ISA Transactions 43 (2004) 361–376

ISA
TRANSACTIONS®

Flexible-structure control: A strategy for releasing input constraints

Leonardo L. Giovanini*

Industrial Control Centre, University of Strathclyde, Graham Hills Building, 50 George Street, Glasgow G1 1QE, United Kingdom

(Received 26 July 2002; accepted 5 September 2003)

Abstract

In this paper output unreachability under input saturation phenomenon is studied: under a large disturbance or setpoint change, the process output may never reach the set point even when the manipulated variable has driven to saturation. The process output can be brought back to the set point only by activating an auxiliary manipulated variable. A new control structure for designing and implementing a control system capable of solving this problem is proposed by transferring the control from one variable to another and taking into account the different dynamics involved in the system. The control structure, called *flexible-structure control* due to its ability to adapt the control structure to the operating conditons, is a generalization of the *split-range control*. It can be summarized as two controllers connected through a piecewise linear function. This function decides, based on the value of one manipulated variable, when and how the control structure changes. Its parameters control the interaction between both manipulated variables and leave the capability for handling the balance between control quality and other goals to the operator. © 2004 ISA—The Instrumentation, Systems, and Automation Society.

Keywords: Multi-input systems; PID control; cooperative control; Manipulated variable constraint

1. Introduction

Quite frequently process control engineers face problems in which hard constraints restrict the manipulated variables to a finite operating range. The constraints may also come from process constraints intended to avoid damage to the system or the material being proceeded [1]. There are many control techniques to deal with this problem. They can be divided into two categories: *antiwindup compensation* reduces the adverse effects of the constraints on the closed-loop performance [1–4], and *cooperative control schemes* that use a combination of inputs to achieve the reference [5–9].

Antiwindup methods may work well under the nominal conditions under which the controller was designed. However, it is possible that under a large set-point change, load disturbance, or component failure, the manipulated variable will reach its limit while the system output still cannot reach its set point at the steady state. This phenomenon is known as *output unreachability under input constraint* [9]. It is directly associated with the size of the *operability space* of the system. Sometimes this problem is solved by modifying the control structure of the system, but many times the process itself requires substantial changes. On the other hand, cooperative control schemes solve the output-unreachability problem by activating auxiliary manipulated variables, and many times they introduce a degree of optimality in the solution [8]. Nevertheless, manipulated constraints are

*Tel: +44 141 548 2666; fax: 44 141 552 2487. *E-mail address*: leonardo.giovanini@eee.strath.ac.uk

never completely eliminated because these techniques only solve the output-unreachability problem for the steady state. Therefore during the transient the manipulated variables could temporarily reach their limits and the system becomes uncontrollable during this period. Hence the need for finding a way for broadening controlled operation spaces to provide all the available process flexibility while preserving good performances has stimulated the search for new control structures.

This paper proposed a new control structure called *flexible-structure control*, due to its ability to adapt the control structure to the operating conditons. It is a generalization of the *split-range control* [6] and it can be summarized as two controllers connected through a piecewise linear function. This function decides, based on the value of one manipulated variable, when and how the control structure changes. Its parameters control the interaction between both manipulated variables. They can be selected attending not only to the process structure but also following some optimization criteria.

The paper is organized as follow: in Section 2, the requirements that must satisfy the process to apply the proposed control strategy and some examples are presented. In Section 3 the *flexible-structure control* is proposed. The structure can be summarized as two controllers connected with a piecewise-linear function. The parameters of the nonlinear function control the interaction between both inputs and the steady-state values of each manipulated variable. In Section 4 two tuning procedures for the controller are proposed. The first method is based on the PID controller and it is obtained from an IMC parametrization. The second procedure employes more sophisticated controllers. Section 5 addresses the closed-loop stability analysis. In Section 6 some guidelines for the selection of the decision function are presented. Finally, Section 7 presents results obtained from the application of the proposed algorithm to a linear system. Conclusions are presented in Section 8.

2. Process conditions

Figure 1 shows a sketch of the process structure considered in this work. The first special feature to be noted is that the output variable y may be controlled by either u_1 or u_2 through different dy-

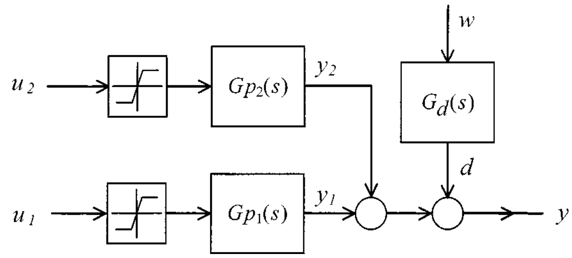

Fig. 1. Basic process structure.

namic elements $Gp_1(s)$ and $Gp_2(s)$. The second feature is that assumed to be u_1 is the *primary manipulated variable* because $Gp_1(s)$ is faster and with smaller time delay than $Gp_2(s)$. A hard constraint might become active at some given extreme values of this variable. If eventually u_1 saturates, u_2 may be used as an *auxiliary manipulated variable* to keep the system under regulation. The process may be also subject to many disturbances. For linear systems, they can be collectively represented by one disturbance d entering the process at the output.

In normal operation, the system is designed so that for any moderate set-point or load disturbance change, the primary manipulated variable u_1 can regulate the process output to achieve a zero output steady-state error working within its working range $[u_{1\,\min},u_{1\,\max}]$, while u_2 is kept unchanged. However, it is possible that under a large set point, load disturbance change, or component failure, there is no $u_1 \in [u_{1\,\min},u_{1\,\max}]$ so that $y(\infty) = r(\infty)$ if u_2 is unchanged, leading to output unreachability.

This problem is frequently found in practice, so it deserves special research. For example, consider the problem of controlling a continuous stirred tank reactor where an irreversible exothermic reaction is carried out at a constant volume employed by Moningred *et al.* [10] to test nonlinear control algorithms. The reactor is cooled by water through a surrounding jacket. The concentration can be controlled either by manipulating the flow rates of the coolant or the process stream. From the process perspective, the use of cooling water is preferred over the process stream. However, the dynamic response of the reactor concentration to a change in coolant flow rate shows a nonlinear behavior. In practice, the nominal flowrates of the process stream is determined in advance, and water is employed as the primary manipulated variable, while the process stream acts as the auxiliary manipulated variable and is normally kept constant.

Another example is the control of the temperature in the reactor [11]. The reactor is cooled by both cooling water flowing through a surrounding jacket and condensing vapor that boils off the reactor in a heat exchanger cooled by a refrigerant. From an energy-saving point of view, the use of cooling water is preferred over the refrigerant due to cost. However, the dynamic response of the plant's temperature to a change in refrigerant flow is faster than a change in cooling water. In practice, the nominal flow rates of both are determined in advance. Water is utilized as the primary manipulated variable, while the refrigerant acts as an auxiliary manipulated variable and it is normally kept constant. If there is a disturbance or setpoint change such that the temperature cannot return to the set point even when the cooling water valve hits its limitation, the flow of refrigerant has to be adjusted to make the temperature reach the desired steady-state value.

Another example can be found in the temperature control of thermical integrated chemical process [12]. A simple configuration of this type of system is given by two heat exchangers in series: one heat exchanger and a service equipment. This arrangement is very common in practice when besides the task of reaching a final temperature target on a process stream there is an extra goal like maximum energy recovery. The heat exchanger is specifically designed for recovering the exceeding energy in the process stream, and the service equipment completes the thermic conditioning through a utility stream [13]. The heat exchanger operates at a constant flowrate that maximizes the energy recovering on nominal conditions and the service is designed to cope with the long term variations on the inlet stream conditions or having the capability of changing stream temperature targets. In the case of a disturbance or set-point change, the steady state may deviate. There could exist a situation in which the effect is so large that the temperature cannot return to the set point even when the service hits its limits. In such a case, the auxiliary manipulated variable, i.e., the flow rate of the heat exchanger, has to be adjusted to make the temperature's steady state reach the desired value.

