, which is usually induced by saturation, is avoided. Computational times for the nonlinearities of the reference governor are modest so that on-line implementation is feasible. Section 8 treats an example where the pointwise-in-time constraint is on the *output* of a closed-loop system. The purpose of the reference governor is to prevent excessive overshoot of the step response. If the state of the controlled process is not available, the reference governor may be implemented with a model of the controlled process. Simulations illustrate the effect of errors in the model.

The following mathematical notation appears: Z^+ is the set of nonnegative integers; R^n and $R^{m \times n}$ are the usual notations for sets of real vectors and matrices; the superscript T denotes

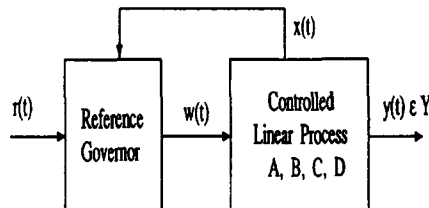


Figure 2. The reference governor and controlled process with constrained output

matrix transpose; subscripts indicate vector components; $|x|$ is the Euclidean norm of $x \in R^n$, l_∞^n is the set of bounded sequences, $\{x(t) \in R^n: t \in Z^+\}$ (when it is clear by context, we write $x \in l_\infty^n$); for $x \in l_\infty^n$, $\|x\|_\infty = \sup\{|x(t)|: t \in Z^+\}$. The boundary, interior, closure and convex hull of a set are denoted respectively by bd , int , cl and co . For $\alpha \in R$ and $X, Y \subset R^n$, $\alpha X = \{\alpha x: x \in X\}$ and $X + Y = \{x + y: x \in X, y \in Y\}$.

The discrete-time, maximal output admissible set is defined in terms of a triple, $A \in R^{n \times n}$, $C \in R^{p \times n}$ and $Y \subset R^p$:

$$O_\infty(A, C, Y) = \{x: CA^t x \in Y, t \in Z^+\} \subset R^n \quad (1)$$

Obviously, $O_\infty(A, C, Y)$ is related to the system $x(t+1) = Ax(t)$, $y(t) = Cx(t)$. The pointwise-in-time output constraint, $y(t) \in Y$, $t \in Z^+$, is satisfied if and only if $x(0) \in O_\infty(A, C, Y)$. Our developments depend strongly on results contained in Reference 10. Several are summarized in the following theorem.

Theorem 1.1

Suppose A is Lyapunov stable (A^t is bounded for all $t \in Z^+$), Y is compact and $0 \in \text{int } Y$. Then, $0 \in \text{int } O_\infty(A, C, Y)$ and $O_\infty(A, C, Y)$ is closed and A invariant ($x \in O_\infty(A, C, Y)$ implies $Ax \in O_\infty(A, C, Y)$). Suppose, in addition, that C, A is observable. Then, $O_\infty(A, C, Y)$ is compact.

2. THE DISCRETE-TIME REFERENCE GOVERNORS

The controlled process and its constraint, shown in Figure 2, are represented by

$$x(t+1) = Ax(t) + Bw(t), \quad x(t) \in R^n, \quad w(t) \in R^m \quad (2)$$

$$y(t) = Cx(t) + Dw(t) \in Y \subset R^p, \quad t \in Z^+ \quad (3)$$

It is assumed hereafter that: (A1) A has its characteristic roots inside the unit disk, (A2) C, A is observable, (A3) Y is compact and is expressed by

$$Y = \{y: f_i(y) \leq 0, i = 1, \dots, s\} \quad (4)$$

where the functions $f_i: R^p \rightarrow R$ are continuous and satisfy $f_i(0) < 0$. These assumptions are consistent with constraint modelling of stable closed-loop systems, such as the one shown in Figure 2. Since constraint satisfaction does not depend on states which are unobservable in y , there is no real loss of generality in (A2). The conditions $f_i(0) < 0$ imply $0 \in \text{int } Y$, so that the assumptions of Theorem 1.1 are satisfied. When (2)–(3) represent the system in Figure 2 the constraint is generally state-dependent ($C \neq 0$), even though Y is a control constraint.

The constraint set, (4), affects the set of constant inputs for which (2)–(3) has an admissible equilibrium solution. Specifically, $w(t) \equiv w_0$ and $x(t) \equiv x_0$ requires $x_0 = (I - A)^{-1}Bw_0$ and $w_0 \in W_0$, where

$$W_0 = \{w: H_0 w \in Y\} \quad (5)$$

$$H_0 = D + C(I - A)^{-1}B \quad (6)$$

By assumption (A1) the inverse of $I - A$ exists.

The static reference governor is a scalar gain, K , which varies between 0 and 1 and depends nonlinearly on the reference command and the state of the controlled process:

$$w(t) = K(r(t), x(t))r(t) \quad (7)$$

Ideally, we want $w(t)$ to be close to $r(t)$. Thus, $K(r(t), x(t)) \in [0, 1]$ should be as large as possible subject to the requirement that $y(\tau) \in Y, \tau \geq t$. The last condition demands a prediction of $y(t)$. In the absence of information on $r(\tau)$ for $\tau > t$, the only sensible strategy is to base the prediction on the behaviour of the system (2) with $w(\tau) = 0$ for $\tau > t$. This prediction involves $O_\infty(A, C, Y): w(\tau) = 0, \tau > t$, implies $y(\tau) \in Y, \tau > t$, if and only if $x(t+1) \in O_\infty(A, C, Y)$. This and the constraint, $y(t) \in Y$, give

$$K(r, x) = \max\{\alpha \in [0, 1]: Ax + Bar \in O_\infty(A, C, Y), Cx + Dar \in Y\} \tag{8}$$

Assume $r \in R^m$ and $x \in O_\infty$. Then $Ax \in O_\infty, Cx \in Y$ and the inclusions in (8) are satisfied for at least one $\alpha \in [0, 1]$. Since Y and $O_\infty(A, C, Y)$ are closed, the maximum exists. Thus, $K: R^m \times O_\infty \rightarrow [0, 1]$ is defined. Moreover, (2), (7) and (8) show that $x(t) \in O_\infty$ implies $y(t) \in Y$ and $x(t+1) \in O_\infty$. Successive application of these results proves (3) is satisfied whenever $x(0) \in O_\infty$.

When does the reference governor allow $r(t) \equiv w(t) \equiv r_0$ and $x(t) \equiv x_0 = (I - A)^{-1}Br_0$? Since $x(t) \in O_\infty(A, C, Y)$, r_0 is admissible if and only if it belongs to $W_0 \cap \{r_0: x_0 \in O_\infty(A, C, Y)\} = W_1$. Often, W_1 is significantly smaller than W_0 , so the static governor seriously restricts the class of admissible constant inputs. Actually, the static governor has a more serious defect.

Consider the following example: $m = n = p = 1, A = -0.9, B = 1, C = 1, D = 0.1, Y = [-1, 1]$. Simple computations show: $O_\infty = [-1, 1], W_0 = W_1 = [-1.596 \dots, 1.596 \dots]$. Let $r_0 = 1$. Then by (8): $K(1, x) = 1 + 0.9x, x \in [-1, 0], = 1, x \in [0, 0.9], = 10 - 10x, x \in [0.9, 1]$. It is easy to see that oscillations can be generated. For example, $x(0) = 0$ generates the sequence $\{x(t): t \in Z^+\} = \{0, 1, -0.9, 1, -0.9, \dots\}$. More generally, it can be shown that $x(t) \rightarrow (I - A)^{-1}Br_0 = 0.5263\dots$ if and only if $x(0) \in (0.102 \dots, 0.908\dots)$. Similarly, if $|r_0| \in [0.19, 1.596\dots]$, there exist initial conditions, $x(0) \in O_\infty$, which lead to oscillations. However, when $|r_0| < 0.19$ it can be shown that $x(t) \rightarrow (I - A)^{-1}Br$ for all $x(0) \in O_\infty$. Thus, the static governor may fail to converge with a constant input, $r(t) \equiv r_0$, unless $|r_0|$ is very small. Convergence to equilibria also fails in examples where $D = 0$.

