Analysis and Design of Multirate-Multivariable Sampled Data Systems

By

Huashan Wang

Submited to the University of London
for the Degree of
Doctor of Philosophy
and for the
Diploma of Membership of Imperial College

May 1987

Department of Electrical Engineering
Imperial College of Science and Technology
University of London

Acknowledgements

I would like to take this opportunity to express my gratitude to my supervisor Dr. D. J. N. Limebeer for his many helps, both technical and moral, during the course of this research.

I am deeply in debt to my family for their support and endurance of my absence. This work is dedicated to my wife Been for her constant encouragement, without which I would not have survived the less pleasant aspects of work and life.

I would also like to thank Xi'an Jiaotong University for the financial support during the first three years. I am grateful for the financial help of the Leech Trust towards the end of this work.

Finally, I have to thank to my colleague students for the many inspiring discussions. They have helped me to recognize myself.

ANALYSIS AND DESIGN OF MULTIRATE-MULTIVARIABLE SAMPLED DATA SYSTEMS

Abstract

In this thesis the problem of analysis and design of certain class of multirate multivariable (MM) sampled data systems is investigated. The overall aim is to embed the study of such systems in a linear time-invariant framework. To achieve this, we first show that a practically very important class multirate controllers can be decomposed as the cascade of LTI and sample-and-hold (SAH) operators. Then we demonstrate that the SAH operators can be approximated by some LTI operators, and that the approximating error can be tightly bounded by another LTI operator. Based on the study of the SAH operators, stability analysis of MM systems can be accomplished by studying the robust stability problem of an LTI system subject to LTI perturbations. Finaly, we propose a design methodology that is based on H^{∞} synthesis theory. This design procedure will result in a digitally implementable controller that gives rise to performances measured on continuous time (or Laplace frequence domain) basis. Numerical methods of the H-infinity optimization are discussed. We have tried our method on an example and believe that there is a potentially important improvement over the conventional discrete time techniques.

List of Symbols

$\in$, $\notin$	in, not in a set
$\supseteq$, $\subseteq$	include, be included as subset
$\forall$, $\exists$	for all, there exists
$\Rightarrow$, $\Leftarrow$ and $\Leftrightarrow$	imply, be implied by and equivalent to
$\mathbb{R}$, $\mathbb{C}$	fields of real and complex numbers
$\mathbb{R}(s)$	field of rational functions with real coefficients
$\mathbb{F}^{m\times l}$	set of $m\times l$ matrices with elements in $\mathbb{F}$
$\mathbb{C}_+$, $\bar{\mathbb{C}}_+$	open (resp. closed) right half of complex plane
$\mathbb{C}_-$, $\bar{\mathbb{C}}_-$	open (resp. closed) left half of complex plane
$L^{\infty}(a, b)$	space of functions ess. bounded on (a,b) where $-\infty \leq a < b \leq \infty$
$L^2(a, b)$	space of functions square integrable on (a,b) where $-\infty \leq a < b \leq \infty$
l^2	space of square summable sequences
$\lVert \cdot \rVert$	norm of functions or the induced norm for operators
A^*	adjoint operator of A
x^{*T}	sampling operation with rate T
$\circ$	operate on
$\perp$	orthogonal to
a.e.	almost everywhere (in Lebesgue measure)

Chapter One
Introduction and Survey

1.0 *Introduction*

A substantial potion of the control literature is devoted to sampled date systems. The importance of this theory stems from the wide use of microcomputers in control systems. Sampled data systems form a relatively complete and independent branch of control theory, in that it has its own methodology and basic techniques. This is due to the most important feature of such systems: the internal mechanism of a sampled data system is discrete as opposed to continuous. Thus we find analysis and design methods for sampled data systems, although sometimes the techniques used are similar to those for continuous time systems.

The overall aim of this thesis is to present a theory which unifies systems analysis and design methods in continuous and discrete time domains.

Since the term "sampled data systems" covers a rather wide range of topics, it is necessary to define the scope of problems discussed in this thesis.

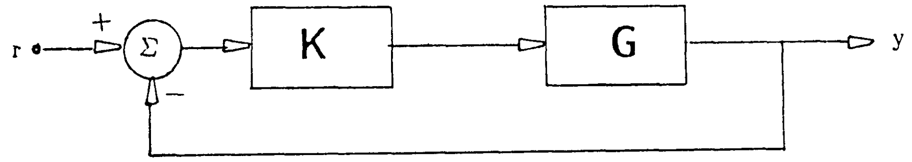


Fig. 1.0.1. Hybrid Feedback System

We will consider the system shown in Fig. 1.0.1. It consists of a controlled system G, and a controller K. G is assumed to be linear and time-invariant and K is a sampled data controller. A precise description of the sampled data controllers that we study in this thesis is given in Chapter Three. Essentially it is a digital device that takes samples from a signal (by a A/D converter), and then sends an output to the actuator (through D/A converter). This kind of controller is sometimes called a hybrid controller. We are particularly interested in the situation when K is multivariable controller with different but fixed sampling rates in various loops. Multirate sampling is the simplest generalization of single rate sampling. For a summary of sampling schemes see [Kalman,D2]. In this thesis we will confine attention to multirate systems only.

There are two main reasons for employing multirate sampling. The first is that onboard computers have finite computing power, so in order to achieve an overall performance, computing resources should be allocated to the loops according to their dynamic properties.

Namely, rapid sampling rates should be used in loops where signals have faster dynamics. There is also an economic consideration, for powerful computer and fast A/D and D/A converters are more expansive. The second reason for the need of multirate sampling is that in certain systems it is desirable to accommodate various devices which have their own operating periods. Typical examples are radar sensors and step motor actuators. In these cases it would be more convenient to let the controller match those devices by having appropriate sampling rates at the interfaces with them.

In the next section we give a brief survey of research work in this field. We are not going to discuss the details in depth, for they can be found in the references quoted. Rather, we will try to demonstrate their basic virtues and limitations, whereby to motivate the search for new directions.

1.1 *A brief survey*

The analysis and design of sampled data systems are two distinct problems. Analysis answers the questions about a system which is completely specified, while design is to specify (part of) a system in order that its behaviour satisfies given requirements. For this reason, we devote two subsections to each of these problems separately.

i) Analysis

The term "analysis" in most cases only refers to stability analysis, thus it is stability that is our main focus. One of the most salient features of a sampled data systems is its time-varying character. It is this time-variance property that makes some of the most effective analysis tools inappropriate.

Discrete time methods

Sampled data systems are a special kind of time-varying systems, i.e. the variance is often periodical in some way, depending on the particular sampling scheme used. For example, if all the sampling rates are the same (not necessary synchronized), then this period is just the sampling period. This is refered to as shift invariant. Stability analysis of this kind of sampled data systems can be simplified by only considering the response of the system at the sampling instants, which can be described by a set of difference equations with constant coefficients. Provided that the samples give adequate information about the actual signal, one may work entirely with the difference equation model. Discrete models are approximations, although they are often good enough for stability purposes.

Stability of single rate systems can be determined by $\mathbb{Z}$-transforms. One may argue that the discrete time approach is a way of avoiding the difficulties associated with time-variance of sampled data systems. This approach is effective, and perhaps this justifies its popularity. It is so successful for single rate systems that people have naturally extended it to the analysis of multirate sampled data systems. The basic idea is to find a single rate system which is equivalent to the multirate system in some sense, and then apply techniques for single rate systems. Methods for analysing multirate systems can be classified according to the way the equivalent single rate systems are derived. For example in [Glasson, D13], they are classified into frequency and time domain methods, and the former is further categorized into frequency and switch decompositions. But they all boil down to the same principle: converting a multirate system into a system that has the same stability properties.

The variety of methods can be classified according to the list below. We will look at some of the important techniques here.

Analysis

- Approximate methods:
  1. Tustin transformation;
  2. PCT [Houpis];
  3. Conic sector [Thompson].
- Exact methods:
  1. Time Domain:
     a. State space [Kalman][Walton];
  2. ($\mathbb{Z}$-transform) Frequency domain:
     a. Switch decomposition [Kranc];
     b. Frequency decomposition [Coffey];

Switch decomposition

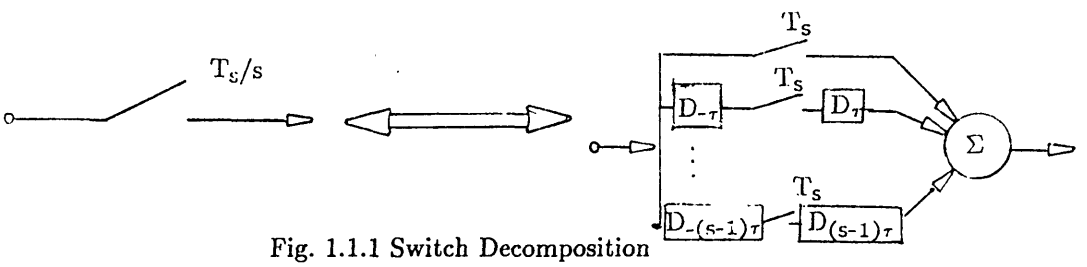


Fig. 1.1.1 Switch Decomposition

Let us suppose that sampling rates in the system are rational multipliers of each other.

This is a ubiquitous assumption for the exact methods, and will be assumed throughout in the discussions of these methods. Suppose also that one of the samplers works at a rate $T=T_s/s$, where T_s is the basic sampling period and s is an integer.

As illustrated in Fig. 1.1.1, this sampler can be "decomposed" into combination of samplers with rate T_s, and advance and retard operators. Advance and retard operators can all be defined in terms of a shift operator D_τ: $x \mapsto y$, as

$$y(t) = D_\tau \cdot x(t) = x(t - \tau)$$

where τ is a real number. Thus when $\tau > 0$, D_τ is a retard operator. If we view a sampler with rate T as an ideal impulse generator, i.e. it generates a sequence of pulses:

$$\sum_{n=0}^{+\infty} e(nT)\delta(t - nT),$$

then the decomposition is the following operation:

$$\sum_{n=0}^{+\infty} e(nT)\delta(t - nT) = \sum_{k=0}^{k=s-1} \sum_{m=0}^{+\infty} e(mT_s + kT)\delta[t - (mT_s + kT)\,] \qquad (1.1.1)$$

Substituting all the samplers in a system by their decomposed versions, we arrive at an equivalent single rate system. Standard techniques of single rate sampled data systems can then be applied. In fact, this technique can be made systematic by introducing the advance and retard vector operators [Boykin, D6].

Frequency decomposition

Frequency decomposition is based on a similar technique. In order to give a simplified explanation of this method, it is necessary to introduce some notations. Let e(s) be the Laplace transform of e(t), then the Impulse Laplace Transform (ILT) of e(t), denoted by e^{*T} where T is the sampling rate, is defined by:

$$e^{*T}(s) = \frac{1}{T} \sum_{n=-\infty}^{+\infty} e(s - 2\pi jn/T)$$

As in (1.1.1) we can rewrite the above as:

$$e^{*nT} = \frac{1}{nT} \sum_{m=-\infty}^{+\infty} e(s - 2\pi jm/nT)$$

$$= \frac{1}{n}\sum_{l=0}^{n-1}\frac{1}{T}\sum_{m=-\infty}^{+\infty} e(s - 2\pi jl/nT - 2\pi jm/T)$$

$$= \frac{1}{n}\sum_{l=0}^{n-1} e^{*T}(s - 2\pi jl/nT). \qquad (1.1.2)$$

In other words, it is possible to express the ILT of a function with sampling rate nT in terms of ILT's with T. We also note that $e^{*T}(s)$ is a periodic function of period $2\pi j/T$. With this preparation, we can derive the Integer Rate Identity (IRI) and Rational Rate Identity (RRI) [Boykin, D5, D6][Jury, D4].

Consider the system shown in Fig. 1.1.2.

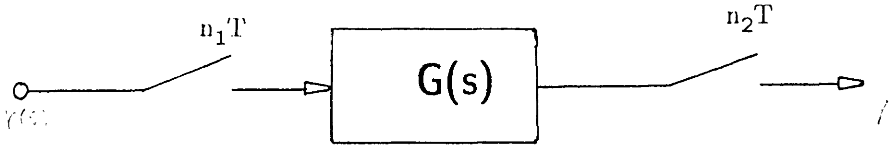


Fig. 1.1.2 Frequency Decomposition

It is desirable to find a multiplication type of relation between y(s) and e(s). If n_1/n_2 is an integer, it immediately follows that:

$$y^{*n_2T}(s) = \left[G(s)\, r^{*n_1T}\right]^{*n_2T}$$

$$= \left[G(s)\right]^{*n_2T} r^{*n_1T}(s) \qquad (1.1.3)$$

for r^{*n_1T} is periodic in $2\pi j/n_2T$. (1.1.3) is refered to as the IRI. In general n_1/n_2 is not integer, and in order to obtain a similar relation to (1.1.3) the following steps are involved. Let $y(s) = G(s)r^{*n_1T}$, and apply (1.1.2) to y^{*n_2T} to obtain an expression that only involves ILT's of y with rate T:

$$y^{*n_2T}(s) = \frac{1}{n_2}\sum_{l=0}^{n_2-1} (y)^{*T}(s - 2\pi jl/n_2T),$$

since the terms $(\cdot)^{*n_1T}$ are periodic in T, and hence are invariant under $(\cdot)^{*T}$ operation. Using this fact and the expression for y(s), we have

$$y^{*n_2T} = \frac{1}{n_2}\sum_{l=0}^{n_2-1} (G\, r^{*n_1T})^{*T}(s - 2\pi jl/n_2T)$$

$$= \frac{1}{n_2}\sum_{l=0}^{n_2-1} G^{*T}(s - 2\pi jl/n_2T)\; r^{*n_1T}(s - 2\pi jl/n_2T) \qquad (1.1.4)$$

(1.1.4) is the RRI. The complexity of these formulas can be seen here. The same procedure must be repeated on all samplers in a system in order to obtain an input-output relation in multiplicative form. A long and tedious calculation is needed to eliminate the intermediate variables [Coffey, D3].

Switch and frequency decomposition methods are dual concepts. They are equivalent in the sense that results obtainable in one can also be obtained via the other [Jury, D4]. However, switch decomposition methods admits more systematic approach in addition to its intuitive clarity.

State space methods

An approach based on state space methods was studied by Kalman *et al* [Kalman et al, D2]. In [Kalman et al, D2] many types of sampled data systems are studied. They pointed out that samplers in a system should entitle additional states, and the dynamics of a system can be analyzed in terms of the propagation of this extended set of state variables. For simplicity we introduce the concept of sampling pattern of a system. It can defined as an array of binary functions of time. Suppose that there are N samplers in a system, then the sampling pattern of the samplers is:

$$P(t) = [p_1(t),\cdots, p_N(t)] \qquad (1.1.5)$$

where

$$p_i(t) \begin{cases} = 1 \text{ if sampler i samples at t;} \\ = 0 \text{ elsewhere.} \end{cases}$$

Obviously, if the sampling rates are rational multiples of each other, P(t) is a periodic function of time. Suppose that a set of state variables is chosen, then according to the dynamics of the system, the states at any given moment can be represented as a transformation of the states at a previous instant. Since sampled data systems are time-varying, the transformation over a fixed length of time is in general variant. But this transformation is a constant over intervals [nT, (n+1)T] if P(t) is periodic: P(t) = P(t + T). Therefore one can describe the propagation of the states at t=nT, n=0, 1, ··· by difference equations with constant coefficients. The method proposed in [Kalman, D2] is briefly outlined as follows: The states of the sample and hold devices evolve only at the discontinuous points of P(t). In between any two such instants, the whole system is governed by the continuous propagation of the states other than those of the samplers. Thus, between those instants states transitions can be obtained.

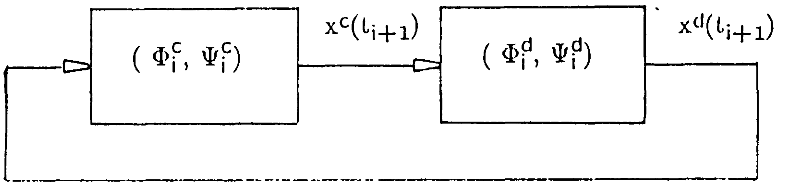


Fig. 1.1.3 General Sampled Data System

To illustrate the idea more clearly, we depict a general sampled data system as in Fig. 1.1.3. It is different in form from [Kalman, D2], but it conveys the same basic message. Suppose that the discontinuous points, or the instants when one of the samplers samples, are t = t_1, ···, t_K, t_{K+1}, ···, where $t_{K+1} - t_1 = T$. As far as the continuous part is concerned, its state transition in (t_i, t_{i+1}) is completely determined by its own states $x^c(t_i)$ and the input to it $x^d(t_i)$, which is constant in this interval. At the instant $t = t_i$, some of the state components of x^d are updated, while others remain the same. Summing up all these, one can formally write:

$$\begin{cases} x^c(t_{i+1}) = \Phi_i^c \cdot x^c(t_i) + \Psi_i^c \cdot x^d(t_i) \\ x^d(t_{i+1}) = \Phi_i^d \cdot x^d(t_i) + \Psi_i^d \cdot x^c(t_i) \end{cases} \tag{1.1.6}$$

or more compactly, by denoting $x \triangleq \begin{bmatrix} x^c \\ x^d \end{bmatrix}$,

$$x(t_{i+1}) = \Phi_i \cdot x(t_i) . \tag{1.1.7}$$

The periodicity of P(t) implies that $\Phi_i = \Phi_{K+i}$. The state propagation over the intervals $(t_{mK+i}, t_{(m+1)K+i})$ is given by

$$x(t_{(m+1)K+1}) = \Phi \cdot x(t_{mK+1}) ,$$

where

$$\Phi \triangleq \prod_{i=1}^{K} \Phi_i . \tag{1.1.8}$$

therefore the state evolution is described by a stationary transformation. From this one can study the stability properties of the system. State space method can also be used for other sampling schemes provided the sampling pattern P(t) is a periodic function. It is important to recognize that the samplers have dynamics.

The boundary between frequency and time domain methods is not clear-cut. For example, the procedure of calculating a $\mathbb{Z}$-transfer function from a continuous time system implicitly includes state transformation. Relations between state space and frequency domain approaches have been studied in [Araki, D19] for a special case where the hybrid controller only realizes a constant gain. We can show that the arguements in [Araki, D19] apply to general situations by incorporating in the separation properties of hybrid controllers, which we will study in detail in chapter 4.

The time variance of sampled data systems can appear in two forms. It is either in the form that relations between variables are not representable as transfer function multiplication, or that transformation over a fixed length of time is not constant. By observing the system variables at specially selected instants, relations between them can be simplified. This is clearly seen in the case of state apace methods. It is not so obvious for frequency decomposition method. The methods discussed up to here give necessary (and sufficient if no dynamic mode becomes hidden after sampling) conditions for closed loop stability. In this sense they are exact methods. Since they only deal with system dynamics at sampling instants, they are also called discrete time methods. It is possible to use approximate methods, which give sufficient conditions only. This is the topic of the next section.

Continuous time methods

Discrete time analysis only describes the dynamical behaviour at sampling instants, and hence do not give a full description of the system dynamics. If discrete time description can be viewed as approximations of the "true" systems, one may also think of other ways of approximation. This is the fundamental philosophy in [Thompson]. Thompson has studied sampled data systems in a different direction. He approached hybrid controllers from a Conic Sector point of view.

The use of conic sector in system analysis has a long history, and the general theoretical treatment in [Safonov] has widened its scope of application. For our purposes, a conic sector can be understood as a collection of operators centered around an operator. More specifically, suppose c and r are two operators defined in some space of functions, a conic sector cone(c,r) of operators is defined as

$$\mathrm{cone}(c,r) \triangleq \{\ K \mid |(K - c)\cdot x| \leq |r\cdot x|\ \forall\ x \in \mathrm{Dom}(\overset{K}{T})\ \}. \qquad (1.1.9)$$

It is a "ball" in the space of operators with a "weighted" radius. Suppose the operator K above is part of a feedback system, then sufficient conditions of stability of this system can be stated in terms of the conic sector that contains K. The remarkable thing about this is that one can embed a "difficult" operator K into a conic sector with "easy" centre and radius in order that the analysis can be greatly simplified. In [Thompson] conic sectors have been found that contain parts of a hybrid controlled feedback system. The centre and radius he has found are LTI operators, therefore he carried out the analysis in Laplace-frequency domain. The main argument there was that a hybrid controller, though time-varying, is essentially linear and time-invariant, i.e. it can be approximated by an LTI system. We can define this approximation as the centre of a conic sector, and define a nominal system by substituting the time-varying hybrid controller with the centre. If the centre chosen is a good approximation to the hybrid controller, then we have reasons to believe that the nominal system should bear the main features of the original system. What must be ensured is that the error of the approximation will not upset conclusions drawn from the analysis of the nominal system. It is especially important to ensure that the stability of the nominal system is the same as that of the original system. To do so, the error of the approximation has to be examined. In order to avoid the confrontation with the time-varying nature of this error, we only seek to assess its effect from an LTI bound on it. "Putting" an operator into a conic sector is the same as finding a bound (radius) on the error for the given approximation (centre). Once a radius is found, one can then apply conic sector theory to check if the "ball" specified by the radius is small enough not to violate stability of the nominal system.

The work of Thompson [Thompson] was only concerned with single rate systems, and the techniques he used are not applicable for finding conic sectors for multirate systems. The results in [Thompson] can be summarized as follows. The type of systems studied in [Thompson] is shown in Fig. 1.1.4.

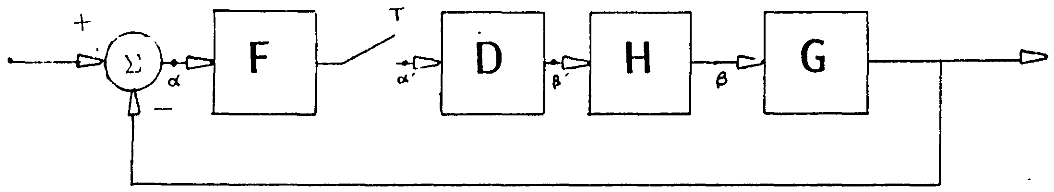


Fig. 1.1.4 Structure of Hybrid Feedback System

The first step is to put part of the system into a conic sector. As was pointed out by

Thompson [Thompson], there are several possibilities as to which part of a system is to be put into a conic sector. We only consider one of these, namely the case of putting the hybrid controller into a conic sector. The main result of [Thompson] is a quantitative specification of the radius of the conic sector that contains a stable hybrid controller with a strictly proper prefilter. The radius involves the compution of several infinite sums, and can be computed when the prefilter is defined. More specifically, a hybrid controller K is inside a conic sector (C, R), where both C and R are LTI, if

$$\sigma_{\min}\,[R(j\omega)] \geq r(\omega)$$

where $r(\omega)$ is a function determined by the prefilter, the algorithm of the internal compution of K and the choice of centre C. Details of the compution can be found in [Thompson]. Stability of the system is guaranteed if G is outside of the conic sector (C, R). The condition for G to be outside the conic sector can be broken into two conditions: (a) the "nominal" system (C, G) is stable and (b) the nominal system is robustly stable with respect to a class of LTI perturbations whose magnitude is bounded by $r(\omega)$. In the study of LTI systems the class of perturbations Δ is typically defined as:

$$\Delta \triangleq \{\, \Delta \mid \Delta \text{ is LTI and } \sigma_{\max}(\Delta)(\omega) \;\leq r(\omega) \,\}.$$

If we regard the C as an approximation to K, and the approximating error as a perturbation to the system, then this perturbation is not a member of Δ. The conic sector theory employed by Thompson gives the theoretical fundation that enables one to view the perturbation just as a member of Δ.

Conic sectors for the hybrid controllers can be used for analysing other properties of a sampled data system as well. Details of this can also be found in [Thompson].

ii) Design

It is possible to make a similar classification of design methods for sampled data systems. The exact methods are those that use exact system models, and result in a controller to be implemented exactly. Typically, exact methods use discrete time models of systems, for it is difficult to use the exact continuous time models, which are time-varying. The approximate methods involve approximation either in the stage of modelling or in the stage of obtaining the realization of controllers. Transformation methods such as w'-domain synthesis belong to this category. Some design methods are listed below.

Design

Approximate methods:

1. Transformation methods:
   a. Tustin transformation (w'-domain);
   b. Frequency response matching [Rattan];
2. PCT [Houpis];

Exact methods:

1. ($\mathbb{Z}$-transform) Frequency domain:
2. Time domain:
   a. LQ optimal design [Amit] [Glasson];

As with the methods of analysis, most of the design techniques for singal rate systems are inappropriate for multirate systems. We will examine some of the methods listed here. We will also discuss the feasibility of modifying these techniques for multirate systems.

Transformation

To use transformation type of methods, a continuous time controller is designed to meet the specifications, usually stated in terms of continuous time behaviour. Following this, the controller is transformed into a set of difference equations (or a $\mathbb{Z}$-transfer function equivalently) so that it can be realized by a digital computer. There are several ways of transforming a continuous time controller into hybrid one, but they have one thing in common, that is they make the input-output behaviour of the hybrid controller be as close to that of the continuous time controller as possible. The simplest way to obtain a hybrid controller from a continuous time controller is via the Tustin transformation. This transformation uses a $\mathbb{Z}$-transfer function $\frac{2}{T}\frac{z-1}{z+1}$ as an approximation to s. A somewhat better approximation can be obtained by Rattan's method [Rattan, D15]. Rattan's method seeks to approximate a continuous time controller by a hybrid controller so that their frequency response are close in the sense of least squares. The term "frequency response" of a hybrid controller is defined by neglecting the side harmonics of the output of a hybrid controller to a sinusoid input in calculating the transfer function [Rattan, D15][Witeback, D11]. One can also do the transformation in such a way that for a specific class of input signals the output signals of the hybrid controller coincide with those of the continuous time controller at the sampling instants. For example one can do "pulse invariant" transformation, so that above property holds true for pulses of fixed width [Aström]. Transformation methods can be used for multirate system designs without additional difficulties (except for one of the invariant transformation). But the obvious problem with transformation methods is that the approximation error is out of control, because nothing is done in the process of continuous time controller design to compensate for this error. Thus the hybrid controlled system is expected to give performance quite different from that of the continuous time design [Glasson, D12]. Since the influence of sampling is ubiquitous in sampled data systems, whatever way of transformation is unable to rid of the errors.

Recently, a so called PCT (Pseudo Continuous Time) has appeared in literature [Houpis et al, D18]. They have pointed out that a hybrid controller is a combination of an ordinary linear time invariant (LTI) system and samplers. By approximating the effects of the samplers by LTI operators and including them in the model of the controlled system, some compensation to the influence of sampling on performance can be achieved in the course of design. This is an improvement, and should lead to superior designs. But it is not possible to model the effects of sampling by means of LTI systems, hence the PCT approach must have its limitations. In fact, the approximation error of the sampler is considerably high in high frequency range. PCT design is similar to transformation methods, and can be used for multirate systems.

Direct Design

Direct methods directly design for the hybrid controllers. No approximation is needed in order to realize the controller. We will discuss the so-called direct digital design first. Design is carried out entirely in discrete time domain. The combination of plant G with the interfaces of the hybrid controller is treated as an integral object and the discrete time input-output relation of it is found. This amounts to breaking the feedback loop at points α' and β', as shown in Fig. 1.1.4. Design is directly done to give the set of difference equations internal to the hybrid controller (or equivalently to design the $\mathbb{Z}$-transfer function D(z)).

Direct digital design methods are in many ways parallel to continuous time systems design methods, both in ($\mathbb{Z}$) frequency domain and via state space. Design specifications must be translated into discrete time domain and the designer analyzes if the specifications are met directly from discrete model. In this regard direct digital methods seem to admit more rigorous approaches. But the translation of specifications is not always possible or accurate. For example, the robustness requirements on the continuous time model is hard to study in discrete time domain. For multirate systems the situation is even worse: the ($\mathbb{Z}$) frequency domain analytical techniques can hardly be used for design. To the knowledge of the author, the only practical methods proposed is via state spaces, i.e. state estimator and LQ regulator approach [Glasson, D12] [Amit].

Before introducing the design methods proposed in [Glasson, D12] and [Amit], we quickly review some basic facts about linear quadratic design. Suppose that the controlled variable is y(t), and the control signal is u(t). In LQ regulator design problem, one seeks to minimize a cost functional

$$J = \int_{0}^{E} [\, y^{T}Qy + u^{T}Ru \,]\, dt$$

subject to the system dynamics given by

$$\begin{cases} \dot{x} = Ax + bu \\ y = Cx \,. \end{cases}$$

In a sampled data system, the control signal is piecewise constant. In single rate systems, let the sampling instants be at t = 0, T, ⋯, then the controlled variable y(t) is determined in the interval [nT, (n+1)T) by x(nT) and u(nT). Suppose that we are interested in a finite time regulation problem in the interval [0, NT]. Substitute these relations into the cost functional, it becomes the form

$$J = \sum_{i=0}^{NT} [\, x^{T}(nT)\hat{Q}x(nT) + x^{T}(nT)\hat{P}u(nT) + u^{T}(nT)\hat{R}u(nT) \,] \quad (1.1.10)$$

subject to

$$x(\overline{n+1}T) = \hat{A}x(nT) + \hat{B}u(nT) \,. \qquad (1.1.11)$$

Note the cross product term in the cost functional. Although this cost functional is dependent of the variables x and u on sampling instants, it is a based on the all-time behaviour of the system. The fact that control is piecewise constant is taken into account. Standard techniques show that the optimal control strategy is

$$u(nT) = -\, K(nT)x(nT),$$

where K(nT) can be obtained by solving a matrix Riccati difference equations. The details of the computation of K(nT) is omitted, only to note that when $N \rightarrow \infty$, K(nT) converges to a steady state limit gain, which can be determined by a algebraic Riccati equation [Franklin].

