Bode's integral theorem for discrete-time systems

C. Mohtadi, MA, DPhil

Indexing terms: Control systems, Control theory, Feedback, Mathematical techniques

Abstract: A paper by Bode has shown the limitations of using a feedback structure in terms of an integral constraint on the sensitivity function for open-loop stable continuous-time systems. The paper by Mohtadi examines and derives equivalent results for discrete-time feedback systems. These integral constraints also provide some guidelines regarding the philosophy of feedback design specifically for sampled-data systems. For example, it is shown that, for all sampled-data control systems, there is a maximum sampling frequency, beyond which little improvement in performance is gained.

1 Introduction

Whether an advocate of any of the multitude of frequency-response approaches for feedback control systems or a worker in the field of the so-called 'optimal' or predictive methodologies, the control engineer must be aware of the limitations imposed by the choice of a feedback structure for control. Frequency, these rules are forgotten leading to designs which have extraordinary properties for the nominal case, and which are extremely sensitive to arbitrarily small perturbations causing instabilities, albeit to the amazement of their designers!

Surprisingly, Bode [1] was the first to recognise and acknowledge this limitation expressed in the form of an integral constraint on the sensitivity function:

In a single-loop feedback amplifier of more than one stage, the average regeneration and degeneration over the complete frequency spectrum is zero (see Reference 1, page 285).

In mathematical terms, this reduces to

$$\int_0^\infty \ln|\sigma|\, d\omega = 0$$

for systems with at least 2-pole roll off, where σ is the sensitivity function [1, 2] and ω is the frequency. For a brief description of sensitivity see Section 2. Loosely speaking, this theorem implies that:

(*a*) we cannot have a sensitivity less than unity at all frequencies using output feedback with finite-bandwidth controllers

(*b*) combined with the open-loop roll-off requirements for stability, the primary cost of feedback is in increased sensitivity at high frequencies.

Subsequently, only a few design methodologies appear to have considered these constraints explicitly [2, 3, 4]. It is only in recent years that Freudenberg and Looze [5, 6, 7] have obtained new integral constraints for the general case of open-loop stable/unstable continuous-time systems. Although these results do not provide us with a specific design methodology, they do explain the reasons for failures of some designs and give insight as to how we should tackle the feedback configuration. In this paper, we attempt to translate these integral constraints to the discrete-time case. Although most of the results are direct analogues of the continuous-time case, there are some differences between the discrete and the continuous counterparts. In addition, it is shown that some of the practices, such as sampling at about 1/10th of the dominant time constant of the process, can be explained satisfactorily using these integrals.

Paper 7114D (C8), first received 31st January and in revised form 13th October 1989

The author is with the Department of Engineering Science, Oxford University, Parks Road, Oxford, OX1 3PJ, United Kingdom

2 Preliminaries

2.1 Poisson's integral theorem

For any function $f(re^{j\theta})$, where $re^{j\theta}$ is a complex variable in polar co-ordinates, we have

$$f(re^{j\theta}) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-1)f(e^{j\phi})\, d\phi}{r^2+1-2r\cos(\theta-\phi)} \tag{1}$$

where $f(.)$ is analytic outside the unit circle and $re^{j\theta}$ is a complex variable outside the unit circle. The function can have zeros on the unit circle. The proof of this via Cauchy's residue theorem is given in Reference 9. The power of this relation is in the fact that the weighted contour integral of $f(.)$ is only related to the value of the function at the chosen point $re^{j\theta}$. This relationship is used in the following Sections to establish the properties of stable closed-loop transfer-functions (i.e. analytic outside the unit circle).

2.2 Sensitivity and complementary sensitivity

Consider the discrete-time two-degrees-of-freedom SISO feedback system of Fig. 1. The first step in the design of

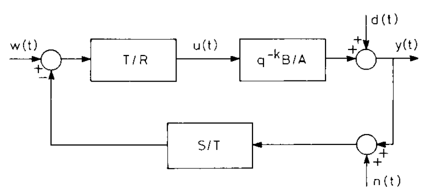

Fig. 1 *A two degrees of freedom pole-placement controller*

the controller is to compensate for the disturbances $d(t)$ and noise $n(t)$ taking into account the required robust stability margins. The servo properties are then adjusted using open-loop (or feedforward) compensation with a prefilter $T(q^{-1})$ [2, 10, 11].

For the purposes of this paper, a standard pole-placement controller, where the only variables are the closed-loop pole-positions, is arbitrarily adopted. The control is of the form:

$$Tw(t) = Ru(t) + Sy(t)$$

where R and S are polynomials in q^{-1}, the backward shift operator, and the plant model is given by

$$Ay(t) = Bu(t-k)$$

where A and B are further polynomials in q^{-1}, k is the plant dead time in samples and the designer only chooses $P(q^{-1})$, where

$$P = RA + q^{-k}BS$$

This fixes the R and S polynomials for the specific case, where $\delta R = \delta B + k$ and $\delta S = \delta A$ and R will always have a zero at 1, corresponding to the integral mode of the controller (δR denotes the degree of the R polynomial). The choice of pole-placement is simply to indicate some of the issues necessary for design, but, as such, is not an ideal one. In most cases, we are interested in the so-called cost of feedback and we have to examine the minimum necessary cost which has to be paid (i.e. minimize the maximum deviations). This necessitates an H^∞ design (see e.g. Reference 12). Here we are only interested in a qualitative analysis and, as such, a pole-palcement design will suffice to demonstrate the basic properties. We have

$$\begin{aligned} y(t) &= -\frac{q^{-k}BS}{RA + q^{-k}BS}\, n(t) + \frac{RA}{RA + q^{-k}BS}\, d(t) \\ &= -\tau(q^{-1})n(t) + \sigma(q^{-1})\, d(t) \end{aligned}$$

where σ is the sensitivity function and τ the complementary sensitivity function (i.e. $\tau + \sigma = 1$ at all frequencies). $d(t)$ is usually used to model low-frequency disturbances; such as load changes, offsets, feed variations, slow environmental changes (e.g. changes in the ambient temperature and pressure), as well as disturbances such as friction and stiction. The regulator is designed to remove these disturbances and therefore we require $|\sigma(q^{-1})|$ to be small at these frequencies. $n(t)$ is usually used to model higher frequency disturbances, such as measurement noise arising from 'bouncy' pressure sensors, poor electrical connections, cross talk etc. We do not wish the control system to react to these disturbances. This implies $|\tau(q^{-1})|$ should be small around these frequencies. $n(t)$ is also used to describe unmodelled dynamics (i.e. the unmodelled dynamics is reflected at the output of the system). Consider the case where the 'real' plant is given by $M(q^{-1})$. Defining $\tilde{M} = M - q^{-k}B/A$ gives