The dynamic reference governor mimics the continuous-time reference governor of Reference 16. It is a first-order, low-pass filter with a bandwidth parameter, K , which depends nonlinearly on $r(t)$, its own state and the state of the controlled process:

$$w(t + 1) = w(t) + K(r(t), x_G(t))(r(t) - w(t)), \quad x_G(t) = \begin{bmatrix} w(t) \\ x(t) \end{bmatrix} \tag{9}$$

We require $K(r(t), x_G(t)) \in [0, 1]$. This implies that $w(t+1)$ is an interpolation between $r(t)$ and $w(t)$ in the sense that it belongs to the line segment joining $r(t)$ and $w(t)$. Ideally, $K(r(t), x_G(t)) = 1$. Then $w(t+1) = r(t)$ and the reference command suffers only a unit delay. When the possibility of constraint violation occurs, $K(r(t), x_G(t))$ is reduced so that $w(t+1)$ is closer to $w(t)$, which has been chosen on the previous time step to be constraint admissible.

The rationale for defining K is best understood by writing (2), (3) and (9) as a single system:

$$x_G(t + 1) = A_G x_G(t) + B_G K(r(t), x_G(t))(r(t) - [I \ 0]x_G(t)), \quad y(t) = C_G x_G(t) \in Y \tag{10}$$

$$A_G = \begin{bmatrix} I & 0 \\ B & A \end{bmatrix}, \quad B_G = \begin{bmatrix} I \\ 0 \end{bmatrix}, \quad C_G = [D \ C] \tag{11}$$

Given $r(t)$ and $x_G(t)$ it is necessary to guarantee that $y(\tau) \in Y$ for all $\tau \geq t$. With no prior knowledge of $r(\tau), \tau > t$, the only safe strategy, based on what we can determine about the

future response of (10), (11), is to require $x(\tau) \in O_\infty(A_G, B_G, Y)$, $\tau \geq t$. Then it is known that an acceptable value of $K(r(\tau), x_G(\tau))$, namely 0, is always available. The largest possible value of $K(r(t), x_G(t))$, subject to our requirements, is determined by

$$K(r, x_G) = \max\{\alpha \in [0, 1]: A_G x_G + B_G \alpha(r - [I \ 0]x_G) \in O_\infty(A_G, C_G, Y)\} \quad (12)$$

Note that A_G is Lyapunov stable, so the results of Theorem 1.1 apply. Thus, $K: R^m \times O_\infty \rightarrow [0, 1]$ is defined. Successive applications of (10) and (12) show that $x_G(0) \in O_\infty$ implies $x_G(t) \in O_\infty$ and $y(t) \in Y$ for all $t \in Z^+$.

The action of the dynamic governor is quite different from the action of the static governor. Consider the equilibria that can be supported by the dynamic governor. Using the notations introduced previously, it is clear that $x_{G0} = [r_0^T \ x_0^T]^T$ must belong to $O_\infty(A_G, C_G, Y)$. From (11), it can be verified that

$$H_t = C_G A_G' \begin{bmatrix} I \\ (I - A)^{-1} B \end{bmatrix} = H_0, \quad t \in Z^+ \quad (13)$$

From this and the definition of $O_\infty(A_G, C_G, Y)$ it follows that $x_{G0} \in O_\infty(A_G, C_G, Y)$ is equivalent to $H_0 r_0 \in Y$. Hence, the set of admissible r_0 is W_0 and there is no additional restriction on the class of constant inputs. Further, as will be seen in Section 4, the convergence difficulties exhibited by the static governor are avoided.

What can be said in general about the behaviour of K in the two governors? Since $0 \in \text{int } Y$ and $0 \in \text{int } O_\infty$, the right sides of (8) and (12) allow $\alpha = 1$ when r , x and x_G are sufficiently small. This and the asymptotic stability of A leads to the following, rather weak, conclusion.

Remark 2.1. For small inputs and initial conditions both governors act as stable linear systems. More specifically, there exists a $c > 0$ such that $\|r\|_\infty \leq c$ and $|x(0)| \leq c$ ($\|r\|_\infty \leq c$ and $|x_G(0)| \leq c$) imply $K(r(t), x(t)) = 1$ ($K(r(t), x_G(t)) = 1$) for all $t \in Z^+$.

Note that the remark does not contradict the results of the above example for the static governor; it applies if $c < 0.19$. It will be shown later that a much stronger result holds for the dynamic governor. Because of its inferior response properties, the static reference governor will not be considered further.

3. IMPLEMENTATION OF THE DYNAMIC REFERENCE GOVERNOR

Practical implementation of the dynamic reference governor requires an algorithm for the evaluation of $K(r, x_G)$. One is available if $O_\infty(A_G, C_G, Y)$ is *finitely determined*,¹⁰ i.e., the infinitely many inequalities appearing in the definition of $O_\infty(A_G, C_G, Y)$ can be replaced by a finite set of inequalities. When $O_\infty(A_G, C_G, Y)$ is finitely determined, the listing of the active inequalities has a special structure:¹⁰ there exist $S^* \subset \{1, \dots, s\}$ and $t_i^* \in Z^+$, $i \in S^*$, such that

$$O_\infty(A_G, C_G, Y) = \{x_G: f_i(C_G A_G' x_G) \leq 0, t = 0, \dots, t_i^*, i \in S^*\} \quad (14)$$

The index set, S^* , and the integers, t_i^* , can be obtained by solving a sequence of mathematical programming problems. Details are given in Reference 10. The computations are often straightforward, even when p , m and n are quite large. For example, when Y is polyhedral the programming problems are linear. Since these computations are only a preliminary step in the determination of $K(r, x_G)$, they are off-line.

Is $O_\infty(A_G, C_G, Y)$ finitely determined? This question is investigated most easily in a different coordinate system. Define

$$U = \begin{bmatrix} I & 0 \\ (I - A)^{-1}B & I \end{bmatrix}, \quad A_g = U^{-1}A_G U, \quad C_g = C_G U \quad (15)$$

Obviously, the indicated inverses exist. It can be verified that

$$A_g = \begin{bmatrix} I & 0 \\ 0 & A \end{bmatrix}, \quad C_g = [H_0 \quad C] \quad (16)$$

The set $O_\infty(A_G, C_G, Y)$ is finitely determined if and only if $O_\infty(A_g, C_g, Y)$ is finitely determined. In fact, S^* and t_i^* , $i \in S^*$, are the same for the two coordinate systems.

Systems of the form (16) are studied in Reference 10. Examples there show that $O_\infty(A_g, C_g, Y)$ is generally not finitely determined. However, a finitely determined approximation of $O_\infty(A_g, C_g, Y)$ always exists. This leads to the following theorem.

Theorem 3.1

Suppose the assumptions (A1)–(A3) are satisfied. Define

$$\hat{C}_G = \begin{bmatrix} H_0 & 0 \\ D & C \end{bmatrix}, \quad Y(\varepsilon) = \{y: f_i(y) \leq -\varepsilon, i = 1, \dots, s\} \quad (17)$$

and assume $0 < \varepsilon < \min\{-f_i(0): i = 1, \dots, s\}$. Then, $O_\infty(A_G, \hat{C}_G, Y(\varepsilon) \times Y)$ is finitely determined and

$$O_\infty(A_G, C_G, Y(\varepsilon)) \subset O_\infty(A_G, \hat{C}_G, Y(\varepsilon) \times Y) \subset O_\infty(A_G, C_G, Y) \quad (18)$$

Proof. First suppose the pair C_g, A_g is observable. Define

$$\hat{C}_g = \begin{bmatrix} H_0 & 0 \\ H_0 & C \end{bmatrix} \quad (19)$$

Then by Theorem 5.1 of Reference 10, $O_\infty(A_g, \hat{C}_g, Y(\varepsilon) \times Y)$ is finitely determined and $O_\infty(A_g, C_g, Y(\varepsilon)) \subset O_\infty(A_g, \hat{C}_g, Y(\varepsilon) \times Y) \subset O_\infty(A_g, \hat{C}_g, Y)$. Working these results backward through the state transformation U proves the theorem. The pair C_g, A_g is observable if and only if $\text{rank } H_0 = m$, a condition which is not implied by our assumptions. For instance, it fails when $m > p$. If $\text{rank } H_0 < m$ the proof is carried through by eliminating the unobservable states. See the comments in Section II of Reference 10 which concern unobservable systems. \square

Remark 3.1. The assumption on ε implies $0 \in \text{int } Y(\varepsilon) \neq \emptyset$. It is also clear that $Y(\varepsilon)$ is compact. Thus, by Theorem 1.1 the sets in (18) are closed, include the origin in their (nonempty) interiors and are A -invariant. They are bounded if and only if $\text{rank } H_0 = m$. They are convex if $Y(\varepsilon)$ and Y are convex.

