Report UPTEC 93024R, Signals and Systems, Uppsala U.

A FREQUENCY DOMAIN APPROACH TO ANTI–WINDUP COMPENSATOR DESIGN

Mikael Sternad

Signals and Systems
Uppsala University
Box 534 SE 751 21 UPPSALA
SWEDEN

Stefan Rönnbäck

Division of Automatic Control
Luleå Univ. of Technology
S-951 87 Luleå, Sweden.

April 1993

Abstract

The effect of input saturation in linear time invariant control systems is addressed. We utilize a controller in polynomial form, with a three degrees of freedom structure. This makes it possible to add a saturation compensator (an anti–windup filter) separately, after tuning of the controller properties for unsaturated signals.

Our aim is that the effect of a saturation event should decay as quickly as possible after desaturation, while the tendency for repeated re–saturations, due to nonlinear oscillations, should be supressed. This is achieved by modifying the relevant loop gain by a safety margin. The modified loop gain is then adjusted, so that it touches the describing function of the saturation nonlinearity. The construction of the safety margin takes the magnitude of exogenous signals into account. The method requires only simple calculations. It can be automated, using a one–dimensional numerical search. Furthermore, the effect of D/A–quantization in digital feedback systems is also controlled by the same antiwindup scheme. The presence of unstructured errors in a linear plant model can be taken into account.

We also provide a discussion on how the frequency domain viewpoint can improve the understanding of previously suggested anti–windup methods, such as the observer–based approach, and the conditioning technique.

Report UPTEC 93024R, Department of Technology, Uppsala University.

1 INTRODUCTION

Linear feedback design is such a common tool, that we often forget to notice the mystery in why such methods do work at all. We adjust linear controllers, and expect them to function in a nonlinear environment. This paper describes a small step towards a more realistic approach. It develops a simple and systematic ways of handling the most common nonlinearity: the saturation of the control signal.

When a controller is adjusted without taking input saturation into account, severe performance loss, large overshoots, or even limit cycle oscillations may result, if the control system saturates. Such problems are sometimes referred to as *windup phenomena*. The reduction of achievable performance due to saturation has been studied by Lozier (1981). Saturation typically happens in regulators during startup, or at major setpoint changes. In high performance servo systems, saturation occurs frequently, since it is normally considered uneconomical to over–dimension the actuator so that saturation events become infrequent.

There do exist methods, such as classical optimal control, or infinite horizon Model Algorithmic Control (Garcia *et.al.* 1989), that take both stability and constraints into account. However, due to the computational difficulty with all such schemes, the effect of input saturation is frequently checked by simulation only.

In PID control, windup problems are of central interest. Many modifications have been suggested, which change the structure of the controller when the control signal saturates. They are called safety nets, or anti–windup modifications. See Hanus (1988), Rundqwist (1991) and Walgama (1991) for good surveys. Two methods for more general control laws are the observer principle by Åström and Wittenmark (1984) and the conditioning technique, see Hanus *et.al.* (1987). These schemes often work well, but neither of them is guaranteed to provide a satisfactory result.

The windup phenomenon has often been interpreted as due to errors in the controller states, in particular the integrator state. Windup problems are, however, not restricted to integrating control laws. They are not really properties of the controller states at all. Instead, windup phenomena are better viewed as problems encountered when a linear system (controller and linear plant) works in closed loop with a static nonlinear feedback (the saturation).

The aim of our present research has been to investigate how far the windup problem can be handled by simple, essentially linear, methods. Anti–windup adjustment should be a natural additional step in any linear feedback design. That situation will, however, become a reality only if simple tools are available, and are adequate. To attain simplicity, an anti–windup modification should not alter the control properties in the linear region. It should be possible to design the linear-range feedback separately, in a first step. In a later step, the properties at saturation are adjusted.

What, then, should be the goal of an anti–windup design? We have focused on a basic compromise, which has previously hardly received attention. Peformance degradation may last for a considerable time after desaturation. It is advantageous if saturation events end as soon as possible, and their effect decays quickly. This may, however, result in deep resaturation and nonlinear oscillations, limit cycles. We search for a way of obtaining both a fast transient and avoidance of limit cycles.

It is hard to address this issue with presently available methods. Several of them, such as the conditioning technique, lack adjustable parameters. When problems occur, they leave no alternative but to modify the linear feedback design. Others, such as the observer–based method, do have adjustable parameters, but there exist no tools for adjusting them in a systematic way.

A novel method is therefore outlined in the present paper. It is based on the following choices of regulator structure, criterion and design tools:

• As *regulator structure*, we consider a controller in polynomial form. It includes the general anti–windup modification of Rönnbäck *et.al.* (1992), by which windup control can be discussed in conventional linear pole placement terms. The control law has three degrees of freedom, as also suggested by Horowitz (1983). We introduce a scalar parameter for adjusting windup properties.

• The *criterion* for adjusting this parameter is that the recovery transient after desaturation should be fast, while nonlinear oscillations should not occur.

• The *tool* for attaining this aim is to place the loop gain on a "sufficient distance" from the describing function of the saturation nonlinearity.

The describing function was used as an antiwindup design tool also in a recent paper by Wurmthaler and Hippe (1991). We use the tool differently from their work, in that we adjust only the anti–windup part, not the whole feedback system. Thus, we avoid loss of performance in the linear range. We also suggest an improved safety margin between the loop gain and the describing function.

The methodology is, in this paper, restricted to linear time–invariant control laws for SISO systems. It can easily be generalized to state feedbacks. The method works for stable or marginally stable systems, with static input nonlinearities which are invertible (except for the saturation). As an added bonus, the effect of D/A-quantization in a digital feedback system is also controlled. Errors in the linear plant model can be taken into account in the anti–windup design.

The paper is organized as follows. The system and regulator structure is presented in Section 2. The use of linear analysis, by regarding the saturation effect as an equivalent disturbance, and of describing function analysis to predict nonlinear oscillations, is outlined in Section 3. These methods are utilized in Section 3.3 to understand the properties of some anti–windup schemes. This discussion complements that of Rönnbäck *et.al.* (1992). In Section 4, our design method is outlined. Key issues are how to introduce the adjustable parameter, and how to construct a safety margin for the loop gain. In Section 5, different anti–windup strategies are compared for three different systems, all controlled by LQG-designed regulators. Section 6 describes how model errors and quantization effects can be handled.

2 A REGULATOR IN POLYNOMIAL FORM WITH WINDUP COMPENSATION

A modification of the controller structure of Figure 1 will be considered. The process subjected to control is assumed linear and time–invariant (LTI), except for a static saturation nonlinearity, representing the actuator constraint.

$$y = \frac{B}{A}(v+\beta) \quad ; \quad y_m = y + \gamma \tag{2.1}$$

$$v = \text{sat}[u]_{v_{\min}}^{v_{\max}} \triangleq \begin{cases} v_{\max} & \text{if} \quad u \geq v_{\max} \\ u & \text{if} \quad v_{\min} < u < v_{\max} \\ v_{\min} & \text{if} \quad u \leq v_{\min} \, . \end{cases} \tag{2.2}$$

Here, y is the scalar controlled output, u and v are control signal before and after saturation, β and γ are process and measurement disturbances, and y_m is the measured signal. The limits $v_{\max}$ and $v_{\min}$ are defined with respect to the working point of the control system, which is here set to zero. If the system contains a general invertible input nonlinearity $f(\cdot)$, we assume that it has been reduced to a simple saturation. This can be done by applying a model of its inverse on a saturated control signal,

$$u' = f^{-1}(\text{sat}[u]_{v_{\min}}^{v_{\max}})$$

and sending the modified signal u' to the actuator $v = f(u')$.

The basic linear controller structure has two degrees of freedom. It is parametrized by polynomials $\{R, S, T\}$, as

$$Ru = -Sy_m + Tw \tag{2.3}$$

where w is the reference for y. Above, A and R are assumed to be monic polynomials, i.e. to have highest degree coefficients equal to 1.

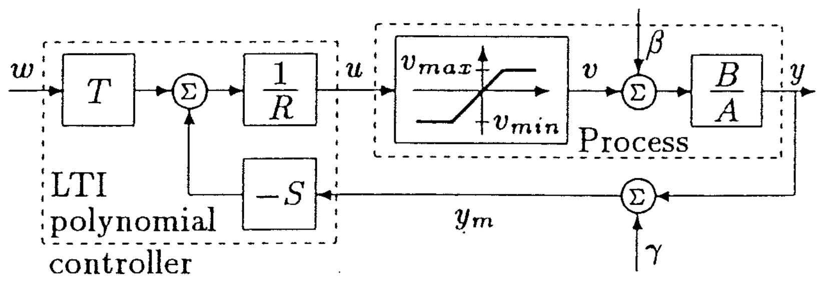

Figure 1: A SISO–process with actuator constraints and a 'polynomial controller' are the main parts of the system.

Both the continuous–time and the discrete–time cases will be treated simultaneously. The reader should think of polynomial arguments as being either the derivative operator p $(= d/dt)$, the forward shift operator q or, when appropriate, the corresponding transform variables s or z. The δ–operator form, advocated by

Middleton and Goodwin (1990), can also be used. Furthermore, the polynomials are presumed to satisfy the following causality/properness restrictions

$$\deg[R] \geq \deg[T] \quad ; \quad \deg[R] \geq \deg[S] \quad ; \quad \deg[A] > \deg[B] \ .$$

We assume B/A to have no hidden modes, and to have all poles inside or on the limit of the stability region. A controller with actuator constraints can then stabilize the system. See e.g. Theorem 6 of Tsirukis and Morari (1992).

We complement (2.3) with a compensation, which becomes active only due to saturation. The technique is called *windup compensation*, and the controller is referred to as an *anti–windup compensator*. It controls the properties of the system during saturation, and the transient effects after leaving saturation. A key to windup compensation is that the true, saturated control signal v is taken into account.

Assume that the saturated output v can be either acccurately measured or accurately reconstructed, using a model of the saturation. Following Rönnbäck *et.al.* (1992), we suggest the use of the regulator structure of Figure 2:

$$\boxed{Fu = (F - PR)v - PSy_m + PTw} \tag{2.4}$$

This structure includes many previously suggested windup compensation schemes as special cases, see Rönnbäck (1992) and Section 3.3 below. In particular, (2.4) reduces to the unmodified control law (2.3) when $F = R$ and $P = 1$.