3. Flexible-structure control

The examples of output unreachability provided in the previous section are only few of many industrial cases. Such cases require changes in the auxiliary manipulated variable. To perform these changes several control strategies have been developed in the past decades. Shinkey [6] proposed the *valve-position control*. This technique uses a maximum signal selector to change u_2 to keep all the actuators from exceeding a preset limit. The algorithm has to wait for the process variable to reach its steady state, then regulate u_2 iteratively until the steady-state error becomes zero. The main drawbacks of the method are that it is time consuming and the results largely depend on the engineer experience.

Another alternative control structure is the *cooperative control*, proposed by Wang *et al.* [9]. Similarly to valve-position control, cooperative control activates u_2 using constant levels which are computed base on a disturbance estimation and output steady-state prediction. The use of these features improves the closed-loop performance by reducing the settling time and simplifying the implementation of this control scheme.

Both techniques solve the output unreachability problem only for the steady-state of the multiple-input single-output (MISO) system with a control structure that remains unchanged. Therefore the transient behavior of the closed-loop system will be poor because the control laws generated by both control strategies do not take account of the dynamic part of Gp_2. Many times the closed-loop response is driven in an open-loop manner during the transient. This situation is significantly important when a fault, like a frozen valve, occurs or during the saturation of the manipulated variables. Hence the above control problem requires an appropriate design that would be able to transfer the control from one variable to another, takes account of the different dynamics involved in the system and, if it is possible, leaves the capability for handling the balance between control quality and other goals to the operator. Hence, under this framework, solving a manipulated constraint problem requires a process engineering approach capable of combining control strategy and process efficiency.

Note that if there is a controller $C_1(s)$ handling $u_1(s)$ to control $y(s)$ at a given set-point value

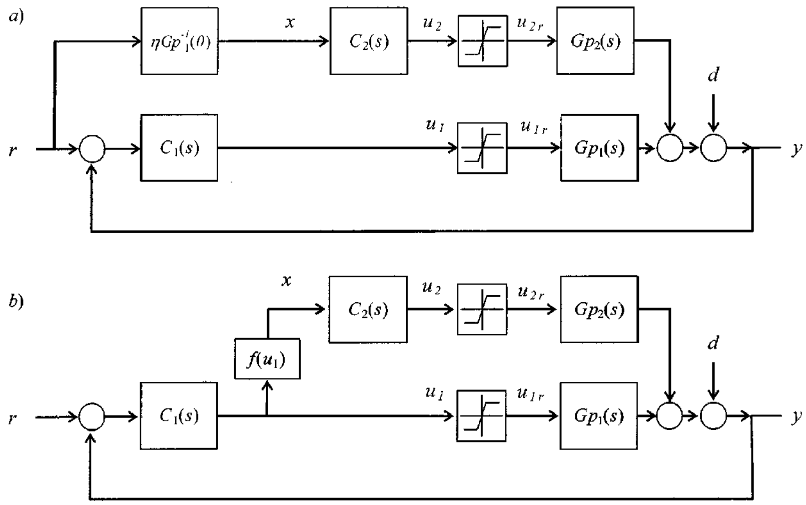

Fig. 2. Block diagram of (a) preventive protection, and (b) reactive protection.

r_0, the stationary value expected for the controlled variable is $Gp_1(0)u_1(0)$, and if an integral mode is present, the manipulated variable goes to $u_1(0)=Gp_1^{-1}(0)r_0$. Let us assume now that just a fraction of this output y can be handled through the manipulated u_1 (this implies that a control constraint is active at a given level), and that an additional capacity can be provided through $u_2(s)$. Then, there are two ways of defining the protection for regulation and operability:

1. *Feedforward or preventive protection*: given a total output steady-state requirement $Gp_1(0)u_1(0)$, a fraction $\eta Gp_1(0)u_1(0)$, $\eta \geqslant 0$, is permanently provided by the process part Gp_2, in order to keep controlled operability.

2. *Feedback or reactive protection*: given the instant manipulated variable $u_1(t)$ at an operating point such that $|u_1(t)|>u_{1\,\max}$, the process part $Gp_2(s)$ provides the complementary output $Gp_1(0)[u_1(t)-u_{1\,\max}]$ to reach the target r_0.

These two types of protections are schematized in Figs. 2(a) and 2(b) respectively, where a second controller $C_2(s)$ is included for better dynamic adjustment of the secondary manipulated variable u_2. It is clear from the block diagrams that in the case of *preventive action*, $C_2(s)$ is a feedforward controller, which does not affect the closed-loop stability, but it cannot protect the system from the effect of nonmeasurable disturbances, uncertainties, and/or faults. If the disturbance $w(s)$, or an estimation $\hat{w}(s)$, and a model of $G_d(s)$ are available, a preventive protection for disturbance can also be included by computing x as follows:

$$x(s)=\eta Gp_1^{-1}(0)r(s)+\eta Gp_1^{-1}(0)G_d(0)w(s).$$

For the second case, Fig. 2(b) shows that the *reactive protection* introduces a new feedback loop that results from combining both controllers, $C_1(s)$ and $C_2(s)$, in a single controller, $C(s)=C_1(s)C_2(s)$. As long as the primary manipulated variable is not saturated, the output is controlled by C_1 and the secondary control loop is no operative because the auxiliar signal $x(t)$ is zero. When $u_1(t)>u_{1\,\max}$, the original control loop remains open and consequently not operative as long as the primary manipulated is saturated. Thus the process part represented by $Gp_2(s)$ is now in charge of the regulation. As soon as the structure of the system changes, due to the saturation of u_1, the structure of the controller is modified from $C_1(s)$ to $C(s)$ following the change of the system and adapting the structure and parameters of the controller to the dynamic of the system. The controller $C(s)$ must include an antiwindup scheme to mitigate the effect of the constraint on u_2, the effect of constraint on u_1 is compensated by the control structure.

The switching element is implemented through a simple nonlinear decision function such that the signal $x(t)$ entering the controller $C_2(s)$ is given by

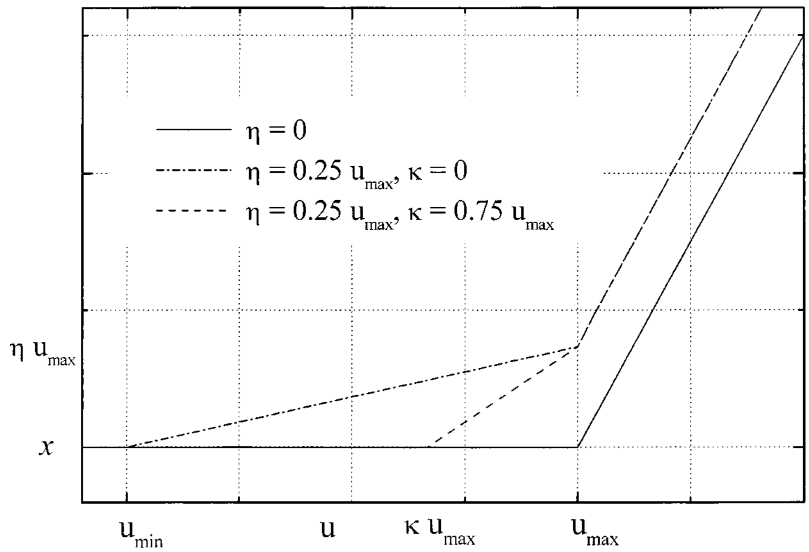

Fig. 3. Plot of the nonlinear switching function for some parameters.

$$f(u_1)=\begin{cases} u_1(t)-u_{1\max} & u_1(t)>u_{1\max}, \\ 0 & u_1(t)\leqslant u_{1\max}. \end{cases} \quad (1)$$

Observe that a similar protection can be developed for a lower constraint but it must actuate on a different process part, let us say $Gp_3(s)$. This means that a third manipulated variable u_3 must be available and become active. In this case the switching function may be written

$$f(u_1)=\begin{cases} 0 & u_1(t)\geqslant u_{1\min}, \\ u_1(t)-u_{1\min} & u_1(t)<u_{1\min}. \end{cases} \quad (2)$$

The switching functions (1) and (2) can be easily implemented using a dead zone of width $u_{1\max}-u_{1\min}$ with a selector connected in series, such that it generates the proper signal for each auxiliary loop.