To simplify notations define

$$O_\infty^\varepsilon = O_\infty(A_G, \hat{C}_G, Y(\varepsilon) \times Y) \quad (20)$$

The approximation of $O_\infty(A_G, C_G, Y)$ by O_∞^ε is good in the sense that the ‘reduction’ of O_∞^ε is no worse than the reduction in $O_\infty(A_G, C_G, Y)$ which occurs when Y is replaced by $Y(\varepsilon)$.

Remark 3.2. Ideally, $Y(\varepsilon) \rightarrow Y$ (in the Hausdorff metric for compact sets) as $\varepsilon \rightarrow 0$. While this property is usually satisfied, it is not guaranteed by assumption (A3). Stronger results are available.¹⁰ If Y is convex, $0 \in \text{int } Y$ and $\varepsilon \in [0, 1]$, it is possible to replace $Y(\varepsilon)$ by $(1 - \varepsilon)Y$ and $O_\infty(A_G, C_G, Y(\varepsilon))$ by $(1 - \varepsilon)O_\infty(A_G, C_G, Y)$

Clearly,

$$O_\infty^\varepsilon = \{x_G: f_i(C_G A_G' x_G) \leq 0, t \in Z^+, f_i([H_0 \ 0]x_G) \leq -\varepsilon, i = 1, \dots, s\} \tag{21}$$

Since O_∞^ε is finitely determined, only a finite number of the inequalities are active:

$$O_\infty^\varepsilon = \{x_G: f_i(C_G A_G' x_G) \leq 0, t = 0, \dots, t_i^*, i \in S^*, f_k([H_0 \ 0]x_G) \leq -\varepsilon, k = 1, \dots, s\} \tag{22}$$

The special structure of the set of active inequalities is a consequence of Theorem 2.4 in Reference 10. The index sets S^* and the t_i^* depend on ε . Usually, the number of inequalities in (22) increases as ε becomes smaller. Thus, in choosing ε , a compromise must be made between the accuracy of the approximation and the complexity of O_∞^ε . Some of the active constraints in (22) may be redundant in the sense that they can be removed from (22) without changing O_∞^ε . See Remark 2.2 in Reference 10. Such removal simplifies the computation of $K(r, x_G)$ which is described below.

By Remark 3.1, O_∞^ε has the same properties as $O_\infty(A_G, C_G, Y)$. Thus it may replace $O_\infty(A_G, C_G, Y)$ in the descriptions of Section 2. In particular, $K: R^m \times O_\infty^\varepsilon \rightarrow [0, 1]$ is defined by

$$K(r, x_G) = \max\{\alpha \in [0, 1]: A_G x_G + B_G \alpha(r - [I \ 0]x_G) \in O_\infty^\varepsilon\} \tag{23}$$

Properties of the resulting reference governor are described in the next section.

Expressions (22) and (23) provide the recipe for evaluating $K(r, x_G)$:

$$\alpha_{ij} = \max\{\alpha \in [0, 1]: f_i(C_G A_G'(A_G x_G + B_G \alpha(r - [I \ 0]x_G))) \leq 0\} \tag{24}$$

$$\alpha_i = \max\{\alpha \in [0, 1]: f_i([H_0 \ 0]'A_G x_G + B_G \alpha(r - [I \ 0]x_G)) \leq -\varepsilon\} \tag{25}$$

$$K(r, x_G) = \min\{\alpha_{ij}; i \in S^*, j = 0, \dots, t_i^* \cup \{\alpha_i; i = 1, \dots, s\}\} \tag{26}$$

By using (11), (15), (16) and $A_G' = UA_G'U^{-1}$, (24) and (25) can be simplified:

$$\alpha_{ij} = \max\{\alpha \in [0, 1]: f_i(CA^{j+1}x + D_{j+1}w + \alpha D_j(r - w)) \leq 0\}, x_G = \begin{bmatrix} w \\ x \end{bmatrix} \tag{27}$$

$$\alpha_i = \max\{\alpha \in [0, 1]: f_i(H_0 w + \alpha H_0(r - w)) \leq -\varepsilon\} \tag{28}$$

$$D_j = D + C(I - A')(I - A)^{-1}B \tag{29}$$

Related expressions provide a simple numerical test for $x_G \in O_\infty^\varepsilon$:

$$x_G \in O_\infty^\varepsilon \Leftrightarrow f_i(CA^j x + D_j w) \leq 0, i \in S^*, j = 0, \dots, t_i^*, f_k(H_0 w) \leq -\varepsilon, k = 1, \dots, s \tag{30}$$

Actually, (27) and (28) are scalar root-finding problems. In most practical applications, where the functions f_i are simple, they have explicit solutions. This is certainly true when Y is a polytope. Then $f_i(y) = \gamma_i^T y - 1, i = 1, \dots, s$, and

$$\begin{aligned} \alpha_{ij} &= \min\{1, (\gamma_i^T D_j(r - w))^{-1}(1 - \gamma_i^T CA^j x - \gamma_i^T D_{j+1} w)\}, \gamma_i^T D_j(r - w) > 0 \\ &= 1, \gamma_i^T D_j(r - w) \leq 0 \end{aligned} \tag{31}$$

$$\begin{aligned} \alpha &= \min\{1, (\gamma_i^T H_0(r - w))^{-1}(1 - \varepsilon - \gamma_i^T H_0 w)\}, \gamma_i^T H_0(r - w) > 0 \\ &= 1, \gamma_i^T H_0(r - w) \leq 0 \end{aligned} \tag{32}$$

Similarly, if f_i is a quadratic form, α_{ij} may be expressed in terms of the real roots of a quadratic equation in α .

A few comments are in order. While the evaluation of (26)–(29) is usually straightforward, $K(r, x_G)$ is a complex, piecewise-defined, nonlinear function. Expressions (31) and (32) illustrate the truth of this statement. When Y is a polytope, $K: R^m \times O_\infty^\epsilon \rightarrow [0, 1]$ is continuous. The proof, based on (26), (31) and (32), is not difficult. The convexity of the polytope does not imply the convexity of K . For real-time implementation of (9) the computation time for $K(r, x_G)$ must be less than T seconds, where T is the sampling interval.

4. RESPONSE PROPERTIES OF THE DYNAMIC REFERENCE GOVERNOR

Our main results on the response properties of the dynamic reference governor are summarized in two theorems. They are stated here, together with remarks pertaining to the nature and significance of their results. Proofs appear in the next section.

The first theorem describes response properties of the reference governor which are independent of the way in which $K(r(t), x_G(t))$ is determined. Specifically, it concerns the system

$$w(t + 1) = w(t) + \kappa(t)(r(t) - w(t)) \tag{33}$$

with $\kappa(t) \in [0, 1]$.

Theorem 4.1

Assume $r \in l_\infty^m$ and $\kappa: Z^+ \rightarrow [0, 1]$. Then, the following results hold. (i) $w(t) \in \text{co}\{w(0), r(0), \dots, r(t - 1)\}$. (ii) $\|w\|_\infty \leq \max\{\|w(0)\|, \|r\|_\infty\}$. (iii) If $r(t)$ has a limit as $t \rightarrow \infty$, then so does $w(t)$.