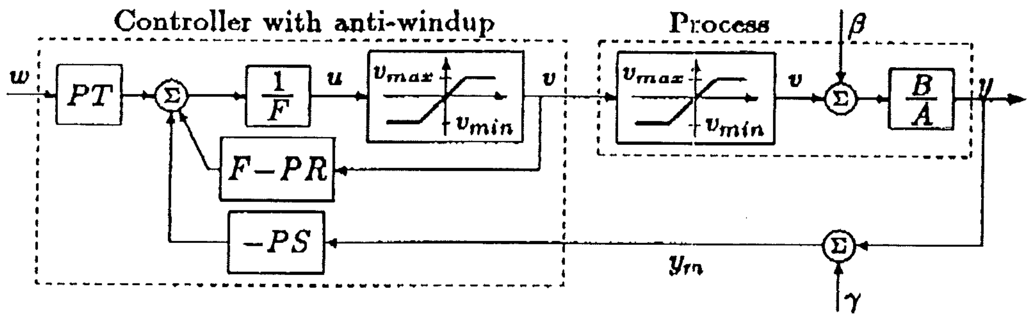

Figure 2: Windup compensation is introduced by complementing the controller in Figure 1 with two stable polynomials F and P. Here, an internal model of the saturation is used to generate v. The "process saturation" can thus be ignored in the calculations.

In the modified controller (2.4), two new polynomials F and P have been introduced. They are denoted *anti–windup polynomials* and are assumed stable. The regulator (2.4) introduces a local feedback around the saturation. In order to avoid algebraic loops, and to preserve causality/properness, F and P must satisfy

$$\deg[F] = \deg[PR] \tag{2.5}$$

and have the same highest degree coefficients.

When the control signal does not saturate, the controller (2.4) provides the same closed–loop dynamics as does (2.3). Substitution of (2.4) into (2.1), using $v = u$, gives the *nominal* (unsaturated) output as

$$P\alpha\, y_{\text{nom}} = P(BTw + BR\beta - BS\gamma) \quad . \tag{2.6}$$

Here,

$$\alpha \triangleq AR + BS \tag{2.7}$$

is the closed loop characteristic polynomial which would have been obtained if (2.3) had been used. The (stable) polynomial P contributes hidden modes. When the control signal saturates, the dynamics change radically. The nonlinearity (2.2) and the expression (2.4) together constitute a variable structure system.

The three degrees of freedom of the controller (2.4) can be utilized as follows.

- The desired feedback properties (stabilization, disturbance rejection and robustness) determine the nominal feedback polynomials R and S. The controller is typically tuned so that the required amplitude of u is not much larger than the limits, in normal operation. The design is made as if (2.1) and (2.3) were used, with $v = u$.
- The polynomial T is then adjusted, with respect to a desired servo response.
- Finally, F and P are utilized for obtaining acceptable behaviour when the system saturates and comes out of saturation.

In the next section, we present some concepts useful for understanding the last step. A design methodology for that step will be outlined in Section 4.

3 TOOLS FOR AN APPROXIMATE ANALYSIS OF WINDUP PHENOMENA

The effect of a saturation may be regarded as an equivalent (time domain) disturbance on the control input. That approach was discussed in Rönnbäck *et.al.* (1992) and is presented in Section 3.1 below. Such a point of view can e.g. be utilized to predict if the recovery after saturation is dominated by slow or oscillative modes. It is, however, difficult to predict the most severe problem: repeated saturations, due to stable or damped limit cycle oscillations. For this, describing function analysis is a useful tool, as is outlined in Section 3.2. It should be emphasized that neither of these methods constitutes an exact way of analyzing the nonlinear system. However, when utilized together, the properties resulting from possible choices of F and P in the regulator (2.4) can be understood in general terms. This is discussed briefly in Section 3.3.

3.1 Saturation seen as an equivalent disturbance

Define

$$\delta \triangleq v - u = \begin{cases} v_{\max} - u & \text{if} \quad u \geq v_{\max} \\ 0 & \text{if} \quad v_{\min} < u < v_{\max} \\ v_{\min} - u & \text{if} \quad u \leq v_{\min} \quad . \end{cases}$$

The saturated signal can then be viewed as the original signal, plus the "disturbance" δ

$$v = u + \delta \quad . \tag{3.1}$$

The transfer functions from δ to all signals in the loop are linear (the nonlinearity is hidden in the fact that δ depends on u in a nonlinear way). Superposition may therefore be used in order to express the output y as

$$y = y_{\text{nom}} + y_\delta \tag{3.2}$$

where y_{nom} is given by (2.6) and y_δ is the contribution from saturation effects. The use of (3.1) in (2.4) gives

$$Fu = F(u + \delta) - PRv - PSy_m + PTw$$

or

$$PRv = F\delta - PSy_m + PTw \quad . \tag{3.3}$$

The use of (3.3) in (2.1) gives (3.2), where y_{nom} is given by (2.6) and where

$$\boxed{y_\delta = \frac{B}{\alpha}\frac{F}{P}\delta \triangleq \mathcal{H}_\delta \, \delta} \tag{3.4}$$

When the system has gone *out of* saturation, $\delta = 0$. The transient of the filter $\mathcal{H}_\delta = BF/\alpha P$ will then determine how fast the saturation effect on y decays. Note that by selecting F and P, we can control these transient dynamics.

When the system is *in* saturation, then $\delta \neq 0$. The "disturbance" δ is then given by

$$\delta = v - u = \frac{P}{F}\left[\frac{\alpha}{A}v_{\text{sat}} - Tw + S\gamma + \frac{BS}{A}\beta\right] \tag{3.5}$$

where v_{sat} denotes either $v_{\max}$ or $v_{\min}$. For example, when $\gamma = \beta = 0$ and w is a large positive step input, the term $-Tw$ gives a negative contribution to δ. It strives to hold the signal u in the saturated region. The term $(\alpha/A)v_{\text{sat}}$ increases δ, i.e. it strives to take the system out of saturation. (This will succeed if $T_{\text{stat}}w < (\alpha/A)_{\text{stat}}v_{\max}$, where $(\cdot)_{\text{stat}}$ denotes the static gain.) The balance between these two effects determines the depth and length of saturation.

3.2 The describing function

The describing function, see e.g. Atherton (1975), is a useful approximative method for analyzing feedback systems which contain a static nonlinearity. The nonlinearity is represented by an amplitude–dependent gain, Y_f, of the main harmonic of the feedback signal. By comparing $-1/Y_f$ to the Nyquist curve, the presence of nonlinear oscillations (limit cycles) is indicated. See Figures 3 and 4.

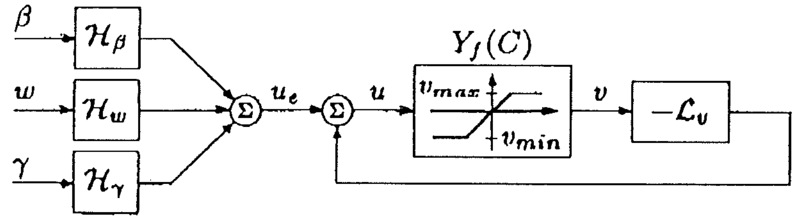

Figure 3: Redrawing of the block diagram of the control system in Figure 2, for the purpose of describing function analysis.

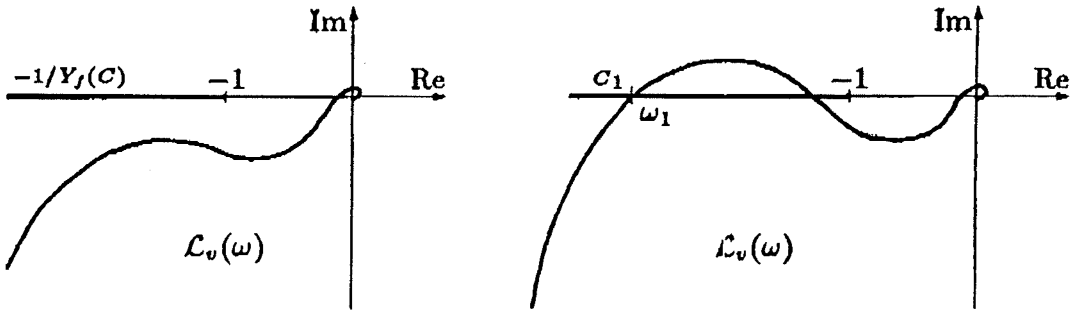

Figure 4: When the Nyquist curve $\mathcal{L}_v$ of the linear system does not cross the describing function of the saturation, stable limit cycles will not occur (left–hand sketch). If they cross (right–hand sketch), an oscillation of frequency $\approx \omega_1$ and amplitude $\approx C_1$ may develop, provided the initial transient is large enough.

The use of describing function analysis, according to Figures 3 and 4, is based on two assumptions. First, $\mathcal{L}_v(\omega)$ should be low–pass, so that higher harmonics are damped. This will mostly be the case. Second, the system should be autonomous, i.e. $u_e = 0$ in Figure 3. This will *not* be true in a realistic situation. The *dual input describing function* can be used to predict the result, if u_e is a sinusoid or a stochastic process. See e.g. Gibson (1963). For small amplitudes of u_e, not much changes, compared to an autonomous system. For larger amplitudes $|u_e| \geq v_{\max}$, the effective gain of the saturation is reduced, as if the endpoint of $-1/Y_f(C)$ moved to the left from -1. See Figures 9.55 and 9.61 in Gibson (1963). Thus, a reasoning based on the single–input describing function ($u_e = 0$) may be somewhat conservative, if exogenous signals have large amplitudes. We have chosen to use it anyway, for simplicity reasons.