$$y(t) = \sigma(q^{-1})\, \frac{1}{1 + \tau\tilde{M}\Big/\dfrac{q^{-k}B}{A}}\, d(t)$$

which implies that the extra term must satisfy the usual Nyquist criterion or the appropriate version of the small gain theorem, if it is nonlinear or time-varying. τ therefore can also be considered as a measure of robustness of the system to unmodelled dynamics, in that large values of $|\tau|$ at high frequencies may cause instability. Note that $|\tau| > 1$ usually implies $|\sigma| > 1$, despite the fact that $\sigma + \tau = 1$, the magnitude of both quantities can be large, as they can both have large imaginary parts. Recall that $|\sigma|$ is the inverse of the distance of the open-loop Nyquist plot from the critical point $(-1, 0)$. Therefore, large values of $|\sigma|$ or $|\tau|$ indicate a poor design. These concepts are traditionally expressed as gain and phase margins, but it can be argued that, under some perverse conditions (e.g. see Fig. 2), $|\sigma|$ is a better indicator. Note

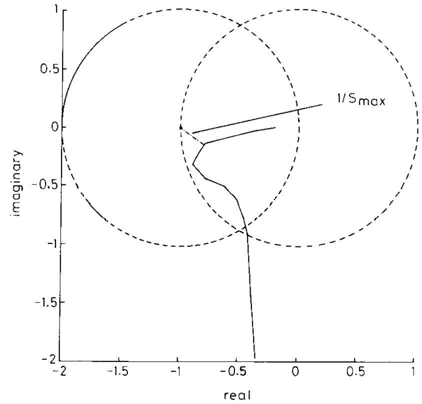

Fig. 2 *Nyquist plot of a system*

that, although we have reasonable gain and phase margins, the distance of the open-loop Nyquist plot from the critical point is quite small.

There are two basic requirements which have to be satisfied:

(*a*) $\sigma + \tau = 1$, we cannot choose these two quantities independently frequency by frequency

(*b*) We require the closed-loop to be internally stable. This means that $\sigma(re^{j\theta}) = 0$ at the open-loop unstable poles of the system (for poles of multiplicity n up to the $n-1$th derivative are also zero) and $\tau(re^{j\theta}) = 0$ at the nonminimum-phase zeros of the system (for zeros with multiplicity m up to the $m-1$th derivative are also zero). This circumvents pole/zero cancellations outside the unit circle.

The first constraint is referred to as the *algebraic* design tradeoff by Freudenberg and Looze [7]. Unfortunately, τ and σ cannot be chosen arbitrarily because of another tradeoff: the *analytic* design tradeoff where the choice of σ at one frequency affects its choice at another. This is the result of requiring P to have roots inside the unit circle (in conjunction with the requirement of internal stability). The remainder of the paper is devoted to this aspect.

3 The basic result

Consider the sensitivity function $\sigma(q^{-1})$. Assuming that the nominal design is stable, we can split σ into two components:

$$\sigma(q^{-1}) = \tilde{\sigma}(q^{-1})\bar{\sigma}(q^{-1})$$

where $\tilde{\sigma}$ is analytic outside the unit circle and $\bar{\sigma}$ is all-pass (i.e. it has a gain of unity at all frequencies up to the Nyquist). From Poisson's integral theorem, we have

$$\ln(\tilde{\sigma}(re^{j\theta})) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-1)\ln(\tilde{\sigma}(e^{j\phi}))\, d\phi}{r^2 + 1 - 2r\cos(\theta - \phi)} \tag{2}$$

for any $r > 1$. Recall that

$$\ln(\bar{\sigma}) = \ln(|\bar{\sigma}|) + j\angle\tilde{\sigma}$$

Note that the integral equation above is valid for any point outside the unit circle. At the NMP zeroes of the system however, $\tau(re^{j\theta}) = 0$. This means that $\sigma = 1$ at these points (this is the requirement of internal stability). This leads to the following theorem:

Theorem 1: For any zero $re^{j\theta}$ of the open-loop transfer function outside the unit circle, the sensitivity function must satisfy the following integral constraints:

$$\ln(|\bar{\sigma}(re^{j\theta})^{-1}|) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-1)\ln(|\sigma(e^{j\phi})|)\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

$$\left.\frac{d^i}{ds^i}\ln(\bar{\sigma}(s)^{-1})\right|_{s=re^{j\theta}} = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-1)}{r^2+1-2r\cos(\theta-\phi)} \left.\frac{d^i \ln(\bar{\sigma}(s))}{ds^i}\right|_{s=e^{j\phi}} d\phi \quad \text{for } i = 1, \ldots, m-1$$

$$\angle\bar{\sigma}(re^{j\theta})^{-1} = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-1)\angle\bar{\sigma}(e^{j\phi})\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

where m is the multiplicity of the zero.

Proof: Follows the proof of Freudenberg and Looze [5]. Equate the real and imaginary parts of the logarithm, recalling that $\sigma = 1$ at the nonminimum-phase zero, its derivatives up to the order $m - 1$ are equal to zero and that $|\sigma|$ is unity evaluated at frequencies round the unit circle.

A similar result can be achieved for the complementary sensitivity function using a similar set of arguments as above. Rewriting τ as $q^{-k}\bar{\tau}\tilde{\tau}$ leads to the second theorem on the complementary sensitivity function.

Theorem 2: For any pole $re^{j\theta}$ of the open-loop transfer function outside the unit circle, the complementary sensitivity function must satisfy the following integral constraints:

$$\ln(|\bar{\tau}(re^{j\theta})^{-1}|) + k\ln(r) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-1)\ln(|\tau(e^{j\phi})|)\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

$$\left.\frac{d^i}{ds^i}\ln(\bar{\tau}(s)^{-1})\right|_{s=re^{j\theta}} = \frac{1}{2\pi}\int_0^{2\pi} \times \frac{(r^2-1)}{r^2+1-2r\cos(\theta-\phi)} \left.\frac{d^i \ln(\bar{\tau}(s))}{ds^i}\right|_{s=e^{j\phi}} d\phi \quad \text{for } i = 1, \ldots, n-1$$

$$\angle\bar{\tau}(re^{j\theta})^{-1} + k\theta = \frac{1}{2\pi}\int_0^{2\pi} \frac{(r^2-1)\angle\bar{\tau}(e^{j\phi})\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

where n is the multiplicity of the open-loop unstable pole.