All of three conclusions depend crucially on the assumption $\kappa(t) \in [0, 1]$. The first two are simple consequences of it. The third, which is used in the proof of the next theorem, is less evident.

Remark 4.1. Result (i) has an interesting application. Suppose $r(t) \equiv r_0 \in R^m$. Then $w(t)$, $t \in Z^+$, belongs to the line segment $\text{co}\{w(0), r_0\}$. Furthermore, because $\kappa(t) \in [0, 1]$, $w(t)$ progresses monotonically away from $w(0)$ toward r_0 . If $\kappa(\bar{t}) = 1$, then $w(t) = r_0$, $t > \bar{t}$.

Remark 4.2. For $w(0) = 0$, (ii) shows that the reference governor is uniformly bounded-input, bounded-output stable with gain bound equal to 1. In Reference 16, bounded-input, bounded-output stability is noted, but no gain bound is given.

Define

$$W_0^\epsilon = \{w: H_0 w \in Y(\epsilon)\} = \{w: f_i(H_0 w) \leq -\epsilon, i = 1, \dots, s\} \tag{34}$$

the set of constant inputs to (2) whose corresponding equilibrium solutions satisfy the constraint $y(t) \in Y(\epsilon)$. From Remark 3.1, W_0^ϵ is closed and $0 \in \text{int } W_0^\epsilon \neq \emptyset$. W_0^ϵ is bounded if and only if $\text{rank } H_0 = m$. If $Y(\epsilon) = (1 - \epsilon)Y$, $W_0^\epsilon = (1 - \epsilon)W_0$.

Theorem 4.2

Consider the system (2), (9) with K defined by (23). Assume: $0 < \epsilon < \min\{-f_i(0): i = 1, \dots, s\}$,

$x_G(0) \in O_\infty^\varepsilon$, and $r \in l_\infty^m$. Then the following results hold. (i) For all $t \in Z^+$: $w(t) \in W_0^\varepsilon$, $x_G(t) \in O_\infty^\varepsilon$ and $y(t) \in Y$. (ii) Suppose $r(t) \rightarrow r_0$ and there exists $\hat{t} \in Z^+$ such that $r(t) \in W_0^\varepsilon$ for all $t \geq \hat{t}$; then, there exists $\hat{t} \in Z^+$ such that for all $t \geq \hat{t}$, $K(r(t), x_G(t)) = 1$. (iii) Suppose $r(t) \rightarrow r_0$ and $r_0 \notin W_0^\varepsilon$; then, $K(r(t), x_G(t)) \rightarrow 0$ and $w(t) \rightarrow w^* \in bdW_0^\varepsilon$. (iv) Suppose $r(t) \equiv r_0 \notin W_0^\varepsilon$; then, there exists $\hat{t} \in Z^+$ such that for all $t \geq \hat{t}$, $K(r(t), x_G(t)) = 0$ and $w(t) = w^* = w(0) + \alpha^*(r_0 - w(0))$, where $\alpha^* = \max\{\alpha \in [0, 1]: w(0) + \alpha(r_0 - w(0)) \in W_0^\varepsilon\}$.

Remark 4.3. Conclusions of the theorem become stronger as $\varepsilon > 0$ becomes smaller. Then, W_0^ε becomes larger and better approximates W_0 . Usually, the number of inequalities needed to specify O_∞^ε increases as $\varepsilon \rightarrow 0$ and a compromise on the choice of ε is necessary.

Remark 4.4. A natural choice for the initial state of the reference governor is $w(0) = 0$. Then, (11) and (21) show $x_G(0) \in O_\infty^\varepsilon$ if and only if $x(0) \in O_\infty(A, C, Y)$.

Remark 4.5. Suppose W_0^ε is compact ($\text{rank } H_0 = m$). Then, $\|w\|_\infty \leq c_w$, where $c_w > 0$ is independent of $x_G(0)$ and $\|r\|_\infty$. Thus, for $w(0) = 0$, the l_∞^m response of the reference governor has a 'saturation' character: $\|w\|_\infty \leq \|r\|_\infty$, $\|w\|_\infty \leq c_w$.

Remark 4.6. Suppose the closed-loop system in Figure 1 is modelled by (2) and (3) and is combined with the reference governor. Then, because of (i), the saturation constraint is inactive and the closed-loop system functions as a linear asymptotically stable system. Furthermore, Remarks 4.2 and 4.5 can be extended to any outputs of the closed-loop system, such as $z(t)$. For example, the zero-state response satisfies $\|z\|_\infty \leq k_z \|r\|_\infty$ and $\|z\|_\infty \leq c_z$ where k_z and c_z are positive constants. It should be emphasized that as an error measure $z(t)$ relates plant variables to $w(t)$, not $r(t)$. Error measures based on $r(t)$ are large when $r(t) - w(t)$ is large.

Remark 4.7. The conclusion in (ii) is strong, considering that it applies to a complex nonlinear system. Since $w(t) = r(t-1)$ for $t > \hat{t}$, the ultimate response of the overall system is close to that of the linear controlled process. Certainly, (ii) implies convergence to the equilibrium corresponding to r_0 : $w(t) \rightarrow r_0$ and $x(t) \rightarrow x_0 = (I - A)^{-1} B r_0$. Corresponding prior results^{10,15,16} on the error and reference governors are much weaker. They depend on the assumption that the governors are always inactive, as in Remark 2.1. The results in (ii) are global in the sense they apply for all convergent sequences, $\{r(t): t \in Z^+\}$, whose elements ultimately belong to W_0^ε and for all $x_G(0) \in O_\infty^\varepsilon$. By (21) it follows that $w(0) = 0$ and $x(0) \in O_\infty(A, C, Y)$ imply $x_G(0) \in O_\infty^\varepsilon$.

Remark 4.8. The assumption in (ii) is about as weak as it can be. Clearly, the result in (ii) cannot hold if $r_0 \notin W_0^\varepsilon$, for then result (iii) would be violated. Note that the assumption in (ii) is satisfied if $r(t) \equiv r_0 \in W_0^\varepsilon$ or if $r(t) \rightarrow r_0 \in \text{int}W_0^\varepsilon$. Thus, for $r_0 \in W_0^\varepsilon$, the only case not handled by (ii) is the one in which $r(t) \rightarrow r_0 \in bdW_0^\varepsilon$ and $r(t) \notin W_0^\varepsilon$ for infinitely many $t \in Z^+$. For this case, it can be shown that $w(t) \rightarrow w^* \in bdW_0^\varepsilon$, but there is no guarantee that $w^* = r_0$.

Remark 4.9. Result (iii) shows that convergence to an equilibrium solution occurs even though $r_0 \notin W_0^\varepsilon$. The stronger result in (iv) requires $r(t) \equiv r_0$.

Remark 4.10. The theorem applies under our overriding assumptions (A1), (A2) and (A3). Thus, Y need not be convex.

5. PROOFS OF THEOREMS

Proof of Theorem 4.1. By (33) and $\kappa(t) \in [0, 1]$, $w(t+1) \in \text{co}\{r(t), w(t)\}$. Repeated application of this property proves (i). From (i), $|w(t)| \leq \max\{|w(0)|, |r(0)|, \dots, |r(t-1)|\}$, which proves (ii). We prove (iii) for $m=1$. This result, applied to the components of $w(t)$, proves (iii) for $m>1$.