Consider the describing function Y_f of a saturation nonlinearity, with unit linear gain and symmetrical saturation limit $v_{\max} = -v_{\min}$. Assume $u_e = 0$. If C is the amplitude of a sinusoidal u, then $Y_f(C) = 1$ for $C \leq v_{\max}$ and

$$Y_f(C) = \frac{2}{\pi}\left(\arcsin\left(\frac{v_{\max}}{C}\right) + \frac{v_{\max}}{C}\sqrt{1 - \frac{v_{\max}^2}{C^2}}\right) \qquad C > v_{\max} \ . \tag{3.6}$$

The function $-1/Y_f(C)$ starts at -1 and approaches $-\infty$ as $C \to \infty$. The loop gain $\mathcal{L}_v$ is obtained by opening the loop at u in Figure 2. The use of (2.1) in (2.4) gives

$$Fu = (F - PR)v + PTw - PS\gamma - PS\frac{B}{A}\beta - PS\frac{B}{A}v \ .$$

Thus, (cf Figure 3)

$$u \;=\; -\mathcal{L}_v\, v + \mathcal{H}_w\, w + \mathcal{H}_\gamma\, \gamma + \mathcal{H}_\beta\, \beta \tag{3.7}$$

where

$$\mathcal{H}_w = \frac{PT}{F} \quad ; \quad \mathcal{H}_\gamma = -\frac{PS}{F} \quad ; \quad \mathcal{H}_\beta = -\frac{PS}{F}\frac{B}{A} \tag{3.8}$$

and where the loop gain $\mathcal{L}_v$ is

$$\boxed{\mathcal{L}_v = \frac{P}{F}\frac{\alpha}{A} - 1} \tag{3.9}$$

Above, rational functions are denoted by script letters. Note the difference between (3.9) and the loop gain

$$\mathcal{L} = \frac{SB}{RA}$$

of the feedback (2.3) and system (2.1). Within the degree restriction (2.5), the polynomials P and F provide complete authority over the properties of the loop gain $\mathcal{L}_v$. In particular, note that the choice $F = \alpha$ and $P = A$ (if A is stable) gives $\mathcal{L}_v = 0$. *This would completely eliminate the risk for nonlinear oscillations.* However, according to (3.4), the choice $P = A$ may lead to unacceptable transients after the system leaves saturation, if the open–loop system has slow or oscillatory modes. As has already been emphasized, a compromise often has to be made between a fast decay of the transient and a low tendency for repeated saturations.

3.3 Some specific choices of anti–windup polynomials

Let us briefly investigate what the expressions (3.4) and (3.9) reveal about specific cases of the general controller structure (2.4).

No windup compensation.

In the case of *no feedback around the saturation*, the controller (2.4) reduces to (2.3), and we have

$$F = R \quad ; \quad P = 1 \implies y_\delta = \frac{BR}{\alpha}\delta \quad ; \quad \mathcal{L}_v = \frac{\alpha}{RA} - 1 = \frac{SB}{RA} \ . \tag{3.10}$$

For high–gain regulators S/R, the loop gain will often cross the negative real axis, as in the right–hand part of Figure 4. Limit cycle oscillations result. For an illustration, see Example 1 in Section 5. Regulators with unstable denominators R, or multiple integrators, may also give rise to oscillations. Badly located (stable) zeros of R may also result in unacceptable transient behaviour of y_δ.

The observer–based method of Åström and Wittenmark.

In cases where F is specified and $P = 1$, the dynamics of a saturation observer in a state–space controller realization corresponds to the zeros of F. See Walgama (1991) or Walgama and Sternby (1990). An observer–based modification of (2.3) was originally suggested in Åström and Wittenmark (1984). In discrete time, a frequently used special case is the *deadbeat observer.* With $nr \triangleq \deg R$, it corresponds to

$$F = q^{nr} \quad ; \quad P = 1 \implies y_\delta = \frac{Bq^{nr}}{\alpha}\delta \quad ; \quad \mathcal{L}_v = \frac{\alpha}{q^{nr}A} - 1 \ . \tag{3.11}$$

A deadbeat observer simply corresponds to feedback of the saturated control signals in the controller recursion

$$q^{nr}u = (q^{nr} - R)v - Sy_m + Tw \ .$$

Deadbeat anti–windup often works well. Regulators with unstable poles can be allowed. The transient y_δ after desaturation has the same time constants as the unsaturated feedback system. However, problems may result when high gain feedbacks are used on a high order system. Such feedbacks increase the bandwidth significantly. They result in a large phase difference between the closed loop characteristic polynomial α and the open-loop denominator A. If

$$|\arg(\mathcal{L}_v + 1)| = |\arg(\alpha(e^{i\omega})) - \arg(e^{i\omega nr}A(e^{i\omega}))| \ > \ \pi \tag{3.12}$$

at some frequency, the describing function will be crossed, or encircled, and limit cycles may occur. A case where deadbeat anti–windup leads to limit cycles (but where the use of no windup compensation works well!) is discussed in Example 3 of Section 5. Deadbeat anti–windup is also unsatisfactory when α contains zeros corresponding to slow modes, since this leads to a slow transient y_δ. The use of an observer polynomial $F \neq q^{nr}$, which cancels these modes, improves this situation.

The model–based approach of Irving and Internal model control.

As mentioned in Section 3.2 above, the risk for limit cycle oscillations can be eliminated completely for open–loop stable systems with the choice

$$F = \alpha \quad ; \quad P = A \implies y_\delta = \frac{B}{A}\delta \quad ; \quad \mathcal{L}_v = 0 \ . \tag{3.13}$$

This controller can be expressed as

$$Ru = -S\left[y_m - \frac{B}{A}\delta\right] + Tw \ . \tag{3.14}$$

Since, due to (3.2) and (3.13), the term $(B/A)\delta$ disappears in (3.14), the feedback signal is modified so that the saturation is eliminated, as seen from the controller. The method is called *model–based anti–windup*. It was suggested by Irving (in a somewhat more general form), and is mentioned in Hanus (1988). This controller makes no effort to modify the saturation dynamics. It is not "aware" of any saturation. Thus, the decay of y_δ after saturation is governed by the open–loop dynamics. That, of course, may leave much to be desired, as is illustrated by Example 2 in Section 5. Since the model–based scheme is open–loop by nature, one could argue that it does not constitute any windup compensation at all.

Internal model control IMC, see Morari and Zafiriou (1989), has similar windup properties. In IMC, $(B/A)v$ is subtracted from y_m inside the controller. This also leads to open–loop desaturation dynamics $y_\delta = (B/A)\delta$ and zero loop gain $\mathcal{L}_v = 0$.

The conditioning technique of Hanus.

The conditioning technique, see Hanus (1988) and Hanus *et.al.* (1987), can also be seen as a special case of the present framework. See Rönnbäck *et.al.* (1992) or Walgama and Sternby (1990). Let the polynomial T have leading coefficient t_o, and let the characteristic polynomial α contain a factor T_2 of T as observer modes

$$T/t_o = T_1 T_2 \quad ; \quad \alpha = \alpha_1 T_2 \ .$$

Let $\deg T = \deg R$. Then, the conditioning technique corresponds to the use of (2.4), with

$$F = T/t_o \quad ; \quad P = 1 \implies y_\delta = \frac{BT_1}{\alpha_1}\delta \quad ; \quad \mathcal{L}_v = \frac{\alpha_1}{AT_1} - 1 \ . \tag{3.15}$$

Compared to the use of deadbeat anti–windup, a faster transient y_δ is in general obtained, since observer poles T_2 do not affect y_δ. The risk of nonlinear oscillations is often increased, due to this speed–up. In Example 2 in Section 5, deadbeat anti–windup works well, while the use of conditioning results in limit cycle oscillations.

4 SYSTEMATIC ANTI–WINDUP DESIGN

Let us now outline a new class of methods for selecting the anti–windup polynomials F and P in (2.4). They are based on the following principles.

• The structure of F and P is chosen with regard to the properties of *both* (3.4) and the loop gain (3.9). The structure contains a single tuning parameter c.

• The parameter c is adjusted so that the loop gain $\mathcal{L}_v$, plus a safety margin, *touches the describing function of the nonlinearity*, if possible. This gives the best transient y_δ, for the given choice of controller structure and of safety margin.

The methods can be used in conjunction with any describing function, not only that of a limiter. The choices of controller structure and of safety margin which presently seem most promising are discussed in Section 4.1 and 4.2, respectively.

4.1 The choice of antiwindup polynomials

We now aim at selecting a suitable structure for F and P, which depends on a scalar parameter c in an intuitively attractive way. It is of advantage to have the alternative $F = \alpha, P = A$ as one member of the admissible set, since this choice guarantees the absence of limit cycles. A value of c for which some norm of $\mathcal{H}_\delta$ in (3.4) is minimized is useful as another alternative. The following optimization problem turns out to be a suitable formalization of these requirements:

Select monic and stable polynomials F and P, with degrees (2.5), which minimize the $\mathcal{H}_2$–criterion

$$\begin{aligned} I &= \| \mathcal{H}_\delta \|_2^2 + c \, \| (\mathcal{L}_v + 1)^{-1} - 1 \|_2^2 \\ &= \left\| \frac{B}{\alpha}\frac{F}{P} \right\|_2^2 + c \left\| \frac{A}{\alpha}\frac{F}{P} - 1 \right\|_2^2 \quad . \end{aligned} \tag{4.1}$$

When $c = 0$, (4.1) represents the minimization of the energy in the impulse response of $\mathcal{H}_\delta$. When $c \to \infty$, we place high priority on the second term, which is minimized by $\mathcal{L}_v = 0$. That extreme case will guarantee absence of limit cycles. Intermediate values of c provide a compromise between these two (often conflicting) requirements. For $c > 0$, the structure of the last term *heavily penalizes closeness of $\mathcal{L}_v$ to the critical point* -1. (This would not be the case if e.g. $\| \mathcal{L}_v \|_2^2$ were used instead.) This property helps us to keep a safety margin to $-1/Y_f(C)$ in the region of low amplitudes C.

The solution to the optimization problem (4.1) is given by the following result.

Theorem 1.

If α is stable and monic, the criterion (4.1) is minimized by

$$\boxed{F \; = \; \alpha \; = \; AR + BS} \tag{4.2}$$

resulting in

$$y_\delta = \frac{B}{P}\delta \quad ; \quad \mathcal{L}_v = \frac{P - A}{A} \tag{4.3}$$

and P given by the stable and monic solution to the spectral factorization

$$\boxed{rPP^* \; = \; BB^* \; + \; cAA^*} \tag{4.4}$$

Here, r is scalar, while B^* denotes $B(z^{-1})$ in discrete time, and $B(-s)$ in continuous time □

Proof: The result follows from straightforward $\mathcal{H}_2$ minimization with respect to a feedforward filter F/P. It is demonstrated for the discrete–time case in Appendix A, using the variational method of Sternad and Ahlén (1993a). The continuous–time proof is analogous. A somewhat complicating feature in the minimization is that *both* of F and P are required to be monic and stable.

The spectral factorization (4.4) is of the same type as the one used for designing LQ feedback control laws in polynomial form (without integration). See e.g. Åström and Wittenmark (1984), or Sternad (1991). The polynomial P will be stable for all $c > 0$ (and also for $c = 0$, if B has no zeros on the stability limit). The degree condition (2.5), $\deg[F] = \deg[PR]$, is fulfilled for all $c > 0$, since $\deg[P] = \deg[A]$. When $c = 0$, P will be a stabilized version of B; thus, $\mathcal{H}_\delta = B/P$ will be an all–pass link, with norm $\| \mathcal{H}_\delta \|_2^2 = r$.