Proof: Immediate as with the sensitivity function.

Note that, in each case, the integral relations have to be satisfied for every nonminimum-phase zero or unstable pole of the system. To ensure internal stability, both theorems have to be satisfied.

4 Interpretations and extensions

In this Section, the properties and consequences of the integral results of the preceding Section will be examined.

4.1 Effect of the weighting

Consider the magnitude integral of theorem 1, the term $\ln|\sigma|$ is weighted by

$$\frac{r^2-1}{r^2+1-2r\cos(\theta-\phi)}$$

where $re^{j\theta}$ is a nonminimum-phase zero of the system. Fig. 3 shows the variation of this weight with r and ϕ. θ was arbitrarily chosen as $\pi/4$.

Note that the peak magnitude is at $\pi/4$ as ϕ varies, and also that the sharpness (i.e. bandwidth) of the weight decreases significantly as r approaches unity. This implies that $\ln|\sigma|$ is weighted heavily and its value contributes to the integral, when

(i) $\phi \simeq \theta$ (i.e. at frequencies corresponding to those of the NMP zeros)

(ii) $r \to 1$ (i.e. when the NMP zeros are close to the unit circle).

A natural consequence of this observation is that if, for one reason (e.g. performance requirements), we force $|\sigma|$ to be small at these critical frequencies, then the price paid at the other frequencies where these weights are small is such that $|\sigma|$ would have to be some orders of magnitude larger than unity. This distance of NMP zeros, if not taken into account, could lead to very poor feedback systems. It is, therefore, important to relax requirements on σ within the bandwidth of the weight. The following example clarifies this point. Consider the system:

$$(1 - 0.9q^{-1})y(t) = (-1 + 1.1q^{-1})u(t-1)$$

with two sets of pole positions $\{0.5, 0.5\}$ and $\{0.8, 0.8\}$. The solid line in Fig. 4 shows the variation of the log sensitivity with frequency, of the poles at 0.5, and the broken lines are for the case with the poles at 0.8. Clearly, relaxing the bandwidth requirement improves the sensitivity at higher frequencies.

The bandwidth of the weight for NMP zeros on the real axis can be approximated by (using statistical analogies)

$$\theta_{BW} = \frac{1}{\pi}\int_0^{\pi} \frac{\theta(r^2-1)}{r^2+1-2r\cos(\theta)}\,d\theta$$

The bandwidths for different values of r given in Table 1.

Table 1: Variation of bandwidth with zero position on the real axis

r	θ_{BW}, rad
1.05	.1462
1.15	.3257
1.25	.4537
1.50	.6712
2.00	.9147

4.2 Pole-zero cancellations outside the unit circle

It is well known that pole-zero cancellations outside the unit circle lead to internally unstable loops. However, it is sometimes hard to quantify the effect of close pole-zero cancellations outside the unit circle. Theorem 1 gives an indication of the weighted average effect of these close cancellations. Recall that the value of the integral is equal to $\ln|\bar{\sigma}(e^{j\phi})^{-1}|$ and $\bar{\sigma}$ has all of the open-loop unstable poles as zeros. Clearly, this quantity approaches infinity as the cancellations become exact. The bounds on the magnitude of the sensitivity function clarify this point further. This is yet another reason why such cancellations should be avoided. Consider the system:

$$(1 = 0.9q^{-1})(1 - 1.1q^{-1})y(t) = 0.1(1 - 1.11q^{-1})u(t - 1)$$

Two cases are considered:

(i) pole-placement with the approximate common factors

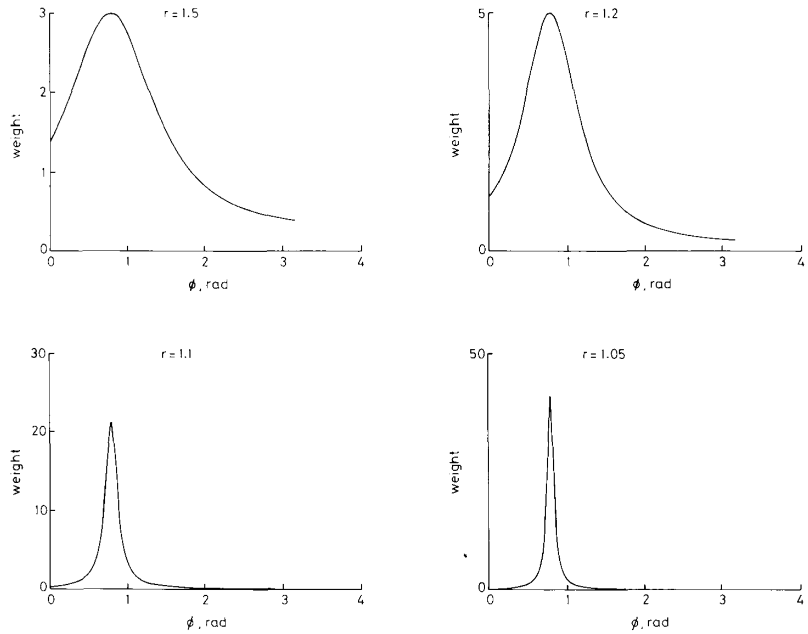

Fig. 3 *Variation weights with r and ϕ*

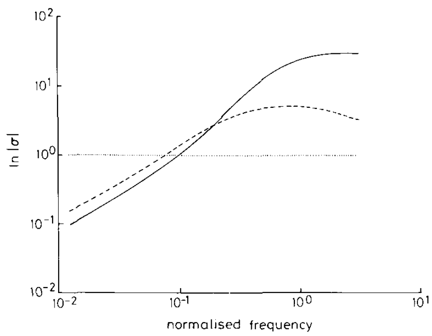

Fig. 4 *Variation of sensitivity with pole position*

(ii) pole-placement without the approximate factor present.

The solid line in Fig. 5 shows the variation of the sensitivity when the common factor is present, for a choice of two poles at 0.8 and the dashed line for the case when the factor is absent. Note the increase in sensitivity by more than an order of magnitude over the entire frequency range!

4.3 Bode's integral theorem for discrete-time systems

The integral constraint of theorem 1 is valid for all NMP zeros of the system, including those at infinity (i.e. due to time delay). As, for discrete-time system, k is at least 1, we have $r \to \infty$ for this NMP zero. Also, as $|\sigma|$ is also symmetric about $\phi = \pi$, we have the following corollary of theorem 1.