In spite that a nonlinearity is introduced in the closed loop to transfer the control from Gp_1 to Gp_2, the stability properties of the system remain unaffected because, as was explained previously, there is no interaction between both control loops. When one manipulated variable is active, said that the output is controlled with u_1, the other is inactive and vice versa.

Functions (1) and (2) represent the *feedback protection* exclusively, however, a single structure can be developed to also include the *feedforward protection* into the control loop. It is achieved by modifying the switching function (1) through the inclusion of the linear term

$$\eta u_1(t), \quad \eta\geqslant 0,$$

for the active range of u_1. This term defines a permanent and increasing level of protection, when the control variable $u_1(t)$ approaches the constraints. For the upper limit case, function (1), the switching function is given by

$$f(u_1)=\begin{cases} u_1(t)-(1-\eta)u_{1\max}, & u_1(t)>u_{1\max}, \\ \eta u_1(t), & u_{1\min}\leqslant u_1(t)\leqslant u_{1\max}, \\ 0, & u_1(t)\leqslant u_{1\min}. \end{cases} \quad (3)$$

If $\eta=0$ the *split-range control* [6] is recovered, the decision function (3) becomes Eq. (1), and the auxiliary variable u_2 is used just to cover output demands when u_1 saturates only. When $\eta>0$ the auxiliary variable u_2 is used to prevent the saturation of u_1, by providing the fraction ηr of $y(t)$ while $u_1<u_{1\max}$. Both manipulated variables are simultaneously acting on the system, but with different gains: Kp_1 for u_1 and ηKp_2 for u_2, where Kp_i $i=1,2$ is the gain of $Gp_i(s)$, respectively. This fact leads to interactions between both control loops that can upset each other, resulting in a deterioration of the closed-loop performance com-

pared with the previous situation.[1] However, it might be a desirable feature since it tends to keep the fastest loop working in a wider range. When $u_1(t) \geqslant u_{1\,\max}$ the control system changes the structure and only Gp_2 controls the system output. Therefore the closed-loop performance will deteriorate a bit more. The gain of the system jumps from ηKp_2 to Kp_2, producing a peak in the control signal u_2 if ideal PID's are employed. In the case of $\eta = u_{2\,\max}/u_{1\,\max}$ both controllers work simultaneously, the fastest control loop is active in the whole operational range maintaining a good control quality in the entire operational space.

A second parameter can be included in the above formulation to start the preventive protection from a given value $u_1(t) = \kappa$ in ahead, instead of doing it from $u_{1\,\min}$ as indicated by Eq. (3), i.e.,

$$f(u_1) = \begin{cases} u_1(t) - \overline{u}_{1\,\max} + \eta(u_{1\,\max} - \kappa), & u_1(t) > u_{1\,\max}, \\ \eta(u_1(t) - \kappa), & \kappa \leqslant u_1(t) \leqslant u_{1\,\max}, \\ 0, & u_1(t) < \kappa. \end{cases} \tag{4}$$

Figure 3 shows a sketch of this function for different parameter values. Note that a similar function can be used for saturations at the lower constraint. Both parameters of the decision function, η and κ, can be used as tuning parameters to satisfy additional process goals than control ones like process efficiency.

Remark 1. *Regarding the work of the actuators, it is interesting to observe that for $\eta = 0$ the control action is executed over a divided range* [6,14], *that is, the second actuator starts moving once the first one saturates. For positive values of η, the control variable is executed over a common range, that is, both actuators work together until one of them saturates. The amplitude of the common range is handled through the parameter κ while the intensity is given by η.*

The capability of transferring the control from one input to another provides an implicit fault-tolerant capabilities to flexible-structure control. This fact means when a major fault in the actuator of u_1 happens, like a actuator frozen at a given a value, the flexible structure is able to transfer the control of the system output to u_2 without any extra information than a measurement of the system output y. In the case of a minor fault, like a loss of a actuator sensibility, it can lead to an output unreachability problem that is overcome as has been explained in the previous paragraph.

3.1. Smith predictor for flexible-structure control

Time delay is a common feature in most of the process models. Control of systems with dominant time delay are notoriously difficult. It is also a topic on which there are many different opinions concerning PID control. For open-loop stable processes, the response to command signals can be improved substantially by introducing *dead time compensation* [15]. The *dead time compensator*, or Smith predictor, is built by implementing a local loop around the controller with the difference between the model of the process without and with time delay [see Fig. 4(a)].

Now, the structure of the Smith predictor is modified to consider the proposed control structure. In this case a second loop is introduced to represent the effect of u_2 over the system output. This fact leads to a new structure that include the models of both process ($\tilde{G}p_1$ and $\tilde{G}p_2$) and the constraints in the manipulated variables [see Fig. 4(b)]. The nonlinearities are required to follow the changes in the structure of the system. This structure works for time delays with different values as we can see in the example.

One can notice that a Smith predictor can be coupled with a robust controller (H_∞, *QFT*, robust pole placement, etc.) to cope with parametric variations.

4. Controller design and tuning

For designing and tuning the controllers involved in the flexible structure it is necessary to analyze each control condition separately: (i) the first control condition—or control structure—is when $C_1(s)$ is in charge of regulation of $y(t)$, i.e., $\eta = 0$ and $u_1(t)$ is not saturated. (ii) The second control structure is defined by the secondary loop only, that is, $\eta = 0$ and $u_1(t)$ is saturated; the controller in this case is the combination $C(s)$

[1] If Gp_1 is much faster than Gp_2, C_1 will be able to compensate the effect of the interactions and the closed-loop performance will not be affected. However, if Gp_1 is not fast enough, the closed-loop response will show an overshoot and an increment of the closed-loop settling time due to the interaction between both control loops.

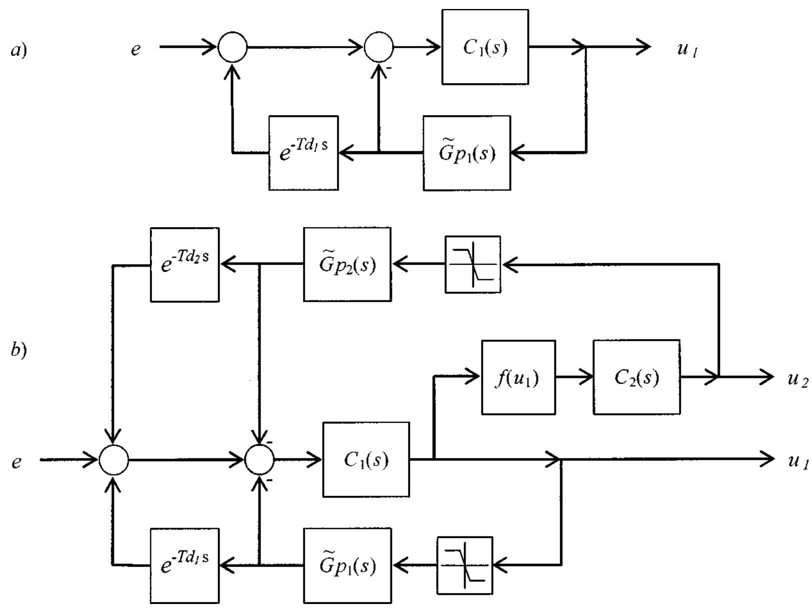

Fig. 4. Block diagram of Smith predictor (a) for normal system, and (b) for cooperative control.

$=C_1(s)C_2(s)$. (iii) The third control condition appears when including preventive protection, i.e., $\eta>0$ while $u_1(t)$ is not yet saturated.

The above decomposition of the problem indicates that $C_1(s)$ must be adjusted for high quality control when the process part $Gp_1(s)$ handles the regulation, i.e., $\forall t: u_{1\,\min} \leqslant u_1(t) \leqslant u_{1\,\max}$, and this is essentially the traditional tuning problem for a single feedback loop. When $u_1(t)>u_{1\,\max}$, the process part $Gp_2(s)$ must provide the complementary effect on the controlled variable, which means that $C_2(s)$ must be combined with $C_1(s)$ such to obtain the best possible performance.