There do, of course, exist other alternatives for selecting P. A simpler choice, which avoids the need to solve any spectral factorization, is the use of the poles of A, shifted radially by a factor c:

$$P(z) = z^{na} + a_1 c z^{na-1} + \ldots + a_{na-1} c^{na-1} z + a_{na} c^{na} \qquad (4.5)$$

where $na \triangleq deg[A]$. This choice gives a good performance if A is sufficiently well–damped and the zeros of B are not "badly" located, and thus need not be canceled by P. A continuous–time variant is the use of $A(s)$, with poles shifted a distance c into the left half–plane: $P(s) = A(s + c)$. Both the use of (4.4) and (4.5) may, however, give transients $y_\delta = \mathcal{H}_\delta \delta$ which have some overshoot. If this is deemed unacceptable, one may use P–polynomials of order na with real multiple poles

$$P(z) \;=\; (z - c)^{na} \quad \text{or} \quad P(s) \;=\; (s + c)^{na} \;. \qquad (4.6)$$

This choice was suggested in Rönnbäck *et.al.* (1992). Nevertheless, one must be aware that when (4.6) is used, for some systems nonlinear oscillations might occur for any value of c.

4.2 The choice of safety margin for the loop gain

If the loop gain $\mathcal{L}_v$ crosses the negative real axis to the left of -1, limit cycle oscillations may occur, for sufficiently large initial transients. If $\mathcal{L}_v$ does not cross the axis, but come close to it (as in the left–hand part of Figure 4), then it may result in repeated resaturations, due to a damped nonlinear oscillation. The frequency of the oscillation approximately corresponds to the value of ω at which $\mathcal{L}_v(\omega)$ is closest to the describing function. We should therefore strive for some sort of *safety margin* between the loop gain and the describing function.

Wurmthaler and Hippe (1991) have investigated a continuous–time state–space method (it can be shown to be equivalent to the conditioning technique, see Rönnbäck and Sternby (1993)). Based on many simulation experiments, they suggest that the loop gain should avoid a cone with opening angle $\varphi = 40^o$ around the describing function, see Figure 5. In other words,

$$|\arg(\mathcal{L}_v + 1)| < 140^o \;. \qquad (4.7)$$

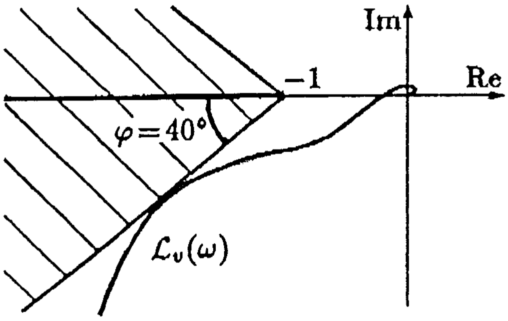

Figure 5: Safety margin as suggested by Wurmthaler and Hippe (1991).

Wurmthaler and Hippe (1991) strive to attain this goal by modifying the whole feedback. In the present scheme, we adjust the anti–windup polynomials only. The parameter c is adjusted until $\mathcal{L}_v$ touches the cone at some point, i.e. $|\arg((P-A)/A)| = 140^o$ for some ω. In our experience, this method often works well, but it also has several drawbacks.

- For some important types of systems, it is impossible to avoid violating the safety margin (4.7). Plants containing double integrators are one example.
- The safety margin seems inadequate around the point -1.[1].
- The safety margin (4.7) is not related to the expected magnitude of the excitations. In frequency regions where disturbances have low amplitude, there is no danger in $\mathcal{L}_v$ coming close to $-1/Y_f(C)$. Even intersections between the loop gain and the describing function could be allowed, if all intersections correspond to excitation amplitudes, C, which never appear in the practical system.

It seems reasonable that the expected excitation magnitude should be taken into account. We have obtained a suitable safety margin by heuristic means.

Assume a symmetrical saturation, and assume that the maximal RMS values of γ and β are known. Their exact spectral distribution is, however, not known. Define the excitation function

$$f(\omega) = \frac{[\| \mathcal{H}_w(\omega) \|_2^2 \, \| \mathcal{W}(\omega) \|_2^2 \, w_{\max}^2 + \| \mathcal{H}_\beta(\omega) \|_2^2 \, \beta_{\max}^2 + \| \mathcal{H}_\gamma(\omega) \|_2^2 \, \gamma_{\max}^2]^{1/2}}{|v_{\max}|} \tag{4.8}$$

where $\beta_{\max}$ and $\gamma_{\max}$ are the maximal RMS values. The function (4.8) indicates the excitation, weighted by the transfer functions (3.8) of the closed–loop system. The signals γ and β are treated as white, with maximum (i.e worst case) power.

[1]This may be less of a problem. By using P from (4.4), the loop gain will tend to keep a distance from -1. This is also the case in Wurmthaler and Hippe (1991), due to their choice of LQ feedback.

The spectral density of the reference, $\| \mathcal{W}(\omega) \|_2^2 \, w_{\max}^2$, may e.g. be white, with RMS value $w_{\max}$, or represent step inputs, with $\mathcal{W}$ being an integrator and $w_{\max}$ being the maximal step amplitude[2].

We now add and subtract a phase margin $\Psi(\omega)$, specified below, to the loop gain.[3] Two loop gain safety limits are then obtained, see Figure 6.

$$\mathcal{L}_{\Psi 1}(\omega) \; = \; |\mathcal{L}_v(\omega)| e^{i[\arg(\mathcal{L}_v(\omega)) - \Psi(\omega)]} \quad ; \quad \mathcal{L}_{\Psi 2}(\omega) \; = \; |\mathcal{L}_v(\omega)| e^{i[\arg(\mathcal{L}_v(\omega)) + \Psi(\omega)]} \; . \quad (4.9)$$

The general properties of the function $\Psi(\omega)$ should be intuitively clear; the higher the possible excitation $f(\omega)$ at ω, the larger safety margin is reqired. There exist many functions which could be used. When working with systems containing at most two integrators, we have obtained excellent results with

$$\Psi(\omega) = \frac{1}{2} \max[\, 0 \, , \arctan(^{10}\log(f(\omega))] \; . \quad (4.10)$$

This function maps $f(\omega) \in]0, \infty[$ onto $\Psi(\omega) \in]0^o, 45^o[$. Higher margins than 45^o are hardly ever required, cf. Wurmthaler and Hippe (1991). We have chosen not to allow negative values of Ψ, since they could lead to intersections between $\mathcal{L}_v$ and the describing function[4].

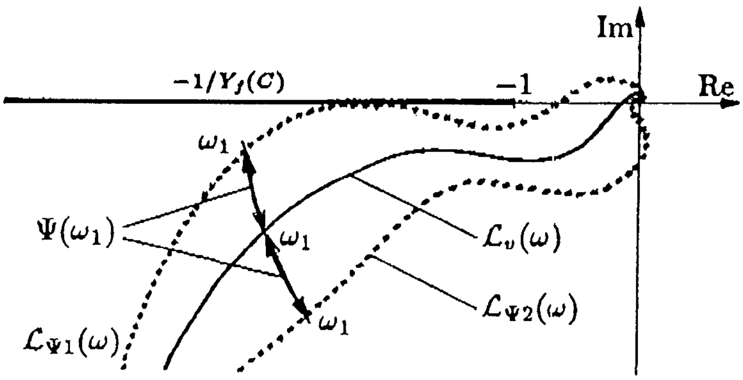

Figure 6: The loop gain $\mathcal{L}_v(\omega)$, the safety limits $\mathcal{L}_{\Psi j}(\omega)$, the phase margin $\Psi(\omega)$ and the describing function $Y_f(C)$. The P–polynomial in (2.4) is adjusted optimally when either one of $\mathcal{L}_{\Psi j}(\omega)$ touches (but does not intersect) $-1/Y_f(C)$.

The properties of this anti–windup design depend on the selected values $w_{\max}, \beta_{\max}$ and $\gamma_{\max}$, as was intended. Since approximations are involved, and the structure

[2]If information about the spectral distribution of the disturbances γ or β is available, this can of course be included. If the saturation is asymmetrical, one could substitute $\min(|v_{\min}|, |v_{\max}|) + \varepsilon$, where $\varepsilon > 0$ to prevent division by zero, for $|v_{\max}|$ in (4.8).

[3]The addition of an amplitude margin would be ineffective; it would shift $\mathcal{L}_v$ mainly along the direction of the describing function, not perpendicular to it Normally, a phase decrease margin is of most interest. For loop gains located in the upper half of the complex plane, a phase *increase* margin is, however, more critical.

[4]For systems containing more that two integrators, such intersections should, however, be allowed, since they are inevitable in some cases. The aim is then to make them occur at amplitudes C which never appear in the system of interest. Thus, oscillations are never excited.

of (4.10) is somewhat arbitrary, these values should be regarded as tuning knobs rather than exact modeling tools. The general design rule is that, if any one of $w_{\max}, \beta_{\max}$ or $\gamma_{\max}$ is increased, then the safety margin will be increased, at relevant frequencies. Thus, a P closer to A will be selected[5].

Let us summarize the class of methods. The simplest variant is to use $F = \alpha$ and P given by (4.5) or (4.6), and then tune the parameter c by utilizing the fixed margin (4.7). Only trivial calculations, and the plotting of the loop gain $\mathcal{L}_v$, is required. A somewhat more elaborate scheme is specified below.

Anti–windup design algorithm.

1. Utilize the controller structure (2.4). Determine R, S, T from linear design considerations.
2. Use $F = AR + BS$, and choose P based on e.g. (4.4)[6]. Define an allowed parameter range $c \in [c_{\min}, c_{\max}]$. Initialize c so that $P \approx A$. Determine reasonable maximal RMS values for the reference w, the disturbance β and the noise γ.
3. Compute the margin $\Psi(\omega)$ from (3.8), (4.8) and (4.10), and the loop gain safety limits $\mathcal{L}_{\Psi i}(\omega)$ from (4.9).
4. Check for intersections between any of $\mathcal{L}_{\Psi i}(\omega)$ and the describing function of the nonlinearity.

By repeating Steps 3 and 4, change the parameter c in the direction of a "faster" $\mathcal{H}_\delta$, until one of the limits $\mathcal{L}_{\Psi j}(\omega)$ touches (but does not intersect) the describing function. If no such bound can be found, (i.e. if all $c \in [c_{\min}, c_{\max}]$ are allowed), we terminate by using the "fastest" P, using $c = c_{\min}$ in (4.4). A Matlab m–file which performs this optimization can be obtained from the authors.