Corollary 1: For all closed-loop stable discrete-time feedback systems, the sensitivity function has to satisfy the following integral constraint:

$$\sum_{i=1}^{m} \ln |\beta_i| = \frac{1}{\pi} \int_0^{\pi} \ln |\sigma(e^{j\phi})| \, d\phi$$

where β_i are the open-loop unstable poles of the system and m is the total number of these poles.

Proof: By substitution in the integral of theorem 1.

This result is the discrete analogue of the result of Freudenberg and Looze [5], and, in effect, is the Bode integral theorem for a general sampled-data system. Recall that $\phi = \omega h$, where h is the sample time and ω is the frequency in radians per second. It is then easy to see that, as $h \to 0$, the result of corollary 1 converges to the continuous-time result (this requires a change of variable from ϕ to ω). However, there are some points which are different between the two results:

(*a*) The upper limit of frequency in the integral of corollary 1 is π/h, the Nyquist frequency. This is because, implicit in the discrete-time design, it is assumed that there are no signals above this frequency floating in the loop. Our ideal requirement is therefore, that $|\sigma| = 1$ for all frequencies above the Nyquist

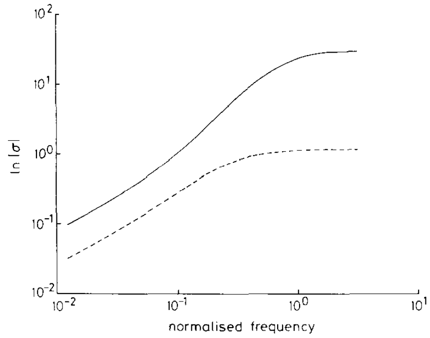

Fig. 5 *Approximate common factors outside the unit circle*

(*b*) As the upper frequency is π/h for all $h > 0$, this implies that if for some frequency $|\sigma| < 1$, then for the other range $|\sigma| > 1$ and its value cannot be arbitrarily close to unity, as only a finite bandwidth is available, unlike the continuous-time case, where $|\sigma|$ can be spread over a wide range of frequencies (other constraints not taken into account)

(*c*) It is well known that the 'optimal' controller using full-state feedback in continuous time is such that $\ln|1 - \tau| < 0$ at all frequencies [13]. The recent results in loop transfer recovery [14, 15] attempt to asymptotically achieve this property by observer state feedback for minimum-phase systems. The net result of observation (*b*) is that these recovery methods cannot be achieved in discrete time: the approximation becomes better as the sampling time decreases. Any serious application of these methods should therefore consider the effects of sampling

(*d*) In the standard result, for the continuous-time case, at least a 2-pole roll-off is required which is not required in the discrete-time result. This is due to the presence of the zeroth-order hold whose frequency response is given by $he^{-j\omega h/2} \sin(\omega h/2)/(\omega h/2)$. Freudenberg and Looze, in their paper [6], show that for time-delay systems the 2-pole roll-off requirement can be relaxed to a single-pole roll off. The zeroth-order hold accounts for a time delay of $h/2$ and the roll off. Our observation here, therefore, is in direct agreement with their result.

4.4 Bounds on the sensitivity function

A typical sensitivity function plot against frequency may look something like that of Fig. 6. The maximum value of the sensitivity function σ_{max} is a measure of how good (or poor) the performance of a control system is. It is useful to find some bounds on this maximum, to consider such concepts as achievable performance in a particular design.

Consider Fig. 6, assume that, for some performance requirements, we want:

(*a*) $|\sigma| < \alpha$, for frequencies below ω_1

(*b*) $|\sigma| \simeq 1$, for frequencies above ω_2.

The 'real' sensitivity function may be that given in Fig. 6, but we make the simplifying assumption that $\ln|\sigma|$ is the following the broken line. Clearly, S_m is a lower bound to S_{max}. Using theorem 1, it is easy to show that

$$S_m = |\tilde{\sigma}(re^{j\theta})^{-1}|^{2\pi/\Omega_2 - \Omega_1} \left|\frac{1}{\alpha}\right|^{\Omega_1/\Omega_2 - \Omega_1}$$

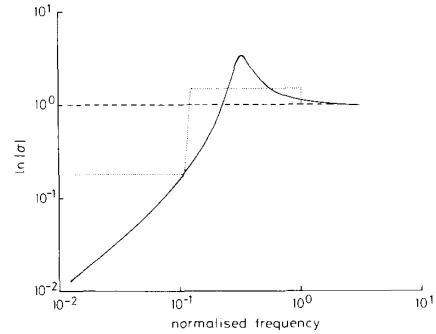

Fig. 6 *Typical sensitivity function*

where Ω_i is given by

$$\Omega_i = \int_{-\omega_i}^{\omega_i} \frac{(r^2 - 1)\, d\phi}{r^2 + 1 - 2r\cos(\theta - \phi)}$$

For a number of NMP zeros, clearly, the maximum lower bound, due to the appropriate zero, is of interest. But simply to obtain guidelines about the value of S_{max}, we will consider the case of infinite zeros (due to delays) or the direct consequence of corrolary 1, which gives

$$S_m = \prod_i |\beta_i|^{\pi/h(\omega_2 - \omega_1)} \left|\frac{1}{\alpha}\right|^{\omega_1/\omega_2 - \omega_1}$$

This implies that S_m will become large when

(i) $|\beta_i|$ are large (i.e. large gains are necessarily to move highly unstable poles back into the stability region)

(ii) $\omega_2 - \omega_1$ is small (i.e. we are asking for a rapid roll off at high frequencies or the crossover frequency is too large)

(iii) α is small (i.e. we require too tight a tracking (or rejection) requirement at low frequencies).

It is also instructive to examine the role of sampling on the magnitude of S_m. For simplicity, assume that there are no requirements on sensitivity at high frequencies (i.e. $\omega_2 = \pi/h$), this gives

$$\ln(S_m) = \frac{1}{\pi f_s/\omega_1 - 1} \ln(1/\alpha)$$

where f_s is the sampling frequency. A plot of $\ln(S_m)/\ln(1/\alpha)$ against f_s/ω_1 is given in Fig. 7.

As seen, there appears to be a point of diminishing returns for $f_s \simeq 3\omega_1$ or $f_s \simeq 18$ times the crossover frequency. Most importantly, this simple analysis implies that, although the maximum value of sensitivity will probably always decrease as f_s increases, there is very little point in sampling the output of a system too quickly. What it also implies is that there is every justification in not sampling too slowly, as the performance degrades exponentially.