Glasson studied the infinite time horizon regulation problem for multirate systems. He pointed out that in the case of multirate sampling, the optimal gains, instead of converging to a stationery gain, propagate into a periodic sequence of gains. The period of this sequence is the same as that of the sampling pattern of the system. Suppose this period is MT, then, the controls are given by:

$$u[(nM + i)T] = - K(i)x[(nM + i)T] \text{ , } i=1, \cdots, M; n=1, 2, \cdots . (1.1.12)$$

Glasson has given detailed discussion on how to transform the weightings in a continuous time cost functional into those for the multirate systems. He also gave some consideration to the numerical solutions of periodic Riccati equations.

Amit studied the computational aspects of multirate system in greater details. In [Amit,] a wider class of systems are studied, namely the multiple order sampling, to which multirate sampling is a special class. The design methods of [Amit] is also based on LQ control theory, and the basic results are in the same region as [Glasson, D12]. He has shown that an LQ control problem of a multirate system can be transformed into one of a single rate system, together with the cost functional. The way the equivalent single rate system is obtained is basicly the same as was shown in [Kalman, D2]. There are some notational differences however. For example, in [Amit] the equivalent single rate system has the same number of states but with an enlarged set of control variables, while in [Kalman, D2] states of the sample and hold devices are introduced explicitly. Amit also showed that the first of the sequence of optimal gains K(1) for the multirate system is the same as the stationery optimal gain for the equivalent single rate system. Thus, as he pointed out, the optimal gains for the multirate system can be obtained easily because it is easy to solve the Riccati equations associated with a single rate system. Once the first optimal gain K(1) in the sequence is computed, the rest can be calculated recursively from the first, which is shown in both [Glasson, D12] and [Amit].

The methods in [Amit] and [Glasson, D12] are exact methods, but they have the additional advantage of being able to convert a continuous time specification into one that can be used directly for multirate design. Apart from the limitation of the LQ design methodology, these methods are satisfactory. There is, however, one point that is not fully justified in these works. It has been mentioned in both [Glasson, D12] and [Ami], that the first step of design should be to determine the weightings for a continuous time cost functional. Then these weightings are transformed into equivalent weightings for the sampled data system. The equivalence is in the sense that both cost functionals put the same proportion of emphasis on the variables. But it is not clear if it is desirable to minimize this equivalent cost functional, because the effect of sampling might require to minimize a different cost functional with a different proportion of emphasis in order to achieve an optimal performance.

Comment: We will briefly summarize the survey of analysis and design methods discussed. The intention is to motivate our study, rather than giving a verdict on these methods. For a more comprehensive survey on this subject, see [Walton, D11]. First of all, the approximate methods are not interesting to us, for what we are seeking for is a rigorous approach. As far as analysis is concerned, the discrete time methods, whether in frequency of state space domain, all give necessary and sufficient conditions for closed loop stability. Perhaps the complexity of use is the main point to note. For this reason the frequency decomposition technique seems less convenient than switch decomposition. Also, due to the work of Coffey et al, switch decomposition can be made systematic by the use of vector operator [Coffey, D3]. State space methods are powerful in that they are applicable to a large class of sampled data systems. The biggest problem of these methods, we believe, is that the sampling rates are treated not only as system parameters but also as part of structural factors. In other words, changes in the sampling rates will result in a "new" system, hence the same procedure for analysing such a system has to be repeated for every new set of sampling rate. A second weak point of these discrete time methods is that they are incapable of dealing with unstructured variations in the continuous time part of the systems. In this regard, the conic sector approach in [Thompson] is more appealing, since it offers a unified approach to sampled data system. Conic sector method does not suffer from dimension growth, and can handle robustness analysis easily. But it only gives sufficient conditions for stability.

The decomposition methods for multirate system analysis are not suitable for design purposes, because design goals can hardly be translated into those for the equivalent single rate system, apart from stability. The state space methods, on the other hand, have been successfully used for design, namely the LQG methods. The key point is that an equivalent cost functional can be obtained. Conic sector methods as studied in [Thompson] has not been used for design purposes, and the reason for this is that Thompson's conic sector can be computed only after the controller has been given.

It is time to find out what we can do. We are interested in a unified approach to sampled data systems, both in continuous time versus discrete time and analysis versus design. None of the methods discussed so for satisfies this requirement. Furthermore, we would like to have more flexibility in design than LQ methodology could offer, most importantly, the ability to treat robustness issue as an integral part of design. All these point to the direction of conic sector, or perturbation approach, for in this framework sampling systems is viewed as a continuous time device.

1.2 *Our point of view*

The basic philosophy we will adopt is this: if the sampling rates in a system are reasonably high, the system should behave similarly to a LTI system. Thus the following questions are raised:

(1) What kind of hybrid controllers are similar to LTI systems?

(2) In what way are they similar to LTI systems?

(3) Can one identify the parts of a hybrid system that make it not like a LTI systems?

(4) Can this part be separated from the rest of the hybrid system?

The answers to (1) and (2) are given in [Thompson] for single rate systems. For multirate systems, it depends on the sampling scheme used. We will show that for the practically useful schemes, the answers to (1) and (2) can be fully given. Answers to (3) and (4) are affirmative for the class of most commonly used hybrid systems. For this class of multirate systems, we can separate the "trouble-making" parts from the rest of a hybrid system, i.e. to separate a multirate hybrid controller into LTI and time-varying parts. The time varying parts are shown to be sample-and-hold (SAH) operators, to them a whole chapter is devoted. It turns out we can find quite tight conic sectors to bound these SAH operators, although the concept of conic sector is not essential here. The important point is: we can approximate the action of a hybrid system by substituting the SAH operators with their LTI approximations. We can also find bounds on the errors. Furthermore, because of the separation property of this class of multirate controllers, both the approximation and the error bounds are independent of the LTI part of the hybrid system. In short, a hybrid controlled feedback system is just an ordinary LTI system with some SAH operators inserted in the loop. In contrast to [Thompson], we only seek to put the non-LTI elements in a hybrid controller into conic sectors, instead of the whole controller. We will see this is why our method works for multirate systems.

As an additional application of the separation property, one can readily apply the switch decomposition techniques to get necessary conditions of closed loop stability.

This philosophy enables one to use continuous time LTI techniques to sampled data systems. And in this sense we consider our approach a unified methods for linear systems.

1.3 *Theoretic Fundations*

Here are some points concerning the theoretic framework we will work in. The basic motto is to make the theory as rigorous as possible and yet not to fall into mathematical complexity. The ultimate goal is to present a working method for analysing and designing multirate systems. We feel that the following two aspects are particularly important.

Stability Theory

Stability is of prime importance in both analysis and design. Since sampled data systems are time-varying, it is convenient to use operator theoretic terminology to describe their dynamic behaviour. Although the systems under study are time-varying, we eventually use LTI system theory to estimate their stability and other properties. The bridge over the gap between time-varying and time-invariant is the perturbation theory for linear operators. Thompson has used conic sector method to achieve this, and the theory used in [Thompsom] was from [Safonov]. For our purpose, a conic sector cone(c, r) defined in [Thompson] is a neighbourhood of the operator c such that every element in this neighbourhood has a relatively bounded difference from c. But one has considerable flexibility in using perturbation theory of Kato [Kato] in the case where c is not stable (whose domain is not the whole space). We found that all the results that can be obtained by conic sector method are also derivable by perturbation method, as far as linear systems are concerned. Conic sector provides an instructive conceptual framework and a compact presentation of robust stability, so whenever it is appropriate we also use terms form conic sector theory.

H^{∞} system design

In [Safonov, et al], a design method is presented that makes part of a system inside (or outside) of a conic sector. Our basic design methodology can be summarized as making the LTI part of a hybrid controlled system outside of the conic sectors containing the SAH operators, although the idea came from Dr. D. J. N. Limebeer in a different form. H^{∞} theory is used to form a systematic design procedure. It is a natural consequence of the preceeding analysis rather then an arbitrary choice. The steps leading to this choice is briefly as follows:

(1) A hybrid controller (of the kind that is most commonly used) can be separated as composition of SAH and finite dimensional LTI operators, and the only designable part of a hybrid controller is the LTI part;

(2) Stability of a multirate hybrid system can be determined in two steps: first, the nominal stability which is completely characterized by some transfer functions; second, robust stability that the nominal system must possess to resist the effects of sampling. Only a subset of the nominal-stabilizing controllers have this additional property. This in turn can be expressed as a conic sector condition. Therefore, a stabilization problem of a sampled data system can be formulated as a robustness stabilization problem of an appropriate LTI system.

(3) A design should achieve performance while at the same time guarantee stability; also performance measured with nominal system should not deteriorate too much when true "perturbed" version of the system is switched in. A sufficient condition for this can be expressed in terms of the L^{∞} norm of certain operator. H^{∞} design is the right tool to achieve the minimization of this norm.

1.4 *The main contributions of this thesis*

We have looked into three aspects of multirate sampled data systems: (a) structure of general multirate hybrid controllers; (b) analysis, mainly of stability, of multirate systems; (c) design methodology for such systems. These aspects may be of independent interests, but they are parts of a unified approach of linear sampled data systems. The most important conclusion of this thesis is that continuous time approach to hybrid controlled systems, first studied in [Thompson], is valid and can be effective.

The effectiveness of this approach hinges on the structural analysis of a general class of multirate systems. About this, we believe we have achieved the following original results:

(1) Characterization of the input-output behaviour of two practical class of sampling schemes, namely the input and output triggered schemes. We have shown that the variable part of a hybrid controller is entirely specified by an LTI system;

(2) For a hybrid controller to have a separable structure (defined in chapter 4) it is necessary that the realization scheme is both input and output triggered. In this case, the input-output relation can be decomposed as the cascade of SAH operators and a LTI systems.

Since the SAH operators are very important to our exposition, we studied their properties in details. Among the interesting features of SAH operators, we have shown in particular:

(1) SAH operators can be approximated effectively by LTI operators;

(2) When constrained to a special class of signals, the optimal approximation to an SAH operator S_T is h_T. h_T is in general a good approximation if the sampling rates are practically high;

(3) Quantitative criterion are given to estimate the effectiveness of the approximations and to indicate the class of signals for which the approximation assumes the smallest and the

greatest errors.

(4) Two bounds are found. One is valid for finite bandwidth signals satisfying the Shannon sampling condition. The other is valid for all signals produced by passing L^2 functions through an LTI filter with strictly proper transfer function. The general validity of the second bound makes it suitable for design purposes, because we must not pose any assumption on the signals before the controller is defined.

We have used LTI approximation to SAH and bound on the error to analyze the stability of multirate systems. In such a method, only LTI system techniques are needed, e.g. Nyquist array and frequency dependent singular values. We have shown that scaling can reduce the typical conservatism of Small Gain type of tests, though a complete solution to the optimal scaling problem is still unsolved.

The most important potential of the approach presented in this thesis is, we believe, a systematic design procedure for multirate sampled data systems. As opposed to LQ approach to sampled data system design, the influence of sampling can be accounted for by a trade-off procedure. This trade-off procedure will result in not only a balance between performance optimization and stability, but also a balance between nominal performance and degradation due to sampling. The following have been achieved concerning the stabilization of multirate systems:

(1) The stabilization problem of multirate system has been formulated as an H^∞ minimization. The bounds on the SAH operators are shown to be the best choices as weight functions.

(2) Optimal scaling is introduced to reduce the conservatism inherent of the Small Gain theorem. A solution to a suboptimal scaling is proposed and solved, and numerical procedure developed.

(3) Stability of the designed system can be predicted from the optimum of the H^∞ minimization.

A pure stabilization formulation will not result in a satisfactory design. This is well demonstrated by numerical examples. The significance of the stabilization problem as posed in an H^∞ framework is that useful information can be gained from this process. Apart form the scaling matrix, one can also inspect how demanding the stabilization task is. This is subsequently used in the trade off of an overall performance. Trade-off between several performance requirements is a difficult problem, which ultimately accounts for the complexity of weights selection for H^{∞} optimization. We have developed a general algorithm for making balance between two competing targets, where one of them is a "hard" requirement, i.e. it must satisfy certain prescribed criterion. This problem is casted as a constrained optimization. In the case of multirate system design, the constraint reflects the requirement for stability and robust performance. The following issues are investigated:

(1) Characterization of the feasible region for the constrainted optimization problem;

(2) A two-step one-dimensional binary search scheme is proposed to solve the constrainted optimization problem without calculating derivatives. An approximate solution is presented to ease the computation demand;

Once the balance is found, we arrive at a standard H^{∞} optimization problem. It is sometimes justifiable to use suboptimal solutions to this problem as is indicated in the examples. Post-processing to get the multirate controller is straightforward.

At this stage it is not possible to judge the full potential of the methods proposed in this thesis, for substantial experience has yet to be gained. In particular, the performance of an H^{∞} design is up to the choice of "correct" weights, which is more experience dependent than theory oriented. We have succeeded in transforming a sampled data system design into the context of a LTI system synthesis. Therefore, successful application of H^{∞} methods to practical situations will make the method of this thesis a useful addition to the collection sampled data system design methodologies.

Chapter Two
Stability Theory

2.0 *Introduction*

This chapter contains some background material on the stability theory of dynamic systems. This theory is particularly well developed for LTI systems. In this case Nyquist type of stability tests are widely used in systems analysis. But for time-varying systems, there is lacking of something that is both general enough and easy to use. This difficulty has motivated a lot of research into this area.

A large body of research in this area has made use of functional analysis and operator theory [Davis, M2, M4]. There are at least two reasons behind this. Firstly, many physical systems have an input-output behaviour that can be conveniently described as an operator. Secondly, operator theory itself is well established. Many of the concepts of engineering systems, such as stability, time-invariance and causality, can be rigorously stated in operator theoretic terms, and many classical results in operator theory can be used in analysing systems. It is particularly true for stability of systems.

Mathematically speaking, the stability of a system is equivalent to the invertability of certain operators describing the interconnection of the system. Since the concept of an operator embraces a wide range of situations, by studying the invertability properties of the operators of interests more insight can be gained. It is these ideas which will concern us in this chapter. We begin by defining the spaces which are suitable for the purposes of our study. Systems are then defined as operators on these spaces. The stability of a system is then defined in terms of the domains of certain operators describing the input-output mapping. We will then state some results concerning the stability of systems described by certain operators. Particular emphasis is put on the stability properties of systems under perturbations. The terminology, approach and main results here are taken from [Kato]. Some of the similarities of the results in [Safonov] and [Thompson] to those of [Kato] are discussed. Finally, we will point out why the foregoing theory can be used to lay the foundation to a frequency domain method for sampled data system, and indeed any system that can be efficiently approximated by LTI operators.

2.1 *Definition and Terminology*

In this section terminology and definitions are introduced which will be needed later in this chapter as well as in the rest of the thesis.

i) space and operators

We need function spaces $L^2[\mathbb{C}^n; (0, +\infty)]$ and $L^2[\mathbb{C}^2; (-\infty, +\infty)]$ for signals.

Specifically:

$$L^2[C^n;\ (-\infty,\ +\infty)] \triangleq \{\ x(t) : (-\infty,\ +\infty) \mapsto C^n \mid \int_{-\infty}^{+\infty} \bar{x}^T(t)\cdot x(t)dt < \infty\ \}$$

and L^2 $[C^n;\ (0,\ +\infty)]$ is defined as a subspace of $L^2[C^n;\ (-\infty,\ +\infty)]$ of functions which are (almost everywhere) zero for $t<0$. When $n = 1$, $L^2(-\infty,\ +\infty)$ and $L^2[0,\ +\infty)$ are used. We will use a prefix **R** for spaces of real functions. For example, $\mathbf{R}L^2(0,\ +\infty)$ is the space of real functions that are square integrable on $(0,\ +\infty)$. L^2 is a Hilbert space with an inner product defined by :

$$< x, y > \triangleq \int_{-\infty}^{+\infty} \bar{x}^T(t)\cdot y(t)dt$$

together with a norm:

$$|x| \triangleq \sqrt{< x, x >}\ .$$

This Hilbert space is denoted as $\mathbf{H}[C^n;\ (-\infty,\ +\infty)]$. If the dimension of a space A is finite we use dim(A) to represent the dimension. If X is a subspace of H then $X^\perp$ denotes the orthogonal complementary space of X in H:

$$X^\perp \triangleq \{x \in H \mid <x, m> = 0\ \forall\ m \in X\ \}.$$

If two spaces H_1 and H_2 are given, we define the product of the two spaces as:

$$H_1 \times H_2 \triangleq \{\ (x, y) \mid x \in H_1 \text{ and } y \in H_2\}\ .$$

The inner product and hence also the norm of elements on a product space are induced from the individual spaces:

$$<(x, y), (u, v)> \triangleq <x, u> + <y, v>$$

and

$$|(x,y)| \triangleq \sqrt{\|x\|^2 + \|y\|^2}\ .$$

A linear manifold M in a space is a set satisfying:

$$\forall\ x, y \in M,\ \alpha,\ \beta \in C\ \Rightarrow\ \alpha x + \beta y \in M.$$

A closed linear manifold is a subspace.

We need the concept of an extended space. The definition below is from [Safonov], but we restrict it only to a simple case.

Definition 2.1.1: Let x_τ denote the truncated version of x:

$$\begin{aligned} x_\tau(t) &\triangleq x(t) \text{ if } |t| \leq \tau \\ &\triangleq 0 \quad \text{elsewhere.} \end{aligned}$$

The extended space of X, denoted as X_e, is the collection of functions x such that $x_\tau \in X \ \forall \ \tau$.

Because we have in mind the applications in stability analysis, the following definition of an operator is adopted:

Definition 2.1.2: Let X and Y be function spaces. A linear operator T form X to Y is a linear mapping from a linear manifold in X into another linear manifold in Y. These linear manifolds are called the domain and range of the operator, denoted Dom(T) and Rang(T) respectively.

Since we will only need linear operators, all operators will be assumed linear throughout.

Definition 2.1.3: An operator is bounded if there exists a constant $C < \infty$ such that

$$||T \cdot x|| \leq C||x|| \quad \forall \ x \in \text{Dom}(T).$$

For a bounded operator T its domain Dom(T) can be extended to cl{Dom(T)}, the closure of Dom(T), by

$$T \cdot x \triangleq \lim_{n \rightarrow \infty} T \cdot x_n$$

where $\{x_n\} \subset \text{Dom}(T)$ and $x_n \rightarrow x$. A norm can be defined for bounded operators:

$$||T|| \triangleq \sup_{x \in \text{Dom}(T)} \frac{||T \cdot x||}{||x||}$$

and it is called the induced norm from the space they are defined on. The induced norm from L^2 is sometimes called the infinite norm.

Definition 2.1.4: The graph of an operator from X to Y, denoted as G(T), is a subset of X×Y:

$$G(T) \triangleq \{ (x,y) \subseteq X{\times}Y \mid x \in \mathrm{Dom}(T) \text{ and } y \in \mathrm{Rang}(T) \}.$$

Definition 2.1.5: An operator is said to be closed if G(T) is a closed set in X×Y .

Projections are an important class operators. An operator P is called an oblique projection if it satisfies that

$$P{\cdot}P = P,$$

and is called an orthogonal projection if it is further self-adjoint.

ii) Systems and operators

Many system theoretic concepts have counterparts in operator theory. In this section we will attempt to make some of these connections clear. We will first try to connect the concept of a system to that of an operator. A physical system is an entity whose output is completely determined by the input signal and the initial conditions. For linear systems the output is a linear function of the input and initial conditions. Such a system is shown in Fig. 2.1.1.

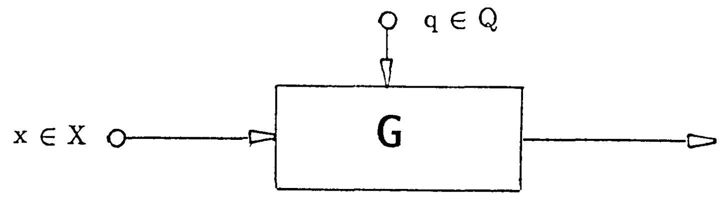


Fig. 2.2.1 General Linear System

Let X be the set of all input signals and Q the initial conditions, then the systems is described by a linear function G:

$$G\colon X{\times}Q \mapsto Y$$

We assume that the input signals form a Hilbert space. An engineering interpretation of this could be that we confine our attention to signals with finite energy. For certain signals in X and under some initial conditions the system will produce an output with finite energy too, therefore one can define an operator from some linear manifold of X×Q to another manifold in Y. We shall call it the operator describing the system, although it only specifies the "well behaved" part of the system. We can also see from this that one must specify the definition domain of the operator, for it is part of the description of the system.

What follows is a discussion of the role of the initial conditions. To do this we need the concept of controllability. A system defined as above is said controllable if for any initial condition in Q one can choose an input function such that the output function is identically zero after some finite time. For controllable systems we can assume that the initial conditions are always zero because their effects can be perfectly cancelled by a proper choice of input. In this case, We simplify the problem by considering the operator G: $X \mapsto Y$ only.

Sometimes it is desirable to consider systems defined on the whole time scale $t \in (-\infty, +\infty)$. The operators describing these systems are defined on $X \mapsto Y$ where both X and Y are function spaces on $(-\infty, +\infty)$, and the initial condition at $t = -\infty$ is assumed to be zero.

Definition 2.1.6: A system is said time-invariant if the operator G describing the system satisfies:

$$G \circ D_\tau \cdot x = D_\tau \circ G \cdot x \;\; \forall\; x \in \mathrm{Dom}(G)$$

where D_τ is the pure delay operator:

$$[D_\tau \cdot x](t) \triangleq x(t - \tau).$$

In other words G commutes with delay operator. The operator is also said time-invariant.

For linear time-invariant (LTI) systems an invaluable tool of analysis is the Laplace transformation. Whenever it makes sense time-invariant operators also commute with differential operator D:

$$D \cdot x(t) \triangleq e(t)$$

where x(t) is absolutely continuous: $x(t) = \int_0^t e(\theta)d\theta$, for some $e \in L^2[0, \infty)$. This property has profound implications. In fact this alone is enough to lead to the conclusion [Power]: there exists a G(s) such that for any $(x, y) \in \mathrm{Graph}(G)$

$$y(s) = G(s)x(s)$$

where x(s) and y(s) are the Laplace transforms of x and y, respectively. G(s) is called the transfer function of the systems and we also use it for the operator. It should be noted that systems defined on half time scale are not time-invariant in the above sense. For instance if we are only interested in the positive times, then in general operators do not commute with D_τ if $\tau < 0$. However when they do commute with D_τ for $\tau \geq 0$ and when the initial condition is zero, it is still possible to relate the input-output by a transfer functions defined on C_+ only. These systems are also called time-invariant.

Definition 2.1.7: A system is causal if for any $x \in \mathrm{Dom}(G)$ with $x(t) = 0$ for $t \leq \tau$, $y(t) = [G \cdot x](t) = 0$ for $t \leq \tau$.

It is easy to determine the domain of an LTI operator describing a causal system from its transfer function. We know that if $x(t) = 0$ for $t \leq 0$, and $x \in L^2$, then $x(s)$ is analytic in C_+. If $y = G \cdot x$ is to be in $L^2[0,\infty)$, $y(s)$ must be also analytic in C_+. But $y(s) = G(s)x(s)$, so $x(s)$ must be such that it cancels all the singularities of $G(s)$ in C_+. If $G(s)$ admits an inner-outer factorization [Rudin]: $G(s) = N(s)D^{-1}(s)$, where both $N(s)$ and $D(s)$ are analytic in C_+, then the domain of G is given by

$$\mathrm{Dom}(G) = \{ x \in X \mid x(s) = D(s)e(s), \text{ e is analytic in } C_+\}.$$

Now we can define the stability of a system.

Definition 2.1.8: A system is said stable if the operator $G: X \mapsto Y$ describing the system has dense domain, i.e. $\mathrm{cl}(\mathrm{Dom}(G)) = X$. In addition a system is said to have finite bound if

$$|G| \triangleq \sup_{\substack{x \in X \\ x \neq 0}} \frac{|G \cdot x|}{|x|} = \; < \infty.$$

A system is said to have extended stability if there exists a constant $C < \infty$ such that

$$|(G \cdot x)_\tau| \leq C\, |x_\tau| \;\forall\; x \in X_e.$$

The following lemma shows the relation between these concepts .

Lemma 2.1.9: If a causal system G is stable and has finite bound, then it also has extended stability. Furthermore:

$$\inf \{ C \mid |(G \cdot x)_\tau| \leq C\, |x_\tau| \;\forall\; x \in X_e.\} = |G|.$$

Proof:

Firstly, for any $\tau > 0$, and $\forall\; x \in X_e$, we have that

$$\|(G\cdot x)_\tau\| \leq \|G\| \, \|x_\tau\|$$

hence G is extendedly stable. Now suppose that there is a constant $C < \|G\|$ such that $\|(G\cdot x)_\tau\| \leq C \, \|x_\tau\|$ for all $\tau > 0$ and $x \in X_e$. Take an $x \in L^2[0, \infty)$, we have $\underset{\tau\to\infty}{\text{l.m.t.}} \|x_\tau\| = \|x\|$, hence

$$\|G\cdot x\| = \underset{\tau\to\infty}{\text{l.m.t.}} \|(G\cdot x)_\tau\| \leq C \underset{\tau\to\infty}{\text{l.m.t.}} \|x_\tau\| = C \, \|x\|$$

or $\|G\| \leq C$, a contradiction. □

As for the relation to asymptotical stability, we first state:

Lemma 2.1.10: A system G is stable ⇒ for any $x \in L^2$, $y = G\cdot x$ satisfies:

$$\lim_{m,n\to\infty} \int_m^n |y(t)|^2 dt = 0.$$

Proof:

It follows from the fact that $y \in L^2$. □

This is not exactly asymptotic stability unless we further assume that every y(t) is piecewise continuous function. The following lemma gives an idea of when this will happen:

Lemma 2.1.11: Suppose that G is defined by

$$y(t) = G\cdot x \triangleq \int_{-\infty}^{+\infty} G(t\text{-}\theta)x(\theta)d\theta$$

and $G(\cdot) \in L^2$ is a bounded function on $(-\infty, +\infty)$, then y(t) is continuous in t for all $x \in L^2$.

Proof:

It is clear that

$$|y(t_1) - y(t_2)| \leq \int_{-\infty}^{+\infty} |[G(t_1 - \theta) - G(t_2 - \theta)]x(\theta)|d\theta$$

$$\leq \left[\int_{-\infty}^{+\infty} |G(t_1 - \theta) - G(t_2 - \theta)|^2 d\theta \cdot \int_{-\infty}^{+\infty} |x(\theta)|^2 d\theta \right]^{\frac{1}{2}}$$

It is known [Rudin, p196] that translation D_τ: $e(t) \rightarrow e(t - \tau)$ is a continuous mapping from R to $L^2(L^p, 1 \leq p < \infty)$, we conclude that $|y(t_1) - y(t_2)| \rightarrow 0$ as $|t_1 - t_2| \rightarrow 0$. For LTI systems this is the case if the transfer function is strictly proper. □

2.2 *Mathematical description of stability*

In this section we study some of the basic aspects of stability theory for linear systems. Necessary and sufficient conditions of closed loop stability are given in terms of certain invertability conditions. Results of this kind are well understood for LTI systems, and it is of interests to generalize these results to a more general class of systems. An example is included to demonstrate how the concepts of LTI systems are generalized. Then we preceed to study the continuity property of a feedback system, i.e. the robustness issue. The study of continuity properties paves the ways to the use of perturbation methods for non-LTI systems that are, in an appropriate sense, close to LTI ones.

2.2.1 *Necessary and sufficient conditions of stability*

We assume throughout this section that all the operators under discussion are closed. This is connected to the fact that operators described by proper transfer functions are closed. Under this assumption a stable system is described by a bounded operator with full domain. Before going on discussion, some elementary properties about all stable systems are briefly outlined here. Let the symbol $\mathfrak{S}_{(X,Y)}$ denote the collection of all linear stable systems as well as the operators describing them, from X into Y. The subscript (X, Y) is normally omitted when there is no danger of confusion. It is well known that $\mathfrak{S}$ is a (Banach) space, i.e. it is closed under linear operations (with the induced norm from X and Y).