In the following subsections two approaches for designing and tuning controllers $C_1(s)$ and $C_2(s)$ are presented. One is based on IMC parametrization of the controllers, the other is based on cancelation design criteria. In the IMC design the order of the process will be constrained to one and two, leading to PI and PID controllers. In the cancelation design, the constraint in the order of the systems is removed and the controllers can be designed by any controller design techniques.

4.1. IMC design

For simplicity, let us assume stable plants, such that first- or second-order plus time delay models are adequate for describing the dynamics. Then, the IMC strategy provides a rapid parametrization of traditional PI or PID controllers [16,17]. If two PID algorithms with time delay compensation are proposed for controllers $C_1(s)$ and $C_2(s)$, recall that they work in series, i.e., the outlet of $C_1(s)$ is the input to $C_2(s)$ through the switching function $f(u_1)$. The following few hypotheses and practical reasons allow the selection of some important terms and the elimination of others:

1. The double integration term is not necessary since offset elimination is required for set-point changes only.

2. A single integral mode is necessary just in $C_1(s)$, because offset elimination is desired under any working condition and this controller is the one which is always active.

3. The control system structure assumes that $Gp_1(s)$ is faster and with smaller time delay than $Gp_2(s)$; this could be a main argument for selecting $u_1(s)$ as *primary* manipulated variable, but it also suggests that if a derivative term is desired, this should be in $C_2(s)$, i.e., the slower plant dynamics.

These arguments support the selection, for instance, of $C_1(s)$ as a PI controller,

$$C_1(s)=K_{C_1}\left(1+\frac{1}{T_{I_1}s}\right), \tag{5}$$

and $C_2(s)$ as a PD controller,

$$C_2(s)=K_{C_2}\frac{(1+T_{D_2}s)}{(1+T_F s)}. \tag{6}$$

Hence the combination $C(s)=C_1(s)C_2(s)$ results to be a PID controller,

$$C(s)=K_C\left(1+\frac{1}{T_I s}+T_D s\right)\frac{1}{(1+T_F s)}, \tag{7}$$

whose parameters are

$$K_C=K_{C_1}K_{C_2}\left(1+\frac{T_{D_2}}{T_{I_1}}\right), \tag{8a}$$

$$T_I=T_{I_1}+T_{D_2}, \tag{8b}$$

$$T_D=\frac{T_{I_1}T_{D_2}}{T_{I_1}+T_{D_2}}. \tag{8c}$$

The forms of Eqs. (8a)–(8c) lead to the following adjustment procedure:

1. Approach the dynamics relating $y(t)$ and $u_1(t)$ with a first-order plus time-delay transfer function

$$Gp_1(s)\cong Kp_1\frac{e^{-T_{d1}s}}{\tau_{p1}s+1}.$$

2. Use the IMC [16] parametrization to define the parameters of controller $C_1(s)$, which is based on the Smith predictor structure, i.e.,

$$K_{C_1}=\frac{2\tau_{p1}}{Kp_1\lambda_1}, \tag{9a}$$

$$T_{I_1}=\tau_{p1}. \tag{9b}$$

3. Approach the dynamics relating variables $y(t)$ and $u_2(t)$ with a second-order plus time-delay model,

$$Gp_2(s)\cong Kp_2\frac{e^{-T_{d2}s}}{(\tau_{p1}s+1)(\tau_{p2}s+1)}, \tag{10}$$

where one of the time constants is arbitrarily made equal to τ_{p1}, the time constant determined for the model $Gp_1(s)$. This condition is just a convenient way to get consistent individual and combined adjustments for $C_1(s)$ and $C(s)$, respectively. Notice that this still leaves three parameters, Kp_2, τ_{p2}, and T_{d_2}, to adjust the dynamic model $Gp_2(s)$ to the correspondent physical data.

4. Follow the IMC parametrization procedure for $Gp_2(s)$, which gives a PID controller based on a Smith predictor structure—the combined controller—whose parameters K_C, T_I, and T_D are calculated by the following relationships:

$$K_C=\frac{\tau_{p1}+\tau_{p2}}{Kp_2\lambda_2}, \tag{11a}$$

$$T_I=\tau_{p1}+\tau_{p2}, \tag{11b}$$

$$T_D=\frac{\tau_{p1}\tau_{p2}}{\tau_{p1}+\tau_{p2}}. \tag{11c}$$

Comparing relationships (8a)–(8c) with (11a)–(11c) and taking in account Eqs. (9a) and (9b), the parameters of $C_2(s)$ are given by

$$K_{C_2}=\frac{\tau_{p1}}{K_{C1}Kp_2\lambda_2}=\frac{Kp_1\lambda_1}{Kp_2\lambda_2}, \tag{12a}$$

$$T_{D_2}=\tau_{p2}. \tag{12b}$$

Observe the procedure leaves three parameters, λ_1,λ_2, and T_F, for adjusting both controllers to achieve robust performance of the closed-loop system. Since both transfers are connected through an interactive control scheme, the tuning must be coupled, because the uncertainties of each model, $lm_1(s)$ and $lm_2(s)$, affect both controllers simultaneously. If there are uncertainties, the system output $y(s)$ controlled by the flexible-structure control is given by

$$y(s)=\{\tilde{G}p_1(s)[1+lm_1(s)]+C_2(s)\tilde{G}p_2(s)\times[1+lm_2(s)]\}u_1(s).$$

Using the approximation (10) for $\tilde{G}p_2(s)$ and the tuned the procedure proposed in this section for C_2, based on IMC design procedure,

$$C_2(s)=\{\tilde{G}p^*(s)\}_+^{-1}f_2(s),$$

the system output is given by

$$y(s)=Gp_1(s)\{[1+lm_1(s)]+f_2(s)\times[1+lm_2(s)]\}u_1(s).$$

From this expression we can identify the overall uncertainty that will suffer $C_1(s)$,

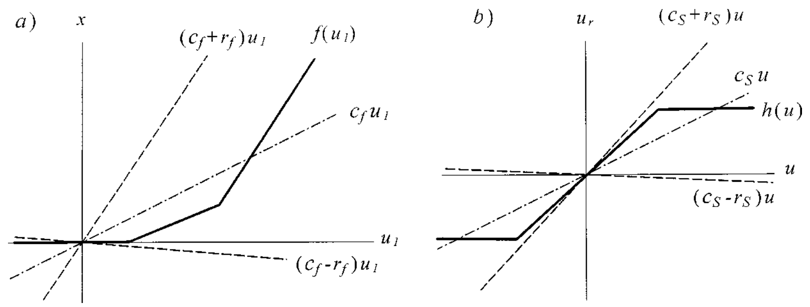

Fig. 5. Nonlinearities and its approximations for stability analysis: (a) decision function, and (b) saturation.

$$lm(s)=lm_1(s)+f_2(s)lm_2(s).$$

Then, the last expressions lead to the following tuning procedure for the controllers parameters:

1. Tune the controller $C_2(s)$ to obtain the robust stability of Eq. (16), modifying λ_2, for $\overline{lm}_2(s)$ as if $C(s)$ is an ideal PID.

2. Tune the controller $C_1(s)$ to obtain the robust stability of Eq. (15), modifying λ_1, for $\overline{lm}(s)$.

3. Finally, the parameter T_F is determined such that the sensitivity function is attenuated in the high-frequency range.

4.2. Cancellation design

As a general rule and from some of the above arguments it can be concluded that selecting the combined controller $C(s)$ as being of equal or higher order than $C_1(s)$ will always give a realizable $C_2(s)$. Hence any available design and adjustment procedure can be followed to completely define $C_1(s)$ and $C(s)$ for controlling $Gp_1(s)$ and $Gp_2(s)$, respectively, as if they were not related to each other. Then, the second controller $C_2(s)$ is determined by

$$C_2(s)=C(s)C_1^{-1}(s). \tag{13}$$

There are no stability problems in this design technique because the zeros and poles of $C_1(s)$ and $C(s)$ are stable and known.