[5]Note that the choice of P does not only affect the loop gain. It also, via (3.8) and (4.8), affects the phase margin $\Psi(\omega)$. Thus, even when a change of the parameter c does not alter the shape of $\mathcal{L}_v$, we may still find a reasonable tradeoff. This would not be the case with the fixed margin (4.7). The linear feedback design also influences the anti–windup design, through (4.2) and via the presence of S in (3.8)

[6]Other constructions of P can be used, such as (4.5) or (4.6). The initialization of c is different in (4.6) ($c =$ "slow value").

5 EXAMPLES

The different anti–windup strategies, introduced in Section 3 and 4, will be compared using three different systems, all controlled by LQG–designed feedbacks R, S, T. For a discussion of LQG design using polynomial methods, see e.g. Sternad (1991). In all examples, we utilize a symmetric nonlinearity, with $v_{\max} = 10$, $v_{\min} = -10$.

Example 1.

For a discrete–time system, with four real poles and two zeros close to the unit circle, a high gain linear feedback has been designed. The polynomials are given in Appendix B. No integration is used in the feedback. The loop gain $\mathcal{L} = BS/AR$ intersects the describing function of the nonlinearity. Thus, limit cycle oscillations result when no antiwindup precaution is used (Figure 7a). In this case, deadbeat antiwindup works well (Figure 7b). The safety margin $\mathcal{L}_{\Psi 1}$ slightly intersects the describing function, but the output y and the saturated input u behave well[7]. The conditioning technique gives almost the same result as deadbeat antiwindup, since the observer dynamics is fast.

Figures 7c,d demonstrate the use of the novel method, with $F = \alpha$ and P from (4.4). When the parameter c is chosen too small, the loop gain comes too close to the describing function. This is indicated by the dotted safety margin in Figure 7c, which intersects the describing function. Damped nonlinear oscillations result. An even smaller value of c would result in limit cycles. When c is adjusted correctly, so that the safety margin $\mathcal{L}_{\Psi 1}$ just touches the describing function, we obtain a good response y and u (Figure 7d). It is similar in this example to that with deadbeat antiwindup (Figure 7b). The use of the fixed margin (4.7) would have resulted in a similar design in this example. In case the overshoots are undesirable, they could be avoided by choosing P according to (4.6) instead of (4.4).

Example 2

The model considered here is oscillative. It represents the belt tension dynamics of a Coupled Electric Drives laboratory process, sampled with a period of 20 ms. For a description of an early version of the process, see Wellstead (1979). A good linear model for one operating point was obtained by system identification, and a belt tension regulator (R, S, T) was designed by LQG optimization. See Appendix B.

Figure 8a exemplifies the result of using model–based anti–windup. Since the open–loop dynamics is oscillative, desaturation transients become oscillative, as (3.13) predicts. The result of using deadbeat anti–windup (Figure 8b) is better.

When the conditioning technique is used (Figure 8c), this results in a limit cycle oscillation. In contrast, the novel method, with $F = \alpha$ and P from (4.4), works

[7] In all figures, the safety margin $\Psi(\omega)$ is computed from the largest step size of w utilized in the simulations: $\beta_{\max} = 0, \gamma_{\max} = 0$ and $\mathcal{W} = 1/(e^{i\omega} - 1)$, with $w_{\max} = 80$ in example 1, $w_{\max} = 40$ in example 2 and $w_{\max} = 25$ in example 3.

well (Figure 8d). The transients resulting from the use of both (4.5) and (4.6) are considerably worse, in this example.

Example 3.

We now consider a model of the temperature dynamics in a chemical analysis instrument, developed by Pharmacia Biosensor in Uppsala, Sweden. In Wågman (1989), the sampled model was obtained by a prediction error identification method, and a temperature controller (R, S, T) with integration was designed by LQG methods. See Appendix B.

To the surprise of the control system designers, limit cycle oscillations were obtained when deadbeat antiwindup was utilized (Figure 9b), while the system was stable without antiwindup (Figure 9a). These effects can, however, be understood by the behaviour of the corresponding loop gains.

When the suggested anti–windup method is utilized, with P either from (4.4) or (4.5), the saturation behaviour becomes satisfactory (Figures 9c and 9d).

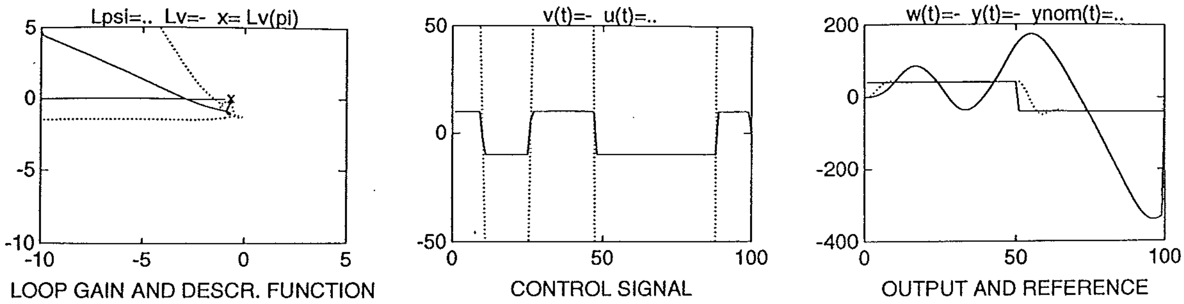

Figure 7a: Example 1, with *no feedback around the saturation* (3.10): $F = R$, $P = 1$. The loop gain intersects the describing function, so we obtain limit cycle oscillations.

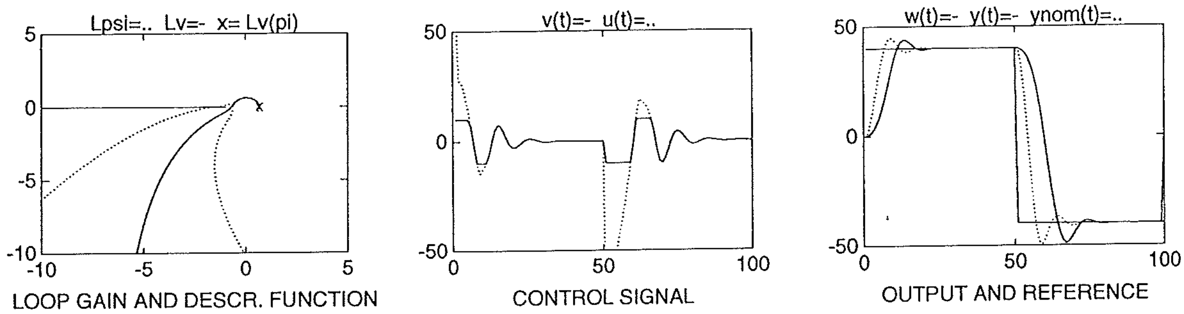

Figure 7b: Example 1, with *deadbeat antiwindup* (3.11): $F = q^{nr}, P = 1$. The step response is satisfactory.

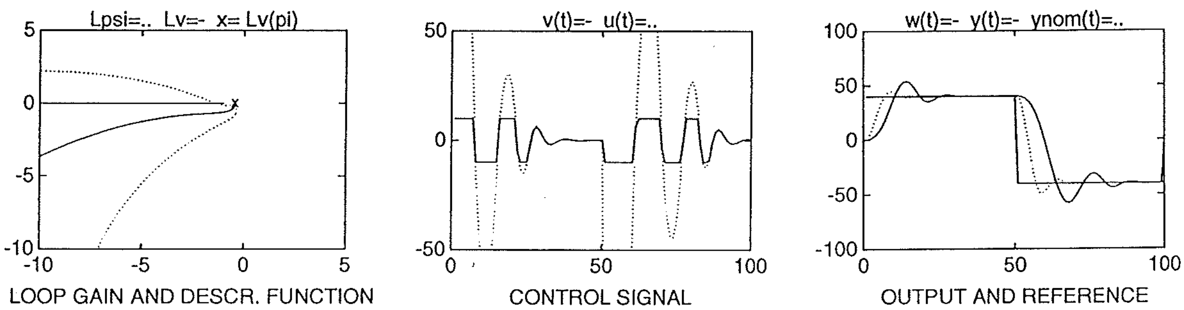

Figure 7c: Example 1, with $F = \alpha$ and P from (4.4), using $c = 0.02$. The loop gain has come too close to the describing function, as indicated by the safety margin. There are damped nonlinear oscillations in the input v.

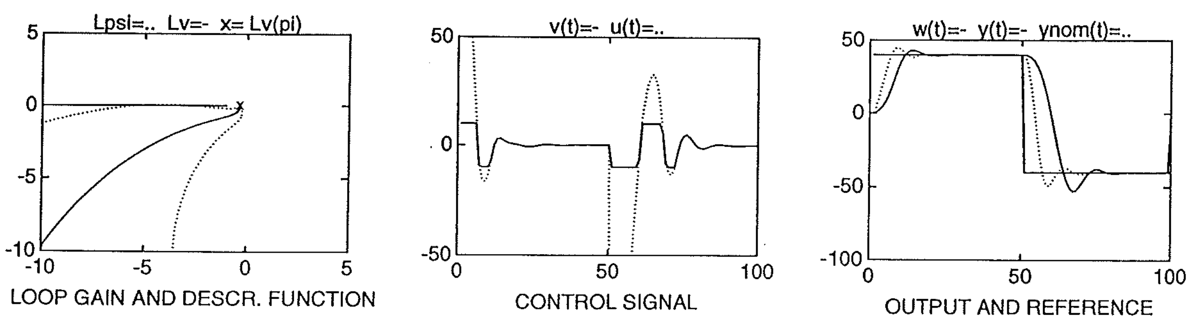

Figure 7d: Example 1, with $F = \alpha$ and P from (4.4), using $c = 0.10$. This is the value for which one of the safety margins (dotted) touches the describing function. The loop gain avoids the point -1 more than in Figure 7b, but the step responses are rather similar.