To verify this property on the actual S_{max}, a set of discrete-time pole-placing controllers were designed for the following systems:

$$\frac{1}{1+10s}, \frac{e^{-2s}}{1+10s} \text{ and } \frac{1-s}{1+10s}$$

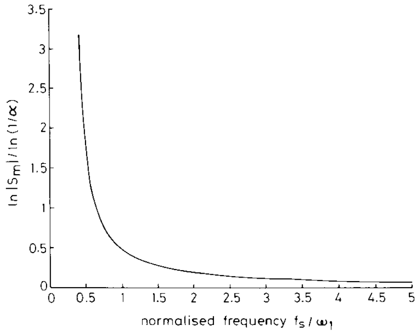

Fig. 7 *Max log sensitivity against sampling frequency*

at the following sample rates: 0.1, 0.2, 0.5, 1, 2, 5 and 10 Hz. In each case, an integrating pole-placer was designed and the desired closed-loop poles were 2 poles at $s = -1/3$. Fig. 8 shows the variation of the maximum value of sensitivity with frequency.

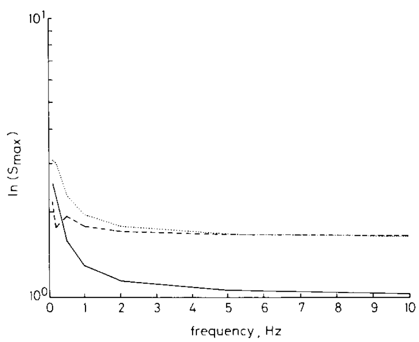

Fig. 8 *Variation of maximum sensitivity with sampling frequency*

Consider the system with a time constant of T. The plant is, at most, speeded up by a factor of 3 in most real process control applications. This means that $\omega_c \simeq 3/T$ which, using the rough rule above, gives an $f_s \simeq 9/T$ or a sampling interval 1/9th of the dominant time constant of the system. This is probably the real reason for the practitioners' rule of thumb. Sidi [20] shows that, from a classical design aspect, the ratio of the crossover frequency to the sampling frequency for minimum-phase systems with no delay is given by

$$\omega_c = \frac{\omega_s}{\pi} \tan^{-1} (2^{-GM/12\alpha})$$

where GM is the gain margin in decibels and $(1 - \alpha)\pi$ is the phase margin in radians. For a gain margin of 12 dB and a phase margin of 45°, the ratio of the sampling to crossover frequency is about 8. For more realistic situations where there are time delays and NMP zeros, this ratio can perhaps not be reduced to less than about 20, for reasonable values for desired gain and phase margins.

4.5 Effect of NMP zeros on achievable performance

Typically, discretised systems have two types of zeros:

(*a*) zeros due to the presence of an actual zero in the process

(*b*) zeros due to the sampling operation.

Unfortunately, unlike the poles, the exact position of neither of the zeros can be predicted, but suffice it to say that zeros due to the inverse response of the system to a step (i.e. zeros in the right-half s-plane) appear in the right-half Z-plane outside the unit circle (for the usual sample-rates), and those due to sampling frequently appear in the left-half Z-plane and frequently on the real axis (for a thorough discussion of the zeros see Reference 16).

Recall that we require $|\sigma|$ to be small at low frequencies and $|\sigma| \simeq 1$ at ω close of π/h. As was discussed earlier because of the presence of the NMP zeros on the positive real axis, it is not possible to have small $[\sigma]$ at low frequencies without paying any penalty at higher frequencies. This is because the weight of the integral is around the same frequencies as that of the NMP zero. A sensible design would therefore require relaxation of the tight requirements on σ within the bandwidth of the weight. This implies that these zeros impose real constraints on the feedback system independent of the design method applied. The sampling zeros, on the other hand, should not impose any difficulty in the design, as $\ln|\sigma| \simeq 0$ round the appropriate frequencies, where the weights are large. Any problem arising, therefore, is a function of the design and is not inherent in the control problem, unlike the case of 'real' NMP zeros.

A simple rule of thumb for a reasonable design is to choose

$$\omega_c h = 2 \tan^{-1} \left(0.5 \frac{q_z - 1}{q_z + 1}\right)$$

where q_z is an NMP real zero in the RHP, see Reference 17 for a detailed discussion. Recall that the usable bandwidth of the weight is given in Table 1. Using those values in conjunction with the idealised variation of the sensitivity function gives (as a first approximation)

$$\omega_c h = \frac{\theta_{BW} \ln (S_m)}{\ln (S_m) - \ln (\alpha)}$$

Table 2 shows the variation of the two estimates of $\omega_c h$ with the zero position on the real axis: the values for the $\omega_c h$ (Int) are obtained assuming $\alpha = 0.1$ and $S_m = 2$. It can be seen that once again Horowitz's rule can be verified using the integral constraints.

Table 2: Variation of $\omega_c h$ with zero position

r	$\omega_c h$ (Hor)	$\omega_c h$ (Int)
1.05	0.0244	0.0338
1.15	0.0697	0.0754
1.25	0.1110	0.1050
1.50	0.1993	0.1553
2.00	0.3303	0.2116

4.6 Effect of the dead time on the sensitivity function

Unlike the NMP zeros, dead time does not appear to influence the value of the integral in Theorem 1 directly: this is counterintuitive. To examine the effect of zeros, the following first-order system plus time delay is considered:

$$(1 - 0.9q^{-1})y(t) = 0.1u(t - k) \quad \text{where} \quad 1 < k < 5$$

A pole-placement controller with a closed-loop pole situated at 0.8 was designed for each value of k. The sums of the logarithms of the absolute values of the open-loop unstable poles are given in Table 3.

Table 3: Variation of $\Sigma \ln |\beta_i|$ as k (time delay) varies

k	$\Sigma \ln \lvert\beta_i\rvert$
1	—
2	0.0953
3	0.1740
4	0.2398
5	0.2956

It can be seen that, even for this simple model, as k increases, so does the value of $\Sigma \ln |\beta_i|$, which then indirectly puts a constraint on the maximum value of the sensitivity function. This ia a particularly simple example, but this phenomena is quite common, whereby the controller designed using pole-placement design becomes unstable as k increases, and the limits derived from corollary 1 impose a lower bound on the maximum value of the sensitivity function. This effect can also be seen from theorem 2, where the constraints on the complementary sensitivity were imposed. It is also relatively easy to show that, for a system with N_z nonminimum-phase zeros, a controller with $N_z - 1$ unstable poles situated in appropriate positions can satisfy the constraints on sensitivity and complementary sensitivity [7].