5. Stability analysis

A general stability analysis for the flexible control system can be performed using two different frameworks: (i) a stability analysis of the resulting nonlinear system [18], or (ii) a stability analysis of switched linear system [19]. In the first case the nonlinearities involved in the system (saturations and decision function) are approximated through a conic sector and then the resulting linear system is analyzed using robust control tools. In the second approach, the stability of each linear system is guaranteed and then the stability of the switching between them is analyzed using describing function techniques [21].

Since the nonlinearities involved in the flexible-structure control, the decision function, and the saturations, satisfied the sector nonlinearity condition [20]

$$\left|\frac{h(u)-cu}{u}\right|^2<r^2-\varepsilon, \quad \forall u\neq 0,\varepsilon>0, \tag{14}$$

the nonlinearities can be approximated by (see Fig. 5)

$$h(u)\approx(c\pm r)u,$$

where c is the gain of the linear model and r is the uncertainty of the linear model. Then, the controllers are tuned to obtain the robust stability of the closed-loop system for the overall uncertainties [17,18]: the model and the approximation errors r. This approach leads to an over conservative tune due to the consevatiness of the process gain, introduced by the approximation (14). Therefore the resulting closed-loop performance will be poor and the response will be sluggish.

The stability analysis of the flexible-structure control from the switching system point of view implies the stability analysis of every single or combined loop and the stability analysis for a general switching sequence between these loops.

In a first step, the stability of each closed-loop system is studied. While $u_1(t)<\kappa$ the characteristic equation for the primary control loop includes

$Gp_1(s)$ and $C_1(s)$ only, i.e., the stability condition may be written as follows:

$$1+C_1(s)Gp_1(s)\neq 0, \quad \forall s\in C^+, \tag{15}$$

where C^+ stands for the right-side complex plane. The second important stability condition is for the secondary loop which works through the manipulated u_2 when $u_1(t)>u_{1\,\max}$,

$$1+C_2(s)C_1(s)Gp_2(s)\neq 0, \quad \forall s\in C^+. \tag{16}$$

Finally, for the intermediate case when $\eta>0$ and $\kappa<u_1(t)<u_{1\,\max}$, two control paths coexist simultaneously: one through u_1 and the other through u_2. Then, the condition takes the form

$$1+C_1(s)Gp_1(s)+\eta C_1(s)C_2(s)Gp_2(s)\neq 0, \quad \forall s\in C^+. \tag{17}$$

The first two conditions can be satisfied sequentially, Eq. (15) while adjusting controller $C_1(s)$, Eq. (16) while adjusting controller $C_2(s)$. Hence the stability problem of Eq. (17) is automatically satisfied for the nominal system.

Theorem 5.1. *Given a two-inputs and one-output system with stable transfer function between the inputs,* $u_1(s)$ *and* $u_2(s)$, *and output,* $Gp_1(s)$ *and* $Gp_2(s)$, and $\eta\geqslant 0$, *the closed-loop system resulting from the application of the flexible structure control is stable if the following conditions are satisfied*:

$$1+Gp_1(s)C_1(s)\neq 0, \quad \forall s\in C^+,$$

$$1+Gp_2(s)C_2(s)C_1(s)\neq 0, \quad \forall s\in C^+.$$

Proof: Given the closed-loop equation

$$1+C_1(s)Gp_1(s)+\eta C_1(s)C_2(s)Gp_2(s)=0, \tag{18}$$

both transfer functions

$$Gp_1(s)=K_1\frac{e^{-T_{d_1}s}}{\tau_1 s+1}, \tag{19a}$$

$$Gp_2(s)=K_2\frac{e^{-T_{d_2}s}}{(\tau_1 s+1)(\tau_2 s+1)}, \tag{19b}$$

and the controllers $C_1(s)$ and $C_2(s)$ tuned according with the procedure described in the previous section characteristic equation of the closed-loop system is given by

$$1+\frac{(\tau_1 s+1)}{K_1\lambda_1 s}\frac{K_1}{\tau_1 s+1}+\eta\frac{(\tau_1 s+1)}{K_1\lambda_1 s}\frac{K_1\lambda_1}{K_2\lambda_2}\times\frac{(\tau_2 s+1)}{(T_F s+1)}\frac{K_2}{(\tau_1 s+1)(\tau_2 s+1)}=0. \tag{20}$$

Operating with this equation, the characteristic equation is

$$\lambda_1\lambda_2 T_F s^2+\lambda_2(\lambda_1+T_F)s+(\lambda_2+\lambda_1\eta)=0, \tag{21}$$

and the poles of the closed-loop system p are given by

$$p_{1,2}=-\frac{\lambda_1+T_F}{2\lambda_1 T_F}\pm\sqrt{\frac{(\lambda_1-T_F)^2}{4\lambda_1^2T_F^2}-\frac{\eta}{\lambda_2 T_F}}. \tag{22}$$

The stability only depends on the real part of these poles, therefore it is clear that the stability of the closed-loop system only depends on the values of the controllers parameters (λ_1,λ_2 and T_F) and the parameter η of the decision function.

Remark 2. *If* $\lambda_2 T_F\gg\eta$ *the closed-loop poles will be approximately located at* $p_1=\lambda_1^{-1}$ and $p_2=T_F^{-1}$. *However, if*

$$\eta>\lambda_2\frac{(\lambda_1-T_F)^2}{4\lambda_1^2 T_F},$$

λ_1 *and* T_F *can be used to fix the damping ratio of the closed-loop response and* λ_2 *and* η *can be employed to fix the natural frequency of the closed-loop response*.

Finally, the stability analysis of the switching sequence implies the analysis of the following systems:

$$1+C_1(s)Gp_1(s)+S(s)\,\eta C(s)Gp_2(s)=0, \tag{23a}$$

$$1+S(s)C_1(s)Gp_1(s)+C(s)Gp_2(s)-S(s)(1-\eta)C(s)Gp_2(s)=0, \tag{23b}$$

$$1+S(s)C_1(s)Gp_1(s)+[1-S(s)]C(s)Gp_2(s)=0, \tag{23c}$$

that represent all possible changes in the system. In these expressions $S(s)$ is the Fourier transform of the switching function $S(t)$ given by

$$S(t)=\frac{1}{2}[1-g(t)], \tag{24}$$

where $g(t)$ is a scalar signal that only assumes the value -1 or $+1$. Eq. (23a) represents the changes in the system when $u_1(t)>\kappa$, therefore the second loop becomes active. The second equation, Eq. (23b), represents the change in the system when $u_1(t)>u_{1\max}$, then only the second loop controls the system output. Finally, Eq. (23c) represents the direct transition from $u_1(t)<\kappa$ to $u_1(t)>u_{1\max}$, this fact means that the main loop becomes inactive at the same time that the secondary loop is activated.

For switching sequences slower than the system dynamic, the stability of the overall closed-loop system is guaranteed [19]. The stability of an arbitrary switching sequence between these systems can be analyzed using describing function techniques and harmonic balance explained by Leith *et al.* [19] First, the nonlinearities are approximate through a Fourier series,

$$S(t)=f_0+f_1\cos(\omega t+\phi)+h.o.t,$$

then, due to the low-pass characteristic of the system $S(t)$ may be approximate,

$$S(t)\approx f_0+f_1\cos(\omega t+\phi).$$

For the controllers resulting from the IMC design, the transfer function of the equivalent linear system of Eqs. (23a)–(23c) are given by

$$H_1(s)=-(f_1)/\{\lambda_1\lambda_2T_Fs^2+\lambda_2(\lambda_1+T_F)s+(\lambda_2+\eta\lambda_1f_0)\}, \tag{25a}$$

$$H_2(s)=-\{f_1[\lambda_2T_Fs+(\lambda_2-\lambda_1)]\}/\{\lambda_1\lambda_2T_Fs^2+\lambda_2(\lambda_1+T_Ff_0)s+[(1-\eta-f_0)\lambda_1+f_0\lambda_2]\}, \tag{25b}$$

$$H_3(s)=-\{f_1[\lambda_2T_Fs+(\lambda_2-\lambda_1)]\}/\{\lambda_1\lambda_2T_Fs^2+\lambda_2(\lambda_1+T_Ff_0)s+[(1-f_0)\lambda_1+\lambda_2]\}. \tag{25c}$$

Finally, the resulting transfer functions are analyzed using the method of harmonic balance to determine the stability of the switching sequence. The harmonic balance method predicts instability everywhere the magnitude of the Bode plot of the resulting linear system exceeds unity.