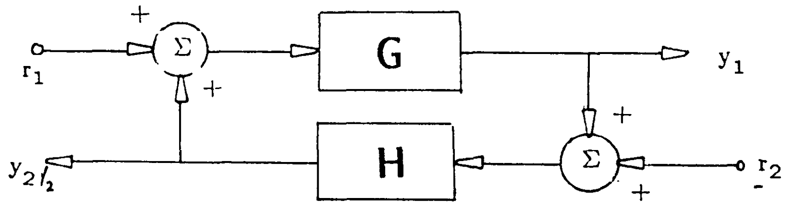


Fig. 2.3.1 Canonical Feedback Structure

The standard configuration of feedback systems discussed here is shown in Fig. 2.3.1. G and H are linear operators with domains Dom(G) and Dom(H) respectively. The input signal to this system is the pair (r_1, r_2) and the output is (y_1, y_2). The input-output mapping is defined through the following system equations:

$$\begin{cases} y_1 = G\cdot(r_1+y_2) \\ \\ y_2 = H\cdot(r_2+y_1) \end{cases} \tag{2.2.1}$$

The explicit relations between (r_1, r_2) and (y_1, y_2) are:

$$\begin{cases} y_1 = (I\text{-}G\cdot H)^{-1}\cdot G\cdot r_1 + (I\text{-}G\cdot H)^{-1}\cdot G\cdot H\cdot r_2 \\ \\ y_2 = (I\text{-}H\cdot G)^{-1}\cdot H\cdot G\cdot r_1 + (I\text{-}H\cdot G)^{-1}\cdot H\cdot r_2 \end{cases} \tag{2.2.2}$$

supposing the inversions are all well defined. This leads to

Lemma 2.2.1: [Vidyasagar, 1984] [Desor, 1980] The feedback system (2.2.1) defines a stable operator $(r_1, r_2) \rightarrow (y_1, y_2)$ if and only if:

(a) (I-G·H) and (I-H·G) are one-to-one, and

(b) $(I\text{-}G\cdot H)^{-1}G$, $(I\text{-}G\cdot H)^{-1}G\cdot H$, $(I\text{-}H\cdot G)^{-1}H\cdot G$ and $(I\text{-}H\cdot G)^{-1}H \in \mathfrak{S}$.

(a) is a kind of well-posedness condition; while (b) states that four operators have to be all stable. In order to make the notation simple we introduce in variables: $r \triangleq (r_1, r_2)^T$ and $y \triangleq (y_2, y_1)^T$, and define

$$K \triangleq \begin{bmatrix} 0 & H \\ G & 0 \end{bmatrix}.$$

Then (2.2.1) becomes

$$(I\text{-}K)\cdot y = K\cdot r. \tag{2.2.1'}$$

The system is stable if $(I\text{-}K)^{-1}K \in \mathfrak{S}$. Since $(I - K)^{-1}K = (I - K\ )^{-1} - I$, we conclude that the system is stable if $(I\text{-}K)^{-1} \in \mathfrak{S}$. So we have shown that stability is just an invertability condition, i.e. the operator $(I - K)$ must be one-to-one and of dense range. In order to pursue the investigation further, the following definitions are introduced [Kato]:

Definition 2.2.2: Let T be an operator from X to Y, and let $\mathrm{Ker(T)} \subseteq X$ denote kernel space of T. The nullity and deficiency of T are defined to be the numbers:

$$\mathrm{Nul}(T) \triangleq \dim(\mathrm{Ker}(T))$$

and

$$\mathrm{Def}(T) \triangleq \dim(\mathrm{Rang}^{\perp}(T)).$$

In terms of Nul and Def, the above discussion can be summarized as

Lemma 2.2.3: The system defined in Fig. 2.3.1 is stable if and only if

$$\mathrm{Nul}(I - K) = \mathrm{Def}(I - K) = 0.$$

Proof:

$\mathrm{Nul}(T) = 0 \Longleftrightarrow$ T is one to one hence invertable. $\mathrm{Def}(T) = 0 \Longleftrightarrow \mathrm{Rang}(T) = \mathrm{Dom}(T^{-1})$ is dense. T^{-1} is closed since T is (by hypothesis), therefore we conclude that T has a bounded inverse from Closed Graph Theorem. □

The real problem is that the nullity and deficiency of an operator are not easy to compute. However for the class of LTI operators whose transfer functions are rational functions of s, this is very simple. The following example illustrates the basic ideas.

Example 2.2.4: Let G(s) be a real rational scaler function that has neither pole nor zero on $j\omega$-axis, and has C_+-poles $\{p_i\}_{i=1}^{M}$ and C_+-zeros $\{z_i\}_{i=1}^{N}$ with corresponding multiplicities m_i and n_i, respectively. It is also assumed that G(s) is proper. Suppose that G(s) is the transfer function of a causal operator G defined on $X = RL^2[0,\infty)$ into X itself. We can show that

(a) $\mathrm{Nul}(G) = 0$;

(b) $\dim(\mathrm{Dom}^{\perp}(G)) = \sum_{i=1}^{M} m_i$;

(c) $\mathrm{Def}(G) = \sum_{i=1}^{N} n_i$.

Proof:

(a) Suppose that $G \cdot x = 0$. This implies, by Paserval identity, that

$$\int_{-\infty}^{+\infty} |G(j\omega)|^2 |x(j\omega)|^2 d\omega = 0.$$

Since $G(j\omega)$ is non-zero $\forall \omega$, $x(j\omega) = 0$ a.e., or $x(t) = 0$ a.e. as a consequence of the uniqueness property of Fourier transform.

(b) Let G(s) have a coprime factorization over the ring of stable proper rational functions: $G(s) = N(s)D^{-1}(s)$, where both N(s) and D(s) are analytic in C_+. Then inner-outer factorizations can be performed on N and D, and the inner factors ϕ of N and θ of D are both rational functions:

$$\theta(s) = \prod_{i=1}^{M} \left[\frac{(s-p_i)}{(s+p_i)}\right]^{m_i}$$

and

$$\phi(s) = \prod_{i=1}^{N} \left[\frac{(s-z_i)}{(s+z_i)}\right]^{n_i} .$$

If a function e(t) is to lie in the domain of G, it must cancel all the poles $\{p_i\}_{i=1}^{M}$ of G(s). In general a function $x \in \mathrm{Dom}(G)$ if and only if

$$x(s) = D(s)e(s)$$

where $e \in X$. We shall now consider the complementary space of Dom(G). Let e be a function in X, the following calculations show:

$$<D\cdot x, e> = \frac{1}{2\pi j}\int_{-\infty}^{+\infty} D(j\omega)x(j\omega)\bar{e}(j\omega)dj\omega$$

$$= \frac{1}{2\pi j}\int_{-\infty}^{+\infty} e(-j\omega)D(j\omega)x(j\omega)dj\omega \quad \text{(e(t) is real function)}$$

$$= \frac{1}{2\pi j}\int_{\Gamma} e(-s)D(s)x(s)ds \qquad \text{(Γ is $j\omega$-axis from -$j\infty$ upwards)}$$

It is known that if a function is analytic in a strap covering the imaginery axis then it can be (up to a constant) uniquely decomposed into the sum of two functions analytic in C_+ and C_- respectively. Thus:

$$e(-s)D(s) = g_+(s) + g_-(s)$$

where $g_+(s)$ is analytic in C_+. The integrand $g_+(s)x(s)$ is analytic in C_+ hence the integral represent the inner product of a causal and anticausal functions, which is zero. This leads to

$$<D\cdot x, e> = \frac{1}{2\pi j}\int_{\Gamma} g_{-}(s)x(s)ds$$

$$= <g_{-}(-s), x(s)>.$$

Now since $g_{-}(-s)$ is a function in $L^2[0,\infty)$, so the inner product with any x in X is zero if and only if $g_{-}(s) = 0$. This is true only when e(-s)D(s) is analytic in C_{+}, in other words all the C_{+}-poles of e(-s) are to be canceled by D(s). In general, functions of the following form

$$e(s) = \sum_{i=1}^{M}\sum_{k=1}^{m_i} \frac{\alpha_{i,k}}{(s+p_i)^k} \tag{2.2.3}$$

have this property. In fact we have:

Lemma 2.2.5: The functions in form of (2.2.3) are the only ones that satisfy:

$$<D\cdot e, x> = 0 \ \forall\ x \in X.$$

Proof:

Note that e must be the sum of functions of the form

$$\frac{p(s)}{(s+p_i)^k}$$

where p(s) is analytic in the whole plane, in order that e(s) only has poles at p_i's. This is to say that

$$p(s) = \sum_{n=0}^{+\infty} a_n s^n \ \forall s.$$

But on the other hand e(s) must be bounded in C_{+} since it is the Laplace transform of an L^2 function [Duren, p191], so $a_n = 0$ for $n \geq k$, i.e. p(s) is a polynomial of order less than k. Functions of this form are in the space

$$sp\{ \{\frac{1}{(s+p_1)^{k_1}}\}_{k_1=1,\dots,m_1} ;\dots; \{\frac{1}{(s+p_M)^{k_M}}\}_{k_M=1,\dots,m_M} \}.$$

whose dimension is obviously $\sum m_i$. □

(c) Analysis can be carried out in the same way for the range of G, which is given by $\phi\cdot X$ [Rudin]. So we have proved the claims. □

Any $e(t) \in L^2(-\infty, +\infty)$ can be decomposed as

$$e(t) = e_+(t) + e_-(t)$$

where $e_+(t)$ is identically zero for $t \leq 0$, and is called the causal projection of e. Since it is a useful concept in later chapters, we give a special symbol P_+ to denote the projection operator. Similarly $e_-(t)$ is called the anti-causal projection, and the mapping $e \mapsto e_-$ is defined as P_-. It is well known that $e_+(s)$ is analytic in C_+, and so is $e_-(s)$ in C_-.

We have seen that the deficiency and nullity are completely determined by the poles and zeros of the transfer function for LTI operators. If in system Fig. 2.3.1, K is LTI, then the conditions for stability are :

(a) $1 - K(j\omega) \neq 0 \ \forall \omega$ and

(b) $[1 - K(s)]^{-1}$ analytic in C_+ .

There are effective methods of testing if condition (b) is satisfied from information about $K(j\omega)$ alone, for example the Nyquist criterion. But when K is not an LTI operator and thus can not be described by a transfer functions, Nul(K) and Def(I − K) have to be determined by other means.

2.2.2 *Robustness of stability*

It is important to study the tolerance of the stability property to variations in the open loop operators. The first issue is to specify what is meant by a small variation of an operator.

Definition 2.2.6: A weighted ball centred at A with radius R, where A and R are operators in the same space such that Dom(R) $\supseteq$ Com(A), is defined as a collection of operators

$$\mathrm{Ball}(A, R) \triangleq \{B: \mathrm{Dom}(B) \supseteq \mathrm{Dom}(A) \mid \| (A - B)\cdot x\| \leq \| R\cdot x \| \ \forall \ x \in \mathrm{Dom}(A) \}.$$

This is a generalization of the conic sector defined in [Thompson]. We have lifted the requirement that A and B are stable. In other words this ball consists all the relatively bounded operators to A [Kato].

The use of an operator R instead of a constant as the radius reduces conservatism since more information about the difference between B and A is made use of [Safonov].

We will substitute the subsystem H in Fig 2.3.1 with a member in the R-weighted ball centred at H and study the stability of the system. All the members in Ball(H, R) are in the form of H + ΔH, where $\mathrm{Dom}(\Delta H) \supseteq \mathrm{Dom}(H)$ and $|\Delta H\cdot x| \le |R\cdot x|$ $\forall x \in \mathrm{Dom}(H)$. We have

Lemma 2.2.7: Suppose that the system (G, H) is stable. Let $H + \Delta H \in \mathrm{Ball}(H, R)$. Then (G, H+ΔH) is stable if

$$|R\cdot G\cdot(I - H\cdot G)^{-1}| < 1. \tag{2.2.4}$$

Proof:

The assumption that (G, H) is stable implies

$$\mathrm{Dom}(H) \supseteq \mathrm{Rang}[G(I - HG)^{-1}] \cap \mathrm{Rang}[(I - G\cdot H)^{-1}]$$

and since $\mathrm{Dom}(R) \supseteq \mathrm{Dom}(H)$ the use of norm in (2.2.4) is justified. Denote by $\tilde{H}$ the operators H + ΔH. We want to show that the systems (G, $\tilde{H}$) are all stable. It can be seen that the following calculations make sense:

$$I - \begin{bmatrix} 0 & \tilde{H} \\ G & 0 \end{bmatrix} = \left[I - \begin{bmatrix} 0 & H \\ G & 0 \end{bmatrix} - \begin{bmatrix} 0 & \Delta H \\ G & 0 \end{bmatrix} \right]$$

$$= \left[I - \begin{bmatrix} 0 & \Delta H \\ 0 & 0 \end{bmatrix} \begin{bmatrix} I & -H \\ -G & I \end{bmatrix}^{-1} \right] \begin{bmatrix} I & -H \\ -G & I \end{bmatrix}$$

$$\triangleq P(I - K)$$

so if $P^{-1} \in \mathfrak{S}$, then (G, $\tilde{H}$) is stable, for $(I - K)^{-1}$ is stable. Further simple calculations lead to

$$P = \begin{bmatrix} I - \Delta H\cdot G(I - HG)^{-1} & -\Delta H\cdot(I - GH)^{-1} \\ 0 & I \end{bmatrix},$$

and all the entries in P have full domains due to the assumption on Dom(ΔH). By making reference to the identity

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{bmatrix}, \text{ for invertable A and C,}$$

we conclude that if the (1,1) entry of P is invertable and have dense domain, then it will follow that P^{-1} is stable. But (2.2.4) is sufficient for this, since

$$\left|\Delta H\cdot G(I - HG)^{-1}x\right| \le \left|RH(I - HG)^{-1}x\right| \quad \forall x$$

or $\left|\Delta HG(I - HG)^{-1}\right| < 1$. □

Conditions in the above lemma can be relaxed. From the proof we have seen that it is enough to have the following set of conditions:

(1) $\mathrm{Dom}(R) \supseteq \mathrm{Rang}\{ G(I - H\cdot G)^{-1}\}$;

(2) $\mathrm{Dom}(\Delta H) \supseteq \mathrm{Rang}\{ (I - G\cdot H)^{-1}\} \cup \mathrm{Rang}\{ G\cdot(I - H\cdot G)^{-1} \}$;

(3) $|\Delta H\cdot x| \le |R\cdot x| \ \forall x \in \mathrm{Rang}\{ G\cdot(I - H\cdot G)^{-1}\}$.

In other words, the assumption $\mathrm{Dom}(\Delta H) \supseteq \mathrm{Dom}(H)$ is not needed.

Remark 2.2.8: Conditions in (2) on the domains of ΔH come from the requirement of internal stability. Though the input r_2 may be fictitious, stability of the mappings from r_2 are necessary in order to guarantee interenal stability. This is easy to understand from a transfer function point of view, because there can be cancelations of unstable poles between G and H. However, if we are only interested in the stability of the mapping from r_1 to y_1, which is the case we will encounter later, we can simplify the conditions further. This is summarized in the following theorem.

Theorem 2.2.9: Suppose that the operator $(I - GH)^{-1}G$ is stable. Let $\tilde{H} \triangleq H + \Delta H \in \mathrm{Ball}(H, R)$. Then $(I - G\cdot\tilde{H})^{-1}G$ is stable if

(1) $\mathrm{Dom}(\Delta H) \supseteq \mathrm{Rang}[G\cdot(I - H\cdot G)^{-1}]$,

and

(2) $\left|R\cdot G\cdot(I - H\cdot G)^{-1}\right| < 1$.

Proof:

First we note that $(I - G\cdot H)^{-1}G = G\cdot(I - H\cdot G)^{-1}$, thus

$$G\cdot(I - \tilde{H}\cdot G)^{-1} = G\cdot[(I - HG) - \Delta HG]^{-1}$$

$$= G\cdot(I - HG)^{-1}[I - \Delta H\cdot G\cdot(I - H\cdot G)^{-1}]^{-1}.$$

The stability assumption on the system (G, H) and (1) ensures that the operator in the bracket has full domain of definition. (2) implies that this operator is bounded away from zero, therefore it is boundedly invertàble [Kato, p190].

Remark 2.2.10: The assumptions on the domain of $\tilde{H}$ is still restrictive, because, in the case of LTI operators, it requires that $\tilde{H}$ should have the same C_+-poles as those of H. If we recall the Nyquist criterion for stability, it is clear that what is needed is that $\tilde{H}$ should have the same number of C_+-poles as that of $\tilde{H}$, but not necessarily on the same locations. The combined effects of having the same number of C_+-poles and closed $j\omega$-axis behaviour ensure that the two operators are close to each other. We seem to have defined the "topology" on the space of operators in a rather conservative way, i.e. it is so strong that some of the "small" changes are excluded as the candidates as members of the neighbourhood. Can we define a weaker topology that would still make the stability of systems a continuous function? The answer is yes. Actually the weakest topology that possesses this property is defined in [Kato]: the Gap between operators. We deplore some of this below for the interests of generality. We begin by working in a general operator setting, for the sake of simplicity. As we will see, in the case of frequency responses some further detailed study is necessary.

Definition 2.2.11: The gap between two closed subspaces M and N of H is defined as:

$$\Theta(M,N) \triangleq \max\Big(\sup_{\substack{x\in N \\ |x|=1}} \text{dist}(x, M), \sup_{\substack{x\in M \\ |x|=1}} \text{dist}(x, N) \Big). \qquad (2.2.5)$$

Theorem 2.2.12: Let P_M and P_N denote the orthogonal projections onto M and N respectively, then

$$\Theta(M, N) = \max\left(|(P_M - P_N)P_M|, |(P_M - P_N)P_N| \right)$$

$$= |(P_M - P_N)| \qquad (2.2.6)$$

If $\tilde{P}_M$ and $\tilde{P}_N$ are oblique projections onto M and N then

$$|(P_M - P_N)| \le |(\tilde{P}_M - \tilde{P}_N)|.$$

Proof:

See [Kato, p56].

We can see from here that it is possible to use norm to compute the gaps, provided one can get an expression for the projection operators. This is done in the next theorem. But first we need

Lemma 2.2.113. Suppose that T is an operator from X to Y that can be factorized as:

$$T = ND^{-1}$$

where N: E↦Y and D: E↦X have dense domains in E, and where D is one-to-one and Rang(D) = Dom(T); then the graph $G(T) \subseteq X \times Y$ is given by:

$$G(T) = \begin{bmatrix} D \\ N \end{bmatrix} E.$$

Proof:

For any $e \in X$ we can see that $D \cdot e \in \mathrm{Dom}(T)$ so let $x = D \cdot e$. We have:

$$N \cdot e = N \cdot D^{-1} x = T \cdot x,$$

or

$$\begin{bmatrix} D \\ N \end{bmatrix} E \subseteq G(T).$$

On the other hand if $(x, y) \in G(T)$, just let $e = D^{-1} \cdot x$ (because $x \in \mathrm{Dom}(D^{-1})$), and clearly

$$\begin{bmatrix} D \\ N \end{bmatrix} \cdot e = (x, y) \in \begin{bmatrix} D \\ N \end{bmatrix} E$$

$\Rightarrow$

$$\begin{bmatrix} D \\ N \end{bmatrix} E \supseteq G(T).$$

So we conclude $\begin{bmatrix} D \\ N \end{bmatrix} E = G(T)$. □

Theorem 2.2.14: Suppose that M is a closed subspace of a Hilbert space H and is defined by

$$M \triangleq F \cdot X \triangleq \begin{bmatrix} D \\ N \end{bmatrix} \cdot X$$

where both D and N are closed operators and have dense domains in X. Also assume that D is one-to-one. Let P_M denote the projection operator onto M. Then we have:

$$P_M = \begin{bmatrix} D \\ N \end{bmatrix} \Delta^{-1}[D^* \ N^*] \qquad (2.2.7)$$

where

$$\Delta \triangleq F^*F = [D^*D+N^*N]$$

has dense range and hence has an inverse with dense domain.

Proof:

It is clear that P_M has range G(F), and what remains to be verified is that P_M as defined in (2) is an orthogonal projection operator. But this is easily done because we can see that $P_M \cdot P_M = P_M$ and $P_M^* = P_M$. In order to prove that Δ has dense range we assume the opposite and deduce that there will be an $y \in X$ such that

$$<F^*F \cdot x, y> = 0 \ \forall \ x \in X$$

$\Rightarrow$

$$|F \cdot y| = 0, \text{ if we choose } x = y.$$

But this can only be true if y = 0, because of the assumption that F is one-to-one. This proves the theorem. □

In order to apply the above theory to frequency response operators it is important to define a suitable space of functions. If the space is chosen to be $L^2(-\infty,\infty)$, it is difficult to represent an operator as a coprime factorization of operators with dense domains, but if the space is $L^2[0,\infty)$ we run into the problem of the adjoint operators being not representable as transfer functions. Here we choose to work with $L^2[0, \infty)$, so the formula for the gap between will involve the computation of norms of an operator that is a mixture of Laurent and Toeplitz. Although it is not clear how to numerically calculate such a norm, an upper bound can found in terms of the coprime factors only. Let T(s) have a coprime factorization

$$T(s) = N(s)D^{-1}(s)$$

such that both N(s) and D(s) are analytic in C_+. It is seen that if the function T(s) represents an operator T, then T has range $N(s) \cdot H^2$, and its graph is given by

$$G(T) = \begin{bmatrix} D(s) \\ N(s) \end{bmatrix} \cdot H^2.$$

Let $\Delta(s) \triangleq [D^T(\text{-}s)D(s) + N^T(\text{-}s)N(s)\,]^{-1}$, and we know that $\Delta(s)$ can factorized as $\alpha(s)\alpha^T(\text{-}s)$, where $\alpha(s)$ is analytic in C_+. We assert that the orthogonal projector onto G(T) is

$$\begin{bmatrix} D(s) \\ N(s) \end{bmatrix} \alpha(s)\ P_+ \cdot \alpha^T(\text{-}s)\ [D^T(\text{-}s),\ N^T(\text{-}s)\ ]. \qquad (2.2.8)$$

$$= \begin{bmatrix} D(s) \\ N(s) \end{bmatrix} \alpha(s) \cdot \mathbf{T}_\psi$$

where $\mathbf{T}_\psi$ is the Toeplitz operator with symbol $\psi = \alpha^T(\text{-}s)\ [D^T(\text{-}s),\ N^T(\text{-}s)\ ]$. Note that although the operation involves functions not in $L^2[0,\infty)$, the final effect is well defined entirely in the space $L^2[0,\ \infty)$. To verify that (2.2.8) is indeed the projection we are looking for, we first note that it maps into the graph of T. It is also easy to show that $P^2{=}P$, so it is a projection. To see it is an orthogonal projection we need to prove that (2.2.8) defines a self-adjoint operator. In the derivation below we take the liberty of switching between $L^2(\text{-}\infty,\infty)$ and $L^2[0,\infty)$. A more rigorous justification via complex analysis is also possible but tedious. We have

$$< \begin{bmatrix} D(s) \\ N(s) \end{bmatrix} \alpha(s)P_+ \cdot \alpha^T(\text{-}s)[D^T(\text{-}s),\ N^T(\text{-}s)]x(s),\ y(s) > \qquad (\text{in } L^2[0,\infty)$$

$$= <P_+ \cdot \alpha^T(\text{-}s)[D^T(\text{-}s),\ N^T(s)]x(s), \alpha^T(\text{-}s)[D^T(\text{-}s),\ N^T(s)]y(s) > \quad (L^2(\text{-}\infty,\infty)\ )$$

$$= <\alpha^T(\text{-}s)[D^T(\text{-}s),\ N^T(s)]x(s), P_+ \cdot \alpha^T(\text{-}s)[D^T(\text{-}s),\ N^T(s)]y(s) > (L^2(\text{-}\infty,\infty)\ )$$

$$= <x(s),\ \begin{bmatrix} D(s) \\ N(s) \end{bmatrix} \alpha(s)P_+ \cdot \alpha^T(\text{-}s)[D^T(\text{-}s),\ N^T(\text{-}s)]y(s) > \quad (\text{back in } L^2[0,\infty)\ ),$$

i.e. the expression in (2.2.8) is self-adjoint. However, the involvement of P_+ has made it is impossible to represent the operator defined in (2.2.8) as a transfer function.

Some bounds on the gaps are given below which enable one to get around the problem of $\mathbf{T}_\psi$. We will not use the results presented in [El-Akkary] because they involve the calculation of the gap between domains, which in general is difficult to compute. First we have

lemma 2.2.15: Let T be closed and P_T the orthogonal projector onto the graph of T. Suppose that R satisfies: $R \cdot R = I$ and $R = R^*$. Then the orthogonal projector onto $R \cdot T$ is

$$P_{RT} = \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} \cdot P_T \cdot \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix}.$$

If R is a general non-singular matrix, then

$$\tilde{P}_{RT} = \begin{bmatrix} I & 0 \\ 0 & R^{-1} \end{bmatrix} \cdot P_T \cdot \begin{bmatrix} I & 0 \\ 0 & R^{-1} \end{bmatrix}.$$

defines an oblique projection.

Proof:

If $(x, y)^T \in G(T)$ then $(x, Ry)^T \in G(RT)$. Thus the above defines an operator mapping into $G(RT)$. It is easy to see that P_{RT} thus defined satisfies i) $P_{RT}P_{RT} = P_{RT}$ and ii) $P^*_{RT} = P_{RT}$. So it is the one we want. In the case R is only non-singular the second expression is easily seen to satisfy $\tilde{P}^2_{RT} = \tilde{P}_{RT}$. □

Lemma 2.2.16: Let A and B be closed, and let $T = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$. Suppose the orthogonal projections onto $G(A)$ and $G(B)$ are P_A and P_B. We have

$$P_T = \psi \cdot \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \cdot \psi$$

where ψ is unitary.

Proof:

If $(x_1, y_1) \in G(A)$ and $(x_2, y_2) \in G(B)$, then $(x_1, x_2; y_1, y_2) \in G(T)$. Take a point in the appropriate space $(u_1, u_2; v_1, v_2)$, i.e. $P_A(u_1, v_1) = (x_1, y_1)$ etc. If we define a permutation matrix ψ as:

$$\psi \triangleq \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix}$$

and use the definitions of P_A and P_B we can see that (3.y) maps $(u_1, u_2; v_1, v_2)$ into $(x_1, x_2; y_1, y_2)$. Its other properties for an orthogonal projection is easily checked. □

These two lemmas lead to the conclusion:

Theorem 2.2.17: The gap between $\begin{bmatrix} 0 & B_i \\ A_i & 0 \end{bmatrix}$, i=1,2 is given by $\max\{ \|P_{B_1} - P_{B_2}\|, \|P_{A_1} - P_{A_2}\| \}$.

Proof:

It follows from the preceding lemmas and the relation

$$\begin{bmatrix} 0 & B \\ A & 0 \end{bmatrix} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

that the projection onto the corresponding graph is

$$\gamma\begin{bmatrix} P_{B_i} & 0 \\ 0 & P_{A_i} \end{bmatrix}\gamma^T \text{ , } i = 1, 2$$

where γ is an unitary operator. So the gap is

$$\left\|\begin{bmatrix} P_{B_1} - P_{B_2} & 0 \\ 0 & P_{A_1} - P_{A_2} \end{bmatrix}\right\|$$

and the theorem follows. □

The situation we are facing requires to find the gap between operators of the form

$$\begin{bmatrix} I & B_i \\ A_i & I \end{bmatrix},$$

which is more complicated, for the graphs of them can not be obtained as orthogonal compositions of simple graphs. We could use the formula [Kato]:

$$\text{gap}\{ A+S, A+T \} \leq 2\left(1+\|A\|^2\right)^{\frac{1}{2}}\text{gap}\{S, T\}$$

as was done in [El-Akkary], but this is conservative, for no structural information of A, which is I in our case, is used. Although it is hard to find the orthogonal projection onto the graphs of operators in the form (I + K), it is possible to find an oblique one. Note that if $(x, y) \in G(T)$, then $(x, x+y) \in G(I+T)$. Let

$$\pi = \begin{bmatrix} I & 0 \\ I & I \end{bmatrix}$$

we find that the operator $\pi \cdot P_T \cdot \pi^{-1}$ maps a pair (x, y) into G(I+T). But it is obviously a projection. Using this we get

$$\begin{aligned} \text{gap}(I+T_1, I+T_2) &= \| \pi(P_{T_1} - P_{T_2})\pi^{-1}\| \\ &\leq \|\pi\|\|P_{T_1} - P_{T_2}\|\|\pi^{-1}\| \\ &= \frac{3+\sqrt{5}}{2}\|P_{T_1} - P_{T_2}\|, \qquad (2.2.9) \end{aligned}$$

which is an improvement over the inequality quoted from [Kato]. Having in mind the situation where only part of the system is under perturbation, we can reduce the bound even further. Suppose in the foregoing analysis only the operator A is perturbed. Some simple calculation shows that

$$\text{gap}\left\{\begin{bmatrix} I & B \\ A_1 & I \end{bmatrix}, \begin{bmatrix} I & B \\ A_2 & I \end{bmatrix}\right\} = \|\Phi_1(P_{A_1} - P_{A_2})\Phi_2\|$$

where

$$\Phi_1 = \begin{bmatrix} I & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix}^T$$

and

$$\Phi_2 = \begin{bmatrix} 0 & I & 0 & 0 \\ -I & 0 & I & 0 \end{bmatrix},$$

hence the gap is less than

$$\|\Phi_1\|\|(P_{A_1} - P_{A_2})\|\|\Phi_2\|$$

$$\leq 2\|(P_{A_1} - P_{A_2})\| \tag{2.2.10}$$

$$= 2\text{gap}\{\ A_1,\ A_2\ \}.$$

Next a bound is given for the gaps between operators that have transfer function representations. Suppose that T_1 and T_2 are closed and their transfer functions have inner-outer factorization:

$$T_1(s) = N_1(s)D_1^{-1}(s)$$

$$T_2(s) = N_2(s)D_2^{-1}(s)$$

where $N_i(s)$ and $D_i(s)$ are analytic in C_+. As we have seen that the graph of T_i is $G(T_i) = L_i \cdot E$, where

$$L_i(s) \triangleq \begin{bmatrix} D_i(s) \\ N_i(s) \end{bmatrix}.$$

Let S_1 be the unit ball in $G(T_1) \triangleq \{(x,y) \mid \|x\|^2 + \|y\|^2 = 1\ \}$. By definition we have

$$\mathrm{gap}(T_1,\, T_2) \triangleq \max\{\ \vec{\delta}(G(T_1),G(T_2)),\ \vec{\delta}(G(T_2),G(T_1))\ \}$$

and

$$\begin{aligned}
\vec{\delta}(G(T_1),G(T_2)) &\triangleq \sup_{u \in S_1} \mathrm{dist}(u,\, G(T_2)) \\
&= \sup_{u \in S_1} \inf_{e'} |u - L\cdot e'| \\
&= \sup_{L_1\cdot e \in S_1} \inf_{e'} |L_1\cdot e - L_2\cdot e'| \\
&\leq \sup_{L_1\cdot e \in S_1} |L_1\cdot e - L_2\cdot e| \\
&= \frac{\|L_1 - L_2\|}{\sup\limits_{|L_1 e|=1} |e|}
\end{aligned} \tag{2.2.11}$$

The last term can be evaluated easily via transfer functions of L_i's :

$$\|L_1 - L_2\| = \max_{\omega} \bar{\sigma} \begin{bmatrix} D_1 - D_2 \\ N_1 - N_2 \end{bmatrix}$$

and

$$\sup_{|L_1 e|=1} |e| = \min_{\omega} \bar{\sigma}^{-1} \begin{bmatrix} D_1 \\ N_1 \end{bmatrix}. \tag{2.2.12}$$

Similarly for $\vec{\delta}(G(T_2),G(T_1))$.