An important class of problems much encountered in chemical engineering is the class of first-order systems plus time delay:

$$G(s) = \frac{Ke^{-s\Delta}}{1 + sT}$$

It is instructive to examine the variation of the maximum value of the sensitivity function and the frequency, where $|S| > 1$ (f_c), with the sampling frequency and the desired pole location. Two special cases are considered: state-dead-beat control (i.e. where all the closed-loop poles are fixed at the origin) and the case where the closed-loop characteristic equation is set to be the discrete equivalent of $T_0^2 s^2 + 1.414 T_0 s + 1$ (i.e. the poles are fixed at Butterworth positions). T and K are set to unity, in both cases, and T_0 is set to 0.3 in the second case. Δ varies as $\{0.0, 0.2, 1.0, 5.0\}$. Fig. 9 shows the variation of S_{max} and f_c with the sampling frequency for a state-dead-beat controller.

Note that:

(*a*) For the minimum-phase system (solid line, no delay) S_{max} does not go to zero as sampling frequency increases. This is because we require the presence of the integrator in the loop

(*b*) In general, S_{max} increases as the sampling frequency increases. This is the main reason why dead-beat control should only be used at slow sample rates

(*c*) The best ratio of sampling frequency to f_c is about 6. This implies that, if we need to overcome disturbances up to f_c, we should at least sample the system an order of magnitude faster.

For the second case, we consider the situation where $T_0 = 0.3$, as in Fig. 10. This is a somewhat more realistic situation. Here, we have:

(i) S_{max} for the case with no time delay tends to unity as the sampling frequency increases, because, unlike the case above, we are not pushing the integrator and the open-loop pole to faster locations, with the increase in the sampling frequency

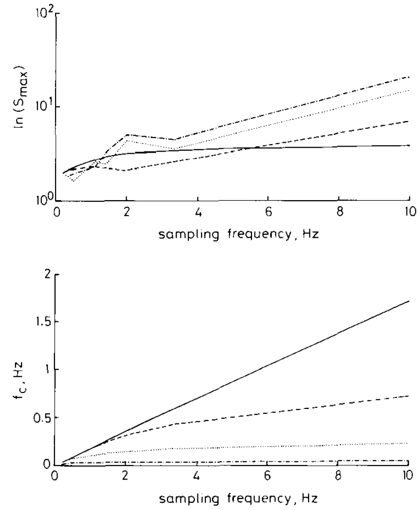

Fig. 9 *Variation of S_{max} and f_c with sampling frequency for state dead-beat control*

—— delay = 0
- - - - delay = 0.2
······ delay = 1.0
-·-·-· delay = 5.0

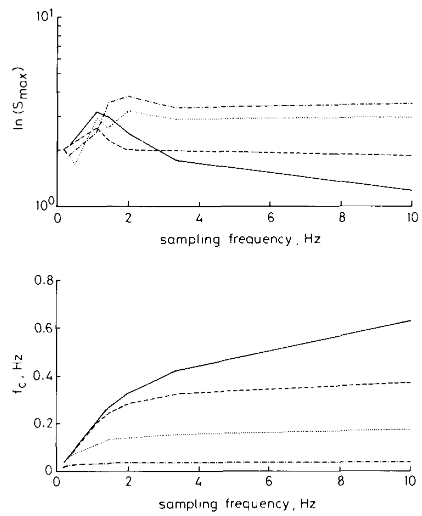

Fig. 10 *Variation of S_{max} and f_c with sampling frequency for pole-placement control $T_0 = 0.3$*

—— delay = 0
- - - - delay = 0.2
······ delay = 1.0
-·-·-· delay = 5.0

(ii) In the other cases, S_{max} tends to its final value by about a 4 Hz sample rate. This implies that there is very little point in sampling the system any faster

(iii) The ratio of the sampling frequency to f_c is, at best, around 10, confirming our previous observations

(iv) For time-constant dominated systems (i.e. $T > \Delta$), sampling the system any faster than $5/T$ or $10/T$ is pointless, and this gives an f_c of around $1/2T$. For delay-dominated systems (i.e. $\Delta > T$), on the other hand, sampling rates above $2/\Delta$ to $5/\Delta$ are unnecessarily fast, if we desire to have reasonable values for S_{max}. These observations tally very closely with the 'rules of thumb' which are frequently quoted in the literature.

4.7 Bounds on the complementary sensitivity function

Similar to the bounds on the sensitivity function, we may approximate the complementary sensitivity function as in Fig. 11. Using the same approach as before, we have

$$\tau_m = |\tilde{\tau}(re^{j\theta})^{-1}|^{2\pi/\Omega_2-\Omega_1} \left|\frac{1}{\alpha}\right|^{2\pi-\Omega_2/\Omega_2-\Omega_1} |r|^{k/\Omega_2-\Omega_1}$$

where τ_m is the maximum and α is the minimum value of the complementary sensitivity function, the rest of the definitions are exactly the same as before. The value of τ_m (i.e. the size of the resonance peak) will be large if:

(*a*) $|\tilde{\tau}|^{-1}$ is large, evaluated at the unstable poles of the system (i.e. when there are NMP zeros close to these locations)

(*b*) k is large and/or the unstable pole is fast

(*c*) α is small (i.e. too fast a roll off is required)

(*d*) $(\Omega_2 - \Omega_1)$ is small (i.e. the Q factor of the resonance will be large).

4.8 Poles and zeros inside the unit circle

The integral constraints above do not tell us anything about the constraints imposed (if any) by the pole-zero locations inside the stability boundary. Again, using the Poisson's integral theorem [9], we state the following two theorems:

Theorem 3: For any zero $re^{j\theta}$ of the open-loop transfer function inside the unit circle, the sensitivity function must satisfy the following integral constraints:

$$\ln(|\tilde{\sigma}(re^{j\theta})^{-1}|) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(1-r^2)\ln(|\sigma(e^{j\phi})|)\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

$$\frac{d^i}{ds^i}\ln(\tilde{\sigma}(s)^{-1})\Big|_{s=re^{j\theta}} = \frac{1}{2\pi}\int_0^{2\pi} \times \frac{(1-r^2)}{r^2+1-2r\cos(\theta-\phi)} \frac{d^i \ln(\tilde{\sigma}(s))}{ds^i}\Big|_{s=e^{j\phi}} d\phi \quad \text{for } i = 1, \ldots, m-1$$

$$\angle\tilde{\sigma}(re^{j\theta})^{-1} = \frac{1}{2\pi}\int_0^{2\pi} \frac{(1-r^2)\angle\tilde{\sigma}(e^{j\phi})\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

where m is the multiplicity of the zero.

Proof: The proof follows the proof of theorem 1, but $\tilde{\sigma}$ is analytic inside the unit circle.