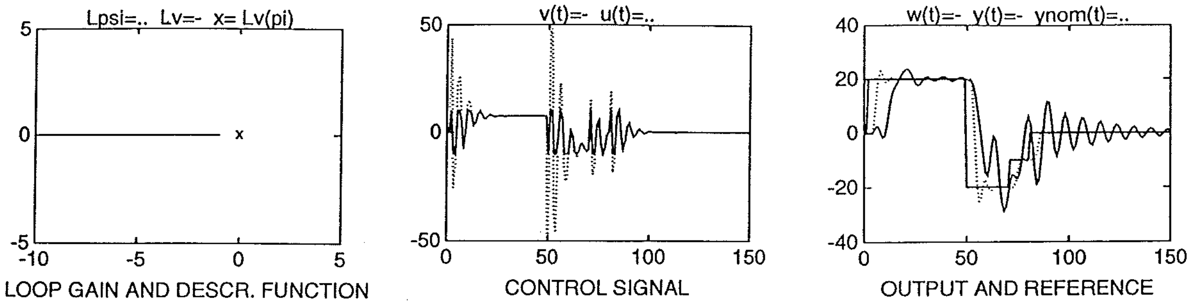

Figure 8a: Example 2, with *model-based anti-windup* (3.14): $F = \alpha, P = A$. Since the open–loop system is oscillatory, the same is true for the transient $y_\delta(t) = y(t) - y_{\text{nom}}(t)$ after desaturation.

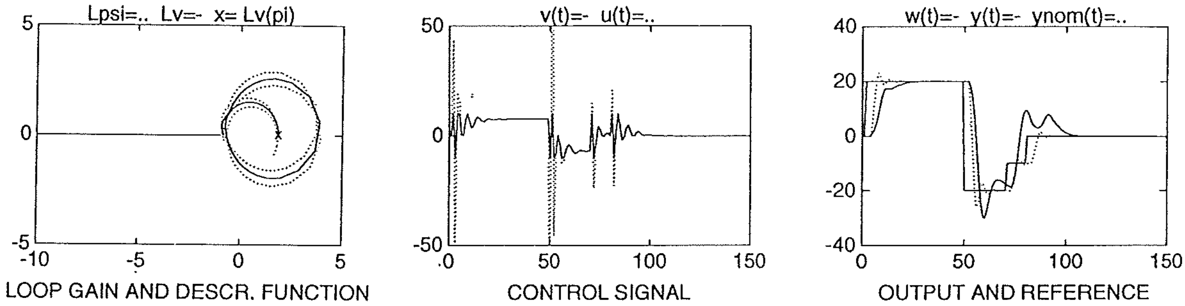

Figure 8b: Example 2, with *deadbeat antiwindup* (3.11): $F = q^{nr}, P = 1$. Since F does not cancel the slow modes of $\mathcal{H}_\delta$, these modes are noticeable in the output.

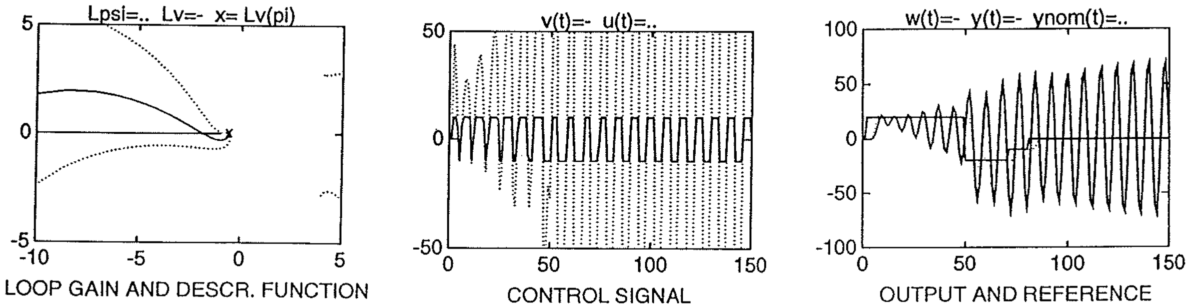

Figure 8c: Example 2, with use of the *conditioning technique* (3.15): $F = T_2/t_0$, $P = 1$. The loop gain intersects the describing function, so we obtain limit cycle oscillations.

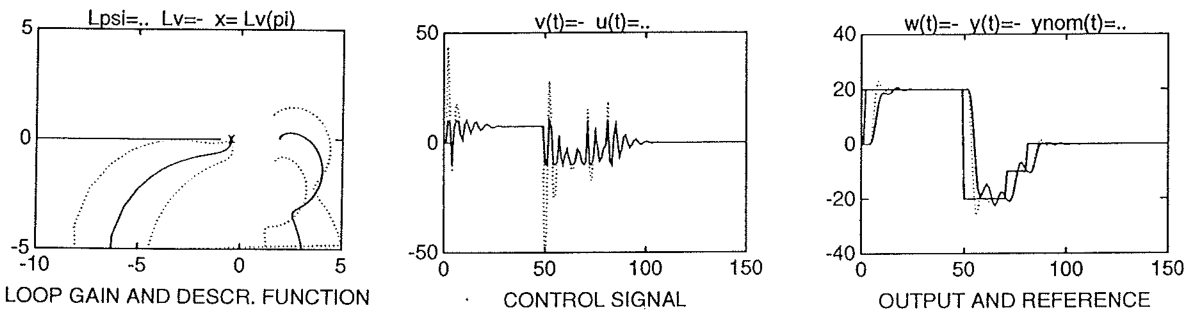

Figure 8d: Example 2, with $F = \alpha$ and P from (4.4), using $c = 0.25$. This is the value for which one of the safety margins (dotted) touches the describing function. The transient $y_\delta(t) = y(t) - y_{\text{nom}}(t)$ after desaturation looks good. The saturated control signal $v(t)$ has some oscillative tendencies, mainly because the control signal would be oscillatory even without saturation.

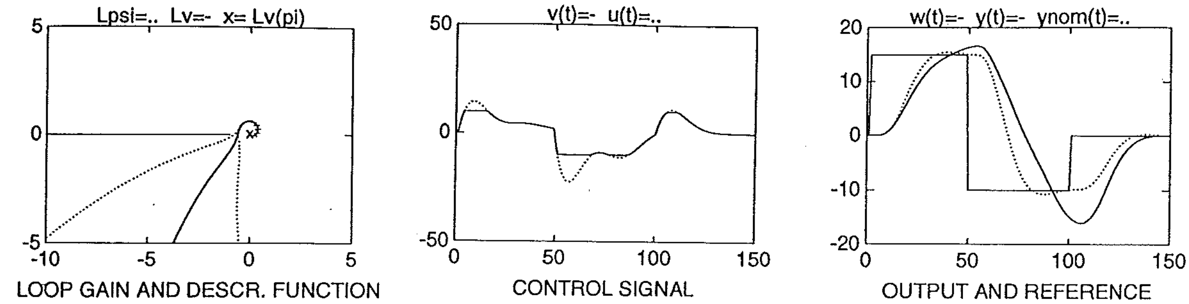

Figure 9a: Example 3, with *no feedback around the saturation* (3.10): $F = R$, $P = 1$. The system is stable, but the desaturation transient is far too slow.

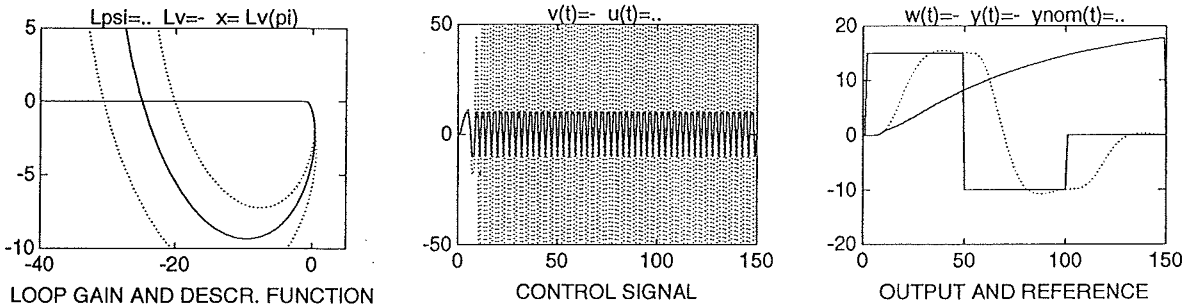

Figure 9b: Example 3, with *deadbeat antiwindup* (3.11): $F = q^{nr}, P = 1$. The loop gain intersects the describing function, so we obtain limit cycle oscillations.

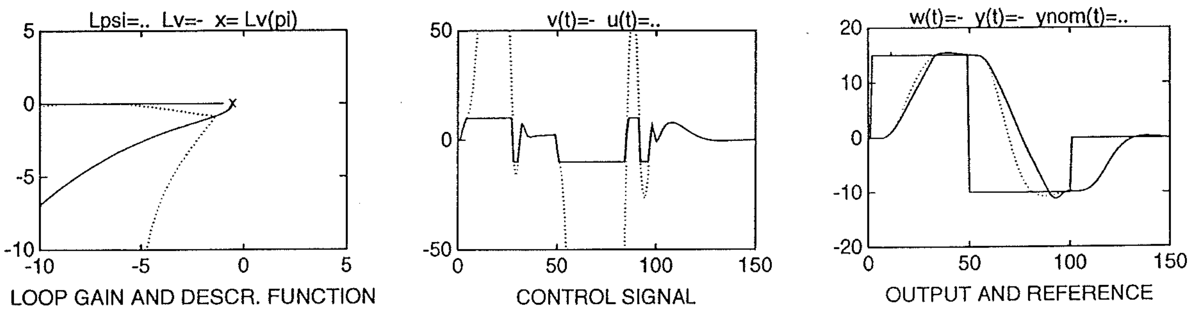

Figure 9c: Example 3, with $F = \alpha$ and P from (4.4), using $c = 0.0003$. One of the safety margins (dotted) touches the describing function of the nonlinearity. The saturated input and resulting output behaves satisfactory.

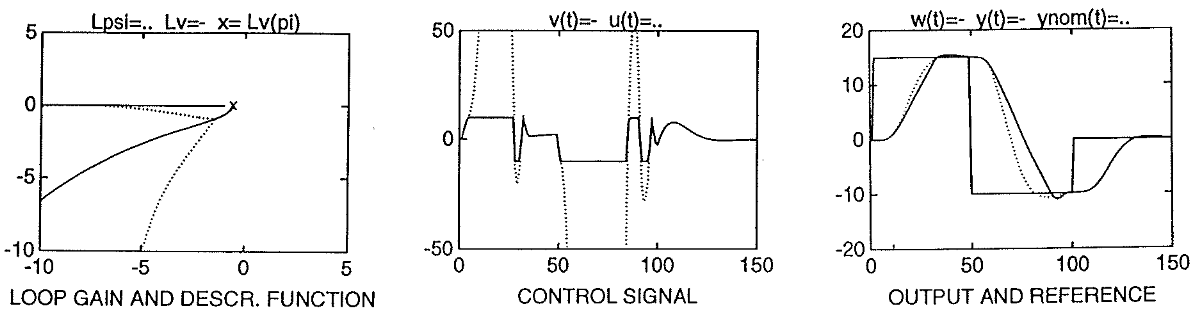

Figure 9d: Example 3, with $F = \alpha$ and P from (4.5), using $c = 0.32$. When this P–polynomial is tuned, using the safety margin, we obtain in this example essentially identical properties as with the use of P from (4.4) (Figure 9c).