We note that (2.2.12) is always greater than one.

From the above formula we have

$$\mathrm{gap}(T_1,T_2) \leq \beta \sqrt{\|D_1 - D_2\|^2 + \|N_1 - N_2\|^2}$$

where β is a constant less than 1, which is in agreement with the results in [Vidyasagar, 1984].

Having found a way to compute the gap between two operators we can give a set of conditions that the gap between $\tilde{H}$ and H must satisfy in order that the system is stable. This has been treated in detail in the book by Kato [Kato], and we quote the results there in a modified form.

Theorem 2.2.18: For the system in Fig. 3.1.1, if K and $\tilde{K}$ are closed and $(I - K)^{-1}$ is stable, then the operator $(I - \tilde{K})^{-1}$ is also stable if

$$\text{gap}(K,\tilde{K}) \leq \frac{2}{3+\sqrt{5}} \left[1 + \|(1 - K)^{-1}\|^2\right]^{\frac{1}{2}}.$$

In the case only H is perturbed the condition becomes:

$$\text{gap}(H,\tilde{H}) \leq \frac{1}{2} \left[1 + \|(1 - K)^{-1}\|^2\right]^{\frac{1}{2}}.$$

Proof:

Apply theorems in [Kato, p205], and use the lemmas developed above the conclusion follows. □

2.3 *Use LTI operators for stability analysis*

We have seen that LTI operators, which can be represented as transfer functions, may be treated effectively. Also we have studied the stability behaviour of a system in the neighbourhood of another system whose stability properties are known. This leads us to the idea of studying the stability properties of linear time varying systems which are close to LTI systems. In this section we attempt to make some general observation on this issue.

2.3.1 *The effectiveness of approximation*

In order to apply the ideas discussed above, the operators we wish to study via LTI operators must be "close" to LTI operators, otherwise only trivial results can be obtained. So the first thing is to give a criterion by which we may judge the extent to which a given operator can be represented by an LTI one. We notice that for any bounded operator there is a trivial approximation to it such that the error of approximation is bounded: a constant, for example 0. In this case $|(T - 0)\cdot x| = |T\cdot x|$. We may regard 0 as the worst possible approximation to a given operator, and hence we require that any less trivial choice T_0 should at least make $|(T - T_0)\cdot x| < |T\cdot x|$ for some x.

Definition 2.3.1: An approximation T_0 to T is said to be effective if $\exists\ x \in \text{Dom}(T) \cap \text{Dom}(T_0)$ such that

$$|(T - T_0)\cdot x| < |T\cdot x| . \qquad (2.3.1)$$

On the set of x where (2.3.1) is satisfied T is effectively approximated by T_0 . We are interested in the cases where an effective LTI approximation can be found. We can also see that the set mentioned above is in fact a subspace. Suppose such a space is M, with an orthogonal projection P_M onto it, then

$$\left|(T - T_0)\cdot P_M\cdot x\right| < \left|T\cdot P_M\cdot x\right| \ \forall\ x \in \mathrm{Dom}(T). \qquad (2.3.2)$$

Whether an operator has an effective LTI approximation depends on its character, although how to find such approximations is another question. We will see that for sample-and-hold (SAH) operators of next chapter the frequency response of signals is the important factor that decides if good LTI approximation exists. The following two numbers can be used to measure the effectiveness of an approximation:

$$\alpha_M \triangleq \inf_{\substack{P_M\cdot x\in \mathrm{Dom}(T)\\ Tx\neq 0}} \frac{\left|(T - T_0)\cdot P_M\cdot x\right|}{\left|T\cdot P_M\cdot x\right|} \qquad (2.3.3a)$$

$$\beta_M \triangleq \sup_{\substack{P_M\cdot x\in \mathrm{Dom}(T)\\ Tx\neq 0}} \frac{\left|(T - T_0)\cdot P_M\cdot x\right|}{\left|T\cdot P_M\cdot x\right|}\,. \qquad (2.3.3b)$$

α_M and β_M represent the best and worst directions for T_0 to approximate T. If β_M is small enough we might regard the approximation as uniformly good.

The α and β defined are functions of T_0, and on this basis we can talk about how to choose T_0 to optimize the criterion. A sensible definition for optimality is for β to be minimized. Another definition that we will actually use is

$$\left|(T - T_{opt})\cdot P_M\right| = \inf_{T_0\in LTI} \left|(T - T_0)\cdot P_M\right|$$

2.3.2 *Bound on the approximation error*

It is necessary to find a relative bound for the approximation error in order to use perturbation theory. Although the choice of T_0 is arbitrary, the task of finding a bound for the error is difficult. It is harder when we want the bound to be tight. We can argue as follows: since T_0 has taken away the LTI part of T, and the remaining part is very different from an LTI operator, and hence it is not easy to relate it to LTI operators. Existence of such a bound is no problem since there is always a trivial bound, i.e. $R = \left|T - T_0\right|\cdot I$. The key issue is how to reduce conservatism. The trivial bound is not tight because it doesn't take into account any information about the directions on which the error may assume varying values. We observe that in Theorem 2.2.9 it is not required that R should behave like ΔH. What is required is that on every direction R dominates ΔH in gain. The fact that R has varying gains at different directions is the reason why conservatism can be reduced. The ideal situation would be that $|\Delta H \cdot x| = |R \cdot x|$ for all x, but this is the case only when R is merely a rotated version of ΔH, which is unlikely to happen for an LTI R. We can also define a number to measure the quality of the bound:

$$\gamma \triangleq \sup_{\substack{x \in \mathrm{Dom}(\Delta H) \\ Rx \neq 0}} \frac{|\Delta H \cdot x|}{|R \cdot x|} . \qquad (2.3.4)$$

It is clear that $\gamma \leq 1$, and the bigger γ is the tighter is the bound. The most conservative direction is the one on which γ attains maximum.

Chapter Three
The Sample and Hold Operator

3.0 *Introduction*

It is well known that any sampled data systems must contain a combination of sample and hold actions, and that it is these actions that make such systems time-varying. This chapter is devoted to the study of the properties of sample and hold.

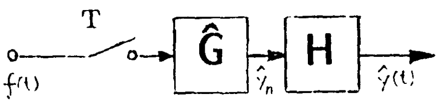


Fig. 3.0.1 a

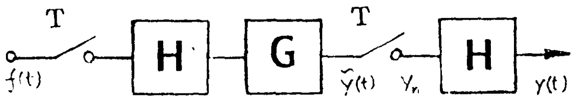


Fig. 3.0.1 b

Before describing the general mathematical properties of this operation, We first study a typical "hybrid digital device" as in Fig. 3.0.1a. Its function is to take samples from a continuous signal f(t) at instants $\{nT\}$, $n = 0, 1, 2, ...$, record the sampled data $f_n \triangleq f(nT)$, and perform some linear operations on them. Then it generates a sequence of data $\hat{y}_n$ in synchronization with the sampling operations at the input. The relation between $\hat{y}_n$ and f_n is a linear convolution defined as

$$\hat{y}_n = \sum_{k=0}^{n} g_{n-k} f_k$$

where $\{g_n\}_{i=0}^{\infty}$ is a fixed sequence. If the $\mathbb{Z}$-transforms of g_n, $\hat{y}_n$ and f_n are $\hat{G}(z)$, $\hat{y}(z)$ and $f(z)$, then the above relation can also be expressed as

$$\hat{y}(z) = \hat{G}(z)f(z) \ .$$

On the other hand, we can also construct a system as in Fig. 3.0.1b, where G is a continuous time system with transfer function G(s). If the samples of the output of this system are taken at $t = T, 2T, \cdots$, etc, a sequence of numbers $y_n \triangleq y(nT)$ is obtained. From an external point of view, we have two systems, each maps a continuous time function into a piece-wise constant function. The questions is, when the digital system is given, can one find a continuous system as in Fig. 3.0.1b to emulate the external behaviour of the digital system? How about the other way round?

The following proposition establishes the relation between continuous time and discrete time mappings:

Proposition 3.0.1: (i) For any proper rational function $\hat{G}(z)$, if it does not have the factor $\frac{1}{z}$ (which represents a pure delay), there exists a proper rational function G(s) such that

$$\mathbb{Z}\{G(s)h_T(s)f(e^{sT})\}|_{s=-\frac{1}{T}\ln(z)} = \hat{G}(z)f(z),\ \forall f(\cdot)$$

where

$$h_T(s) \triangleq \frac{1}{sT}(1 - e^{-sT})$$

and $\mathbb{Z}\{\cdot\}$ denotes the operation of

$$\mathbb{Z}\{A(s)\} \triangleq \sum_{n=-\infty}^{+\infty} A(s - j\omega_s n)\ ,\ \omega_s \triangleq \frac{2\pi}{T}\ . \tag{3.0.1}$$

(ii) For any given proper rational function G(s) the operation $\mathbb{Z}\{h_T(s)G(s)\}$ defines a rational function $\hat{G}(z)$ so that the two systems in Fig. 3.0.1a and 3.0.1b have the same external behaviour.

Note: The requirement that $\hat{G}(z)$ does not have factors z^{-1} can be dropped if we allow G(s) to have factors of e^{-sT}.

The proof of this proposition is given in the appendix to this chapter. The operation $\mathbb{Z}\{\cdot\}$ is also called the $\mathbb{Z}$-transform though this term normally refers to the operation on a sequence of numbers. We point out that the function G(s) in (i) is not necessary a real rational function even when $\hat{G}(z)$ is. More detailed discussion is included in Appendix A. Now let us study the input-output mapping of the system in Fig. 3.0.1a under the assumption that $\hat{G}(z)$ satisfies Proposition 3.0.1 for some G(s). Note that the same symbols are used for both the time domain functions and their Laplace or $\mathbb{Z}$-transforms. Thus

$$\begin{aligned}\hat{y}(s) &= h_T(s)\cdot\hat{y}^{*T}(s)\ , \qquad (\text{here } \hat{y}^{*T} \triangleq y(z)|_{z=\exp(-sT)}\,) \\ &= h_T(s)\cdot[\ \hat{G}(z)f(z)]|_{z=\exp(-sT)} \\ &= h_T(s)\mathbb{Z}\{\ G(s)h_T(s)f(e^{-sT})\ \}\end{aligned}$$

Introduce in symbol S_T to designate the operation defined below:

$$(S_T\cdot e)(s) \triangleq h_T(s)[\mathbb{Z}\{e(s)\}]. \tag{3.0.2}$$

We then have

$$y(s) = S_T \cdot G \cdot S_T \cdot f(s). \qquad (3.0.3)$$

The dots in (3.0.3) are carefully put down to indicate the functional operations rather than merely multiplications. This indicates that the input-output behaviour of system Fig. 3.0.1a can be externally emulated by Fig. 3.0.1b.

The operation defined in (3.0.2) is in frequency domain. For it to have a time domain meaning certain conditions have to be imposed on the kind of signals it operates on. We will do this in a function space context next section. For this reason we define the time domain SAH transformation:

Definition 3.0.2: Suppose e(t) is a function of t, continuous at t=0,T,2T,..., then the SAH transformation $\hat{e}(t)$ of e(t) is a function given by

$$\hat{e}(t) \triangleq e(nT) \qquad t \in [nT, (n+1)T)$$

or

$$= \sum_{n=0}^{\infty} e(nT) \mathbb{1}_{[nT,\ (n+1)T)}(t),$$

where $\mathbb{1}_A(t)$ is the characteristic function of a set A:

$$\mathbb{1}_A(t) = 1 \text{ if } t \in A$$
$$= 0 \text{ if } t \notin A .$$

The continuity assumption on e(t) is a technicality, but it is not unreasonable. For a physical sampler can only measure the values of a signal at t = NT if it is continuous in some small neighbourhood of nT. This assumption will simplify analysis, and is always assumed.

3.1 *SAH operators*

In the sequel we shall analyze SAH operation in the context of function spaces. Of special interests are the questions concerning domain of this operation. It is clear that SAH does not map $L^2[0,\infty)$ into $L^2[0,\infty)$, as the the example below demonstrates:

Example 3.1.1: Let $f(t) = \sum_{n=0}^{\infty} \alpha(n(t\text{-}nT))$ for $t \geq 0$, where

$$\alpha(t) = t + \frac{T}{2} \text{ if } t \in [\frac{\text{-}T}{2}, 0)$$

$$= -t + \frac{T}{2} \quad \text{if } t \in [0, \frac{T}{2})$$

$$= 0 \qquad \text{elsewhere.}$$

It is easily seen that

$$\int_0^{+\infty} |f(t)|^2 dt \le \frac{2}{3} T^3 \sum_{n=1}^{\infty} \frac{1}{n^3} < \infty$$

or $f \in L^2[0, \infty)$. But the SAH transformation of $f(t) = T \ \forall t$, i.e. $S_T \cdot f \notin L^2$. □

In order to study the SAH transformation in the context of function spaces, one has to identify those functions that, after the operation by SAH, yield elements in $L^2[0, \infty)$. Another concern is that we would like the definition of the SAH operation to be in agreement with (3.0.2). Since Fourier transformation from $L^2[0, \infty)$ to $L^2(-\infty, \infty)$ is a Hilbert space isomorphism, one can say little about the pointwise property of a function from its transformation alone. But the operation of SAH is defined only in relation to the samples at a very small set of points, so sometimes it does not reflect the global property of a function adequately. This is why we can not identify S_T with the time domain SAH transformation beforehand. Some smoothness condition must be present to avoid ambiguity of the frequency-time domain correspondence.

Definition 3.1.2: Class **S** functions are those $x(t) \in L^2[0,\infty)$ which satisfy

$$\mathcal{L}^{-1}\{ h_T(s)\sum_{n=-\infty}^{+\infty} x(s\text{-}j\omega_s n)\} = \sum_{n=0}^{+\infty} x(nT)\mathbb{I}_{[nT,\overline{n+1}T)}(t) \text{ a.e.} \qquad (3.1.1)$$

where $\mathcal{L}^{-1}$ denotes the inverse Laplace transform.

For class **S** functions, frequency representation of SAH operation is precisely (3.0.2). Hence from now on the symbol S_T is used for SAH when the functions under discussion are in **S**. But first we have to find out what kind of functions are in **S**.

Lemma 3.1.3: If $x(t) \in L^2[0,\infty)$ satisfies the following conditions, it is in **S**:

(i) $x(nT) \in l^2$, i.e. $\sum|x(nT)|^2 < \infty$;

(ii) $x(s)$ is such that $\sum_{n=-\infty}^{+\infty} x(s+j\omega_s n)$ converges in C_+ to an analytic function $\Gamma(s)$, in L^2 sense;

(iii) The inversion formula for x holds true for t=nT:

$$\frac{1}{2\pi}\int_{-j\infty+c}^{j\infty+c} x(s)e^{snT}ds = x(nT)\ ,\ c>0$$

Proof:

From (i) we know that $S_T \cdot x \in L^2[0,\infty)$, thus it makes sense to define its Fourier transform (in the sençe of Plancherel). It is easy to see that

$$\mathcal{L}\{S_T \cdot x(t)\} = \frac{1-e^{-sT}}{s}\sum_{n=0}^{+\infty} x(nT)e^{-snT}$$

here the convergence is in the sense of L^2. (ii) says that $\Gamma(s)$ is a periodic function so it can be expanded as Fourier series:

$$\Gamma(s) = \Gamma(\sigma+j\omega) = \sum_{m=-\infty}^{+\infty} C_m(\sigma)e^{-j2\pi m\frac{\omega}{\omega_s}},\ \sigma>0$$

$$= \sum_{m=-\infty}^{+\infty} C_m(\sigma)e^{-j\omega mT}$$

where

$$C_m(\sigma) = \frac{T}{2\pi}\int_{-\omega_s/2}^{\omega_s/2} \Gamma(\sigma+j\omega)e^{j\omega mT}d\omega$$

$$= \frac{T}{2\pi}\int_{-\infty}^{+\infty} x(\sigma+j\omega)e^{j\omega mT}d\omega$$

by (iii)

$$= Te^{-\sigma mT}x(mT).$$

In other words we have established that (in L^2 sense):

$$\sum_{n=-\infty}^{+\infty} x(s+j\omega_s n) = \sum_{m=-\infty}^{+\infty} C_m(\sigma)e^{-j2\pi m\frac{\omega}{\omega_s}},\ \sigma>0$$

$$= T\sum_{m=0}^{+\infty} x(mT)e^{-smT},$$

Multiplying on both sides the factor $\frac{1-e^{-sT}}{sT}$ leads to the assertion. □

But the question still remains of under what condition the frequency domain summation converges. It suffices for our purposes to have

Lemma 3.1.4: If x(s) is analytic in C_+ and $|x(\sigma + j\omega)| = O(\omega^{-1-\delta})$ as $\omega \to \infty$, $\delta > 0$, then

$$\sum_{n=-\infty}^{+\infty} x(s - j\omega_s n)$$

is analytic in C_+

Proof:

There exists a constant α such that when n is big enough

$$|x(s - j\omega_s n)| \leq \frac{\alpha}{n^{1+\delta}} \ \forall \ \omega \in [-\tfrac{\omega_s}{2}, +\tfrac{\omega_s}{2}].$$

Since

$$\sum_{n=0}^{+\infty} \frac{1}{n^{1+\delta}} < \infty$$

we conclude that in every band $\{ n\omega_s - \frac{\omega_s}{2} \leq \mathrm{Im}(s) \leq n\omega_s + \frac{\omega_s}{2}\} \bigcap C_+$ the sum converges uniformly. But every term in the sum is analytic, so must be the sum.

□

Note that ordinary L^2 functions do not possess this property unless they are filtered. The above lemma simply says that if the filter has a rational and strictly proper transfer function, then it produces signals in **S**. To see this recall that for a function x in $L^2[0,\infty)$ its transfer function satisfies [Duren]:

$$x(s) = O(|s|^{-\frac{1}{2}}) \text{ for } \mathrm{Re}(s) > 0 \text{ and } |s| \to \infty,$$

so after multiplication by a factor of $O(|s|^{-1})$ and taking radical limit to $j\omega$-axis, the condition in the above lemma is seen met.

3.2 *Approximation to SAH operators*

The frequency domain description of the SAH operator provides a way of approximating them by LTI operators or transfer functions. S_T maps a function, whose Laplace transform is e(s), into

$$f(s) = h_T(s)\sum_{n=-\infty}^{+\infty} e(s\text{-}j\omega_s n).$$

f(s) contains the frequency shifts, which is something that an LTI operator can never do. But the effect of $h_T(s)$ is a low-pass filter, so the high frequency harmonics in f(s) are "filtered out". Intuitively those terms in the above expression for large n's should not have significant contribution to f(s), and the main part of f(s) is based on e(s). This hints that it is possible to capture the basic characteristics of SAH operation by an LTI operator.

An almost immediate choice for this purpose is h_T, meaning simply to ignore the side harmonics of f(s) outside of $\Omega \triangleq [-\frac{\omega_S}{2}, \frac{\omega_S}{2}]$. Note that h_T is, when interpreted as an operator, a convolution with kernel $\mathbb{1}_{[0,T)}(t)$. We will see later that h_T is in fact a good choice with justifications given below.

The first question we ask is whether this approximation is effective, i.e. if

$$\left|(S_T - h_T)\cdot x\right| < \left|S_T\cdot x\right| \;\forall\; x \in \mathbf{S}\,.$$

But this is in general not the case, unless the following condition is met:

$$\left|\sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} x(j\omega - j\omega_S n)\right| < \left|\sum_{n=-\infty}^{+\infty} x(j\omega - j\omega_S n)\right| \;\forall\; \omega \in \Omega\,. \tag{3.2.1}$$

In other words h_T as an approximation is effective for those functions whose spectra are well concentrated in low frequencies. One would naturally think that the typical signals that satisfies (3.2.1) would be of narrow bandwidth, in particular the ideal low pass signals: $|x(j\omega)| = 0$ for $\omega > \frac{\omega_S}{2}$. But none of the $L^2[0, \infty)$ functions possesses this property [Slepian]. On the other hand, narrow frequency band is thought to be related to the smoothness of functions, and it is a widely accepted notation to talk about the "low frequency part" of a signal. It is rather inconvenient that this part of an $L^2[0, \infty)$ signal goes out of the space $L^2[0, \infty)$. For this reason we will for a while use $L^2(-\infty,\infty)$, because the Fourier transform of this space admits the frequency truncation operations that we are accustomed to. S_T can be defined on $L^2(-\infty, \infty)$ in the obvious way.

Since every function x in $L^2[0, \infty)$ is the projection of some $y \in L^2(-\infty, \infty)$ on the positive times, i.e. $x = P_+\cdot y$, it follows that approximation to $S_T\cdot x$ can be done via $S_T\cdot y$. But the latter is easier to analyze and gives more insight too.

Let B_T be a subspace of $L^2(-\infty, \infty)$ defined as:

$$B_T \triangleq \{\, x \in L^2(-\infty, \infty) \mid x(j\omega) = 0 \;\forall\; \omega \notin \Omega \,\}\,. \tag{3.2.2}$$

Let P_T be the orthogonal projection from $L^2(-\infty, \infty)$ onto B_T. It is easy to see that P_T is an LTI operator with transfer function

$$P_T(j\omega) = \mathbb{1}_\Omega(\omega)\,. \tag{3.2.3}$$

However it is not causal. If a signal in $L^2[0,\infty)$ is fed into this "filter" the result is a smoother function which is not identically zero for $t < 0$. None of the functions in $L^2[0,\infty)$ lies in B_T [Slepian], but those close to it are the ones with relatively little high frequency harmonics, and they should not be much different from their projections onto B_T as far as the parts in the positive time are concerned. Also the quantity $|P_{-}\circ P_T\cdot x|$ should be small for a "smooth" function x in $L^2[0, \infty)$. Based on the above observations, we have reasons to believe that for smooth functions in the above sense, LTI approximation of S_T can be given by one that is effective on B_T. On the range of B_T strong results can be obtained concerning the optimal LTI approximation of S_T. In fact we have

Theorem 3.2.1: The solution to the optimization below

$$\inf_{A \in LTI} |(S_T - A)\cdot P_T\cdot x|$$

is $A_{opt} = h_T$. 'The optimality is independent of x.

Proof:

$$|(S_T - A)P_T\cdot x|^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |(S_T - A)P_T\cdot x(j\omega)|^2 d\omega$$

$$= \frac{1}{2\pi}\int_{-\frac{1}{2}\omega_s}^{\frac{1}{2}\omega_s} |h_T(j\omega) - A(j\omega)|^2 |x(j\omega)|^2 d\omega + \text{...(independent of A)}$$

It is obvious that when A is chosen to be h_T, minimization is achieved. □

We can now estimate the quality of this approximation in the way outlined last chapter. i.e. the numbers

$$\alpha \triangleq \inf_{P_T x \neq 0} \frac{|(S_T - h_T)P_T x|}{|S_T P_T x|}$$

and

$$\beta \triangleq \sup_{P_T x \neq 0} \frac{|(S_T - h_T)P_T x|}{|S_T P_T x|} .$$

The range of P_T is the class of signals which can be recovered from their sampled data, so S_T is an invertable operator when restricted to this subspace, or

$$\left[S_T\circ P_T\right]^{-1} = h_T^{-1}\circ P_T. \qquad (3.2.4)$$

This is well known in literature of sampled data systems. In order to compute the numbers α and β we perform the following calculations

$$|S_T P_{\Gamma} \cdot x|^2 \triangleq \frac{1}{2\pi} \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \left| \sum_{n=-\infty}^{+\infty} (P_T \cdot x)(j\omega - j\omega_s^{n}) \right|^2 d\omega$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \sum_{n=-\infty}^{+\infty} |(P_T \cdot x)(j\omega - j\omega_s)|^2 d\omega$$

$$= \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} \int_{(n-\frac{1}{2})\omega_s}^{(n+\frac{1}{2})\omega_s} |h_T(j\omega)|^2 |x(j\omega - j\omega_s)|^2 d\omega$$

$$= \frac{1}{2\pi} \int_{-\frac{1}{2}\omega_s}^{\frac{1}{2}\omega_s} \left[\sum_{n=-\infty}^{+\infty} |h_T(j\omega + j\omega_s)|^2 \right] |x(j\omega)|^2 d\omega$$

$$= \frac{1}{2\pi} \int_{-\frac{1}{2}\omega_s}^{\frac{1}{2}\omega_s} |x(j\omega)|^2 d\omega = |P_T \cdot x|^2. \tag{3.2.5}$$

In much the same way we have

$$|(S_T - h_T) P_T \cdot x|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \left| \sum_{n \neq 0} (P_T \cdot x)(j\omega - j\omega_s) \right|^2 d\omega$$

$$= \frac{1}{2\pi} \int_{-\frac{1}{2}\omega_s}^{\frac{1}{2}\omega_s} \left[1 - |h_T|^2(j\omega) \right] |x(j\omega)|^2 d\omega.$$

So we conclude

$$\alpha = \inf_{\omega \in \Omega} \left[1 - |h_T(j\omega)|^2 \right]^{\frac{1}{2}} = 0, \tag{3.2.6}$$

and

$$\beta = \sup_{\omega \in \Omega} \left[1 - |h_T(j\omega)|^2 \right]^{\frac{1}{2}} = \sqrt{1 - (\tfrac{\pi}{2})^2} \doteq 0.77 \ . \tag{3.2.7}$$

So in the range of P_T, h_T is an LTI approximation to S_T with the worst case relative error 77%. The two numbers are not attainable, though arbitrarily close, by functions in L^2. Functions which are almost constant would give a ratio close to α, while those that are close to sinusoid with period $\frac{\omega_s}{2}$ tend to give β. In a practical sampled data system, the sampling rate should be high enough so that the important part of the frequency spectra of the signals to be sampled is well inside of Nyquist band Ω [Franklin]. In light of this, h_T is in general a good LTI approximation to S_T.

3.3 *Bound on the error*

Once the choice of the LTI approximation to S_T is made, one needs to know how much error it will introduce. A bound must be found which specifies in how big a neighbourhood of the approximation the original operator lies. One can then assess to what extend the approximation can be trusted. This is done in this section.

Define $\Delta_T = S_T - h_T$. We need to find an LTI operator R_T such that

$$|\Delta_T \cdot x| \leq |R_T \cdot x| \quad \forall\ x \in \mathrm{Dom}(\Delta_T). \tag{3.3.1}$$

From the analysis of last section we know that

$$<\Delta_T \cdot P_T \cdot x,\ R \cdot P_T \cdot x> = 0\ \forall x$$

whenever R is LTI. This implies that the nature of Δ_T is quite different from any LTI operator, at least on Rang(P_T). It should be expected as Δ_T is the residual of S_T after the LTI part is taken out. It is impossible to visualize Δ_T in the form of a Bode diagram. The bound we are seeking should not be interpreted as having dominating gains in all frequencies. The lemma below gives a simple result concerning the "gain" behaviour of Δ_T on Rang(P_T).