6. Choice of η and κ

In this control scheme η represents the amount of u_2 requires to prevent the saturation of u_1 while κ is the value of u_1 at which the protection begins. However, both parameters can be employed to satisfy some additional objective than control objectives.

If the objective is to obtain *a good transient behavior*, Gp_1 must be kept active in the whole operating range. Hence the parameters of the decision function (4) must be fixed to

$$\eta=\frac{u_{2\max}}{u_{1\max}}, \tag{26a}$$

$$\kappa=0. \tag{26b}$$

This selection implies the use of u_2 to prevent the saturation of u_1 for any value. This might be a desirable feature for the control system since it keeps the fastest loop working in order to maintain a better control quality. This protection is clearly done at the expense of u_2. In the opposite case, if the objective is to *minimize the amount of* u_2 *employed* to control the system, Gp_2 must be kept inactive as much as possible. Therefore the parameters of the decision function must be fixed to $\eta=0$. This selection implies the use of the auxiliary variable to cover output demands when u_1 saturates only. Finally, if the objectives are a combination of previous ones, Gp_1 must be kept active in the range such that it can handle the most frequent changes. Therefore the parameters of the decision function (4) are given by

$$\eta\geqslant\frac{1}{Kp_2}(\max\{\mathrm{var}[r(t)],\mathrm{var}[d(t)]\}-Kp_1u_{1\max}), \tag{27a}$$

$$\kappa\geqslant u_{10}, \tag{27b}$$

where u_{10} is the steady-state value of u_1 for the nominal set-point value. This selection implies the use of u_2 to prevent the saturation of u_1 for the most frequent changes only when $u_1>\kappa$.

To explain these concepts, let us consider the case of two heat exchangers in series: one heat exchanger and a service equipment with a direct

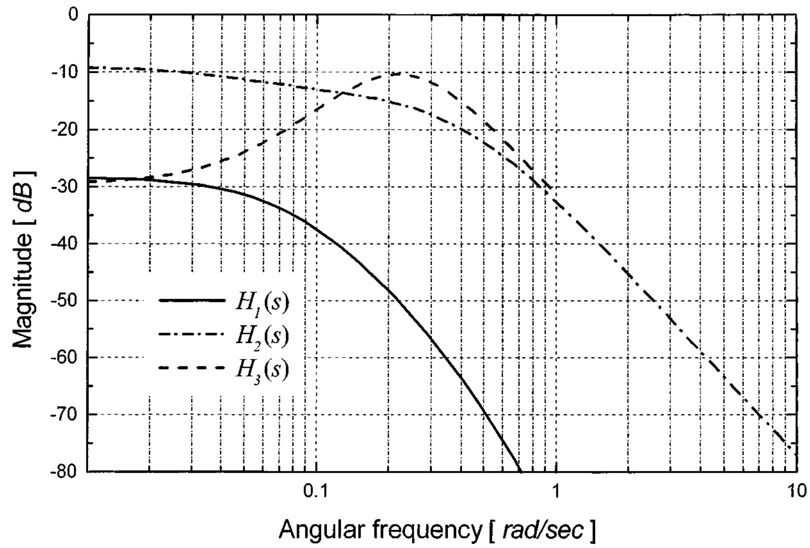

Fig. 6. Bode plots associated with approximated analysis.

bypass on the controlled stream. In this configuration the service equipment works intermittently such that recovered energy is maximized. In this problem, the parameters η and κ correspond to the service energy level required to avoid the controllability problems, which implies a loss of process efficiency, and the level at which the protection begins, respectively. If the objective is to obtain a good dynamic behavior, the parameters must be set to $\eta=1$ and $\kappa=0$, respectively. In the opposite case, if the objective is to maximize the amount of energy recovered, the parameters must be fixed to $\eta=0$. Finally, if the objective is to maximize the amount of energy recovered while keeping good closed-loop performance, the parameters must be fixed to $0<\eta<1$ (according to the disturbances and set-point variance) and $\kappa \geqslant u_{10}$, where u_{10} is the steady-state value u_1 for the nominal set-point value.

7. Simulation and results

In this section a simulation example is considered to show the effectiveness of the control scheme proposed in this work. We consider a linear system previously used by Wang *et al.* [9] to evaluate the cooperative control [9]. The system is given by

$$Gp_1(s)=\frac{e^{-4s}}{(2s+1)^2}\cong\frac{e^{-5s}}{2.75s+1}, \tag{28}$$

$$Gp_2(s)=\frac{2e^{-10s}}{(6s+1)(17s+1)}, \tag{29}$$

and the disturbance $d(s)$ are modeled by

$$G_d(s)=\frac{1}{5s^2+s+1}. \tag{30}$$

The primary and the auxiliary manipulated variables u_1 and u_2 are constrained to $u_{1,2}\in[-1,+1]$.

The controllers will be developed using a Smith predictor structure to compensate the time delays of the process. First, the controller $C_1(s)$ is designed using the formulas (9a) and (9b) and obtain a response without offset. This means that $C_1(s)$ is a PI controller, whose parameters are

$$K_{C_1}=\frac{2.75}{\lambda_1}, \tag{31a}$$

$$T_{I_1}=\tau_{p1}=2.75. \tag{31b}$$

To tune the controller $C_2(s)$, we approximate Gp_2 using Eq. (10). The result is

$$Gp_2(s)\simeq\frac{2e^{-10s}}{(2.75s+1)(18.75s+1)}. \tag{32}$$

First, we design the combined controller $C(s)=C_1(s)C_2(s)$ that controls this model. In this application, a PID controller is adopted for $C(s)$.

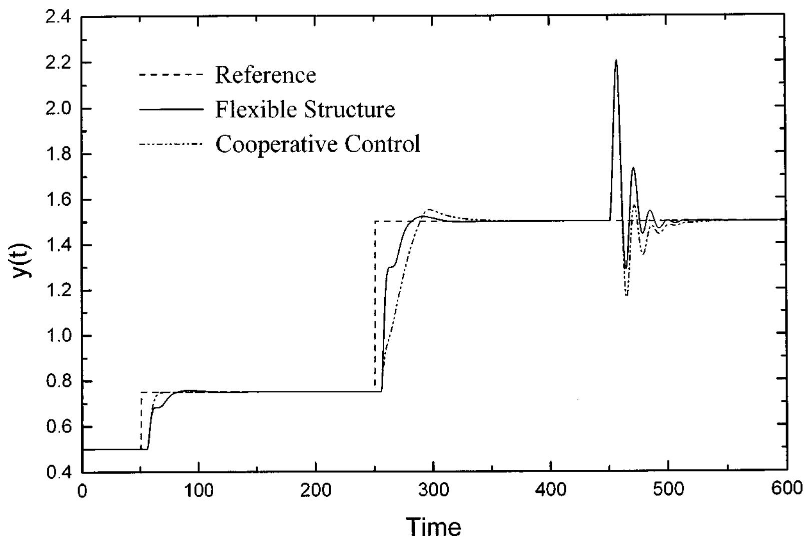

Fig. 7. Set-point responses obtained using flexible structure and cooperative control.