6 QUANTIZATION EFFECTS AND MODEL ERRORS

6.1 The control of quantization effects.

When digital controllers are used in servo systems, the finite resolution of analog–digital (A/D) and digital–analog (D/A) converters influence the achievable precision. Feedback systems may amplify quantization effects. The controlled signal may also settle in a stable nonlinear oscillation around the reference value. Describing function analysis can be used to predict such phenomena.

Consider the describing function of a roundoff quantizer, with $2h$ quantization steps of magnitude d, symmetrical with respect to $u = 0$. It is zero for $0 < C < d/2$ and

$$Y_f(C) = \frac{4d}{C}\sum_{i=1}^{n}\sqrt{1-\left[\frac{2i-1}{2C}d\right]^2} \qquad \frac{2n-1}{2}d \; < \; C \; < \; \frac{2n+1}{2}d \; . \tag{6.1}$$

See e.g. Åström and Wittenmark (1984). This function is real-valued, with smallest value zero and largest value $4/\pi \approx 1.27$. Thus, $-1/Y_f(C)$ covers the range $-\infty$ to -0.78. If the control signal saturates when $C > hd$, the describing function changes into that of a saturation for larger C, i.e. $Y_f(C) \to 0$ when $C \to \infty$. From this, the following conclusions can be drawn.

- The regulator structure (2.4) can be used for controlling the effect of not only saturation, but also of quantization, of the control signal. A condition is that the feedback v in (2.4) is the saturated *and quantized* control signal.
- In (3.4), one may use
$$\delta = \delta_s + \delta_q$$
where δ_s is the "disturbance" due to saturation, and δ_q is *quantization noise.* It has rectangular distribution and variance $d^2/12$.
- The analysis in Section 3 and the synthesis procedure in Section 4 remain valid. The only difference is that we should refer all results to a describing function starting in -0.78 instead of in -1. By tuning F and P, we obtain a control system which suppresses the amplification of D/A quantization noise, under the constraint that nonlinear oscillations are avoided.

The situation is different with respect to *quantized measurements* y_m. It is straightforward to show that the polynomials F and P are of no help if unacceptable nonlinear oscillations occur; the design of R and S will then have to be modified.

6.2 Unmodelled dynamics

Errors are unavoidable in linear models. The model errors can be large in some situations of particular interest in an anti–windup design. An example is the transfer between operating points with differing linearized dynamics, if one nominal linear model is used for all points. Assume that the model (2.1) is corrupted by an unknown additive model error

$$\frac{B}{A} = \frac{B_o}{A_o} + \Delta\mathcal{G} \tag{6.2}$$

where B_o/A_o denotes the nominal model and $\Delta\mathcal{G}$ is assumed to be stable. Define the nominal characteristic polynomial

$$\alpha_o = A_oR + B_oS \ . \tag{6.3}$$

Furthermore, assume that a (deterministic or probabilistic) frequency domain bound

$$|\Delta\mathcal{G}(\omega)| \leq |L(\omega)| \tag{6.4}$$

is known for the model error.

If the uncertainty set (6.4) is not too large, the design method of Section 4 can be modified so that nonlinear oscillations are avoided for the whole set. Assume that R and S can be chosen so that the linear dynamics is stable for the whole set (6.4). Thereby choosing $F = \alpha_o$, the set of possible (true) loop gains becomes

$$\begin{aligned} \mathcal{L}_\Delta &\triangleq \frac{P}{\alpha_o}\frac{(AR+BS)}{A} - 1 = \frac{PR}{\alpha_o} + \frac{PS}{\alpha_o}\left(\frac{B_o}{A_o} + \Delta\mathcal{G}\right) - 1 \\ &= \frac{P}{A_o} - 1 + \frac{PS}{\alpha_o}\Delta\mathcal{G} = \mathcal{L}_{vo} + \frac{PS}{\alpha_o}\Delta\mathcal{G} \ . \end{aligned} \tag{6.5}$$

The expression above is affine in the model error $\Delta\mathcal{G}$. It is therefore easy to compute the largest phase shift which can be caused by an admissible model error. This occurs when $(PS/\alpha_o)\Delta\mathcal{G}$ is perpendicular to $\mathcal{L}_{vo} = P/A_o - 1$, pointwise in the complex plane. For $P \neq A_o$, the maximal phase shift is therefore given by

$$\varphi_\Delta(\omega) = \arctan\frac{|PS/\alpha_o||L(\omega)|}{|P/A_o - 1|} \ . \tag{6.6}$$

In the anti–windup design, we simply use this extra safety margin. Thus, (4.9) is modified into

$$\mathcal{L}_{\Psi j}(\omega) = |\mathcal{L}_{vo}(\omega)|e^{i[\arg(\mathcal{L}_{vo}(\omega)) \mp\varphi_\Delta(\omega)\mp\Psi(\omega)]} \quad , \quad j = 1,2 \ . \tag{6.7}$$

When computing the excitation function $f(\omega)$ from (4.8), the worst case gain of B/A should be used in $\| \mathcal{H}_\beta \|_2^2 = \| PS/F \|_2^2 \| B/A \|_2^2$, if a worst case design is desired. It is perhaps more realistic to assume a soft, statistical, bound in (6.4), rather than a hard, guaranteed, bound. We then have to accept a higher risk of unsatisfactory performance of the control system.

There do, of course, exist other possible objectives for optimizing F, P in (2.4) with respect to the set (6.4). A further modification could be to exchange the use of the nominal model in (4.1) for a minimization of the $\mathcal{H}_2$–norm, *averaged* over the set (6.4). The design method for robust feedforward control presented in Sternad and Ahlén (1993b) can be utilized to solve this problem.

7 CONCLUSIONS

We have presented a frequency domain approach to anti–windup compensator design. It utilizes the loop gain, modified by a phase margin $\Psi(\omega)$. That principle is somewhat analogous to the adjustment of linear controllers for unsaturated systems based on phase margin considerations. We have found the methodology to be easy to work with, and to function reliably in simulation examples. The windup problem can indeed be handled well by using simple, essentially linear, concepts.

The describing function was, in general, found to be a remarkably reliable indication of system behaviour, even though it was doubtful if the basic assumptions on which its use is based were strictly fulfilled. However, the method can certainly be improved. We conclude by indicating a number of restrictions, and directions for further research.

- It is straightforward to generalize the method to regulators with multiple measurements, such as state feedbacks. The main principles are outlined in Rönnbäck and Sternby (1993). As long as we have only a single saturated control signal v, both $\mathcal{H}_\delta$ and $\mathcal{L}_v$ remain scalar functions. Generalization to systems with multiple saturated control signals would be more complicated.
- Other variants of the saturation margin (4.8)–(4.10) could certainly be conceived. It would be of interest to compare them with the present suggestion.
- If the saturated signal v utilized by the controller (2.4) is measured directly, it might be corrupted by noise. In cases when it is obtained by using a model of an input nonlinearity $f(\cdot)$, errors in that model could corrupt the result. A useful extension would be to quantify the performance degradation resulting from these two problems.
- Generalization to systems with poles strictly outside the stability limit is of interest. There are two problems here: the presence of the open loop denominator A in $\mathcal{L}_v$, and the fact that stability can be guaranteed only for limited disturbance amplitudes.
- Actuator signals are sometimes not only limited in amplitude, but also limited in rate. The presence of both amplitude and rate constraints would complicate the design considerably.

Finally, since our method is based on a model of the plant, it is not suitable in situations where no model is available. This is e.g. often the case when PID regulators are adjusted. The requirement for a model should not, however, be surprising. The more performance and safety we demand from a design methodology, the more knowledge and insight will the method require from us.

APPENDICES

A Proof of Theorem 1.

Consider the discrete–time case, and let the polynomials be expressed in backward shift operator form, to obtain correspondence with the notation in Sternad and Ahlén (1993a,b). Thus,

$$A(z^{-1}) \;=\; 1 + a_1 z^{-1} + \ldots + a_{na} z^{-na} \quad ; \quad A_*(z) \;\triangleq\; 1 + a_1 z + \ldots + a_{na} z^{na}$$

etc. In the time domain, q^{-1} is substituted for z^{-1}. Define the (artificial) signals

$$y_\delta(t) = \frac{B}{\alpha} m(t) \quad ; \quad z(t) = \frac{A}{\alpha} m(t) - e(t) \quad ; \quad m(t) = \frac{F}{P} e(t) \tag{A.1}$$

where $e(t)$ is white, with unit variance. Minimization of (4.1) then corresponds to minimization of the LQG criterion

$$J \;=\; E[\, y_\delta(t)^2 + c z(t)^2 \,] \tag{A.2}$$

with respect to stable monic polynomials F and P. Any admissible alternative to the signal $m(t)$ can be expressed as

$$m'(t) \;=\; \frac{F'}{P'} e(t) \;=\; \frac{F}{P} e(t) + \eta(t) \quad ; \quad \eta(t) \;=\; q^{-1} \mathcal{M} e(t) \tag{A.3}$$

where $\eta(t)$ is a variational term. Note that the numerator F' is preserved monic if and only if this term includes a pure delay q^{-1}. The rational function $\mathcal{M}$ is stable, and is such that F' has stable zeros, but otherwise it is arbitrary. When (A.3) is substituted into (A.1), this results in the modified signals

$$y_\delta(t) = y_{\delta o}(t) + \Delta y_\delta(t) \quad ; \quad z(t) = z_o(t) + \Delta z(t)$$

where $y_{\delta o}(t)$ and $z_o(t)$ result from the use of (A.1), while $\Delta y_\delta(t)$ and $\Delta z(t)$ are caused by the variation $\eta(t)$. The criterion can then be expressed as

$$J \;=\; J_o + 2J_1 + J_2$$

where

$$J_o = E[y_{\delta o}^2 + c z_o^2] \quad ; \quad J_1 = E[y_{\delta o} \Delta y_\delta + c z_o \Delta z] \quad ; \quad J_2 = E[\Delta y_\delta^2 + c \Delta z^2] \;\; .$$