Suppose $r(t) \rightarrow r_0 \in R$. For $\varepsilon > 0$, define $\hat{t}(\varepsilon)$ so that $|r(t) - r_0| \leq \varepsilon$ for all $t \geq \hat{t}(\varepsilon)$. Given $\varepsilon > 0$, there are two possible outcomes: there exists $\tilde{t}(\varepsilon) \geq \hat{t}(\varepsilon)$ such that $|w(\tilde{t}(\varepsilon)) - r_0| \leq \varepsilon$; $|w(t) - r_0| > \varepsilon$ for all $t \geq \tilde{t}(\varepsilon)$. Suppose the second outcome holds and $w(\tilde{t}(\varepsilon)) > r_0 + \varepsilon$. Then because $\kappa(\tilde{t}(\varepsilon)) \in [0, 1]$ and $r(\tilde{t}(\varepsilon)) \leq r_0 + \varepsilon$ it follows that $r_0 + \varepsilon < w(\tilde{t}(\varepsilon) + 1) \leq w(\tilde{t}(\varepsilon))$. Repeating this argument shows $w(t)$, $t \geq \tilde{t}(\varepsilon)$, is monotone and hence $w(t) \rightarrow w^* \geq r_0 + \varepsilon$. Similarly, $w(\tilde{t}(\varepsilon)) < r_0 - \varepsilon$ implies $w(t) \rightarrow w^* \leq r_0 - \varepsilon$. Thus, we have conclusion (1): $w(t) \rightarrow w^*$ and $|w^* - r_0| \geq \varepsilon$. If the first outcome holds, (33), $\kappa(\tilde{t}(\varepsilon)) \in [0, 1]$ and $|r(\tilde{t}(\varepsilon)) - r_0| \leq \varepsilon$ imply $|w(\tilde{t}(\varepsilon) + 1) - r_0| \leq \varepsilon$. Repeating this argument leads to conclusion (2): there exists $\tilde{t}(\varepsilon)$ such that $|w(t) - r_0| \leq \varepsilon$ for all $t \geq \tilde{t}(\varepsilon)$. Thus, for each $\varepsilon > 0$ there are the two possible conclusions. If (1) is satisfied for some $\varepsilon > 0$, (iii) is proved and $w^* \neq r_0$. If (1) is not satisfied for all $\varepsilon > 0$, then (2) implies $w(t) \rightarrow r_0$.

Proof of Theorem 4.2. By (10) and (23), $x_G(t) \in O_\infty^\varepsilon$ implies $x_G(t+1) \in O_\infty^\varepsilon$. Thus, $x_G(t) \in O_\infty^\varepsilon$, $t \in Z^+$. The remaining conclusions in (i) follow from (21), (4) and (34).

Part (iii) of Theorem 4.1 and the asymptotic stability of A imply $x_G(t) \rightarrow x_G^* = [(w^*)^\top (x^*)^\top]^\top$, where $x^* = (I - A)^{-1} B w^*$ and, by (i) and the closure of W_0^ε , $w^* \in W_0^\varepsilon$. From (13), $C_G A_G^j x_G^* = H_0 w^*$. Thus,

$$f_i(C_G A_G^j x_G(t)) = f_i(y_j(t)), \quad y_j(t) = H_0 w^* + C_G A_G^j (x_G(t) - x_G^*) \quad (35)$$

Clearly, $y_j(t) \in Y$, $t \in Z^+$, and $y_j(t) \rightarrow H_0 w^* \in Y(\varepsilon)$. Since the f_i are continuous and $f_i(H_0 w^*) \leq -\varepsilon$, $i \in \{1, \dots, s\}$, there exists $\tilde{t} \in Z^+$, which is dependent on $\{r(t) : t \in Z^+\}$, $x_G(0)$ and ε , such that $f_i(C_G A_G^j x_G(t)) < 0$ for all $t \geq \tilde{t}$. Suppose $t \geq \tilde{t}$ and let $x_G = x_G(t)$ and $r = r(t)$ in (24). Then from $f_i(x_G(t+1)) < 0$, (10) and (26), it follows that $K(r(t), x_G(t)) = \alpha_{ij}$ if and only if $\alpha_{ij} = 1$. Thus for $t \geq \tilde{t}$, the α_{ij} contribute nothing to (26) and it may be replaced by

$$K(r, x_G) = \min\{\alpha_i : i \in \{1, \dots, s\}\} \quad (36)$$

Now consider the proof of (ii). For $t \geq \hat{t} = \max\{\tilde{t}, \hat{t}\}$, $r(t) \in W_0^\varepsilon$ and (5.2) applies. Let $w = w(t)$ and $r = r(t)$ in (28). Then, $\alpha_i = 1$, $i \in \{1, \dots, s\}$ and $K(r(t), x_G(t)) = 1$ for all $t \geq \hat{t}$.

Theorem 4.1, (i) and Remark 4.4 imply $w(t) \rightarrow w^* \in W_0^\varepsilon$. Thus, for $r_0 \notin W_0^\varepsilon$, $w^* \neq r_0$. Letting $t \rightarrow \infty$ in (9) proves the first result in (iii). Suppose, contrary to the second result in (iii), that $w^* \in \text{int } W_0^\varepsilon$. Then, by (13) and (34), $f_i(C_G A_G^j x_G^*) = f_i(H_0 w^*) < -\varepsilon$, $i \in \{1, \dots, s\}$. Because of the strict inequalities, there exist $\hat{\varepsilon} > \varepsilon$ and $\hat{t} \in Z^+$ such that for all $t \geq \hat{t}$, $f_i(C_G A_G^j x_G(t)) \leq -\hat{\varepsilon}$ and $f_i([H_0 \ 0]x_G(t)) \leq -\hat{\varepsilon}$. Let $x_G = x_G(t)$ and $r = r(t)$ in (24) and (25). Since $\hat{\varepsilon} > \varepsilon$ and the f_i are continuous, there exists an $\hat{\alpha} > 0$ such that $K(r(t), x_G(t)) > \hat{\alpha}$ for all $t \geq \hat{t}$. This, contradicts $K(r(t), x_G(t)) \rightarrow 0$ and the proof of (iii) is complete.

Let w^* be given by the expression in (iv). It is obvious from Remark 4.2 that for all $t \in Z^+$, $w(t) \in \text{co}\{w(0), w^*\} \subset \text{co}\{w(0), r_0\}$. Again, for $t \geq \hat{t}$, $K(r(t), x_G(t))$ is determined by (36). Hence, by (28), $w(\hat{t}+1) = w^*$. Since further progress toward $r_0 \neq w^*$ is impossible, $K(r(t), x_G(t)) = 0$, $t \geq \hat{t} = \hat{t} + 1$. \square

Remark 5.1. The assumption $\varepsilon > 0$ is used in the proof of (ii), (iii) and (iv) and cannot be relaxed even if $O_\infty(A_G, C_G, Y)$ is finitely determined.

6. ALTERNATIVE DYNAMIC GOVERNORS

A natural alternative to (9) is the dynamic governor introduced by Tan²⁶:

$$w(t+1) = \lambda w(t) + K(r(t), x_G(t))(r(t) - \lambda w(t)) \quad (37)$$

where $0 < \lambda < 1$ and λI replaces I in (10) and in the upper left block of A_G in (11). As in (9) $K(r(t), x_G(t)) = 1$ implies $w(t+1) = r(t)$. Since A_G is asymptotically stable, $O_\infty(A_G, C_G, Y)$ is finitely determined and the approximation discussed in Section 3 is not needed. Unfortunately, this advantage is overcome by the same disadvantages which haunt the static governor: a reduced set of admissible equilibrium solutions²⁶ and the possibility of steady-state oscillations. The underlying difficulty is in Theorem 4.1; simple examples show that it fails unless $\lambda = 1$. If $1 - \lambda$ is small, simulation studies²⁶ show that the behaviour of (37) is often acceptable and quite similar to (9). However, (37) appears to have no real advantage over (9).

Another alternative to (9) is to replace the scalar valued K by a diagonal matrix and then determine its diagonal elements, K^i , $i = 1, \dots, m$, by solving the following, on-line optimization problem: minimize $|w(t+1) - r(t)|^2 = |(I - K)(r(t) - w(t))|^2$ subject to the constraints $0 \leq K^i \leq 1$, $i = 1, \dots, m$ and $A_G x_G(t) + B_G K(r(t) - [I \ 0]x_G(t)) \in O_\infty^\varepsilon(A_G, C_G, Y)$. For this alternative, it turns out that parts (ii) and (iii) of Theorem 4.1 and parts (i), (ii) and (iii) of Theorem 4.2 remain valid. The scheme has not been tried. Conceivably, its greater flexibility could offer superior performance.