Lemma 3.3.1: There exists an LTI operator R such that $\forall\ x \in L^2(-\infty, \infty)$

$$||\Delta_T \cdot P_T \cdot x|| = ||R \cdot P_T \cdot x||.$$

Proof:

We know that

$$[\Delta_T \cdot e](j\omega) = h_T(j\omega) \sum_{\substack{n \neq 0 \\ n=-\infty}}^{+\infty} e(j\omega - j\omega_s n), \quad \omega_s \overset{\Delta}{=} \frac{2\pi}{T}.$$

If $e = P_T \cdot x$, then

$$\left| \sum_{\substack{n \neq 0 \\ n=-\infty}}^{+\infty} e(j\omega - j\omega_s n) \right|^2 = \sum_{n \neq 0} |\, e(j\omega - j\omega_s n)|^2 ,$$

thus

$$||\Delta_T \cdot P_T \cdot x||^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \sum_{n \neq 0} |e(j\omega - j\omega_s n)|^2 d\omega$$

$$= \frac{1}{2\pi} \sum_{n \neq 0} \int_{(n-\frac{1}{2})\omega_s}^{(n+\frac{1}{2})\omega_s} |h_T(j\omega)|^2 |x(j\omega - j\omega_s)|^2 d\omega$$

$$= \frac{1}{2\pi} \int_{-\frac{1}{2}\omega_s}^{\frac{1}{2}\omega_s} \Big[1 - |h_T|^2(j\omega)\Big] |x(j\omega)|^2 d\omega.$$

$$\triangleq |R \cdot P_T \cdot x|^2,$$

where R has transfer function R(s) that satisfies

$$|R(j\omega)|^2 = 1 - |h_T(j\omega)|^2 \ \square \tag{3.3.2}$$

This states that although Δ_T is not similar to any LTI operator, one can find an R that has the same gain as Δ_T for all elements in Rang(P_T). The transfer function that can be solved from (3.3.2) can then be used as a bound on Rang(P_T), and it is the tightest one. Another observation we may make is that R is a high pass filter, which is also expected, for h_T approximate S_T in low frequencies only. A bound on the range of P_T only will not suffice for our purpose. Nevertheless, (3.3.2) gives a picture of the dominant trend of the error.

The only restrictions we may impose on the class of signals is that they are generated by passing L^2 functions through filters with a strictly proper transfer functions. This is in fact a subclass of absolutely continuous functions.

In the calculations below some of steps, though intuitively plausible, are not mathematically sound. A more detailed proof with justification for the steps is included in the Appendix to this chapter.

What we are looking for is an operator bound that is valid for a general class of functions. Formally we have

$$2\pi ||\Delta_T \cdot x||^2 = \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \left| \sum_{\substack{n \neq 0 \\ n=-\infty}}^{\infty} x(j\omega - j\omega_s n) \right|^2 d\omega$$

$$= \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \left| \sum_{\substack{n \neq 0 \\ n=-\infty}}^{\infty} h_T(j\omega - j\omega_s n) h_T^{-1}(j\omega - j\omega_s n) x(j\omega - j\omega_s n) \right|^2 d\omega$$

$$\leq \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \Big(\sum_{n\neq 0} |h_T(j\omega - j\omega_s n)|^2\Big)\Big(\sum_{n\neq 0} \big|h_T^{-1}(j\omega - j\omega_s n)x(j\omega - j\omega_s n)\big|^2\Big)d\omega$$

Using the identity

$$\sum_{n=-\infty}^{+\infty} |h_T(j\omega - j\omega_s)|^2 \equiv 1$$

and defining

$$\Gamma^2(j\omega) = |h_T(j\omega)|^2(1 - |h_T(j\omega)|^2)$$

the above calculation can be continued as

$$= \int_{-\infty}^{+\infty} |\Gamma(j\omega)|^2 \sum_{\substack{n\neq 0\\ n=-\infty}}^{\infty} |h_T^{-1}x(j\omega - j\omega_s n)|^2 d\omega$$

$$= \sum_{\substack{n\neq 0\\ n=-\infty}}^{\infty} \int_{-\infty}^{+\infty} |\Gamma(j\omega + j\omega_s n)|^2 |h_T^{-1}x(j\omega)|^2 d\omega$$

Let $\sum_{n\neq 0} |\Gamma(j\omega + j\omega_s n)|^2 |h_T^{-1}(j\omega)|^2$ be denoted by $|R(j\omega)|^2$, we have

$$||\Delta_T\cdot x||^2 \leq \frac{1}{2\pi}\int_{-\infty}^{+\infty} |R(j\omega)|^2 |x(j\omega)|^2 d\omega = |R\cdot x|^2.$$

Exchange of the order of summation and integration has been made several times on the assumption they are allowed. The conclusion is indeed correct if $x \in \mathrm{Dom}(D)$, where D is a kind of differential operator defined below; also an explicit expression of R can be given. These are summarized in a theorem.

Definition 3.3.2: Let $x \in L^2$ with Laplace transform $x(s)$ be defined in a vertical strap covering $j\omega$-axis. Let Dom(D) be defined as the class of functions x that satisfies

$$\int_{-\infty}^{+\infty} \omega^2 |x(j\omega)|^2 d\omega < \infty.$$

D is defined in frequency domain by $D\cdot x(s) \triangleq sx(s)$, i.e. an multiplication by s. Note that this definition of differential operator is compatible with the definition given in chapter two.

Theorem 3.3.3: There exists an LTI operator R_T such that

$$||\Delta_T \cdot x|| \leq ||R_T \cdot x|| \quad \forall\ x \in \mathrm{Dom(D)}. \tag{3.3.3}$$

The transfer function of R is given by

$$|R_T(j\omega)|^2 = \frac{1}{6}(\omega T)^2 + |h_T(j\omega)|^2 - 1 \tag{3.3.4}$$

It can be verified that the right hand side is positive $\forall\ \omega \neq 0$, therefore an R that is stable and minimum phased can always be defined from (3.3.4). □

A proof of this result involves justifying some of the steps outlined above, as well as computing several infinite summations. It is put into Appendix B to this chapter, for they are merely technical details. Theorem 3.3.3 is stated for an SAH operator that samples at t = 0, T, ..., . In fact, this result is also correct for SAH's that operate at t = δ, T+δ, ..., .

Lemma 3.3.4: If δ is a real number satisfying $0 \leq \delta \leq T$, we use the symbol $S_T(\delta)$ for an SAH operator sampling at t = δ, T+δ, ..., . Define $\Delta_T(\delta) = S_T(\delta) - h_T$, then

$$|\Delta_T(\delta)\cdot x| \leq |R_T \cdot x| \ \forall\ x \in \mathrm{Dom(D)}.$$

Proof:

It is evident that on Dom(D), $S_T(\delta)$, hence $\Delta_T(\delta)$ is well defined. We also note the relationship $S_T(\delta) \equiv D_\delta \cdot S_T \cdot D_{-\delta}$ and that D_δ commutes with LTI operators. Thus we have $\Delta_T(\delta) = D_\delta \cdot \Delta_T \cdot D_{-\delta}$. Therefore, $\forall$ x $\in$ Dom(D)

$$\begin{aligned} |\Delta_T(\delta)\cdot x| &= |D_\delta \cdot \Delta_T \cdot D_{-\delta} \cdot x| \\ &= |\Delta_T \cdot (D_{-\delta} \cdot x)| \\ &\leq |R_T \cdot D_{-\delta} \cdot x| \\ &= |R_T \cdot x| \end{aligned} \qquad \square$$

In other words, $S_T(\delta)$ is contained in the conic sector cone(h_T, R_T) for all $\delta \in [0, T]$. We can also see that R is asymptotically a factor s, and hence has large gains at high frequencies. Comparison with the bound in (3.3.2) shows some interesting similarity between their transfer functions at low frequencies. Of course the assumption on x has ensured that R·x is in L^2. As to the tightness of this bound, we state

Lemma 3.3.5: The bound specified in (3.3.4) is tight in the sense that it is achievable by functions from Dom(D).

Proof:

We only need to construct a function that actually achieves the bound. Take the function defined below.

$$x(t) = \begin{cases} t & t \in [0, T); \\ 2T - t & t \in [T, 2T) \\ 0 & t \in [2T, \infty), \end{cases}$$

Its Laplace transform is $T^2 h_T(s) h_T(s)$. It follows that $x(j\omega) = T^2 |h_T(j\omega)|^2 e^{-j\omega T}$. Note that the only step in the proof of theorem where the inequality is introduced is the Schwarts inequality. It is apparent that the above choice of x will make the equality hold. It is also clear that the x is in Dom(D). □

Since the term $|h_T(j\omega)|^2 - 1$ is small compared to $\frac{1}{6}(\omega T)^2$, in practice it can be ignored. Therefore R_T is merely the operator $\frac{T}{\sqrt{6}}D$, which is a differentiator for signals generated by passing L^2 functions through strictly proper filters. See Fig. 3.3.1 for an illustration of time domain actions of S_T, h_T, Δ_T and R_T.

3.4 *Miscellaneous Properties of SAH*

Some properties of SAH operators are listed here for later reference of independent interests.

(*i*) *Integral representation*

Define a function

$$\mu_T(t) \triangleq [\tfrac{t}{T} + 1],$$

where [·] denote the maximum integer not exceeding (·). We have that

$$S_T \cdot x(t) = \int_{(t-T,\, t]} x(\tau) d\mu_T(\tau). \tag{3.4.1}$$

It is clear that the value of $S_T \cdot x(t)$ is determined by the values of x(t) in the immediate history (t − T, t]. Generally speaking, a time invariant causal system Φ: $x \mapsto y$ is characterized by a translate invariant measure ϕ of real numbers, i.e.

$$y(t) = \int_0^t x(\tau)\, d\phi(\tau - t)\ .$$

Or, the value of y(t) is the weighted average of the past history of x in [0, t]. This expression of LTI operator gives further insight into the choice of h_T as an approximation to S_T . The action of h_T on x(t) is

$$h_T \cdot x(t) = \int_{t-T}^t x(\tau) d(\frac{\tau - t}{T})\ ,$$

so h_T has the same "forgetting" factor as S_T. Also, it is interesting to see the graphs of the two functions $\frac{\tau}{T}$ and $\mu_T(\tau)$, $0< \tau \leq T$, in Fig. 3.4.1. It appeares that without *a priori* knowledge of x and τ, the best approximation to $\mu_T(\tau)$ is the "fair" measure on [t − T, t].

There can be many time invariant measure approximating $\mu_T(t)$. One of them, for example, is:

$$\psi(t) \triangleq [t - \frac{T}{2}]\mathbb{1}_{[t - T,\ t]}$$

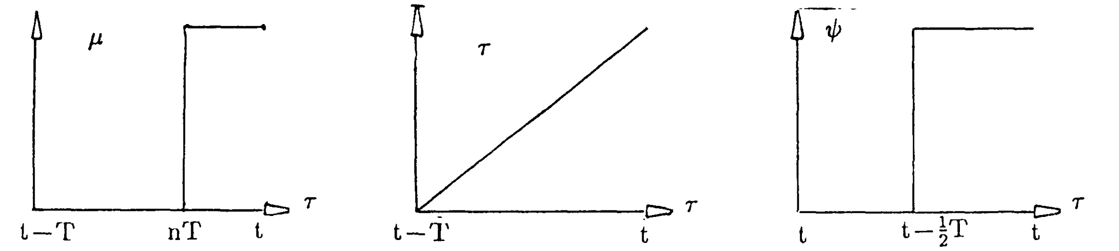


Fig. 3.4.1 Graph of the Measures

(ii) Closedness of S_T

Later we will use the following

Theorem 3.4.1: If G is LTI with transfer function G(s) which is bounded on $j\omega$-axis, and also satisfies

$$|G(j\omega)| = O(\frac{1}{|\omega|}) \text{ as } |\omega| \rightarrow \infty.$$

Then $S_T \cdot G$ is closed.

Proof

We can write

$$S_T = h_T + \Delta_T$$

where Δ_T satisfies

$$|\Delta_T \cdot G \cdot x| \le |R_T \cdot G \cdot x| \ \forall x \in \mathrm{Dom}(G).$$

The assumptions on $G(j\omega)$ together with the property of R_T guarantee that the operator $R_T \cdot G$ is bounded on Dom(G), and hence so is $\Delta_T \cdot G$. Since we know that Dom(G) is closed [El-Arkkary], thus any sequence in Dom(G) that converges must converge to a point in Dom(G). This is enough to ensure that every Cauchy sequence in the graph of Δ_T is convergent in the graph. [Kato] [Chatelin, p89].

(iii) Quantization of delay

S_T does not commute with delay operators D_τ in general unless $T = n\tau$ for some integer n. However when S_T is to operate on a piecewise constant signal with the same period T, delays are quantized. More explicitly

$$S_T \cdot D_\tau \cdot S_T = D_T \cdot S_T, \text{ if } \tau < T.$$

This indicates that a small asynchronization between cascaded sampled data devices can cause big delays.

(iv) Absorbing law

When SAH's with integer ratios are cascaded the slow ones tend to dominate the overall effects:

$$S_T \cdot S_{T/n} = S_{T/m} \cdot S_T = S_T, \text{ n,m being integer.}$$

This will be used for the derivation of a switch decomposition of S_T.

(v) Switch decomposition of SAH operator

In much the same way as doing switch decomposition of a sampler, one can also do it for SAH operators, i.e. to express a SAH with T as SAH's with nT. To accomplish this we first introduce an impulse modulator, I_T whose function is

$$I_T \cdot x(t) = T \sum_{n=0}^{+\infty} x(nT)\delta(t - nT) \tag{3.4.2}$$

where $\delta(t)$ is the Dirac function. It is clear that

$$S_T = h_T \cdot I_T. \tag{3.4.3}$$

Note that here the symbol h_T represent zero-order holder in the conventional sense, i.e. its response to a impulse is $\mathbb{1}_{[0,T)}(t)$, though impulses don't belong to L^2. (One could define a bigger spaces, for instance space of distributions that would include all the first order Dirac distributions, but this is not necessary in our case). Switch decomposition of I_T is:

$$I_T = R_n \cdot \overset{\longleftarrow\ n\ \longrightarrow}{\begin{bmatrix} I_{nT} & & \\ & \ddots & \\ & & I_{nT} \end{bmatrix}} \cdot A_n \tag{3.4.4}$$

where

$$R_n \triangleq \left[I,\ D_T,\ \cdots,\ D_{(n-1)T} \right] \tag{3.4.5}$$

and

$$A_n \triangleq R_n^* = [I,\ D_{-T},\ \cdots,\ D_{-(n-1)T}]^T \tag{3.4.5'}$$

and D_τ is the delay operator. R_n and A_n are sometimes called the retard and advance operators. h_T is related to h_{nT} is following manner:

$$h_T = h_{nT} \cdot F_n \tag{3.4.6}$$

where F_n is LTI whose transfer function is

$$F_n(s) = \frac{n}{\sum_{k=0}^{n-1} e^{-skT}} \tag{3.4.7}$$

which is obtained by use of the identity

$$(1 - e^{-snT}) = (1 - e^{-sT}) \sum_{k=0}^{n-1} e^{-skT}.$$

So we have

$$
\begin{aligned}
S_T &= h_T \cdot I_T \\
&= F_n \cdot R_n \cdot \begin{array}{c} \leftarrow n \rightarrow \\ \begin{bmatrix} S_{nT} & & \\ & \ddots & \\ & & S_{nT} \end{bmatrix} \end{array} \cdot A_n
\end{aligned}
\qquad (3.4.8)
$$

(*vi*) *Non-integer cascade of SAH*

The behaviour of $S_{T_1} \cdot S_{T_2}$ is complicated when the ratio T_1/T_2 is not a integer; it becomes irregular when the ratio is irrational. The fundamental difficulty of multirate systems lies here, and to some extend it indicates the necessity of requirement that the ratios between sampling rates should be integers. The absorbing laws demonstrate the fact that when SAH's are cascaded the slower ones tend to dominate the overall characteristic. This to some extend is also true for any T_1 and T_2 if the faster sampler is placed after the slower one. See Fig. 3.4.2.

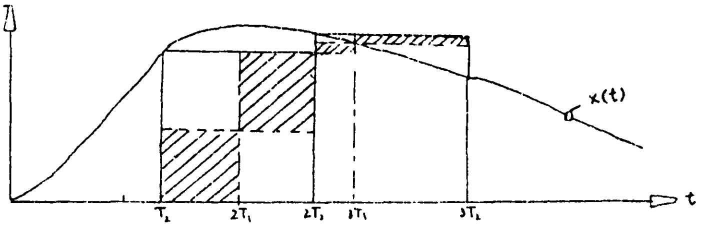


Fig. 3.4.2 Cascaded SAH's

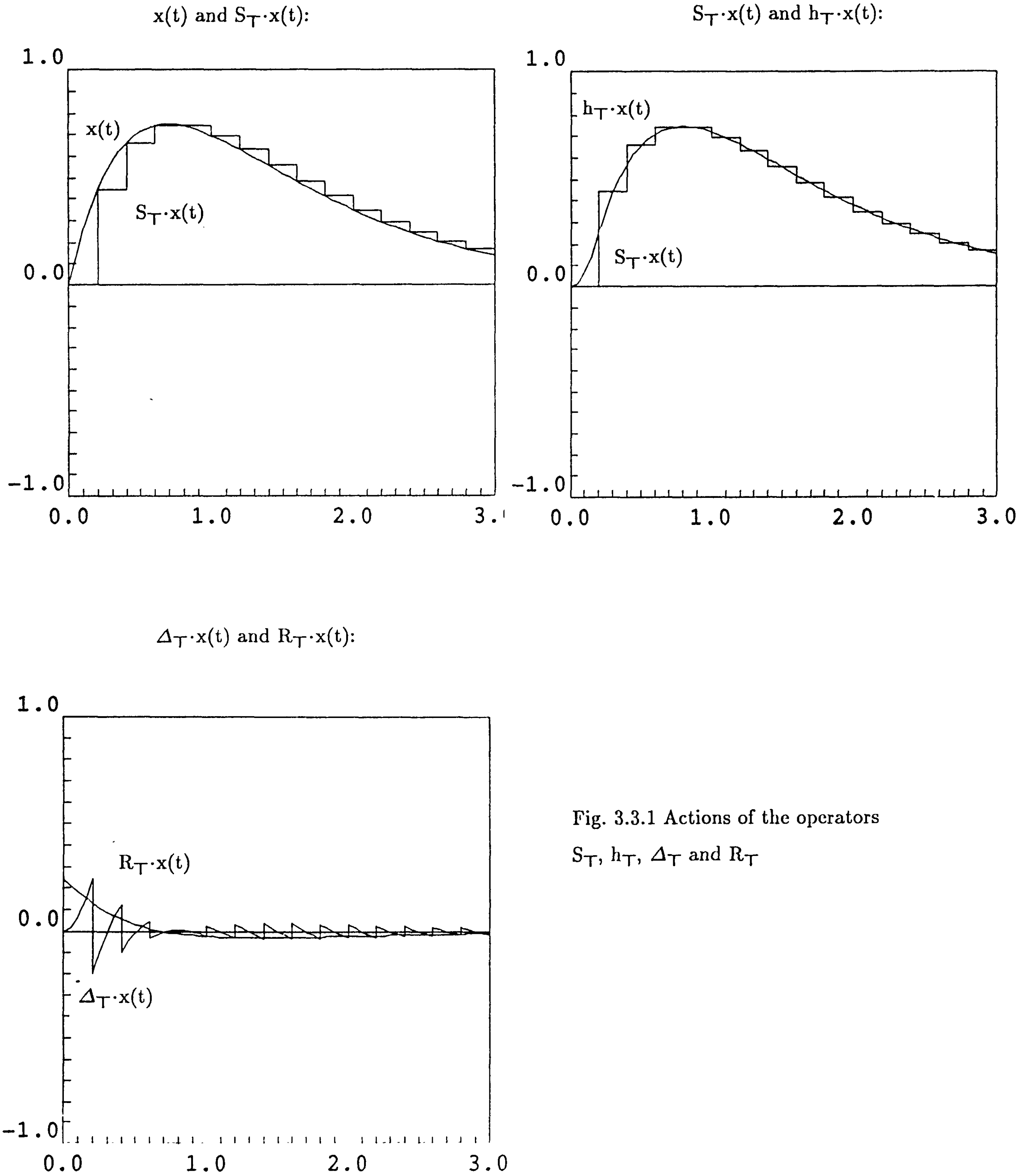
Suppose that $T_1 < T_2$, then $S_{T_1} \cdot S_{T_2}$ can be approximated by

$$
\frac{(1 + D_{T_1})}{2} \cdot S_{T_2}
$$

and the error

$$
\begin{aligned}
&\left| S_{T_1} \cdot S_{T_2} \cdot x - \tfrac{1}{2}(1 + D_{T_1}) \cdot S_{T_2} \cdot x \right|^2 \\
&= \sum_{i=1}^{+\infty} \left[\alpha_i \{ \tfrac{1}{2}(x(nT_2) - x(\overline{n-1}T_2)) \}^2 - (T_1 - \alpha_i) \{ \tfrac{1}{2}(x(nT_2) - x(\overline{n-1}T_2)) \}^2 \right] \\
&= \frac{T_1}{4} \sum_{i=1}^{+\infty} (x(nT_2) - x(\overline{n-1}T_2))^2 \\
&= \frac{T_1}{4T_2} \left| (1 - D_{T_2}) \cdot S_{T_2} \cdot x \right|^2 .
\end{aligned}
$$

This error bound is also valid when the two SAH's are not synchronized at t=0. But the situation is more complicated when $T_1 > T_2$, since it is impossible to predict the amount of error that a random delay of $\tau \in [0, T_2]$ will incur at an arbitrary point of t.



$x(t)$ and $S_T \cdot x(t)$:

$S_T \cdot x(t)$ and $h_T \cdot x(t)$:

$\Delta_T \cdot x(t)$ and $R_T \cdot x(t)$:

Fig. 3.3.1 Actions of the operators S_T, h_T, Δ_T and R_T

Appendix A

Proof of Proposition 3.0.1

Consider the systems in Fig 3.0.1. Let $\hat{G}(z)$ have a minimum state space realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$, i.e. the variables e_n and y_n are related by

$$\begin{cases} z_{n+1} = \hat{A}z_n + \hat{B}e_n \\ y_n = \hat{C}z_n + \hat{D}e_n \, . \end{cases} \qquad (A.1)$$

Suppose that G(s) also have a minimum realization [A,B,C,D] such that

$$\begin{cases} \dot{x}(t) = Ax + BS_T\cdot e(t) \\ y(t) = S_TCx + S_TDS_T\cdot e(t) \end{cases} \qquad (A.2)$$

which leads to

$$\begin{cases} x(\overline{n+1}T) = e^{AT}x(nT) + e^{AT}\int_0^T e^{-A\tau}Bd\tau e(nT) \\ y(nT) = Cx(nT) + De(nT) \, . \end{cases} \qquad (A.3)$$

In order for y(nT) to coincide with y_n when $e(nT) = e_n$, we only need to let

$$C = \hat{C}, \quad D = \hat{D},$$

$$e^{AT} = \hat{A} \text{ and } \int_0^T e^{A\tau}d\tau B = \hat{B}. \qquad (A.4)$$

But in the above we may not get A for any choice of $\hat{A}$, unless $\hat{A}$ does not have zero eigenvalues, or $\hat{G}(z)$ does not have factors z^{-1}. This is in fact sufficient for G(s) to be defined, for the matrix

$$\int_0^T e^{A\tau}d\tau$$

is always non-singular for $T > 0$. To see this we look at the only two possible cases.

(i) A defined from $\frac{1}{T}\ln(\hat{A})$ is invertable. A does not have zero eigenvalues, therefore we can write

$$\int_0^T e^{A\tau}d\tau = A^{-1}[\exp(AT) - I].$$

But exp(AT) does have eigenvalue one, and hence the right hand side of the above expression is invertable.

(ii) A has zero eigenvalues. Then there exist invertable matrix P such that

$$A = P\begin{bmatrix} A_1 & 0 \\ 0 & 0 \end{bmatrix}P^{-1},\ A_1 \text{ invertable.}$$

Thus it follows that

$$\int_0^T e^{A\tau}d\tau = P\begin{bmatrix} A_1^{-1}[\exp(A_1T) - I] & 0 \\ 0 & T \end{bmatrix}P^{-1}$$

From (i) the conclusion follows.

Remark A.1: A obtained via (A.4) is not necessarily real even when $\hat{A}$ is, therefore the G(s) can be complex rational matrix. But the lemma below gives condition under which a real rational $\hat{G}(z)$ gives rise to a real rational G(s).

Lemma A.2: If $\hat{G}(z)$ is real rational and does not have poles on the closed negative real line R^-, then (A.4) defines a real rational G(s).

Proof:

We only need to show that the A matrix defined in (A.4) is real under the hypothesis. For function f(z) analytical in a region Ω and continuous on the boundary $\partial\Omega$ that encircles all the eigenvalues of a matrix X, f(X) is defined by

$$f(X) = \frac{1}{2\pi j}\int_{\partial\Omega} f(z)(zI - X)^{-1}dz.$$

In the present case, f(z) = ln(z), which is analytic on the whole plane except the closed negative real line R^-. If $\hat{A}$ does not have eigenvalues on R^-, it is always possible to choose a region Ω so that $\ln(\hat{A})$ can be expressed in the form of the above integral. It is also known that there exists real invertable matrix Q that transforms A into a block diagonal matrix:

$$Q\hat{A}Q^{-1} = \mathrm{diag}(A_1, \ldots, A_K)$$

where A_i are either real numbers or 2×2 real matrices of the form $\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$. It is thus routine to verify that $\ln(A_i)$ are real matrices. We note that the eigenvalues of A_i are those of $\hat{A}$. Therefore,

$$\ln(\hat{A}) = Q \text{ diag } [\ln(A_1), \ldots, \ln(A_K)] \; Q^{-1}$$

is real. □

Appendix B

Proof of Theorem 3.3.3:

Because of the assumptions imposed on x, i.e. x ∈ Dom(D), we know that

$$2\pi\|\Delta_T \cdot x\|^2 = \lim_{N \to \infty} \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \; | \sum_{\substack{n=-N \\ n \neq 0}}^{N} x(j\omega - j\omega_S n) \; |^2 d\omega$$

For any finite N we can write

$$\int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \; | \sum_{\substack{n=-N \\ n \neq 0}}^{N} x(j\omega - j\omega_S n)|^2 d\omega$$

$$\leq \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 \sum_{\substack{n=-N \\ n \neq 0}}^{N} |h_T(j\omega - j\omega_S n)|^2 \sum_{\substack{n=-N \\ n \neq 0}}^{N} |h_T^{-1} x(j\omega - j\omega_S n)|^2 d\omega$$

$$\leq \int_{-\infty}^{+\infty} |h_T(j\omega)|^2 (1 - |h_T(j\omega)|^2) \sum_{\substack{n=-N \\ n \neq 0}}^{N} |h_T^{-1} x(j\omega - j\omega_S n)|^2 d\omega$$

$$\leq \sum_{\substack{n=-N \\ n \neq 0}}^{N} \int_{-\infty}^{+\infty} |\Gamma(j\omega)|^2 |h_T^{-1} x(j\omega - j\omega_S n)|^2 d\omega$$

$$\leq \int_{-\infty}^{+\infty} \sum_{\substack{n=-N \\ n \neq 0}}^{N} |\Gamma(j\omega + j\omega_S n)|^2 |h_T^{-1} x(j\omega)|^2 d\omega$$

$$\leq \int_{-\infty}^{+\infty} \sum_{\substack{n=-\infty \\ n \neq 0}}^{+\infty} |\Gamma(j\omega + j\omega_S n)|^2 |h_T^{-1}(j\omega)|^2 d\omega$$

$$\leq \int_{-\infty}^{+\infty} |R(j\omega)|^2 |x(j\omega)|^2 d\omega.$$

Since the last integrand is integrable and independent of N, and we have already know that $||\Delta_T \cdot x|| = ||\ (S_T - h_T)\cdot x\ ||$ exists, hence we get

$$||\Delta_T \cdot x||^2 \le \frac{1}{2\pi}\int_{-\infty}^{+\infty} |R(j\omega)|^2 |x(j\omega)|^2 d\omega.$$

The exchanges of integration and summation are legal because N is finite. The appearance of h_T^{-1} seems to be troublesome since it introduces infinitely many poles on $j\omega$-axis, but in effect they are cancelled by the zeros of other terms. The identity

$$\sum_{n=-\infty}^{+\infty} |h_T(j\omega - j\omega_s n)|^2 \equiv 1$$

is used [Thompson]. Another infinite summation we need is

$$\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} |\Gamma(j\omega - j\omega_s n)|^2$$

But since

$$|\Gamma(j\omega)|^2 = |h_T(j\omega)|^2 - |h_T(j\omega)|^4$$

we only need to calculate

$$\sum_{n=-\infty}^{+\infty} |h_T(j\omega - j\omega_s n)|^4.$$

Let $a(s) = h_T(s)h_T(-s)$, and $b(s) = a(s)a(-s)$. Note that

$$\sum_{n=-\infty}^{+\infty} |h_T(j\omega - j\omega_s n)|^4 = \sum_{n=-\infty}^{+\infty} b(j\omega - j\omega_s n).$$

Suppose that b(t) is the Laplace inverse of b(s), then Possion Summation Formula gives

$$\sum_{n=-\infty}^{+\infty} b(j\omega - j\omega_s n) = T\sum_{m=0}^{+\infty} b(mT)e^{-j\omega nT}$$

so if b(t) is known we have an alternative way to calculate the summation, i.e. by the right hand side of the above equality. b(t) can be calculated as follows: Firstly the Laplace inverse of a(s) is given by

$$a(t) = h_T(t)*h_T(-t) = \begin{cases} (T+t)/T^2 & t\in[-T,o) \\ (T - t)/T^2 & t\in[0,T) \\ 0 \text{ elsewhere.} & \end{cases}$$

There is no need to give expression for b(t) because only b(nT) $n = 0, \pm 1, \pm 2, \ldots$, are needed. More calculation reveals that

$$b(0) = \frac{2}{3T},$$

$$b(\pm T) = \frac{1}{6T}$$

and

$$b(\pm nT) = 0, \text{ for } n>1.$$

So it follows that

$$\sum_{n=-\infty}^{+\infty} b(j\omega - j\omega_s n) = T\sum_{n=-\infty}^{+\infty} b(nT)e^{-j\omega nT}$$

$$= \frac{1}{6}e^{j\omega T} + \frac{2}{3} + \frac{1}{6}e^{-j\omega T}$$

$$= \frac{1}{3}(2 + \cos\omega T).$$

This gives

$$\sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} |\Gamma(j\omega - j\omega_s n)|^2 = 1 - |h_T(j\omega)|^2 - \frac{1}{3}(2 + \cos\omega T) + |h_T(j\omega)|^4$$

$$= \frac{1}{3}(1 - \cos\omega T) + (|h_T(j\omega)|^2 - 1)|h_T(j\omega)|^2.$$

Since

$$|h_T(j\omega)|^2 = \frac{2}{(\omega T)^2}(1 - \cos\omega T)$$

we finaly get

$$|R(j\omega)|^2 = \frac{1}{6}(\omega T)^2 + |h_T(j\omega)|^2 - 1. \qquad \square$$

In the case that S_T is prefiltered by h with transfer function

$$h(s) = \frac{\frac{2}{T}}{s+\frac{2}{T}}$$

a slightly tighter bound can be given for $\Delta_T \cdot h$. Because

$$2\pi||\Delta_T\cdot h\cdot x||^2 = \int_{-\infty}^{+\infty} |h_T|^2 \left| \sum_{\substack{n\neq 0\\ n=-\infty}}^{+\infty} h(j\omega - j\omega_s n)x(j\omega - j\omega_s n) \right|^2 d\omega$$

$$\leq \int_{-\infty}^{+\infty} |h_T|^2 \sum_{\substack{n\neq 0\\ n=-\infty}}^{+\infty} |h(j\omega - j\omega_s n)|^2 \sum_{\substack{n\neq 0\\ n=-\infty}}^{+\infty} |x(j\omega - j\omega_s n)|^2 d\omega$$

$$= \int_{-\infty}^{+\infty} \{ |h_T|^2\Gamma(j\omega) - |h_T|^2|h|^2 \} \sum_{\substack{n\neq 0\\ n=-\infty}}^{+\infty} |x(j\omega - j\omega_s n)|^2 d\omega$$

$$= \int_{-\infty}^{+\infty} \{ (1 - |h_T|^2)\Gamma(j\omega) - \Pi(j\omega) \} |x(j\omega)|^2 d\omega$$

where

$$\Gamma(j\omega) \triangleq \sum_{n=-\infty}^{+\infty} |h(j\omega - j\omega_s n)|^2$$

and

$$\Pi(j\omega) \triangleq \sum_{\substack{n\neq 0\\ n=-\infty}}^{+\infty} |h_T\cdot h(j\omega)|^2.$$

So the main task is to calculate the two summations, which involve the same techniques as those used in proof of theorem 3.3.3. The details are as follows.