This means that $C_1(s)$ is a PD controller, whose parameters are given by Eqs. (12a) and (12b),

$$K_{C_2}=\frac{Kp_1\lambda_1}{Kp_2\lambda_2}=\frac{\lambda_1}{2\lambda_2}, \tag{33a}$$

$$T_{D_2}=\tau_{p2}=18.75. \tag{33b}$$

Since there is no uncertainty, the parameters λ_1,λ_2, and T_F are determined to obtain the best possible performance. The values of these parameters are

$$\lambda_1=1.5, \quad \lambda_2=0.56, \quad T_F=30.2. \tag{34}$$

An IMC antiwindup scheme is included in the $C(s)$ controller to compensate for the effect of the constraint in u_2. Finally, the parameters of the decision function (1) were chosen to obtain a good performance, therefore η and κ were fixed to

$$\kappa=0.25, \quad \eta=1.0. \tag{35}$$

Finally, the stability of the system for the switch is analyzed. Fig. 6 shows the Bode plots for the transfer functions (25a)–(25c) for parameters (34) and (35). It is easy to see the resulting linear system does not exceed unity, therefore all possible switching between these systems are stable.

In this example the proposed control structure is compared with the cooperative algorithm proposed by Wang *et al.* [9]. The controller employed by the cooperative schemes is $C_1(s)$, with an IMC antiwindup scheme, to compensate the effect of the constraint in the primary manipulated variable (u_1). The parameters of the cooperative control are observation time T_o and a margin β. The value of these parameters are the same to that one employed by Wang *et al.* to obtain the best performance,

$$T_o=0.275, \quad \beta=2.$$

In order to evaluate the performances of both control schemes, a sequence of reference and disturbance changes are introduce at several times. The set point r was changed in intervals of 200 sec from 0.5 to 0.75 and then steps to 1.5, and finally the disturbance w changes from 0 to 0.5 at 450 sec.

In Fig. 7 we see the responses obtained by both control schemes. This figure shows the superior performance of the proposed scheme when the system needs to modify the value of the auxiliary manipulated variable. The better performance is due to the fact that the proposed scheme takes account of the dynamics of the auxiliary manipulated variable and the main loop is still working. This fact leads to the temporary saturation of the primary manipulated variable of cooperative control during the transition period (Fig. 8), that leads to open-loop behavior of the closed-loop system that deteriorates the closed-loop performance. However, the cooperative control scheme shows a better performance when a change in the auxiliary variable is not needed. It is clear that the interac-

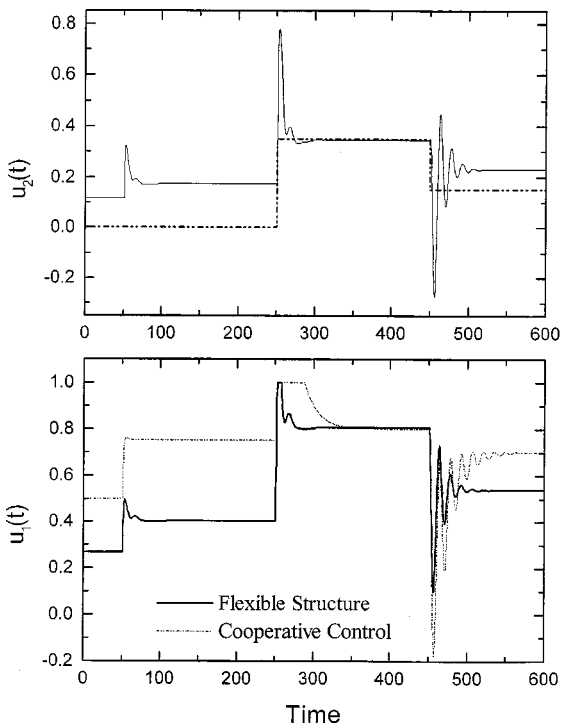

Fig. 8. Primary and auxiliary variables corresponding to set-point and disturbance changes of Fig. 7.

tion between both loops improves the performance of the closed-loop performance when the control is transferred from u_1 to u_2, but it deteriorates the performance when this change is not required.

To address the effects of the protection level η and the start of the protection κ, the proposed algorithm is simulated with different values of them. The results with different η are shown in Fig. 9. It can be seen that the higher the level of η the better the performance, since both transfer functions control the system output. When $\eta=0$ the system output is first controlled by Gp_1 and then by Gp_2. In this case the system output shows a large overshoot and large settling time due to the change of the process structure. The system output is now controlled by the portion of the system that has the slower dynamics and the bigger time delay, therefore closed-loop performance is the poorest.

Fig. 10 shows the control actions associated to the responses of Fig. 9. In this figure it is easy to see that u_2 works to avoid the saturation u_1 or takes full control of the system output, depending on the value of η. If $\eta>0$, u_2 prevents the saturation of u_1 by providing a level ηr of the output. This fact keeps Gp_1 active for a wider output range, therefore good closed-loop responses are obtained. When $\eta=0$, u_2 takes full control of the system output when u_1 is saturated, therefore the slowest dynamic is in charge of the output regulation and poor closed-loop responses are obtained. In these figures we can also see that the auxiliary variable exhibits peaks. This fact is due to the increment in the gain of $C_2(s)$, which is a PD controller.

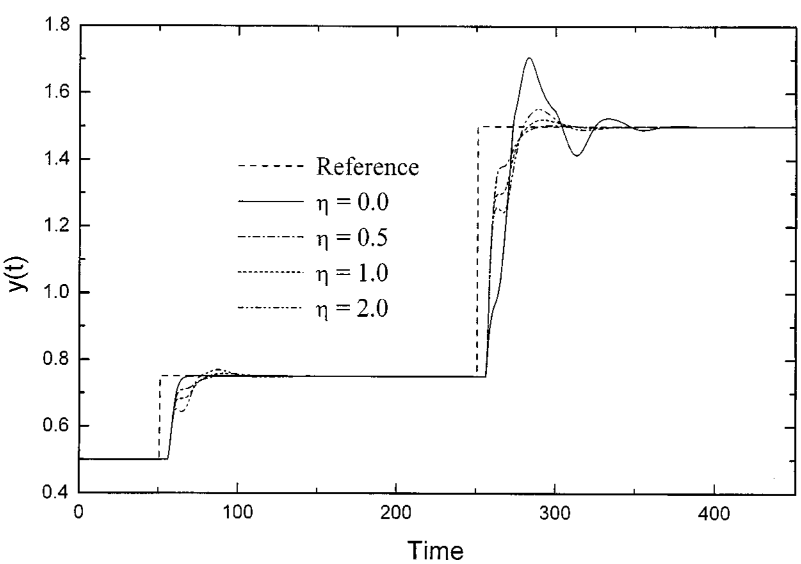

Fig. 9. Set-point responses obtained for different values of η.

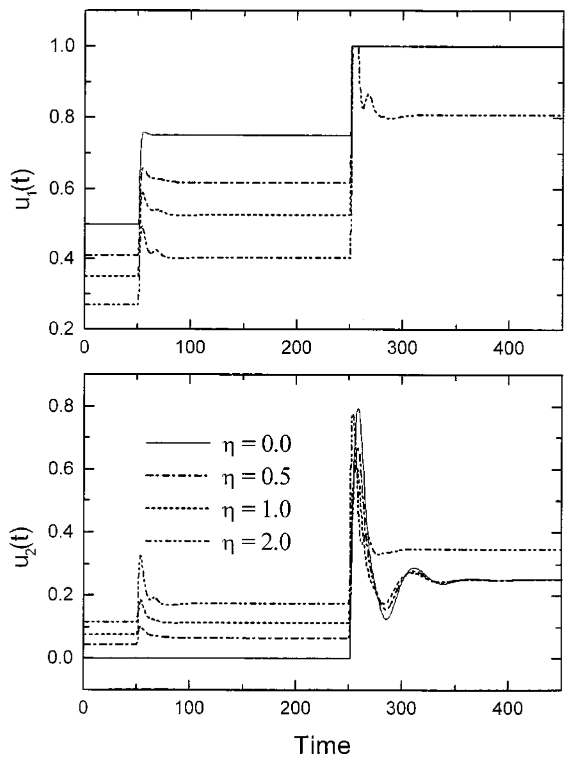

Fig. 10. Primary and auxiliary variables corresponding to set-point and disturbance changes of Fig. 9.