7. A HELICOPTER WITH CONTROL SATURATION

The controlled process is a tenth-order closed-loop system whose plant is a fourth-order linear model of the vertical dynamics of a helicopter.²⁶ State variables are horizontal and vertical velocity (knots), pitch rate (degrees/s), and pitch angle (degrees). Inputs to the plant are collective pitch, u_1 , and longitudinal cyclic pitch, u_2 , measured in degrees. Plant outputs to be controlled are the horizontal and vertical velocities. Zero-order holds with a sampling period, $T = 0.1$ s., are used at each of the plant inputs. The resulting discrete-time model has characteristic roots at: 0.8128, 0.9770, $1.0276 \pm j0.0265$. Clearly, the helicopter is open-loop unstable. Inputs to the controller are the plant outputs and the reference commands. The controller is sixth-order, consisting of an observer-based, stabilizing feedback for the plant and integral control to eliminate steady-state tracking errors. Combining plant and controller dynamics gives a closed-loop system of the form (2)–(3) where $y = u$. The characteristic roots of A are: 0.3331, 0.4282, 0.5959, 0.9502, $0.6871 \pm j0.0390$, $0.5864 \pm j0.4091$, $0.9705 \pm j0.0821$.

The two plant inputs saturate at ± 10 degrees. Thus, $f_i(y) = y_i - 10$, $i = 1, 2$ and $f_i(y) = -y_i - 10$, $i = 3, 4$. O_∞^ε is determined by solving a sequence of linear programming problems.

Table I. The complexity indices for O_∞^ε

ε	t_1^*	t_2^*
0.01	135	159
0.05	88	117
0.1	80	106
0.2	74	97
0.3	56	93

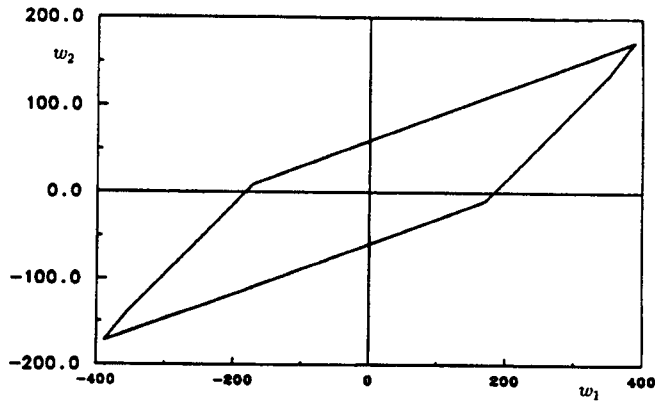


Figure 3. W_0^+ for helicopter example

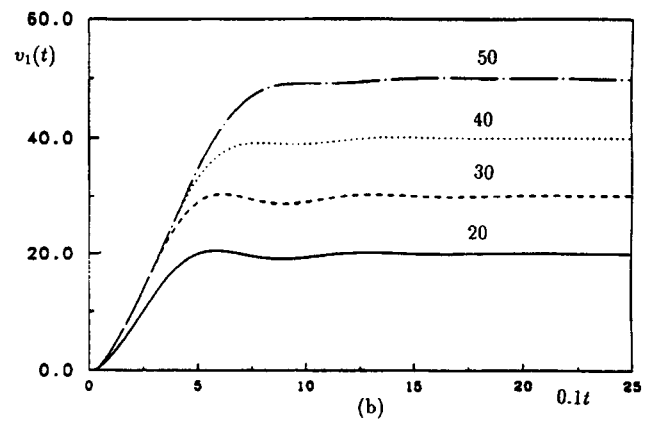
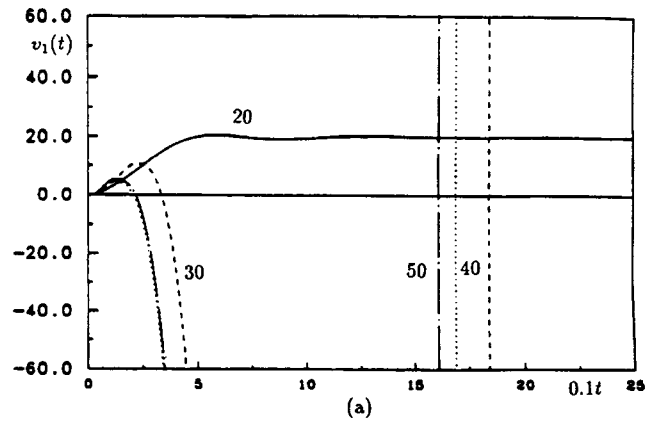


Figure 4. Horizontal velocity versus time in seconds for several values of r_{01} (a) No reference governor. (b) With reference governor

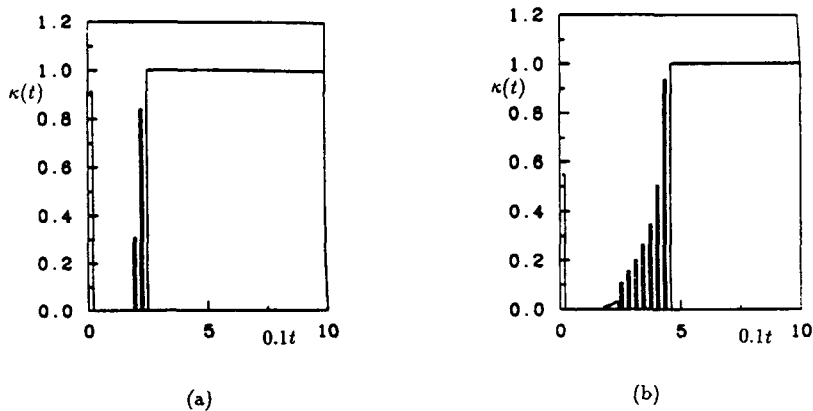


Figure 5. Reference governor gain versus time in seconds. (a) $r_{01} = 30$. (b) $r_{01} = 50$

The dependence of $t_1^* = t_3^*$ and $t_2^* = t_4^*$ on ε is shown in Table I. Large values for t_1^* and t_2^* are not too surprising, since $n = 10$ and several eigenvalues are close to the unit circle. In addressing the compromise of Remark 4.3, we choose $\varepsilon = 0.2$. The corresponding set of admissible constant reference commands, W_0^ε , is shown in Figure 3. Since it is quite large, the reduction in equilibrium feasibility is not of great concern. Note that $W_0^\varepsilon = (1 - \varepsilon)W_0$, so the effect of changing ε is a simple scale change.

$K(r, x_G)$ is determined by (26), (31) and (32). Because $\gamma_3 = -\gamma_1$ and $\gamma_4 = -\gamma_2$, the computational load in (31) and (32) can be reduced by almost a factor of two. On reasonable machines the actual computational time is much less than $T = 0.1$: for an Apollo DN 400t it is less than 10 milliseconds, for an IBM 6000 it is less than 2 milliseconds. This is in sharp contrast to the extremely large times needed for continuous-time error and reference governors.^{15,16}

Response results are shown in Figure 4. The initial state of the controlled process is zero and the input is a step command to the horizontal velocity: $r_1(t) \equiv r_{01}$, $r_2(t) \equiv 0$. The corresponding horizontal output velocity, $v_1(t)$, is plotted. Figure 4 (a) shows what happens when the reference governor is not used and the control variables are allowed to saturate. For $r_{01} = 20$, the system is just below saturation and it behaves as a linear system. For $r_{01} = 30, 40, 50$, saturation is reached; the instability of the open-loop plant comes into play and the response diverges. Figure 4(b) shows the corresponding plots when the reference governor, with initial state zero, is added. The response is excellent: well-behaved, only slowing a bit as r_{01} increases. When $r_{01} = 20$, $\kappa(t) = K(r(t), x_G(t)) \equiv 1$. Its behaviour for $r_{01} = 30, 50$ is shown in Figure 5. As predicted by Theorem 4.2, $\kappa(t) = 1$ for $t \geq \hat{t}$, where \hat{t} depends on r_{01} . While $w_1(t)$ is not shown, its behaviour may be inferred easily from Figure 5: if $r_{01} = 50$, $w_1(0)$ jumps to 34 and stays at that level until $t = 18$ (time = 1.8 seconds) when it begins a slow increase toward 50, arriving there when $t = 46$ and $\kappa(t) = 1$. Since $r_{02} = 0$, $w_2(t) \equiv 0$.