(i) let $\beta(j\omega) = h(j\omega)\cdot h(-j\omega)$, hence

$$\Gamma(j\omega) = \sum_{n=-\infty}^{+\infty} \beta(j\omega - j\omega_s n).$$

It can shown that the inverse Fourier transform of $\beta(j\omega)$ is

$$\beta(t) = \frac{1}{T}\exp(-\frac{2}{T}|t|)$$

and it follows that

$$\Gamma(j\omega) = T\sum_{n=-\infty}^{+\infty} \beta(nT)\exp(-j\omega nT)$$

$$= 1 + \frac{2e^{-2}\cos\omega T - 2e^{-4}}{1 + e^{-4} - 2e^{-2}\cos\omega T}. \qquad \Box$$

(ii) Let $\alpha(j\omega) = (h_T\cdot h)(j\omega)\cdot(h_T\cdot h)(-j\omega)$, and $\alpha(t)$ the inverse Fourier transform of $\alpha(j\omega)$, then we can compute the following

$$\alpha(0) \quad = \frac{1}{2T}(1 + e^{-2})$$

$$\alpha(\pm T) = \frac{1}{4T}(1 - e^{-2})$$

$$\alpha(\pm nT) = e^{-2(n-1)}\alpha(T)$$

and hence

$$\sum_{n=-\infty}^{+\infty} \alpha(j\omega - j\omega_s n) = T\sum_{n=-\infty}^{+\infty} \alpha(nT)\exp(-j\omega nT)$$

$$= \frac{1}{2}\,\frac{1 + 3e^{-4} + (1 - 4e^{-2} - e^{-4})\cos\omega T}{1 + e^{-4} - 2e^{-2}\cos\omega T},$$

and

$$\Pi(j\omega) = \sum_{n=-\infty}^{+\infty} \alpha(j\omega - j\omega_s n) - |h_T\cdot h(j\omega)|^2.$$

Using these equations one can have an explicit expression for an upperbound for $\Delta_T\cdot h$. See Fig B.1 for a comparison. In general, if we know the prefilter characteristics, tighter bounds can be obtained. But the calculations are quite tedious, and the improvement is marginal if we have to use low order approximation to the bound. In the subsequent chapters, we will always use the simplest form of the bounds, i.e. the transfer function $\frac{1}{\sqrt{6}}sT$.

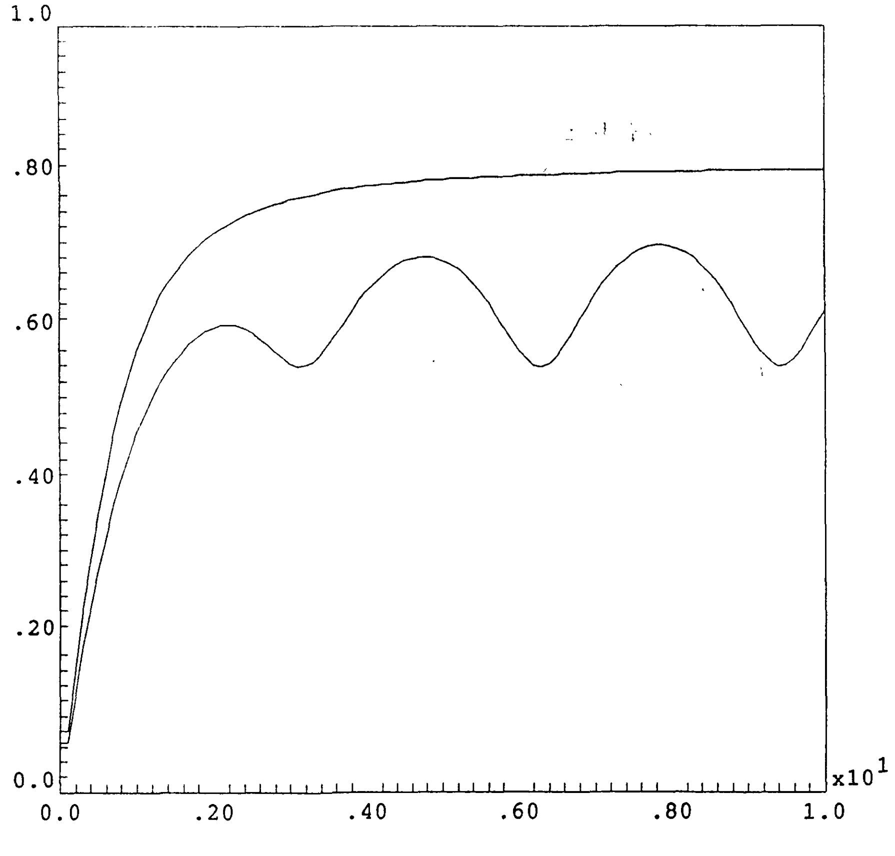


Fig. B.1

Chapter Four
Analysis of Multirate-Multivariable Systems

4.0 *Introduction*

The first aim of this chapter is an input-output characterization of a class of multirate hybrid controllers. The input-output mapping of this class of controllers is shown to be the cascade of an LTI operator and two SAH operators. Based on this, a stability test that only uses LTI techniques will be presented. The idea is that an SAH operator can be approximated by an LTI operator together with an LTI bound for the approximation error, therefore the perturbation methods studied in chapter two can be applied. We will demonstrate the procedure of stability analysis by an example. Following the same idea, we then show that the perturbation method can also be used to estimate certain performance characteristics.

The structural analysis of hybrid controllers gives us insight into how the internal digital implementation of control algorithms influences the external behaviour. Furthermore, the results of this chapter lead to the design philosophy of next chapter.

4.1 *Structure of MM controllers*

We start with the physical structure of a multirate sampled data controller, because we are only interested in the class of controllers that can be practically implemented. A multirate multivariable controller has multiple input and multiple output channels. A computer performs computations on the data acquired from input channels, and sends the results to the output channels. Let the input channels be numbered from j = 1 to n, and the output ones from i = 1 to m. Both the input and the output signals are analog, though the internal ones are digital. See Fig. 4.1.1.

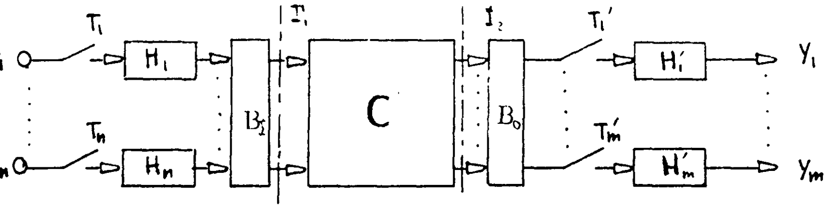


Fig. 4.1.1 Structure of Multirate Hybrid Controller

In Fig. 4.1.1 the control programme performs the following taskes:

i) It fetches data from input bufffers.

ii) It performs finite linear calculations on the data at specified instants; The computation rules are specified by certain given linear difference equations;

iii) It places the results into output buffers.

Communication between the internal software and the outside world is achieved through A/D and D/A converters. An ideal A/D converter can be modelled as the combination of an impulse modulater and a zero-order hold. Clearly, this means that the action of putting data into registers is treated as an integral part of the A/D operation. The sampling and holding at the output represent the action of taking data from memory and putting them into output buffers for given periods. The following assumptions are made throughout:

i) the samplers are ideal

ii) all the samplers work at fixed, but different, rates; the j-th input sampler has rate T_j and the i-th output sampler T_i'; these symbols are also used to denote the sampler themselves.

iii) all the operations are instantaneous.

From the above discussion we can see that the signal flows from e_i's to I_1 and from I_2 to y_j's are easy to describe: they are just SAH operators. In the sequel we will focus on the mapping from I_1 to I_2, and give a mathematical model for it. There are infinitely many ways to programme the software to realize this mapping, thus it is impossible to study all of them. We can only concentrated on the "sensible" ones.

We have already assumed that the calculations are linear, but we have also to specify the way these calculations are carried out and the data they operate on. To facilitate discussion, some fictitious devices are introduced. These devices have nothing to do with the actual implemention of the controller, though their functions can be physically realized.

i) Input buffers: E_j, $j=1,\cdots n$. When T_j samples, it puts the sampled data into E_j. This value will remain the same until the next sample is taken by T_j. Other parts of the hybrid controller can have access to these buffers, but they do not change the values in them.

ii) Computing units: C_{ij}, $i=1,\cdots m$, $j=1,\cdots n$. Each C_{ij} maps a sequence of real numbers e_{ij} into another sequence y_{ij} via the following equations:

$$\begin{cases} x_{ij}(k+1) = A_{ij}x_{ij}(k) + B_{ij}e_{ij}(k) \\ \\ y_{ij}(k) = B_{ij}x_{ij}(k) + D_{ij}e_{ij}(k). \end{cases} \qquad (4.1.1)$$

where x_{ij} is a vector sequence and can be viewed as the internal state of the C_{ij}. The input-output behaviour of this unit, expressed in terms of the $\mathbb{Z}$-transforms of the variables e_{ij} and y_{ij}

is

$$y_{ij}(z) = \hat{C}_{ij}(z)\cdot e_{ij}(z).$$

Without loss of generality, we assume that the computing unit C_{ij} only communicates with the j-th input and i-th output channels. We assume further that all the units operate periodically. In other words, C_{ij} performs the computation (4.1.1) every T_{ij}, for a fixed T_{ij}. When C_{ij} operates, $e_{ij}(k)$ is taken from the input buffer E_j, regardless when the contents of this buffer was updated last time.

iii) Output buffers: Y_i, $i=1,\cdots m$. These are the buffers that store the data to be taken out by the output samplers. Because the whole hybrid controller is linear, the values of the content of these buffers are linear combinations of the outputs of the computing units. In general, we define

$$Y_i(t) = \sum_{j=1}^{n} y_{ij}(t).$$

Note that we have embedded the variables into a universal time scale. So $y_{ij}(t)$ represents the value of the variable y_{ij} at the moment t. Regarded as a function of continuous time, $y_{ij}(t)$ is obviously piecewise constant.

We have to specify the synchronization of the computing units. There may be infinitely many ways of doing this, but not all are meaningful or practical. Our intuition is that the operation of the computing units should be synchronized with either the input or the output samplers, or both if possible. According to the synchronization scheme, the realization of a hybrid controller may be classified into two categories.

a. Input triggered scheme

In this scheme, the computing units are synchronized with the input samplers, i.e, the units C_{ij}, $i=1, \cdots, m$ are called upon immediately after the j-th input buffer is updated. More explicitly, if the sequence of data stored in the buffer E_j is $\hat{e}_j(k)$, with a $\mathbb{Z}$-transform $\hat{e}_j(z)$, then the output of the ij-th computing unit y_{ij} is given by:

$$y_{ij}(z) = \hat{C}_{ij}(z)\hat{e}_j(z).$$

If we attach a time scale to the above expression, by using discrete Laplace transforms, it becomes:

$$y_{ij}(s) = \hat{C}_{ij}(e^{-sT_j})(S_{T_j}\cdot e_j)(s). \qquad (4.1.2)$$

Note y_{ij} is viewed as a variable of continuous time, and e_j is an input signal to the hybrid controller. For simplicity, we assume that there is no delays in the computing units. Therefore from Proposition 3.0.1, there is a rational function $C_{ij}(s)$ such that

$$y_{ij}(s) = S_{T_j}\cdot C(s)\cdot S_{T_j}\cdot e_j(s). \qquad (4.1.3)$$

The presence of SAH transformations clearly indicates the action of sampling data and storing them into memory. A logical diagram of the functioning of C_{ij} can be depicted in Fig. 4.1.2.

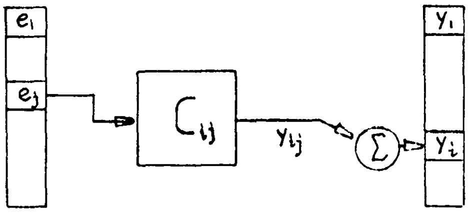


Fig. 4.1.2 Computing Unit

Due to the assumption of linearity, the i-th output can only be the sum of the contributions from C_{ij}, j=1, ···, n. Obviously the most sensible way to do this is to let the i-th output have the most recently updated values from the computing units. But since $T_i^?$ is not synchronized with the operations of C_{ij}, it has to take values that are already stored in memory. This process is conveniently described by SAH operations. In other words

$$y_i(t) = S_{T_i^?}\cdot Y_i(t) = S_{T_i^?}\cdot(\sum_{j=1}^{n} y_{ij}(t)\ ). \qquad (4.1.4)$$

The interface, represented by the presence of SAH operators, are a consequence of the requirement that computations for y_{ij} are synchronized with the input samplers. The obvious drawback of this scheme is that y_{ij} are not immediately accessed by the output samplers, and hence some delays are incured. Since the relation between the j-th input and the i-th output sampling rates is in general non-integer, we can expect the overall mapping of a hybrid controller to be complicated.

b. Output triggered scheme

In this scheme the computed y_{ij} are immediately sent to the output channels. This requires that the computing unit c_{ij}, $j=1, \cdots, n$ are called upon just before the i-th output sampler takes value from buffer Y_i. In this case, the content of Y_i, viewed as a function of continuous time t, is given by:

$$Y_i(t) = \sum_{j=1}^{n} y_{ij}(t)$$

while y_{ij} is defined by:

$$y_{ij}(s) = \hat{C}_{ij}(e^{-sT'_i}) \left[S_{T'_i}\cdot(S_{T_j}\cdot e_j)\right](s). \qquad (4.1.5)$$

Explanation: $S_{T_j}\cdot e_j$ is the contents of input buffer E_j, therefore the action of taking values form this buffer, at rate T_i', is described by $S_{T_i'}\cdot(S_{T_j}\cdot e_j)$. By Proposition 3.0.1, there is a $C_{ij}(s)$ such that the above equation can be rewritten as:

$$y_{ij}(s) = S_{T_i'}\cdot C_{ij}(s)\cdot S_{T_i'}\cdot S_{T_j}\cdot e_j(s). \qquad (4.1.6)$$

Clearly, the i-th output y_i is

$$y_i(t) = S_{T_i'}\cdot Y_i(t) = \sum_{j=1}^{n} y_{ij}(t)$$

In this scheme, the samples taken by the input samplers are not used immediately. It is therefore hard to say if this scheme is better than the input triggered scheme.

For both schemes, we conclude that the input-output mapping of a multirate hybrid controller K, as in Fig. 4.1.1, has the form of

$$K = S_2\cdot C_S\cdot S_1 \qquad (4.1.7)$$

where

$$S_1 \triangleq \begin{bmatrix} S_{T_1} & & \\ & \ddots & \\ & & S_{T_n} \end{bmatrix}, \qquad (4.1.8)$$

and

$$S_2 \triangleq \begin{bmatrix} S_{T_1'} & & \\ & \ddots & \\ & & S_{T_m'} \end{bmatrix}. \qquad (4.1.9)$$

C_s has different expressions according to the scheme of realizations. In the case of input triggered scheme,

$$C_s = \begin{bmatrix} S_{T_1}C_{11} & \cdots & S_{T_n}C_{1n} \\ & \ddots & \\ S_{T_1}C_{m1} & \cdots & S_{T_n}C_{mn} \end{bmatrix}, \tag{4.1.10}$$

while in the output triggered scheme

$$C_s = \begin{bmatrix} C_{11}S_{T'_1} & \cdots & C_{1n}S_{T'_1} \\ & \ddots & \\ C_{m1}S_{T'_m} & \cdots & C_{mn}S_{T'_m} \end{bmatrix}. \tag{4.1.11}$$

Remark 4.1.1: Hybrid controller has the above forms because of the two important assumptions: (1) the computing units operate periodically; (2) these units have time invariant parameters. These assumptions are due to practical considerations. Also, in the case of input triggered scheme, one column of computing units can share one set of state variables, because they share the same input and they operate at the same rate. A similar remark may be made in the case of the output triggered scheme.

The C_{ij}'s in (4.1.10) and (4.1.11) are finite dimensional LTI operators. But they are mixed up with the SAH operators. In order to study the input-output behaviour of a hybrid controller, it is desirable to separate the SAH's from the LTI components. If the finite dimensionality of C_{ij}'s is abandoned, it is easily seen that

$$K = S_2 \cdot \hat{C} \cdot S_1 \tag{4.1.12}$$

where $\hat{C}$ is given by

$$\hat{C}(s) = \left[\hat{C}_{ij}(e^{-sT'_i})\right]_{i=1,\cdots,n;j=1,\cdots,m} \tag{4.1.13}$$

in input-triggered scheme, and by

$$\hat{C}(s) = \left[\hat{C}_{ij}(e^{-sT_j})\right]_{i=1,\cdots,n;j=1,\cdots,m}. \tag{4.1.14}$$

in output triggered scheme. $\hat{C}$ in this case is a multiplication operator, but it is infinite dimensional.

Under further condition on the sampling rates, finite dimensionality C can be obtained. We define a third scheme for the triggering of the computing units as follows. Let each unit C_{ij} have its own clock T_{ij} . As far as the ij-th unit is concerned, its "outside world" are the input and output buffers. The foregoing analysis indicates that the input-output mapping of ij-th unit is

$$S_{T_{ij}} \cdot C_{ij} \cdot S_{T_{ij}}$$

where C_{ij} is an LTI operator, by Proposition 3.0.1. Thus

$$K = \left[S_{T_i} \cdot S_{T_{ij}} \cdot C_{ij} S_{T_{ij}} \cdot S_{T_j} \right]_{i=1,\cdots m; j=1,\cdots,n} . \tag{4.1.15}$$

If every input and output sampling rate is an interger multiple of a $T > 0$, and if we choose $T_{ij} = T$, then by the absorbing law, it is clear that

$$K = S_2 \cdot C \cdot S_1$$

where C is a finite dimensional LTI operator. In this case we say that K is separable as compositions of LTI and SAH operators. This is the most important class of realizations for a hybrid operator, and is the one most commonly studied in the literature of multirate sampling. In the appendix to this chapter we will show that the input-output mapping of a practical hybrid controller is in a separable form only in the above situation.

4.2 *Nyquist stability test*

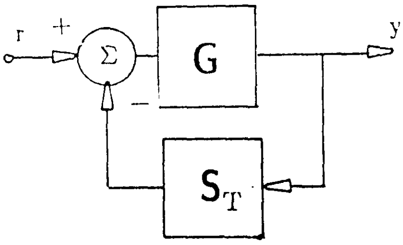


Fig. 4.2.1

In this section we study the stability of the feedback system shown in Fig. 1.0.1 in which G is an LTI system and K a multirate hybrid controller. We begin by studying a single rate system, and then extend the result to multirate systems.

Suppose that G in system Fig. 4.2.1 is LTI, with transfer function G(s). It is easy to write down the relation between the variables:

$$y = G\cdot(r - S_T\cdot y)$$

By applying S_T on both sides, it becomes:

$$(I + S_T\cdot G)\cdot S_T\cdot y = S_T\cdot G\cdot r. \tag{4.2.1}$$

We know that the stability of the system boils down to the question of if the operator $H \triangleq (I + S_T\cdot G)$ is boundedly invertible, giving the fact that it is one-to-one. It would be difficult to give conditions for this, for $(I + S_T\cdot G)$ is time varying. But We can make use of the fact that invertibility is only needed on the space Rang(S_T), for the right hand side of (4.2.1) is always in Rang(S_T). On Rang(S_T), H can be reprensented as transfer function:

$$H|_{\mathrm{Rang}(S_T)}(s) = I + (Gh_T)^{*T}(s)\ . \tag{4.2.2}$$

In fact, this is why $\mathbb{Z}$-transfer function method is effective. Here the SAH operators are used instead of discrete time relations so that it is easier to demonstrate the difficulties of multirate sampling and to compare different approaches to the problem. We know that H (on Rang(S_T)) is boundedly invertible if H(s) does not have zeros in $\bar{C}_+$. Note that H(s) is periodic on any vertical line on the complex plane, i.e., it is completely defined in the zone $\Omega \triangleq \{s: -\frac{\pi}{T} \leq \mathrm{Im}(s) \leq \frac{\pi}{T}\}$. So, in order to determine the invertibility of H, we only need to check if H(s) has any zero in $\Omega\cap\bar{C}_+$. We also observe that the number of poles of H(s) in $\Omega\cap\bar{C}_+$ is the same as that of G(s) in the whole right half plane. Let this number be P_G. Here we impose one condition on the sampling rate T, that it mustn't be such that G(s) has poles on the lines $\mathrm{Im}(s) = \frac{(2n+1)\pi}{T}$, $n=0, \pm1, \cdots$. This is a sensible requirement, because otherwise it may happen that some of the modes of G become hidden after sampling. This condition will be assumed throughout.

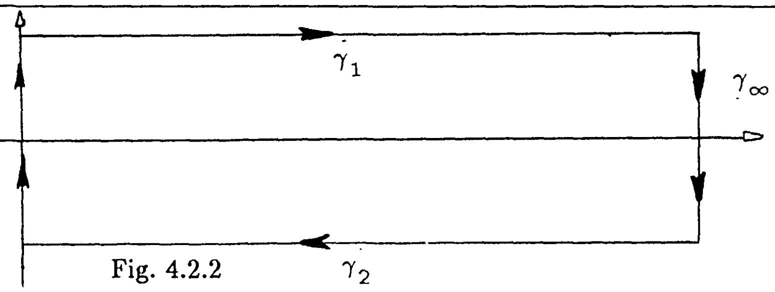


Fig. 4.2.2

Let the boundary of $\Omega \cap \bar{C}_+$ be $\bar{\Gamma} = \Gamma + \gamma_1 + \gamma_2 + \gamma_\infty$, as illustrated in Fig. 4.2.2, where γ_∞ is a vertical segment far enough that all the poles of G(s) lie on its left. Using the fact H(s) is periodic we have

$$\int_{\bar{\Gamma}} \frac{H'(s)}{H(s)} ds = \int_{\Gamma+\gamma_\infty} \frac{H'(s)}{H(s)} ds$$

and hence H(s) does not have zeros in C_+ if and only if

$$P_G = \mathrm{Ind}_{\Gamma'}(0) \qquad (4.2.3)$$

where

$$\Gamma' \triangleq H \circ (\Gamma + \gamma_\infty)$$

and $\mathrm{Ind}_\gamma(0)$ is the number of anti-clockwise encirclement of the point 0 by a curve γ. Note that H(s) is a function of e^{-sT}, therefore if we place γ_∞ further and further into the right half plane, its image under H will shrink into a point. Thus we only need to examine the image of Γ.

The above test is well known in $\mathbb{Z}$-transformation form, but it involves the computation of H(s). However, one would expect that when sampling is sufficiently fast, H(s) should be close to $(I + G(s)h_T(s))$ in low frequency range Γ. So it would be nice to have a stability test of the system in terms of G(s). It is indeed the case if we specify what is meant by "sampling fast".

Theorem 4.2.1: Suppose that G(s) is strictly proper. If the sampling rate T is so high that all the poles of G(s) lie in the zone $\Omega \triangleq \{s: -\frac{\pi}{T} \leq \mathrm{Im}(s) \leq \frac{\pi}{T}\}$, and that

$$\left| \sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} (Gh_T)(s - j\omega_T n) \right| < \left| I + Gh_T(s) \right| \text{ on } \bar{\Gamma}, \qquad (4.2.4)$$

then the condition for stability becomes

$$P_G = \mathrm{Ind}_{\Gamma''}(0) \qquad (4.2.5)$$

where

$$\Gamma'' \triangleq (I + Gh_T) \circ \bar{\Gamma}.$$

Proof:

Let $F(s) = \sum_{n=-\infty}^{\infty} (Gh_T)(s - j\omega_s n)$, clearly $H(s) = 1 + F(s)$. We first show that F(s) is analytic in $\Omega \backslash \{$poles of G(s)$\}$. Pick a point s_0, if G(s) is analytic there, then there can be one of the following cases: either that there is a vertical strap covering s_0 or there are poles of G(s) on the vertical line passing through s_0. In the first instance, choose the boundary Γ of a strap that covers s_0, anti-clockwise oriented, and do the integral

$$\frac{1}{2\pi j}\int_\Gamma \frac{T(Gh_T)(\xi)d\xi}{1-\exp[-(s_0-\xi)]}$$

$$= \sum_{n=-\infty}^{+\infty} \text{Res}\left[\frac{T(Gh_T)(\xi)}{1-\exp[-(s_0-\xi)]}\right]_{\xi=s_0-j\omega_s n}$$

$$= \sum_{n=-\infty}^{+\infty} (Gh_T)(s_0 - j\omega_s n).$$

The integral is well defined since G(s) is strictly proper. On the other hand, the above integral is also

$$\sum_{\text{poles of } G(s)} \text{Res}\left[\frac{T(Gh_T)(\xi)}{1-\exp[-(s_o-\xi)]}\right] = \sum_{i=1}^{P_G} \frac{R_i}{1-\exp[-(s-p_i)]},$$

where R_i is the residue of $Gh_T(s)$ at pole p_i. This is clearly an analytic function.

In the other case, we just need to make some indentations on the integral path and get the same expression as above. In either cases we conclude that F(s) is indeed analytic in $\Omega \backslash \{$poles of G(s)$\}$.

(4.2.4) then says that

$$|H(s) - (1 + G(s)h_T(s))| < |1 + G(s)h_T(s)| \text{ on } \bar{\Gamma}.$$

This implies that the images of $\bar{\Gamma}$ under the two functions $(1 + Gh_T(s))$ and H(s) will enclose the origin the same number of times. We also know that the two functions have the same number of poles in $\Omega \cap \{Re(s) > 0\}$. Thus we conclude that they have the same number of zeros in $\Omega \cap \{Re(s) > 0\}$, from Rouché's theorem applied on curve $\bar{\Gamma}$.□

Remark 4.2.2: (4.2.4) can be used as a criterion for the choices of sampling rates. If it is satisfied, one can think of the sampling, that perturbes $(1 + Gh_T(s))$ to H(s), as fast enough to be trusted. But the condition (4.2.4) is not easy to check, because it must be verified on the whole $\bar{\Gamma}$. So for theorem 4.2.1 to be of any use, we need an efficient way of verifying (4.2.4).