The aim is now to select F/P so that J_1 vanishes. Then, the filter F/P is optimal; no perturbation $\eta(t)$ could improve the performance, since J_o does not depend on $\eta(t)$ and $J_2 \geq 0$. Using (A.1) and (A.3), the term J_1 can be expressed as

$$J_1 = E\left\{\frac{B}{\alpha}\frac{F}{P} e(t)\right\}\left\{\frac{B}{\alpha} q^{-1} \mathcal{M} e(t)\right\} + cE\left\{\left(\frac{A}{\alpha}\frac{F}{P} - 1\right) e(t)\right\}\left\{\frac{A}{\alpha} q^{-1} \mathcal{M} e(t)\right\}.$$

Since α is stable and P must be stable, Parsevals formula can be used, to give

$$J_1 = \frac{1}{2\pi i}\oint_{|z|=1}\left(\frac{BFB_*z}{\alpha P\alpha_*}\mathcal{M}_* + c\frac{(AF-\alpha P)}{\alpha P}\frac{A_*z}{\alpha_*}\mathcal{M}_*\right)\frac{dz}{z} \quad .$$

By inserting the spectral factorization

$$rDD_* \;=\; BB_* + cAA_* \tag{A.4}$$

we obtain

$$J_1 = \frac{1}{2\pi i}\oint_{|z|=1}\frac{(FrDD_* - c\alpha PA_*)}{\alpha P}\frac{1}{\alpha_*}\mathcal{M}_* dz \quad .$$

Now, $J_1 = 0$ is fulfilled if all poles in $|z| < 1$ of the integrand are canceled by zeros. Since $\mathcal{M}$ and α are stable, $(1/\alpha_*)\mathcal{M}_*$ will have poles only in $|z| > 1$. Since P must be stable, all other poles are in $|z| < 1$. Thus, we require

$$FrDD_* - c\alpha PA_* \;=\; L_*\alpha P \tag{A.5}$$

for some polynomial $L_*(z)$. There are three undetermined polynomials in this equation: $F(z^{-1}), P(z^{-1})$ and $L_*(z)$. It is obvious from (A.5) that α must be a factor of $FrDD_*$. Since rDD_* is determined, this can only be fulfilled if

$$F \;=\; \alpha \quad . \tag{A.6}$$

Since α is monic and stable, such an assignment is always admissible. Canceling $F = \alpha$ in (A.5) reduces that expression to

$$rDD_* \;=\; P(cA_* + L_*) \quad .$$

Now, P must be a factor of rDD_*. Since P must be stable and monic, we obtain

$$P \;=\; D \tag{A.7}$$

from which it can also be concluded that $L_*(z) = rD_*(z) - cA_*(z)$. With (A.4),(A.6) and (A.7), the theorem has been proved. Multiply $D(z^{-1}), B(z^{-1})$ and $A(z^{-1})$ in (A.4) by z^{na} and $D_*(z), B_*(z)$ and $A_*(z)$ by z^{-na}. We then obtain the spectral factorization (4.4), in which polynomials are defined as $A = z^{na} + \ldots + a_{na}$, $A^* = z^{-na} + \ldots + a_{na}$ etc. □

B Systems and regulators utilized in Section 5

Example 1:

Discrete–time systems, in backward shift operator form:

$$B(q^{-1}) \;=\; 0.0625q^{-1} - 0.0625q^{-2} + 0.038125q^{-3}$$

$$A(q^{-1}) \;=\; 1 - 3.100q^{-1} + 3.560q^{-2} - 1.796q^{-3} - 0.3360q^{-4} \quad .$$

The system has zeros in $z = 0.5 \pm 0.6i$ and poles in $z = 0.6, z = 0.7, z = 0.8$ and $z = 1$. The utilized regulator has rather high gain:

$$
\begin{aligned}
S(q^{-1}) &= 25.382 - 65.189q^{-1} + 64.672q^{-2} - 29.035q^{-3} + 4.9414q^{-4} \\
R(q^{-1}) &= 1 - 1.8349q^{-1} + 1.4179q^{-2} - 0.5830q^{-3} \\
T(q^{-1}) &= 1.5746 - 0.94478q^{-1} + 0.14172q^{-2} \quad .
\end{aligned}
$$

It results in pole placement in $z_{1,2} = 0.68 \pm 0.46i$, $z_{3,4} = 0.48 \pm 0.24i$, $z_4 = 0.42$, $z_{5,6} = 0.30$. The poles in 0.30 are observer poles. They are canceled by the polynomial T in the reference input.

Example 2:

Discrete–time system, in backward shift operator form:

$$
\begin{aligned}
B(q^{-1}) &= 0.1189q^{-3} + 0.0095q^{-4} + 0.0879q^{-5} \\
A(q^{-1}) &= 1 - 2.9782q^{-1} + 3.8606q^{-2} - 2.4964q^{-3} + 0.6677q^{-4} \quad .
\end{aligned}
$$

The regulator was obtained by LQ pole placement with integration, using an input penalty of 0.01 and an observer with observer dynamics $T_2 = (1 - 0.6q^{-1})^4$:

$$
\begin{aligned}
S(q^{-1}) &= 0.2193 - 2.2156q^{-1} + 4.9970q^{-2} - 4.3614q^{-3} + 1.4167q^{-4} \\
R(q^{-1}) &= 1 - 0.8141q^{-1} + 0.6294q^{-2} - 0.5505q^{-3} - 0.0722q^{-4} - 0.1927q^{-5} \\
T(q^{-1}) &= T_2(q^{-1})/S(1) \quad .
\end{aligned}
$$

The polynomial $R(z^{-1})$ has roots in $z = 1$, $z = 0.16 \pm 0.82i$ and $z = -0.25 \pm 0.46i$.

Example 3:

Discrete–time system, in backward shift operator form:

$$
\begin{aligned}
B &= 0.0099051q^{-5} - 0.0055312q^{-6} + 0.0009423q^{-7} \\
A &= 1 - 2.25837q^{6} + 1.37059q^{5} + 0.10603q^{4} - 0.21743q^{3} \quad .
\end{aligned}
$$

The regulator was obtained by LQ pole placement with integration, combined with an observer with multiple pole in $z = 0.8$, of order 6.

$$
\begin{aligned}
S(q^{-1}) &= 0.00835407 - 0.0186703q^{-1} + 0.0111755q^{-2} + 0.000948547q^{-3} - 0.00178950q^{-4} \\
R(q^{-1}) &= (1 - q^{-1})(1 - 4.597380q^{-1} + 8.849137q^{-2} - 9.118080q^{-3} \\
&\quad +5.301231q^{-4} - 1.648092q^{-5} + 0.21396777q^{-6}) \\
T(q^{-1}) &= 0.274015 - 1.315270q^{-1} + 2.63054q^{-2} - 2.80591q^{-3} + 1.683545q^{-4} \\
&\quad -0.5387347q^{-5} + 0.07183129q^{-6} \quad .
\end{aligned}
$$

REFERENCES

Åström, K.J. and B. Wittenmark (1984)
Computer–Controlled Systems. Prentice–Hall, Englewood Cliffs, NJ. (2nd ed. 1990)

Atherton, D.P. (1975)
Nonlinear Control Engineering Van Nostrand, London, UK.

Garcia C. E., D. M. Prett and M. Morari (1989)
Model predictive control and practice - A survey. *Automatica*, vol 25, pp. 335–348.

Gibson, J.E. (1963)
Nonlinear Automatic Control McGraw–Hill, New York, NJ.

Hanus, R., M. Kinnaert and J–L. Henrotte (1987)
Conditioning technique, a general anti–windup and bumpless transfer method. *Automatica*, vol. 23, pp. 729–739.

Hanus, R. (1988)
Anti–windup and bumpless transfer: a survey. *Proc. of the 12th IMACS World Congress* on Scientific Computation, Paris, France, Part 2, pp. 59–65.

Horowitz, I. (1983)
A synthesis theory for a class of saturating systems. *International Journal of Control*, vol. 38, pp. 169–187.

Lozier J. C. (1981)
On feedback control systems with saturable drives and noisy measurements. *IEEE Transactions on Automatic Control*, vol 26, pp. 321–331.

Middleton R.H. and G.C. Goodwin (1990)
Digital Control and Estimation–A Unified Approach. Prentice–Hall International, London.

Morari, M. and E. Zafiriou (1989)
Robust Process Control. Prentice–Hall, Englewood Cliffs, NJ.

Rundqwist, L. (1991)
Anti–Reset Windup for PID Controllers. PhD Thesis, Lund Institute of Technology, Sweden.

Rönnbäck, S., K.S. Walgama and J. Sternby (1992)
An extension to the generalized anti–windup compensator. *Mathematics of the Analysis and Design of Process Control*, P. Borne et. al. (eds.), Elsevier, North–Holland, 1992, pp. 275–285.

Rönnbäck, S. and J. Sternby (1993)
Anti–windup in state–space controllers. Submitted to the *1993 IEEE Conference on Decision and Control.*

Sternad M. (1991)
Use of disturbance measurement feedforward in LQG self-tuners. *International Journal of Control*, vol 53, pp. 579–596.

Sternad M. and A. Ahlén (1993a)
A novel derivation methodology for polynomial–LQ controller design. *IEEE Transactions on Automatic Control*, vol 38, pp. 116–121.

Sternad M. and A. Ahlén (1993b)
Robust filtering and feedforward control based on probabilistic descriptions of model errors. *Automatica*, to appear, vol 29, no 3, May.

Tsirukis, A.G. and M. Morari (1992)
Controller design with actuator constraints. Proceedings of the *31st Conference on Decision and Control*, Tucson, AZ, December 1992.

Wågman, C. (1989)
Optimal control of a multivariable system. Report UPTEC 89106E, M.S. Thesis, Department of Technology, Uppsala University, Sweden.

Walgama, K.S. (1991)
On the control of systems with input saturation or periodic disturbances. PhD Thesis, Luleå University of Technology, Sweden.

Walgama, K.S. and J. Sternby (1990)
Inherent observer property in a class of anti–windup compensators. *International Journal of Control*, vol. 52, pp. 705–724.

Wellstead, P.E. (1979)
Introduction to Physical System Modeling. Academic Press, London, UK.

Wurmthaler, C. and P. Hippe (1991)
Systematic compensator design in the presence of input saturation. *ECC'91, European Control Conference*, Grenoble, France, July 2–5, 1991, pp. 1268–1273.