Theorem 4: For any pole $re^{j\theta}$ of the open-loop transfer function inside the unit circle, the complementary sensitivity function must satisfy the following integral constraints:

$$\ln(|\tilde{\tau}(re^{j\theta})^{-1}|) + k\ln(r) = \frac{1}{2\pi}\int_0^{2\pi} \frac{(1-r^2)\ln(|\tau(e^{j\phi})|)\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

$$\frac{d^i}{ds^i}\ln(\tilde{\tau}(s)^{-1})\Big|_{s=re^{j\theta}} = \frac{1}{2\pi}\int_0^{2\pi} \times \frac{(1-r^2)}{r^2+1-2r\cos(\theta-\phi)} \frac{d^i \ln(\tilde{\tau}(s))}{ds^i}\Big|_{s=e^{j\phi}} d\phi \quad \text{for } i = 1, \ldots, n-1$$

$$\angle\tilde{\tau}(re^{j\theta})^{-1} + k\theta = \frac{1}{2\pi}\int_0^{2\pi} \frac{(1-r^2)\angle\tilde{\tau}(e^{j\phi})\,d\phi}{r^2+1-2r\cos(\theta-\phi)}$$

where n is the multiplicity of the open-loop stable pole. $\tilde{\tau}$ is analytic inside the unit circle and $\bar{\tau}$ is all pass.

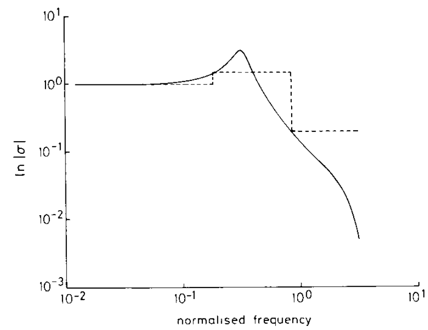

Fig. 11 *Variation of a typical complementary sensitivity function*

Proof: As with theorem 3.

Corrolary 2: For any closed-loop stable discrete-time system:

$$\ln(K) + \sum_{i=1}^{m} \ln|\alpha_i| = \frac{1}{\pi}\int_0^{\pi} \ln|\tau(e^{j\phi})|\,d\phi$$

where $\tau(q^{-1}) = q^{-k}K\prod(1-\alpha_i q^{-1})/P(q^{-1})$, α_i are the zeros of the open-loop system with the first m being nonminimum-phase. P is the closed-loop pole polynomial and is assumed to have unity leading element (i.e. $P = 1 + p_1 q^{-1} + \cdots$).

Proof: By substitution into theorem 4 and evaluating the function at $r = 0$.

This result is similar to Bode's theorem for the sensitivity function, but, as such, does not appear to be as useful, other than in finding some bounds on the magnitude of complementary sensitivity similar to those of the preceding Section. Note that this result is particular to discrete-time systems as the maximum frequency is π/h.

4.8.1 Pole-zero cancellations inside the unit circle: At first sight, it appears that pole-zero cancellations inside the unit circle cause the same degree of difficulty as those outside the unit circle. This is not the case:

$$\tilde{\tau}(q^{-1}) = \frac{(BS)_+ P^*}{(BS)_+^* P}$$

where $(BS)_+$ indicates the minimum-phase zeros and * indicates complex conjugate. Clearly, if there are exact pole-zero cancellations, the value of $\ln|\tilde{\tau}|$ does not change at all. For close cancellations, on the other hand, if this zero does not appear close to one of the poles of the closed loop, the $|\tilde{\tau}|^{-1}$ can become quite large, thus imposing a severe limitation on the complementary sensitivity and sensitivity functions. If, however, one of the closed-loop poles is very close to this approximately cancelled zero, then the cancellation does not impose as severe a difficulty. The following example may clarify the point further. Fig. 12 shows the variation of the

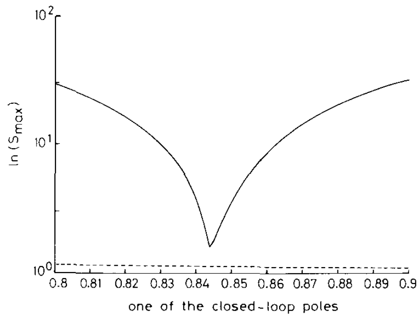

Fig. 12 *Variation of maximum sensitivity with pole-position with closed-loop pole-zero cancellation inside the unit circle*

maximum log sensitivity of a pole-placement design as the pole position varies from 0.8to 0.9. The solid line is the case where a model of the form

$$(1 - 0.9q^{-1})(1 - 0.85q^{-1})y(t) = 0.1(1 - 0.84q^{-1})u(t - 1)$$

is assumed, and the broken line is where a model of the form

$$(1 - 0.9q^{-1})y(t) = 0.1u(t - 1)$$

is considered. Note that, with the closed-loop pole even slightly away from the cancellation, the sensitivity increases by an order of magnitude. It is, therefore, very important to design controllers which automatically consider cancellations and close cancellations should they occur. Clearly, some loss of performance is inevitable, but this should be kept to a minimum.

4.8.2 Selection of the closed-loop poles: Again, from theorem 4, it is quite clear that setting closed-loop poles in the proximity of the open-loop poles makes the value of $\ln|\tilde{\tau}(re^{j\theta})^{-1}|$ reasonably small. Note that this is particularly significant when there are underdamped open-loop poles, as then the weight associated with these, as discussed earlier, is also large, and thus shifting the poles too far may lead to unwanted large sensitivity functions.

Two special cases are worth considering: the minimum-variance or minimum-prototype control where the poles are the open-loop zeros of the process, and the mean-level control where the closed-loop poles are the same as the open-loop poles of the system. For systems with unit delay, we have:

(*a*) $\sigma(q^{-1}) = A\Delta$, for minimum-prototype control

(*b*) $\sigma(q^{-1}) = 1 - q^{-k}B/B(1)$, for mean-level control.

For most values of A and B, it can be shown that the maximum sensitivity of minimum-variance control is higher than the mean-level control. However, for cases where the zeros are close to the point (1, 0), the situation may be reversed. Note that the MV control is only applicable to minimum-phase and ML control to open-loop stable systems. Hence, the comparison is only meaningful for stable and minimum-phase systems. References 18 and 11 show that minimum-prototype and mean-level control are two special cases of generalised predictive control, when the control horizon is set to unity, as the prediction horizon varies from one to infinity. It is conjectured that, for other prediction horizons, the value of sensitivity varies between the two quantities above. It is possible to examine the above quantities to obtain bounds on the sensitivity function of the nominal predictive control loop.