(4.2.3) is a necessary and sufficient condition for closed loop stability, but (4.2.4) and

(4.2.5) is only sufficient. In the following, we discuss extensions of these tests to multirate systems.

To generalize the necessary and sufficient stability test to multirate systems, one needs to construct an equivalent (as far as stability is concerned) single rate system, and apply the techniques for single rate systems. So the focal point is how to transform a given multirate system into a single rate one. The variety of decomposition methods can only be used in cases where, in our terminology, the hybrid controllers are separable. We briefly outline the basic steps of using switch decomposition method for the system in Fig. 1.0.1 The technique is well known, but we believe that its application to the general hybrid controlled configuration is new.

i) Switch Decomposition method

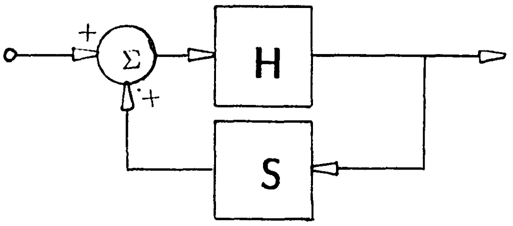


Fig. 4.2.3

The system shown in Fig. 1.0.1 can be transformed into that in Fig 4.2.3, where

$$H \triangleq \begin{bmatrix} 0 & -G \\ \hat{C} & 0 \end{bmatrix}$$

and

$$S \triangleq \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} = \text{diag}\ \{ S_{T_1}, \cdots, S_{T_n}; S_{T'_1}, \cdots, S_{T'_m} \}.$$

Suppose that $T_k = r_k T$ and $T'_j = r'_j T$, for some integers r_k and r_j'. If we take switch decomposition of each SAH, then S becomes

$$S = F \cdot \text{diag}\{\overleftarrow{S_T}, \overset{N}{\cdots}, \overrightarrow{S_T}\} \cdot A \qquad (4.2.6)$$

where $N = \sum_{k=1}^{n} r_k + \sum_{j=1}^{m} r'_j$. F and A can be determined according to (3.4.8). We only need to know that they are LTI. Absorbing F and A into H we obtain an equivalent single rate system, whose stability can now be tested by the method outlined above. An evident problem of this approach is the dimension growth, though it should not be a serious drawback for numerical analysis. What is nice about this test is that it offers a systematic solution, and gives necessary and sufficient condition for stability. Of course both A and F are dependent of the sampling rates.

The reason for using switch decomposition is the following: for single rate systems stability can be deduced from the fact that $(I + G \cdot S_T)$ is boundedly invertible on the range of S_T, which is simply a transfer function. However, the operator $(I + S \cdot H)$ is in general not representable as a transfer function on the range of S.

ii) Perturbation method

It is possible to impose a kind of non-aliasing condition analogue to (4.2.4), which should be satisfied if the dynamical characteristics of G are to be adequately reflected from sampled data. To be more precise let $\bar{S}$ be

$$\bar{S} \triangleq \text{diag} \{ h_{T_1}, \cdots, h_{T_n}; h_{T'_1}, \cdots, h_{T'_m} \}, \tag{4.2.7}$$

and

$$\Delta \triangleq \text{diag} \{ \Delta_{T_1}, \cdots, \Delta_{T_n}; \Delta_{T'_1}, \cdots, \Delta_{T'_m} \}. \tag{4.2.8}$$

We require that the sampling rates are sufficiently high so that the systems $(H, \bar{S})$ and (H, S) are similar. To be more specific, let

$$R = \text{diag} \{ R_{T_1}, \cdots, R_{T_n}; R_{T'_1}, \cdots, R_{T'_m} \}, \tag{4.2.9}$$

and Δ the class of diagonal perturbations, such that if $\delta \in \Delta$, then

$$|\delta \cdot x| \leq |R \cdot x| \ \forall \ x \in \text{Dom}(\Delta).$$

We require that the nominal system $(H, \bar{S})$ to be robust stable for the class of perturbations Δ. Since it is not difficult to verify the stability of $(H, \bar{S})$, what remains to be done is to find an effective method of dealing with the time-variant perturbation Δ. We are particularly interested in solving this this problem in the framework of LTI operators. Hence the results discussed in the two earlier chapters naturally find application in the present situation. The key point here is that although Δ is not LTI, it belongs to Δ which is bounded by the LTI operator R. However in order to use the results in chapter three, we need that H be strictly proper. This is the case only when K is separable, for otherwise $\hat{C}$ is never strictly proper. In the separable case the conditions for closed loop stability are summarised as follows.

Theorem 4.2.3: The system in Fig. 4.2.3 is stable if the the following two conditions are satisfied:

(1) $(H, \bar{S})$ is stable; and

(2) $|R \cdot N| < 1$,

where N denotes the nominal system mapping, given by

$$N = H \cdot (I - \bar{S} \cdot H)^{-1}.$$

Proof:

(1) implies that the nominal system is internally stable for there can not be any cancellation between $\bar{S}_i$ and C or G. Hence N is internally stable. (2) will guarantee that the mapping from (r_1, r_2) to (y_1, y_2) is stable under the perturbation class Δ. The only thing to note here is that the use of small gain theorem is valid, for Δ is well defined on the range of N. Then theorem 2.2.7 can be applied to prove the assertion. □

Comment 4.2.4: The interesting thing about (2) is that it imposes a kind of "envelop" condition on the Bode diagram (of the maximum singular value) of N. From last chapter we have seen that R has transfer function K·s, so for (2) to be satisfied N must have a roll-off rate not less than 20db/dec in high frequencies. Intuitively it makes sense because sampling is especially sensitive to high frequency noise. But what intuition does not tell is the precise rate the high frequency gain should be rendered. The results from last chapter give this information.

Comment 4.2.5: The first step towards analyzing the general non-separable multirate hybrid controller, by the method used above, is to separate the LTI part of such a controller from the rest of its structure. Such a separation is possible, but it does not decompose a hybrid controller into a straightforward cascade of $S_1 \cdot C \cdot S_1$. An attempt to analyze such cases is included in the appendix to this chapter. As will be shown that the non-separable cases are of little interest as far as practical application is concerned.

Example 4.2.6: Suppose that in a hybrid system as Fig. 1 ?? with separable controller the following are defined:

$$G = \frac{1}{s - 0.4},$$

$$K = S_{T_1} \cdot C \cdot S_{T_1}$$

and

$$C = \frac{k}{s + 2}.$$

We will illustrate how to analyze the stability of this system by perturbation method. First, the system is transformed into the configuration in Fig. 4.2.3, with

$$H = \begin{bmatrix} 0 & \frac{-1}{s - 0.4} \\ \frac{k}{s + 2} & 1 \end{bmatrix},$$

$$\bar{S} = \begin{bmatrix} h_{T_1} & 0 \\ 0 & h_{T_2} \end{bmatrix},$$

$$R = \begin{bmatrix} R_{T_1} & 0 \\ 0 & R_{T_2} \end{bmatrix},$$

and

$$N = H \cdot (I - \bar{S} \cdot H)^{-1}.$$

The nominal plant has one unstable pole, (h_T does not have finite poles), and the Nyquist loci is shown in Fig. 4.2.4a,c. Clearly, the nominal system is stable for the indicated sets of parameters. The sampled data system is stable if the maximum singular value of R·N is less than one for all frequencies. The singular value plots are in Fig. 4.2.4b,d, from which we can conclude that for the three sets parameters the system is stable.

These tests, as would be expected, are conservative. To reduce conservatism, scaling can be used. Since the system in Fig. 4.2.3 is equivalent to that in Fig 4.2.5 for diagonal matrices T's with positive entries, one can choose a suitable T to reduce the maximum singular value of R·N.

Idealy, scaling T should be solved from

$$\min_{T \in \mathbb{T}} \sup_{\omega} \bar{\sigma}\, [T(R \cdot N)(j\omega)T^{-1}],$$

but this kind of minimization is not easy to solve. However, it is possible to minimize over $\mathbb{T}$ at the frequency where peak occurs by numerical methods, such as steepest descent procedure.

In this example, the parameters $k=1$, $T_1=1$ and $T_2=0.5$ will correspond to Fig 4.2.4 e,f. The two plots in f illustrate how scaling can reduce conservativeness, for without scaling one may not conclude that the system is stable.

Even with scaling, this method is still conservative. We can see this in the case of single rate sampling. When $T_1 = T_2 = T$, the characteristic polynomial is

$$Z^2 - (e^{-2T} + e^{0.4T})Z + e^{-1.6T} + 1.25(1 - e^{-2T})(1 - e^{0.4T})k .$$

It can be shown that when $k = 1$, the system is stable for $T \leq 1.42$. Using perturbation method, the corresponding estimate is $T \leq 1.0$.

The virtue of this method lies in that it gives an easy and systematic approach to the analysis of stability of multirate systems.

4.3 *Performance*

Following the idea that a multirate hybrid controlled system can be viewed as an LTI system subject to perturbations, one may also analyze system performances by finding a bound on the possible degradation of certain performance index for the nominal system under perturbation Δ. For typical performance requirement, such as tracking error, a system can be transformed into Fig. 4.3.1, where r is a reference signal and y represents the error of tracking. T is entirely LTI, and the approximation errors are lumped into Δ.

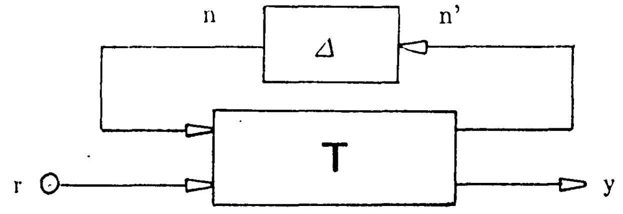
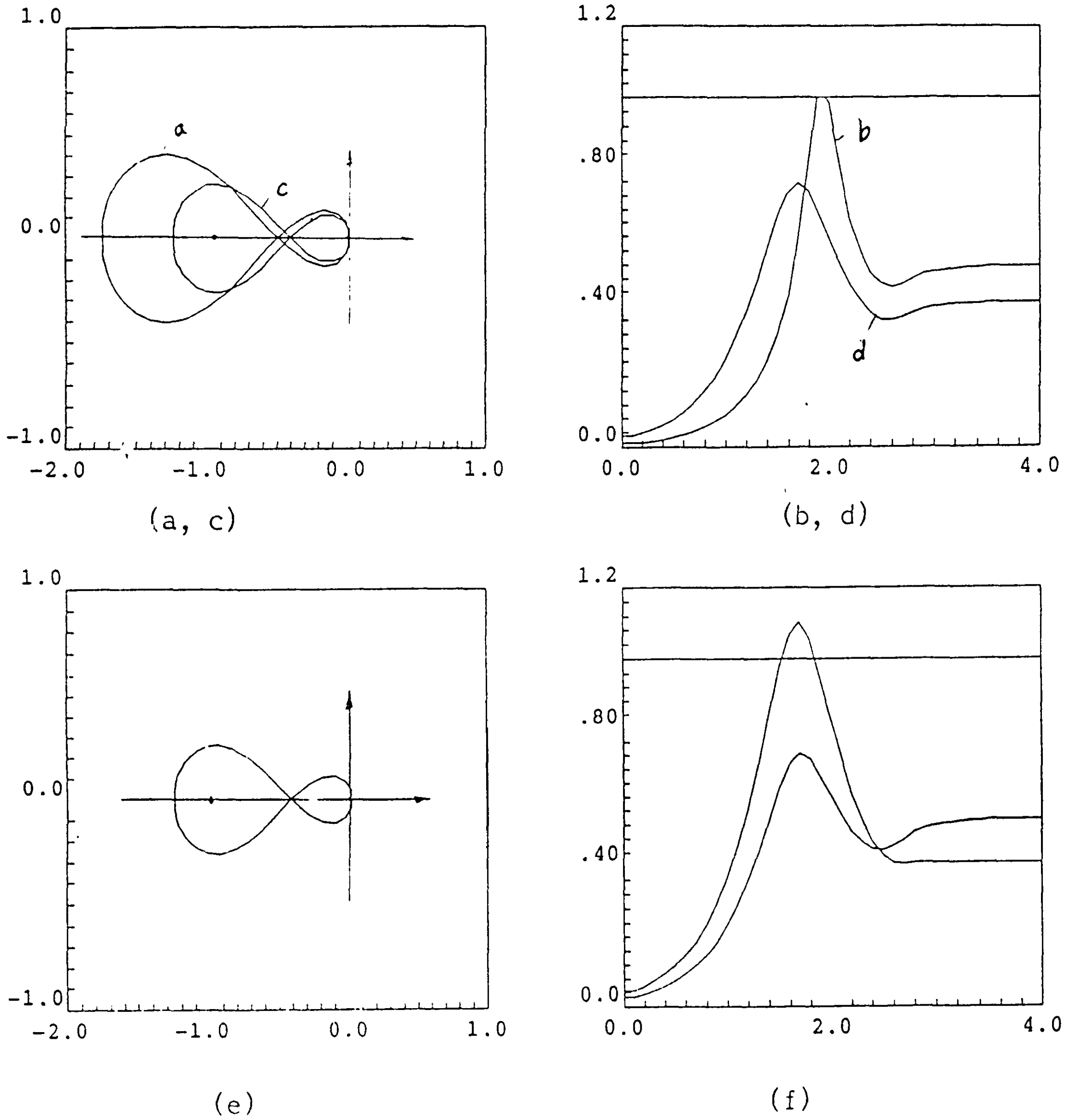


Fig. 4.3.1

Let

$$T \triangleq \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} : \begin{bmatrix} n \\ r \end{bmatrix} \rightarrow \begin{bmatrix} n' \\ y \end{bmatrix} .$$

Nominal performance is defined by $p_n \triangleq |T_{22}|$. Simple calculation shows that



(a, c)

(b, d)

(e)

(f)

Fig. 4.2.4

(a, b): k = 1, $T_1 = 0.5$, $T_2 = 1$;

(c, d): k = 1.4, $T_1 = 0.4$, $T_2 = 0.9$;

(e, f): k = 1, $T_1 = 1$, $T_2 = 0.5$;

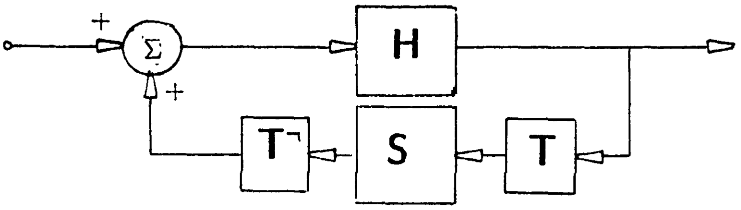


Fig. 4.2.5 Scaling of the perturbation

$$y = [T_{22} + T_{21}\cdot(I - \Delta\cdot T_{11})^{-1}\Delta\cdot T_{12}]\cdot r$$

$$= F_u[T, \Delta]\cdot r$$

where the symbol $F_u[\cdot, \cdot]$ represents a linear fractional mapping for a suitable partitioning of T [Doyèl]. Hence performance under perturbation becomes

$$p \triangleq |F_u[T, \Delta]| \qquad (4.3.1)$$

Because of the presence of Δ, it is not feasible to use (4.3.1) directly to obtain the performance index. Thus it is important to derive a bound on the possible change of the index in terms of the LTI bound on Δ. In [Thompson], an estimate was given in roughly the following manner:

$$p \le |T_{22}| + \frac{||T_{21}|| \cdot ||\Delta T_{12}||}{1 - ||\Delta\cdot T_{11}||}$$

$$\le p_n + \frac{||T_{21}|| \cdot ||R\cdot T_{12}||}{1 - ||R\cdot T_{11}||}. \qquad (4.3.2)$$

Note that $|\Delta\cdot T_{11}| < 1$ is assumed as it is required by robust stability. Obviously, this is a very conservative estimate. For this reason we propose a different estimate, which is also useful for design purposes.

Observe the system in Fig. 4.3.2. T is defined as above, R is an LTI bound on Δ, and α and β are constants.

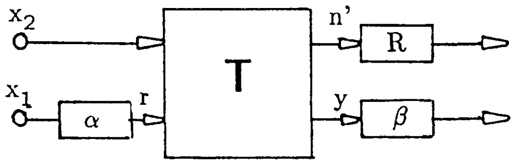


Fig. 4.3.2

First, since the condition $|R\cdot T_{11}| \le 1$ must be satisfied, we can find α and β such that

$$|\hat{T}| \triangleq \left\| \begin{bmatrix} 1 & \\ & \beta \end{bmatrix} \cdot \begin{bmatrix} R & \\ & I \end{bmatrix} \cdot T \cdot \begin{bmatrix} 1 & \\ & \alpha \end{bmatrix} \right\| \le 1 \qquad (4.3.3)$$

To see this we just need to choose α and β small enough. Of course there are many pairs of (α, β) that will satisfy (4.3.3).

Lemma 4.3.1: For any pair of $(\alpha,\ \beta)$ that satisfies (4.3.3), we have

$$p \le \frac{1}{\alpha\beta}. \tag{4.3.4}$$

Proof:

Use the variables defined in Fig. 4.3.2, we have

$$\left| \hat{T}\cdot\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right| \le \left| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right|,$$

and

$$\left| \hat{T}\cdot\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right| = \left| \begin{bmatrix} R & \\ & \beta \end{bmatrix}\cdot\begin{bmatrix} n' \\ y \end{bmatrix} \right| \ge \left| \begin{bmatrix} \Delta & \\ & \beta \end{bmatrix}\cdot\begin{bmatrix} n' \\ y \end{bmatrix} \right|$$

Note that

$$\left| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right|^2 \triangleq |x_1|^2 + |x_2|^2$$

hence

$$|x_1|^2 + |x_2|^2 \ge |\Delta n'|^2 + |y|^2 \quad \forall x_1 \text{ and } x_2.$$

Let $x_1 = \Delta\cdot n'$, i.e. let $y = F_u[\hat{T},\ \Delta]\cdot x_2$, we have

$$\|y\| \le \|x_2\| \ \forall x_2,$$

or

$$\|y\| \le \frac{1}{\alpha\beta}\,\|r\|$$

which implies that

$$p \triangleq \sup_r \frac{|y|}{|r|} \le \frac{1}{\alpha\beta} \qquad \square$$

The foregoing analysis also indicates that if $(\alpha,\ \beta)$ can be found such that $\|\hat{T}\| \le 1$, one can conclude that (1) system is robustly stable and (2) $p \le \frac{1}{\alpha\beta}$. This is because that $\|T_{11}\| \le |\hat{T}|$ regardless what α and β are. It is clear that α and β should be chosen as big as possible to give an accurate estimate of p; it is also evident that it makes difference to select the balance between α and β. A numerical example is given below to show the improvement over the direct estimate in (4.3.2).

Example 4.3.2: Let T be

$$T(s) = \begin{bmatrix} \frac{1}{s+1} & \frac{1}{s+5} \\ \frac{1}{s+6} & \frac{1}{s+2} \end{bmatrix}$$

and the bound on Δ be

$$R(s) = 0.8 \cdot s \ .$$

It is easy to compute that $p_n = |T_{22}| = 0.5$. An estimate of p via (4.3.2) will be

$$p \leq |T_{22}| + \frac{\|T_{12}\| \cdot \|R \cdot T_{21}\|}{1 - \|R \cdot T_{11}\|} = 1.67.$$

If we chose $\alpha = 0.7$ and $\beta = 2.5$, it follows that

$$|\hat{T}| \triangleq \left\| \begin{bmatrix} 1 & \\ & \beta \end{bmatrix} \cdot \begin{bmatrix} R & \\ & I \end{bmatrix} \cdot T \cdot \begin{bmatrix} 1 & \\ & \alpha \end{bmatrix} \right\| = 0.99 < 1$$

hance we know that $p \leq \frac{1}{\alpha \cdot \beta} = 0.57$. In other words the performance degradation is about 14%. The improvement is obvious.

Appendix:

On the separation property of multirate hybrid controllers.

The motivation for decompose the input-output relation of a multirate controller into finite dimensional LTI and SAH operators is clear: it enables one to treat such controller in the framework of LTI systems. In the main text of this chapter, we have shown that in a special case this can be done, i.e. when it is possible to let all the computing units be synchronized with a sampling rate that is the maximum common divisor of all the sampling rates. Naturally one would ask if such separation property also exists for more general case, without the strong assumption that all the sampling rates are in rational ratios of each other.

Such separation exists, though it is perhaps not so helpful or useful. Attempts have been made to overcome this problem, and the only case that is resolved satisfactorily is the one in the main text. Nevertheless, we list the details of an analysis to demonstrate the nature of the problem.

We will first give a separation result. Then we will argue, by means of state space, that this is the best one can hope to do. Finally, we look into the situations that can arise from practical applications, and show that there is no need to force oneself into a dilemma by choosing arbitrary sampling rates.

i) General separation theorem

The goal is to express K as $S_2 \cdot C \cdot S_1$, where C is a finite dimensional LTI operator. In the following, we only derive such a separation for the input triggered scheme. It is similar for output triggered scheme.

Since K is linear, it is in the form of

$$\begin{aligned} K &= [K_1, \cdots, K_n] \\ &= \overset{\longleftarrow\ n\ \longrightarrow}{[I, \cdots, I]} \cdot \begin{bmatrix} K_1 & & \\ & \ddots & \\ & & K_n \end{bmatrix}, \end{aligned}$$

and from the analysis of the structure of such controllers, we know that

$$K_j = \begin{bmatrix} S_{T_1}' S_{T_j} & & \\ & \ddots & \\ & & S_{T_m}' S_{T_j} \end{bmatrix} \cdot C_j \cdot S_{T_j} \triangleq S_j \cdot C_j \cdot S_{T_j},\ j = 1, \cdots, n,$$

and C_j is $[c_{1j}, \cdots, c_{mj}]^T$. Thus we can write K as

$$K = [\overleftarrow{I}, \overset{n}{\cdots}, \overrightarrow{I}]\cdot\begin{bmatrix} S_1C_1S_{T_1} & & \\ & \ddots & \\ & & S_nC_nS_{T_n} \end{bmatrix}$$

$$= \tilde{S}\cdot\tilde{C}\cdot S_1. \qquad (A.1)$$

We note that $\tilde{C}$ is a block diagonal matrix. $\tilde{S}$ is a quite complicated operator, in that it involves entries of the form $S_{T_i}^{?}\cdot S_{T_j}$ about which we don't know much.

So although we can have a separated form for K, it does not help much for we have not found an effective way of dealing with $\tilde{S}$. The above technique for finding such a separation seems primàtive, so one would think if it is the right way of doing it. We have tried many other ways, without any success. What has been sought for is the possibility to let $\tilde{C}$ be in a non-structured form, in other words in the form where

$$\tilde{C} = [C_1, \cdots, C_n] \qquad (A.2)$$

as in the separable cases. We will show below that the separable types are the only cases where it is possible to put $\tilde{C}$ in a non-structured form. We will also see that it is a direct consequence of the requirement that the computing units are invoked periodically with finite rates, and that they have time invariant parameters. If these requirements are droped, and with some additional assumptions that in practice are surely not met, one might get separation always, as we will see later.

A computer can only perform calculation discretely, so there is no way for it to provide a continuous evolution of the state variables as in a continuous time dynamical system. To make it behave like one when observed from the input-output signals, the only way is to let it produce the state variables at the exact moments when they are accessed. Therefore it can be seen that when all the input and output sampling rates are in interger ratio to a T, one just needs to update the state variables periodically in T to conceal the discrete nature of the variables from the input-output samplers. However, when such a T does not exist, then regardless how fast the internal state variables are updated, there is bound to be moments when it is impossible to provide the output sampler with the "true" values of the states.

We can, on the other hand, construct a fictitious algorithm that can emulate a continuous state evolution, as long as the input and output sampling rates are finite. Let C be the continuous time system that we want to emulate, i.e. we want to realize the mapping $S_2\cdot C\cdot S_1$. Suppose that C has a state space realization {A, B, C, D}. States evolution over any time interval $[t_n, t_{n+1}]$ is given by:

$$
\begin{aligned}
x(t_{n+1}) &= e^{A(t_{n+1}\,-\,t_n)}\cdot x(t_n) + \int_{t_n}^{t_{n+1}} \exp(A(t_{n+1} - \tau))B\cdot u(\tau)d\tau \\
&= F(t_{n+1} - t_n)\cdot x(t_n) + H(t_{n+1}, t_n)\cdot u(t_n) \qquad \text{(A.3)}
\end{aligned}
$$

since the inputs are always constant between two consecutive sampling actions. Assume that the computer programme can calculate the matrices $F(t_{n+1}-t_n)$ and $H(t_{n+1}, t_n)$ instantaneously, whenever $t_n < t_{n+1}$ are given. (Is this possible?). The following procedure will emulate C completely: suppose one of the input or output sampler samples at $t = t_n$, record this time; as soon as the next sampling (input or output) happens at $t = t_{n+1}$, calculate F and H and update the state variables using F and H. In doing so the states are always available at the time they are accessed. Hence the input-output mapping of such a K is $S_2\cdot C\cdot S_1$. It is whether the states are updated correctly that decides if the input-output relation should be in a separated form.

Obviously, in input-triggered scheme the states are not updated in the manner we have just described. In fact only the states shared by one column of units are updated correctly. That is why each particular column it is indeed in a separated form. Columns can not share states among themselves. In view of this, it is not surprising to get (A.1) but not anything better. It is not completely right to say that K, in input-triggered scheme, is a multivariable controller. Rather, it is a stack of single-input multi-output controllers piled together. This explains why we have not succeeded in writing K as an m×n LTI controller cascaded between SAH operators.

Because there are advantages from design, analysis and implementation point of views to have K in a separated form, we are led to examine some practical situations where particular pattern of sampling rates are chosen. Basicly, the need for multirate sampling arises under two circumstances. One is when an on-board computer controles a multiloop system and there is a high demand for computation resourses, while the computer has limited capacity. Therefore it is necessary to allocate computation time to different loops according to their dynamics. The other is when the sensor and actuators in a system are operating at different rates, therefore measurement and control must be synchronized with them. In the first instance there is no reason why one should not choose a convenient ratio between sampling rates, and let the state updating rate be the same as the fastest input sampling rate (in input-triggered scheme). For the second situation, unless it so happens that the working rates of the hardware are in irrational ratios, which is an unlikely case, one could always choose a maximum common dividing sampling rate to work with.

We thus conclude that the non-separable case is of little practical interests.

Chapter Five
Design of Multirate-Multivariable Systems

5.0 *Introduction*

In this chapter we study the problem of designing multirate multivariable systems. As we have seen in chapter one, the design of sampled data systems can be classified into exact and approximate methods. Apart from the basic problems associated with the exact method, multirate systems suffers from an additional problems: the growth of dimension. The LQ optimal synthesis methods of Amit [Amit] and Glasson [Glasson] seem to be the best among all, though limited by the fundmental difficulties of LQ methodology. The approximate methods, on the other hand, are more convenient tools for multirate system design. But again it does not offer a rigorous methodology that can provide adequate prediction of the resultant system performance. The decomposition methods do not seem to be feasible as tools for design purposes for several reasons. Perhaps the most important one is that performance specification for the original multirate system can hardly be translated in terms of the equivalent single rate system.

The structural analysis of last chapter offers a new approach to multirate system design. The separation property asserts that a large class of hybrid controllers K can be decomposed into $S_2 \cdot C \cdot S_1$, where C is finite dimensional LTI operator. Also it is apparent that the only designable part of such a controller is C. Furthermore, the correspondance between K (of this class) and C is unique, therefore the design of K is exactly that of C. But K and C have different "outside worlds" to control: the latter must control a system which has SAH operators as part of its dynamics. So the design a hybrid controller K for an LTI system can be transformed into the design of an LTI controller for a time-varying system.

The analysis of SAH operators underlines the basic fact that when sampling rates are chosen appropriately these operators are close to LTI operators. It is thus justifiable to view the operators S_i's as perturbations of $\bar{S}_i$'s. Therefore the system controlled by C is an LTI system with time varying perturbations, and hence the design is essentially a robustness issue. Regarding the effects of sampling as perturbations implies that the high frequency harmonics incured by sampling are undesirable, hence should be restrained. In practice, this has been achieved by either fast sampling or low pass filters. However, the lack of understanding of the nature of these high frequency harmonics encourages excessively fast sampling or narrow band low-pass filters, resulting in waste of computation resources and/or degradation of performance. The merit of our approach to the design of sampled systems is seen in that it gives a rational account of the undesirable effects of sampling.

The most important issue of all system properties is stability, others are considered afterwards. Thus robust stability can be regarded as constraint to the optimization of other performance criterion. From the last chapter, we have seen that such a constraint can be expressed in the form of an upper bound on the roll-off rate of certain loop transfer functions. Constraint of this kind are equivalent to putting upper bound on the weighted H^{∞} norm of these transfer functions. This is where H^{∞} techniques come in.