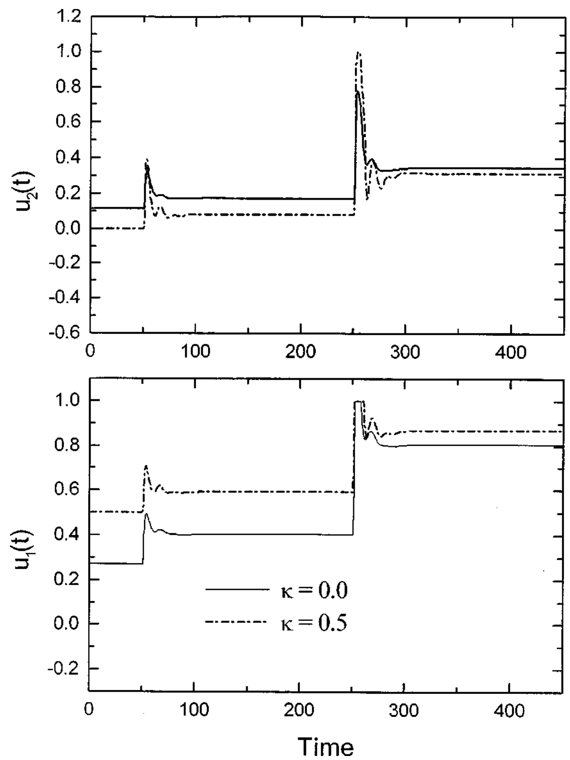

Fig. 12. Primary and auxiliary variables corresponding to set-point and disturbance changes of Fig. 11.

The results with different κ are shown in Figs. 11 and 12. It can be seen that higher the level of κ the smaller the value of u_2, however, similar closed-loop responses are obtained in both cases. When $\kappa = 0$ the protective action is applied all the time. This fact leads to a good closed-loop performance. If $\kappa > 0$ the protective action is only applied when $u_2 \geqslant \kappa$, the closed-loop performance is

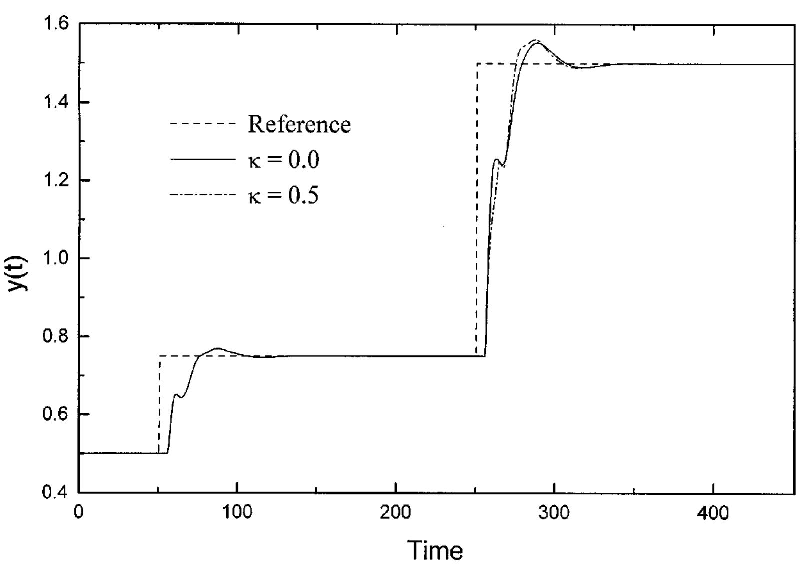

Fig. 11. Set-point responses obtained for different values of η.

similar to the case of $\kappa = 0$. This fact leads to a conservative use of u_2.

8. Conclusions

A new control structure proposed for dealing with constraints on manipulated variables process shows at least two input variables associated to the control target. Flexible-structure control refers to the resulting control system, which switches from one closed loop to another in order to keep controllability. This flexible characteristic of the control system is particularly useful when the optimal operating point locates near a limit constraint of the main manipulated variable. Issues related to design, analysis, and tuning a flexible-structure control system are discussed. The application to a linear system shows the tradeoff between process efficiency and controllability.

Acknowledgments

We are grateful for the financial support for this work provided by the Engineering and Physical Science Research Council (EPSRC) Industrial Non-linear Control and Applications Grant No. GR/R04683/01. Thanks are due to Dr. Hao Xia for valuable discussions and the anonymous referees for the suggestions to improve the paper.

References

[1] Chen, C. and Peng, S., Learning control of process systems with hard constraints. J. Process Control **9**, 151–160 (1999).
[2] Doyle, F., III, An auto-windup input-output linearization scheme for SISO systems. J. Process Control **9**, 213–220 (1999).
[3] Kothare, M., Campo, P., Morari, M., and Nett, C., A unified framework for the study of anti-windup design. Automatica **30**, 1869–1883 (1994).
[4] Wang, Q. and Zhang, Y., A method to incorporate controller output constraints into controller design. ISA Trans. **37**, 403–416 (1998).
[5] Stephanopoulos, G., Chemical Process Control: An Introduction to Theory and Practice. Prentice-Hall, Englewood Cliffs, NJ, 1984.
[6] Shinskey, F., Process Control Systems: Application, Design and Tuning, 3rd ed. McGraw-Hill, New York, 1988.
[7] Gao, Z., Stability of the pseudo-inverse method for reconfigurable control system. Int. J. Control **8**, 469–474 (1991).
[8] Aguilera, N. and Marchetti, J., Optimizing and controlling the operation of heat exchanger networks. AIChE J. **44**, 1090–1104 (1998).
[9] Wang, Q., Zhang, Y., Cai, W., Bi, Q., and Hang, C., Co-operative control of multi-input single-output processes: On-line strategy for releasing input saturation. Control Eng. Pract. **9**, 491–500 (2001).
[10] Morningred, D., Paden, B., and Mellichamp, D., An adaptive nonlinear predictive controller. Chem. Eng. Sci. **47**, 755–765 (1992).
[11] Luyben, W., Process Modelling, Simulation and Control for Chemical Engineers. McGraw-Hill International, New York, 1990.
[12] Marselle, D., Morari, M., and Rudd, D., Design of resilient processing plants: II. Design and control of energy management system. Chem. Eng. Sci. **37**, 259–270 (1982).
[13] Mathisen, K. Integrated Design and Control of Heat Exchanger Networks. Ph.D. thesis, University of Trondheim, Trondheim, Norway, 1994.
[14] Åström, K. and Hägglund, T., PID Controllers: Theory, Design and Tuning, 2nd ed. Instrument Society of America, 1995.
[15] Smith, O., Close control of loops with dead time. Chem. Eng. Prog. **53**, 217–219 (1957).
[16] Chien, I. and Fruehauf, P., Consider IMC tuning to improve controller performance. Chem. Eng. Prog. **86**, 33–41 (1990).
[17] Morari, M. and Zafiriou, E., Robust Process Control, 2nd ed. Prentice-Hall, Englewood Cliffs, NJ, 1989.
[18] Sofanov, M. and Athans, M., A multiloop generalisation of the circle criterion for stability analysis. IEEE Trans. Autom. Control **26**, 415–422 (1981).
[19] Leith, D., Shorten, R., Leithead, W., Mason, O., and Curran, P., Issues in the design of switched linear control systems: A benchmark study. Int. J. Adapt. Control Signal Process. **17**, 103–118 (2003).
[20] Desoer, C. and Vidyasagar, M., Feedback Systems: Input-Output Systems. Academic, New York, 1975.
[21] Khalil, K., Nonlinear System Analysis. Macmillan, New York, 1992.

Leonardo L. Giovanini was born in Argentina in 1969. He received the BSc (Honors) degree in Electronic Engineering from the Universidad Tecnologica Nacional, Regional School at Villa Maria in 1995, and the Ph.D. degree in Engineering Science from Universidad Nacional del Litoral, School of Water Resource in 2000. Since then he has been Lecturer in the Department of Electronic Engineering at Universidad Tecnologica Nacional, Regional School at Villa Maria. Currently, he is working as a research fellow in the Industrial Control Center, University of Strathclyde. His research interests include predictive control for non-linear systems, fault detection and isolation, fault tolerant control and distributed control systems.