8. A SYSTEM WITH AN OUTPUT CONSTRAINT

In this example the reference governor is used to reduce overshoot in the step response of a single-input, single-output closed-loop system. The open-loop transfer function models a proportional-integral compensator and a plant which has gain $c = 1$, a four-step transport delay and a first-order lag: $G^*(z) = 0.15c(1 + 1.7(z-1)^{-1})z^{-4}(1 - e^{-1})(z - e^{-1})^{-1}$. The closed-loop system, with transfer function $G^*(z)(1 + G^*(z))^{-1}$, has an output $y(t)$ and is represented by

(2)–(3). Compensator parameters were chosen subject to a constraint on the velocity constant ($\lim(z - 1)G^*(z), z \rightarrow 1$). Satisfying this constraint makes it impossible to obtain a well damped response to a step change in the reference command. See the oscillatory response in Figure 6.

Suppose physical considerations demand $y(t) \in Y = [y_{\min}, y_{\max}]$. This output constraint can be assured by the dynamic reference governor. It turns out that O_{∞}^{ε} is parameterized in terms of

$$\bar{O}_{\infty}^{\varepsilon} = O_{\infty}(A_G, \hat{C}_G, [-1 + \varepsilon, 1 - \varepsilon] \times [-1, 1]) \tag{38}$$

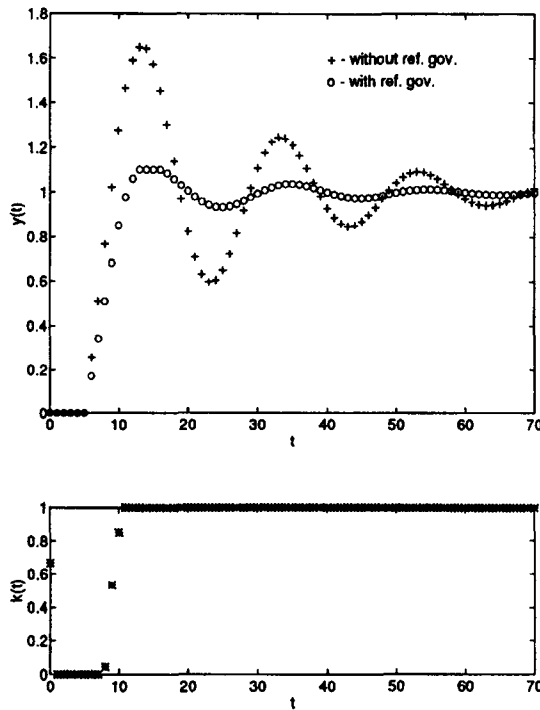


Figure 6. Effect of reference governor on system with $r(t) \equiv 1$ and output constraint $y(t) \in [-0.02, 1.1]$

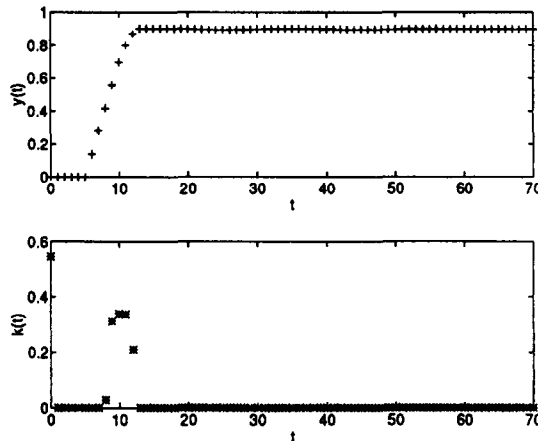


Figure 7. Response of reference-governor system with $r(t) \equiv 1$ and output constraint $y(t) \in [-0.02, 0.9]$

Indeed, straightforward manipulations based on (13), (17), (20) and Remark 3.2 show

$$O_{\infty}^{\varepsilon} = 0.5(y_{\min} + y_{\max})H_0^{-1} \left\{ \begin{bmatrix} 1 \\ (I - A)^{-1}B \end{bmatrix} \right\} + 0.5(y_{\max} - y_{\min})\bar{O}_{\infty}^{\varepsilon} \quad (39)$$

Once $\bar{O}_{\infty}^{\varepsilon}$ is determined, (23) applied to (39) gives (26) and formulas similar to (31) and (32). The advantage of these steps is that y_{\min} and y_{\max} appear explicitly in the formulas. Since the additional time needed for computing $K(r, x_G)$ in terms of these parameters is small, they can be changed on-line as process conditions may require. For the results which follow, $\varepsilon = 0.01$ and the determination of $\bar{O}_{\infty}^{\varepsilon}$ gives $t_1^* = t_2^* = 23$. Also, $w(0) = 0$ and $x(0) = 0$.

The step response with the governor and $Y = [-0.02, 1.1]$ is shown in Figure 6. For $t \geq 11$, $\kappa(t) = 1$. Thus, in a remarkably short time the overshoot is reduced to the specified 10 per cent and the system enters its linear regime. Figure 7 shows what happens when $Y = [-0.02, 0.9]$. Since $r(t) \equiv 1 \notin W_0^{\varepsilon} = [-0.01, 0.89]$, part (iv) of Theorem 4.2 applies and for $t \geq 13$, $y(t) = 0.89$ and $\kappa(t) = 0$. Figure 8 illustrates, with $Y = [-0.02, 1.1]$, the response properties for a very large oscillatory input which converges to 1. Part (ii) of Theorem 4.2 again applies, but 45 time steps are required before $\kappa(t) \equiv 1$.

If the state of the controlled process is not available, or there are errors in the modelling of the controlled process, our results do not apply. A brief empirical investigation of these issues is illustrated in Figure 9. It is assumed that $Y = [-0.05, 1.3]$ and $r(t) \equiv 1$. In the implementation of the reference governor, the state of the controlled process is replaced by the state of a *model* of the controlled process whose input is $w(t)$. If the description of the controlled process and its model are precisely the same and their initial conditions are identical, the reference governor should function as predicted by our theory. This is the case for $c = 1$. The remaining responses in Figure 9 show what happens in the presence of modelling errors. The value of c remains equal to 1 in the model but it takes on the values 0.8 and 1.2 in the controlled process. Both responses are certainly much better than the response with no reference governor (cf. Figure 6). However, the limit $y_{\max} = 1.3$ is violated when $c = 1.2$. The situation suggests several expedient fixes. One is to reduce y_{\max} below the desired value of 1.3. Another is to carry out the nominal design of the reference governor for $c = 1.2$ rather than $c = 1$.