5 Concluding remarks

This paper derives the Bode's integral theorem for discrete-time systems. Many of the standard design guidelines adopted by the control engineer are shown to be a natural outcome of such theorems. Although not a design methodology on its own, it provides further insight into the design of SISO discrete-time control systems: albeit a problem considered to have been completely solved by many authors. The design guidelines derived from these results include:

(*a*) *Choice of sample time:* Sampling faster should always improve the performance, but, once above a certain rate (e.g. an order of magnitude above the desired closed-loop bandwidth), little benefit will result. With time-delay and nonminimum-phase systems, the maximum attainable bandwidth is limited to around $f_c \simeq 1/(2\pi\Delta)$ Hz, where Δ is the time delay, in seconds, and $f_c \simeq s_z/4\pi$, where s_z is the magnitude of the NMP zero of the system on the real axis in the right-half s-plane. The sample rate is typically $10f_c$

(*b*) *Pole-zero cancellations:* Yet another reason was considered as to why such exact or close cancellations outside the stability reagion are harmful. The cancellations inside the unit circle, on the other hand, should not pose any difficulties provided the correct method is adopted

(*c*) *Nonminimum-phase zeros and time delay:* These impose, as is well known, a severe limitation on the achievement of the performance specifications, especially if low sensitivity is required close to the frequencies of these zeros. Time delay, on the other hand, appears to impose an indirect influence on the maximum sensitivity via the restriction on the achievable bandwidth or introduction of unstable poles in the controller transfer function. NMP zeros on the real axis in the left-half Z-plane, however, should not pose any difficulty for a proper design

(*d*) *Bode's theorem for discrete-time systems:* It is shown that there are close similarities between the discrete and continuous-time results. As expected, the 2-pole roll-off requirement is not necessary. However, because of the inherent bandwidth limitation in discrete-time systems, asymptotic loop recovery is not possible in the usual framework. Moreover, time delays are dealt with, without any special considerations.

The extensions of the basic results to the multivariable case are trivial, for any of the multitude of scalar-valued functions of a MIMO system (e.g. characteristic values [19]) provided the branch points outside the unit circle are taken care of, or the determinant or the product of the singular values [7]. It is the implications and interpretations of such results that are unclear. For an excellent exposition of this problem see Freudenberg and Looze [7], they give a detailed discussion of the continuous-time results. The links of these to design guidelines are, however, still very obscure.

6 Acknowledgments

This work was performed while the author was spending a year leave of absence at the Department of Chemical Engineering, University of Alberta, Canada. The author is extremely thankful to Prof. S.L. Shah, for his many helpful comments, and Prof. D.G. Fisher, for numerous discussions on several aspects of this paper. This work forms part of a continuing program of the application of discrete adaptive controllers, which is supported by NSERC of Canada grant 32526, whose funding is gratefully acknowledged.

7 References

1 BODE, H.W.: 'Network analysis and feedback amplifier design' (Van Nostrand, Princeton, 1945)
2 HOROWITZ, I.: 'Synthesis of feedback systems' (Academic Press, 1963)
3 HOROWITZ, I.: 'Quantitative feedback theory', *IEE Proc. D*, 1982, **129**, (6), pp. 215–226
4 FRANCIS, B., and ZAMES, G.: 'An H^∞ optimal sensitivity theory for SISO feedback systems', *IEEE Trans.*, 1984, **AC-29**, (1), pp. 9–16
5 FREUDENBERG, J.S., and LOOZE, D.P.: 'Right half plane poles and zeros and design tradeoffs in feedback systems', *ibid.*, 1985, **AC-30**, (6), pp. 555–565
6 FREUDENBERG, J.S., and LOOZE, D.P.: 'A sensitivity tradeoff for plants with time-delay', *ibid.*, 1987, **AC-32**, (2), pp. 99–104
7 FREUDENBERG, J.S., and LOOZE, D.P.: 'Frequency domain properties of scalar and multivariable feedback systems' (Lecture Notes in Control and Information Sciences, Vol. 104, Springer Verlag, 1988)
8 SUNG, H.K., and HARA, S.: 'Properties of sensitivity and complementary sensitivity functions in single input single output digital control systems', *Int. J. Control* 1988, **48**, (6), pp. 2429–2439
9 CHURCHILL, R.V., BROWN, J.W., and VERHEY, R.F.: 'Complex variables and applications' (McGraw Hill, 1974)
10 ÅSTRÖM, K.J., and WITTENMARK, B.: 'Computer controlled systems' (Prentice Hall, 1984)
11 ÅSTRÖM, K.J., and WITTENMARK, B.: 'Adaptive control' (Addison-Wesley, 1988)
12 O'YOUNG, S.D., and FRANCIS, B.C.: 'Sensitivity tradeoffs for multivariable plants', *IEEE Trans.*, 1985, **AC-30**, (7), pp. 625–632
13 KALMAN, R.E.: 'When is a linear control system optimal?', *J. Basic Eng., Ser. D*, 1964, **86**, pp. 51–60
14 DOYLE, J., and STEIN, G.: 'Multivariable feedback design: Concepts for a classical/modern synthesis', *IEEE Trans.*, 1981, **AC-26**, (1), pp. 4–16
15 ATHANS, M., and STEIN, G.: 'The LQG/LTR procedure for multivariable feedback control design', *ibid.*, 1987, **AC-32**, (2), pp. 105–114
16 ÅSTRÖM, K.J., HAGANDER, P., and STERNBY, J.: 'Zeros of sampled systems', *Automatica*, 1984, **20**, (1), pp. 31–38
17 HOROWITZ, I., and LIAO, Y.K.: 'Limitations of nonminimum-phase feedback systems', *Int. J. Control*, 1984, **40**, (5), pp. 1003–1013
18 CLARKE, D.W., MOHTADI, C., and TUFFS, P.S.: 'Generalized predictive control: extensions and interpretations', *Automatica*, 1987, **23**, pp. 149–160
19 HUNG, Y.S., and MACFARLANE, A.G.J.: 'Multivariable feedback: a quasi-classical approach' (Lecture Notes in Control and Information Sciences, Vol. 40, Springer-Verlag, 1982)
20 SIDI, M.: 'On maximization of gain bandwidth in sampled systems', *Int. J. Control*, 1980, **32**, (6), pp. 1099–1109
21 FRANCIS, B.: 'A course in H^∞ control theory' (Lecture Notes in Control and Information Sciences, Vol. 88, Springer Verlag, 1987)