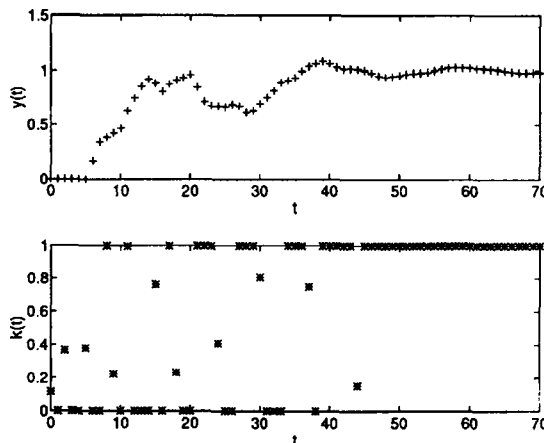


Figure 8. Response of reference-governor system with $r(t) = 1 + 5e^{-0.1t} \cos t$ and output constraint $y(t) \in [-0.02, 1.1]$

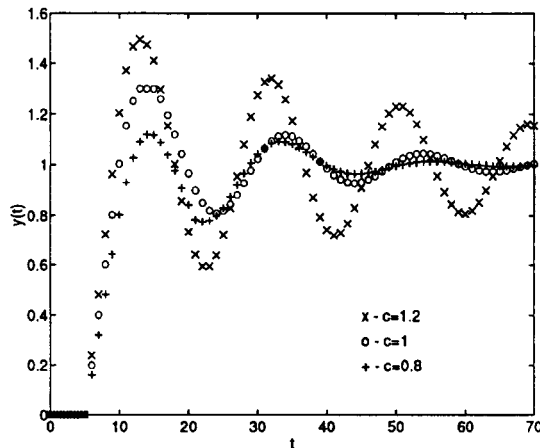


Figure 9. Response of model-based, reference-governor system with $r(t) \equiv 1$ and output constraint $y(t) \in [-0.05, 1.3]$

9. CONCLUSION

Discrete-time reference governors having the configuration shown in Figure 2 have been studied. They enforce the constraint, $y(t) \in Y$, $t \in Z^+$, and thereby avoid undesirable effects of constraint violation such as sluggish response or instability due to actuator saturation. The approach has the distinct advantage that it can be applied to existing, well-designed linear feedback systems. Thus, acceptable large-signal performance can be obtained without sacrificing the small-signal performance of the linear design.

Two governors, one static and the other dynamic, have been considered in detail. The main emphasis is on the dynamic governor because it has superior response characteristics. It is defined by (9), where K is determined by the constraint-predictive property of a maximal output admissible set. For typical constraint sets, K is an easily computed nonlinear function of system state and the reference command. Ease of implementation is a direct consequence of the finite determinability of O_ε^* . Response properties are contained in the theorems and remarks of Section 4. The key result is part (ii) of Theorem 4.2: convergence of the input command to an equilibrium-admissible input implies the existence of $\hat{t} \in Z^+$ such that $w(t) = r(t-1)$ for all $t \geq \hat{t}$. This result is global in character — it applies when there are large reference commands and initial conditions.

Examples confirm the predicted behaviour of the dynamic governor scheme and show that real-time implementations are feasible for systems of moderately high order. State-control constraints other than those imposed by actuator saturation may be treated.

Open questions remain and are being investigated. They include the effects of modelling errors, incomplete measurements of plant state and the treatment of disturbance inputs.

ACKNOWLEDGEMENT

The second author was supported in his work by a François-Xavier Bagnoud doctoral fellowship.

REFERENCES

1. Åström, K. J., and L. Rundqwist, 'Integrator windup and how to avoid it', *Proc. 1989 American Contr. Conf.*, Pittsburgh, PA, 1992, pp. 1693–1698.

2. Åström, K. J., and B. Wittenmark, *Computer Controlled Systems: Theory and Design*, 2nd edn, Prentice Hall, Englewood Cliffs, NJ, 1990.
3. Bitsoris, G., 'Positively invariant polyhedral sets of discrete-time linear systems', *Int. J. Contr.*, **47**, 1713–1726 (1988).
4. Bitsoris, G., 'On the positive invariance of polyhedral sets for discrete-time systems', *Syst. and Contr. Letters.*, **11**, 243–248 (1988).
5. Blanchini, F., 'Feedback control for linear time-invariant systems with state and control bounds in the presence of disturbances', *IEEE Trans. Automat. Contr.*, **AC-35**, 1231–1234 (1990).
6. Campo, F., M. Morari and C. N. Nett, 'Multivariable anti-windup and bumpless transfer: a general theory', *Proc. 1989 American Contr. Conf.*, Pittsburgh, PA, 1989, pp. 1706–1711.
7. Desoer, C. A., and J. Wing, 'A minimal time discrete system', *IRE Trans. Automat. Contr.*, **AC-6**, 111–125 (1961).
8. Doyle, J. C., R. S. Smith and D. F. Enns, 'Control of plants with input saturation nonlinearities', *Proc. 1987 American Contr. Conf.*, Minneapolis, MN, 1987, pp. 1034–1039.
9. Garcia, C. E., D. M. Prett and M. Morari, 'Model predictive control and practice — a survey', *Automatica*, **25**, 335–348 (1989).
10. Gilbert, E. G., and K. T. Tan, 'Linear systems with state and control constraints: the theory and application of maximal output admissible sets', *IEEE Trans. Automat. Contr.*, **AC-36**, 1008–1020 (1991).
11. Gutman, P.-O., and P. Hagander, 'A new design of constrained controllers for linear systems', *IEEE Trans. Automat. Contr.*, **AC-30**, 22–33 (1985).
12. Gutman, P.-O., 'A linear programming regulator applied to hydroelectric reservoir control', *Automatica*, **22**, 533–541 (1986).
13. Gutman, P.-O., and M. Cwikel, 'An algorithm to find the maximal state constraint sets for discrete-time linear dynamical systems with bounded controls and states', *IEEE Trans. Automat. Contr.*, **AC-32**, 251–254 (1987).
14. Hennes, J. C., and J. P. Beziat, 'A class of invariant regulators for the discrete-time linear regulation problem', *Automatica*, **27**, 549–554 (1991).
15. Kapasouris, P., M. Athans and G. Stein, 'Design of feedback control systems for stable plants with saturating actuators', *Proc. IEEE Conf. Decision and Control*, Austin, TX, 1989, pp. 469–479.
16. Kapasouris, P., M. Athans and G. Stein, 'Design of feedback control systems for unstable plants with saturating actuators', *Proc. IFAC Symposium on Nonlinear Control System Design*, Pergamon Press, Oxford, 1990.
17. Keerthi, S. S., and E. G. Gilbert, 'Computation of minimum-time feedback control laws for systems with state-control constraints', *IEEE Trans. Automat. Contr.*, **AC-32**, 432–435 (1987).
18. Keerthi, S. S., and E. G. Gilbert, 'Optimal infinite-horizon feedback laws for a general class of constrained discrete-time systems: Stability and moving-horizon approximations', *J. Opt. Theory and Appl.*, **57**, 265–293 (1988).
19. Mayne, D. Q., and H. Michalska, 'Receding horizon control of nonlinear systems', *IEEE Trans. Automat. Contr.*, **AC-35**, 814–824 (1990).
20. Michalska, H., and D. Q. Mayne, 'Robust receding horizon control of constrained nonlinear systems', *IEEE Trans. Automat. Contr.*, **AC-38**, 1623–1633 (1993).
21. Propoi, A. I., 'Use of LP methods for synthesizing sampled-data automatic systems', *Auto. Remote Contr.*, **24**, 837 (1963).
22. Sontag, E. D., and H. J. Sussmann, 'Nonlinear output feedback design for linear systems with saturating controls', *Proc. IEEE Conf. Decision and Control*, Honolulu, HI, 1990, pp. 3414–3416.
23. Stein, G., 'Respect the unstable', *IEEE Conf. Decision and Control*, Hendrik W. Bode Lecture, Tampa, FL, 1989.
24. Sussmann, H. J., and Y. Yang, 'On the stabilization of multiple integrators by means of bounded feedback controls', *Proc. IEEE Conf. Decision and Control*, Brighton, England, 1990, pp. 70–73.
25. Sussman, H., E. Sontag and Y. Yang, 'A general result on the stabilization of linear systems using bounded controls', *Proc. IEEE Conf. Decision and Control*, San Antonio, TX, 1993, pp. 1802–1807.
26. Tan, K. T., 'Maximal output admissible sets and the nonlinear control of linear discrete-time systems with state and control constraints', Ph.D. dissertation, Aerospace Engineering, University of Michigan, 1991.
27. Teel, A. R., 'Global stabilization and restricted tracking for multiple integrators with bounded controls', *Syst. and Contr. Letters.*, **18**, 165–171 (1992).
28. Wittenmark, B., 'Integrators, nonlinearities, and anti-reset windup for different control structures', *Proc. 1989 American Contr. Conf.*, Pittsburgh, PA, 1989, pp. 1679–1683.