This chapter is concerned with the application of H^{∞} optimization theory to the design of multirate multivariable systems. In section 5.1 some basic facts about H^{∞} synthesis theory are summarized. In section 5.2 we briefly outline some of the computational issues. The emphasis is put on the constraint optimization. Formulation of optimal robustness of multirate sampled data systems is discussed in section 5.3. In section 5.4 we will discuss the minimization of sensitivity (tracking error) under constraint of robust stability. Examples are given to illustrate the numerical aspects of the problems. An attempt is made to give an evaluation of the effectiveness of H^{∞} methods in sampled data systems at the end of this chapter.

5.1 *Outline of H^{∞}* theory

In this section some basic facts about H^{∞} design techniques will be discussed. The material is mainly from [Francis], [Doyle, 1984] and [Safonov et al, 1986]. The system shown in Fig. 5.1.1 is the standard configuration which our discussion is based on.

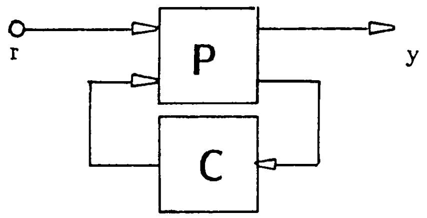


Fig. 5.1.1

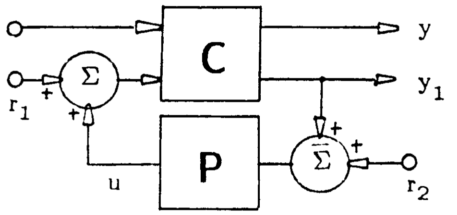


Fig. 5.1.2

In Fig. 5.1.1, P is a given LTI system, with a real proper rational transfer function P(s). It is partitioned conformably with the configuration Fig. 5.1.1 as:

$$P(s) \triangleq \begin{bmatrix} P_{11}(s) & P_{12}(s) \\ P_{21}(s) & P_{22}(s) \end{bmatrix}. \tag{5.1.1}$$

C(s) is also an LTI invariant system, and represents the controller to be designed. It is easy to see that

$$y(s) = \{P_{11}(s) + P_{12}(s)C(s)[I - P_{22}(s)C(s)]^{-1}P_{21}(s)\}\cdot r(s).$$

$$\triangleq F_l[P, C](s)\cdot r(s) \qquad (5.1.2)$$

where $F_l[P, C]$ is the linear fractional map of P over C [Doyle]. Another linear fractional map, $F_u[P, C]$ (U for upper), was introduced in chapter four. The system in Fig. 5.1.1 is said to be internally stable if in Fig. 5.1.2 the mapping from (r, r_1, r_2) to (y, y_1, e) is stable, i.e. all the elements of the transfer matrix are in RH^∞ [Francis]. If there exists C's such that the system is internally stable, P is said stabilizable. A typical statement for an H^∞ design task is: find a C(s) so that:

(i) system Fig. 5.1.1 is internally stable;

(ii) $||F_l[P, C]||$ is minimized.

Suppose that P(s) has a minimal realization with a conformable partition with (5.1.1):

$$P(s) = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix},$$

then the internal stabilizability is equivalent to the condition that the system

$$\begin{bmatrix} A & B_2 \\ C_2 & D_{22} \end{bmatrix}$$

is detectable and stabilizable. Intuitively, this says that all the unstable modes are to be included in the subsystem $P_{22}(s)$. In this case C stabilizes P if and only f it stabilizes $P_{22}(s)$. This is assumed throughout.

A design task stated above is transformed into a solvable form through the following steps:

(i) Q-Parameterization

This is the parameterization of all stabilizing real-rational C's for $P_{22}(s)$, and is achieved through coprime factorizations of $P_{22}(s)$ in the quotient field over the ring of proper and stable real-rational matrices RH^∞. The real-rational requirement will be looked into further in Appendix B. A doubly coprime factorization of P_{22} amounts to finding N_r, M_r, N_l

and $M_l \in RH^{\infty}$, such that

$$P_{22}(s) = N_r M_r^{-1} = M_l^{-1} N_l \,, \tag{5.1.3}$$

together with X_l, Y_l, X_r, and Y_r in RH^{∞} satisfying

$$\begin{bmatrix} X_l & -Y_l \\ -N_l & M_l \end{bmatrix} \begin{bmatrix} M_r & Y_r \\ N_r & M_r \end{bmatrix} = I. \tag{5.1.4}$$

It is known that any proper real-rational stabilizing C(s) is given by the formula

$$\begin{aligned} C(s) &= (Y_r - M_r Q)(X_r - N_r Q)^{-1} \\ &= (X_l - Q N_l)^{-1}(Y_l - Q M_l) \end{aligned} \tag{5.1.5}$$

with a proper real-rational $Q(s) \in RH^{\infty}$. Substituting C(s) defined in (5.1.5) into the system in Fig. 5.1.1, and using the relations in (5.1.4), one can show that

$$\begin{aligned} F_l[P, C] &= P_{11}(s) + P_{12}(I - CP_{22})^{-1} CP_{21}(s) \\ &= T_{11}(s) - T_{12}(s)Q(s)T_{21}(S) \end{aligned} \tag{5.1.6}$$

where $T_{ij}(s)$'s $\in RH^{\infty}$ and are defined by:

$$\begin{aligned} T_{11} &= P_{11} + P_{12} M_r Y_l P_{21} \\ T_{12} &= P_{12} M_r \\ T_{21} &= M_l P_{21}. \end{aligned} \tag{5.1.7}$$

Explicit state space realization for $T_{ij}(s)$'s can be obtained, which is important for computational purposes [Doyle, 1984]. Since the mapping between the proper stabilizing C's and the Q's in (5.1.6) is one to one, the minimization problem $\min_{C} \|T(C)\|$ subject to internal stability is equivalent to that of

$$\min_{Q \epsilon RH^{\infty}} |T_{11} - T_{12} Q T_{21}|. \tag{5.1.8}$$

(5.1.8) is called the model matching problem [Francis, 1986].

There are many possibilities in the choices of M_r, N_r, M_l and N_l and hance T_{ij}'s. It has been shown that particular choice can make T_{12} and T_{21} parts of inner matrices, i.e. there exist $T_{12}^{\perp}(s)$ and $T_{21}^{\perp}(s)$ such that both

$$[T_{12}, T_{12}^{\perp}](s) \text{ and } \begin{bmatrix} T_{21} \\ T_{21}^{\perp} \end{bmatrix}(s)$$

are inner. Since $\| \cdot \|_{\infty}$ is unitary invariant, it follows that

$$\min_{Q \epsilon RH^{\infty}} |T_{11} - T_{12}QT_{21} |$$

$$= \min_{Q \epsilon RH^{\infty}} \left\| \begin{bmatrix} T_{12}^{*} \\ T_{12}^{\perp *} \end{bmatrix} T_{11}[T_{21}^{*}, T_{21}^{\perp *}] - \begin{bmatrix} Q & 0 \\ 0 & 0 \end{bmatrix} \right\|$$

$$= \min_{Q \epsilon RH^{\infty}} \left\| \begin{bmatrix} R_{11} - Q & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \right\| \qquad (5.1.9)$$

Therefore the model matching problem becomes a general distance problem [Chu et al, 1985].

After the formulation of the general distance problem, the rest of the theory concerns the solution to (5.1.9). A special case of the general distance problem is when both $T_{12}(s)$ and $T_{21}(s)$ are square, thus one only needs to solve the Nehari problem:

$$\min_{Q \epsilon RH^{\infty}} \| R - Q \| \qquad (5.1.10)$$

whose solution is well known. The optimal norm is known to be $\|\Gamma_R\|$, where Γ_R is the Hankel operator with symbol R, which is of finite rank for rational R's. For rational R, complete characterization of all the solutions to (5.1.10) is available. Explicit state space formula for Q's that solve (5.1.10) are given in [Glover, 1985]. Francis et al used a different approach based on a Ball and Helton theory [Francis et al, 1984]. Solution to (5.1.10) is important for it is the bottom line of the more general cases.

There is no direct solutions to the general distance problem yet, and it is still an issue for investigation. What is being done is to use a so-called γ-iteration scheme [Chu et al, 1985]. Basic steps are as follows.

i) Two block case

This the case when one of the matrices T_{12} and T_{21} is square. We can assume T_{21}

square without loss of generality. Then the problem is to solve

$$\min_{Q\epsilon RH^{\infty}} \left\| \begin{bmatrix} R_{11} - Q \\ R_{21} \end{bmatrix} \right\| .$$

In order to reduce this problem to the Nehari problem, note the following two equivalent statements:

$$\text{(i)} \qquad \left\| \begin{bmatrix} R_{11} - Q \\ R_{21} \end{bmatrix} \right\| \leq \gamma$$

and

$$\text{(ii)} \qquad \begin{cases} \|R_{21}\| < \gamma \\ \|(R_{11} - Q)\cdot G_o^{-1}\| \leq 1. \end{cases}$$

G_o is the square root of $\gamma^2 I - R_{21}^{*}\cdot R_{21}$ in the following sense: G_o is outer and satisfies $G_o^{*}G_o = \gamma^2 I - R_{21}^{*}R_{21}$. If for a chosen γ it happens that G_o exists and

$$\mu(\gamma) \triangleq \min_{Q\epsilon RH^{\infty}} \|(R_{11} - Q)\cdot G_o^{-1}\|$$

$$= \min_{Q'\epsilon RH^{\infty}} \|R' - Q'\| < 1$$

then for this γ

$$\min_{Q\epsilon RH^{\infty}} \left\| \begin{bmatrix} R_{11} - Q \\ R_{21} \end{bmatrix} \right\| < \gamma.$$

Therefore one can proceed to choose a strictly smaller γ and repeat the above process. This process will converge for there is a lower bound for γ:

$$\gamma \geq \left| \Gamma_{\begin{bmatrix} R_{11} \\ R_{21} \end{bmatrix}} \right| .$$

It has been shown in [Chu et al, 1985] that the function $\mu(\gamma)$ is monotonicly decreasing in the neighbourhood of the optimal γ_{opt}, thus the above procedure will in fact converge to the optimal γ_{opt} if the searching interval is chosen correctly.

(ii) Four block case

In this the case neither T_{12} nor T_{21} is aquare. One has to use the kind of equivalent conditions of above twice to first reduce the problem to a two block problem and then to the

Nehari problem. In doing so, one has to solve two spectral factorizations and one inner-outer factorization [Safonov et al, 1986]. The details are omitted here since they are well known.

A wide range of problems can be formulated into the standard configuration in Fig. 5.1.1. For example the optimal robustness and sensitivity design problems. In the sequel, we will discuss how complicated problems emerge from more realistic design requirements.

5.2 *Robust performance*

The aim of this section is to analyze the robust performance design problem. We will define this problem in the manner suitable for the purpose of sampled data system designs. Then we will propose a procedure to transfer the problem into a form that the standard H^∞ optimization technique of last section is applicable.

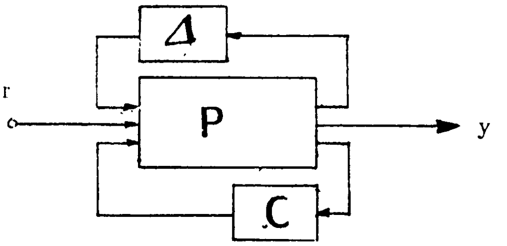


Fig. 5.2.1

Consider the system in Fig. 5.2.1. P(s) is an LTI system, and C is the controller to be designed. The perturbation is lumped into Δ, and is assumed to belong to the class of perturbations Δ satisfying

$$\Delta = \{ \Delta \mid |\Delta \cdot x| \leq |R \cdot x|, \text{ for some LTI R } \}. \qquad (5.2.1)$$

Suppose that P(s) is partitioned according to the configuration of Fig. 5.2.1 as

$$P(s) = [P_{ij}]_{i=1,2,3;\ j=1,2,3.} \qquad (5.2.2)$$

For the convenience of discussion we further denote $P_1(s)$ and P_2 for the submatrices of P:

$$P_1(s) \triangleq [P_{ij}]_{i=1,3;\ j=1,3} ,$$

and

$$P_2(s) \triangleq [P_{ij}]_{i=2,3;\ j=2,3} .$$

For a C(s) we say that the system is robustly stable for the perturbation class Δ if, in addition to internal stability, the system

$$(I - \Delta \cdot F_l[P_1, C])^{-1}$$

is also stable for all $\Delta \in \boldsymbol{\Delta}$. This is sufficiently guaranteed by the condition:

$$\| R \cdot F_l[P_1, C] \| < 1. \qquad (5.2.3)$$

We call$\|$ $F_l[P_2, C]\|$ the nominal performance for a given C. Evidently the performance under the influence of a given Δ is

$$\| F_u[F_l[P, C], \Delta] \|$$

The worst case performance for a given C is defined as

$$\sup_{\Delta \in \boldsymbol{\Delta}} \| F_u[F_l[P, C], \Delta] \|. \qquad (5.2.4)$$

Let $\mathbb{C}$ be the class of stabilizing controllers for P. With the preparation and notation above, we can define the nominal performance optimization (NPO) as

$$\left\{ \begin{array}{l} \inf_{C \in \mathbb{C}} \| F_l[P_2, C] \|, \text{ subject to} \\ \\ \| F_l[P_1, C] \| < 1. \end{array} \right. \qquad \text{(NPO)}$$

NPO is not a satisfactory formulation as it does not take into account the influence of perturbation on the performance. In addition to this drawback, it cannot be transformed into the standard H^{∞} minimization form. It is also desirable to maintain the performance uniformly for the whole perturbation class Δ. Thus we define the robust performance optimization as

$$\left\{ \begin{array}{l} \inf_{C \in \mathbb{C}} \quad \{ \sup_{\Delta \in \Delta} \| F_u[F_l[P, C], \Delta] \| \}, \text{ subject to} \\ \\ \| F_l[P_1, C] \| < 1. \end{array} \right. \qquad \text{(RPO)}$$

But (RPO) is still an unsolved problem, for it is not clear how to take supremum over Δ. We must find an alternative statement that is close to RPO and yet admit a systematic treatment. First we note that the degradation of performance due to perturbation can be bounded in terms of the bound on Δ. Also, the perturbation class $\boldsymbol{\Delta}$ is specified by a uniform bound R on all its members. Therefore the bound of degradation derived in chapter four is valid uniformly for all $\Delta \in \boldsymbol{\Delta}$. In view of this it is suitable to define the following optimization problem:

$$\begin{cases} \min\limits_{\alpha,\beta>0} \dfrac{1}{\alpha\beta}, \text{ subject to:} \\ \min\limits_{C \in \mathbf{C}} \left\| \begin{bmatrix} I & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix} F_l[P, C] \begin{bmatrix} I & 0 \\ 0 & \beta \end{bmatrix} \right\| \leq 1 \end{cases} \tag{P}$$

(P) is suboptimal in the sense that the optimal solution to (P) is suboptimal to (RPO). Since $\frac{1}{\alpha\beta}$ is an upperbound of the robust performance index for any given C, how good (P) is as an approximation to (RPO) depends on how tight the bound $\frac{1}{\alpha\beta}$ is on the index of performance. Example 4.3.6 has shown that this bound is reasonably tight. Therefore we can proceed to solve (P) instead. The question is how to transform (P) into standard H^∞ problem.

Optimization with inequality constraints are normaly dealt with by multipliers. Define

$$N(\alpha, \beta) \triangleq \inf_{C \in \mathbf{C}} \left\| \begin{bmatrix} I & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix} F_l[P, C] \begin{bmatrix} I & 0 \\ 0 & \beta \end{bmatrix} \right\|, \tag{5.2.5}$$

then in principle (P) is equivalent to

$$\min_{\alpha,\beta,\theta} \frac{1}{\alpha\beta} + \theta(N(\alpha, \beta) - 1).$$

where θ is a scaler multiplier. However, in order to solve the above optimization, one needs the derivatives of $N(\alpha, \beta)$, which is neither theoretically verified nor computationally feasible. To avoid derivatives, an one dimensional search scheme is used that, in addition to its simplicity, will have guaranteed convergence. First, some characterization of the feasible region $\Omega \subset [0, \infty)\times[0, \infty)$:

$$\Omega \triangleq \{ (\alpha, \beta) \mid N(\alpha, \beta) \leq 1\} \tag{5.2.6}$$

are needed.

Lemma 5.2.1: (i) Ω in (5.2.6) is radial, i.e. if $(\alpha_0, \beta_0) \in \Omega$, then $(0, \alpha_0)\times(0, \beta_0) \subset \Omega$. A line radiating from (0, 0) into the first quadrant will intersect the boundary of Ω only once.

(ii) There exist $\bar{\alpha}$ and $\bar{\beta}$ such that $\Omega \subset (0, \bar{\alpha})\times(0, \bar{\beta})$.

Proof:

(i) Let $(\alpha_0, \beta_0) \in \Omega$. For $0 < k_i \leq 1$, i=1,2, and define $\alpha = k_1\alpha_0$ and $\beta = k_2\beta_0$. Let also

$$T_0(C) \triangleq \begin{bmatrix} I & 0 \\ 0 & \alpha_0 \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix} F_l[P, C] \begin{bmatrix} I & 0 \\ 0 & \beta_0 \end{bmatrix}.$$

Then we have

$$\left\| \begin{bmatrix} I & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix} F_l[P, C] \begin{bmatrix} I & 0 \\ 0 & \beta \end{bmatrix} \cdot x \right\| \div | x |$$

$$= \left\| \begin{bmatrix} I & 0 \\ 0 & k_1 \end{bmatrix} T_0(C) \begin{bmatrix} I & 0 \\ 0 & k_2 \end{bmatrix} \cdot x \right\| \div | x |$$

$$= \left\| \begin{bmatrix} I & 0 \\ 0 & k_1 \end{bmatrix} T_0(C) \cdot y \right\| \div \left\| \begin{bmatrix} I & 0 \\ 0 & k_2^{-1} \end{bmatrix} \cdot y \right\|$$

$$\leq \ | T_0(C) \cdot y | \div | y |$$

$$\leq \ | T_0(C) |$$

This is true for all stabilizing C's, hence by taking infimum over $C \in C(Q)$ on both sides it follows that

$$N(\alpha, \beta) \leq N(\alpha_0, \beta_0) \leq 1$$

or $(\alpha, \beta) \in \Omega$.

(ii) Let $\bar{\alpha} = \max \{\alpha \mid N(\alpha, 0) \leq 1\}$, and $\bar{\beta} = \{ \beta \mid N(0, \beta) \leq 1\}$. Clearly $\bar{\alpha}$ and $\bar{\beta}$ are finite. We claim that $\Omega \subset B \triangleq (0, \bar{\alpha}] \times (0, \bar{\beta}]$. Because if the pair $(\alpha, \beta) \in \Omega$ but $(\alpha,\beta) \notin B$, then either $\alpha > \bar{\alpha}$ or $\beta > \bar{\beta}$, or both. From (i), $N(\alpha, 0) \leq 1$ and $N(0, \beta) \leq 1$, which is a contradiction to the definitions of $\bar{\alpha}$ and $\bar{\beta}$. □

Lemma 5.2.2: The solution (α, β) to (P) is achieved at the tangent point(s) between the boundary of Ω and the family $\alpha \cdot \beta$ = constants, and the smallest constant is the optima.

Proof:

We use a constructive procedure to prove this lemma.

(0) Choose a pair $(\alpha, \beta) \in \Omega$.

(1) Then, since $N(\alpha t, \beta t)$ is monotonicly increasing and unbounded above for $t \geq 1$, there exists a t_1 such that $N(\alpha t_1, \beta t_1) = 1$, i.e. $(\alpha t_1, \beta t_1)$ is on the boundary of Ω. Let $c_1 = \alpha\beta t_1^2$ and proceed to (2).

(2) Minimize $N(\alpha, \beta)$ on the one dimensional manifold $\alpha\cdot\beta = c_1$. From the previous lemma the search interval is finite. There are two possibilities: One is that the minima achieved is 1, i.e. the curve $\alpha\beta = c_1$ is tangent with the boundary of Ω (possibly at more than one point). In this case the tangent point (α, β) is the optimal pair, the search stops. Otherwise, on the curve $\alpha\beta = c_1$ one will find a point (α, β) such that $N(\alpha, \beta)$ is less than 1. Take this point and go back to (1).

Every time the above process is repeated, the number c_1 is reduced by a strictly non-zero amount. Since the optimum is bounded from below, this process converges. □

Generally speaking, Ω is not convex. Hence there can be more than one solutions. All the solutions are equally good as far as (P) is concerned. From the boundedness of Ω, effective binary search can be used in the two steps outlined above. The geometrical significance of the discussion about Ω and the two-step search scheme can be seen in Fig. 5.2.2.

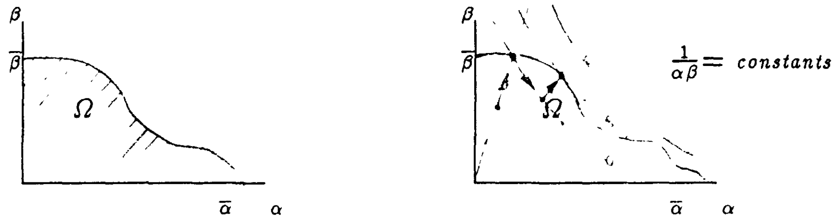


Fig. 5.2.2 The feasible region Ω

The computional burden is considerable, since in every iteration $N(\alpha, \beta)$ has to be evaluated, which is in general a four block H^∞ problem. Computation experience has shown that for most of the well formulated problems, the quantity:

$$\inf_{Q\epsilon RH^\infty} \left\| \begin{bmatrix} R_{11}-Q & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \right\|$$

is quite close to $||\Gamma_R||$ (say less than 10% of error). Thus, after arriving at the general distance problem corresponding to the computation of $N(\alpha, \beta)$, $||\Gamma_R||$ may be used in place of the precise optimum.

5.3 *Stabilizing controllers for sampled data systems*

In this section, we will formulate the stabilization of a multirate sampled data system as an H^{∞} synthesis problem. The purpose of a pure stabilizing design of this section is three-fold. Firstly, we want to demonstrate through examples that the perturbation approach to sampled data systems design is feasible, although theoretically proved in chapter two. Secondly, we will have a better understanding of the nature of the stability constraint, in other words we want to know if there is much room left for other considerations after the stability requirement has been met. Thirdly, and most importantly, we will use the robust stabilization configuration to study the scaling problem in order to reduce conservatism.

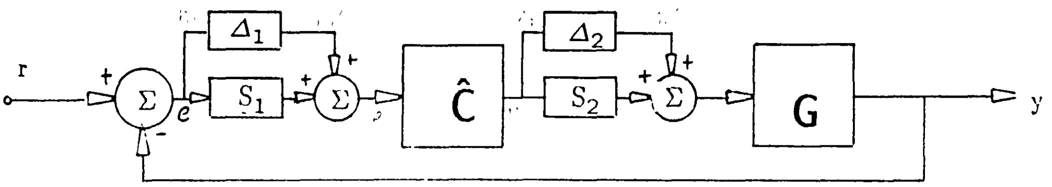


Fig. 5.3.1

Fig. 5.3.1 is a diagram of a hybrid controlled system with separable controller. $\hat{C} = C \cdot F$, where F is a stable and minimum phase filter with relative degree 1. F is included in order that S_2 and hence Δ_2 are well defined. As to the question if there is any restriction to the choice of F, we have:

Lemma 5.3.1. Suppose that F(s) is a stable and minimum phase transfer function with relative degree one. Let C(Q) be the Q-parameterization of all stabilizing controllers for PF. Then $C(Q) \cdot F$ is the parameterization of all strictly proper stabilizing controllers for P.

Proof:

It is evident that for any $C \in C(Q)$, CF is strictly proper. We need to show that all strictly proper C's are in this form. Let $\hat{C}$ be such, i.e. the system $(\hat{C}, P)$ is stable and $\hat{C}$ is strictly proper. Clearly $(\hat{C}F^{-1}, PF)$ is also stable, and $\hat{C}F^{-1}$ is proper. This implies that $\hat{C} \cdot F^{-1} \in C(Q)$, therefore $\hat{C} \in C(Q)F$. □

In view of this, it is not particularly important to choose any special F, for its undesirable influence will be compensated by C(Q). However it is appropriate to choose F sensibly.

Fig. 5.3.1 is transformed into Fig. 5.2.1, with

$$\Delta = \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix}$$

and

$$P_1 = \begin{bmatrix} 0 & G & G\bar{S}_2F \\ 0 & 0 & F \\ I & \bar{S}_1G & \bar{S}_1G\bar{S}_2F \end{bmatrix}.$$

In all our designs, we will always choose F to be the same as $\bar{S}_2$, because this choice will put the right amount of anti-aliasing filtering on the signals going into S_2 while at the same time providing sufficient bandwidth.

It has been shown that the closed loop system in Fig. 5.3.1 is stable if one can choose a stabilizing C such that $||W \cdot F_l[P_1, C]|| \le 1$, where

$$W \triangleq \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \qquad (5.3.1)$$

and both R_1 and R_2 are diagonal matrices defined by

$$R_1 \triangleq \text{diag} \{ R_{T_i}, \}_{i=1}^{m}$$

and (5.3.2)

$$R_2 \triangleq \text{diag} \{ R_{T_j} \}_{j=1}^{n}.$$

Thus, one needs to find a stabilizing C such that $||F_l[P, C]|| \le 1$, with

$$P = \begin{bmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & I \end{bmatrix} \cdot \begin{bmatrix} 0 & G & G\bar{S}_2^2 \\ 0 & 0 & \bar{S}_2 \\ I & \bar{S}_1G & \bar{S}_1G\bar{S}_2^2 \end{bmatrix}. \qquad (5.3.3)$$

Because R_1 and R_2 are non-proper, therefore for P to be proper we require that G be strictly proper. This requirement can be satisfied in practice by the use of filters at the output of the plant, which in general should be done for the sake of reducing aliasing. But we can not use R_i as defined in (5.3.2), for their zeros at s=0 will appear in T_{12} and T_{21} in the corresponding model matching problem. This will make T_{12} and T_{21} rank variant. The existence of an optimal solution is not guaranteed unless T_{12} and T_{21} are of full rank [Francis]. In fact, the program we use for the examples is based on the assumption that P_{12} and P_{21} have full column and row ranks, respectively. To overcome this problem, a small perturbation is introduced so that R_1 and R_2 will have factors $s+\epsilon$ instead of s.

The effect of the minimizing $||W \cdot F_l[P_1, C]||$ is to shape the frequency response of $F_l[P_1, C]$ in such a way that it "avoids" the peak of W. If the optimum is less than one, the C obtained will stabilize the sampled data system (not only the nominal system). Examples have shown that this is indeed the case. But since the envelope on the effect of sampling is conservative, it is expected even the minima fails to be less than one, the system may still be stable. Nevertheless, by inspecting the optimum we can tell if the stabilization has taken much effort for a given set of sampling rates. Since the perturbation here is entirely due to sampling, robust optimization of this section may also give information about whether a set of sampling rates are appropriate. It is evident that the optimum is in general a decreasing function of the sampling rates. But the rates do not influence this optimum in the same way. Thus one can observe which samplers have the dominant effect. The considerations for sampling rate selection are beyond the scope of this thesis. Some general guidance may be found in, say, [Franklin].

Minimizing $||W \cdot F_l[P_1, C]||$ will leave other loop transfer functions in whatever shapes to satisfy the robust stability. As simulations have shown, a pure robust stabilizing design, although gives stable closed loop, offers unacceptable performance.

We have seen in chapter four that scaling can be used to reduce conservatism of the stability assertion, for scaling may bring out structural information of the closed loop (not the perturbation). It should be pointed out that the way we characterize the perturbation class has already taken into account the structure of Δ. The robust stability condition, which is a transformed Small Gain theorem, only involves the maximum singular value of $W \cdot F_l[P_1, C]$, while this singular value is not invariant of a diagonal scaling T as in $TW \cdot F_l[P_1, C]T^{-1}$. Therefore a scaling matrix can reduce conservatism by balancing the singular values. However, the method used in chapter four is of no use here, because we need to scale the transfer functions before the controller is designed, therefore we don't know the frequency at which the singular values assume peak magnitude prior to the design. Also, the meaning of scaling here is different: by scaling the perturbation influences on the system are better identified, and this will lead to a more accurate target. From a different point of view, if scaling helps bring the optimum down, then there will be more room left for other design targets.

Let $\mathbb{T}$ be the class of diagonal matrices with positive numbers on the diagonal, then idealy optimal scaling T should be solved from

$$\min_{T \in \mathbb{T}} \inf_{Q \in H^\infty} \left| T(T_{11} - T_{12}QT_{21})T^{-1} \right| . \qquad (5.3.4)$$

But unless for the first kind of H^∞ minimizations, the above problem is computationally expensive. An approximate solution can obtained by minimizing the lower bound to the optima, i.e. solving T from

$$\min_{T \in \mathbb{T}} \left|\Gamma_{R(T)}\right| , \tag{5.3.5}$$

where $_R(T)$ is defined by the general distance problem corresponding to (5.3.4), which evidently depends on the scaling T. An algorithm is developed to solve (5.3.5) for T, details of which is included in Appendix A. We will assume that such a scaling matrix is computed.

Example 5.3.1: We are interested in a two car three spring system shown in Fig. 5.3.3.

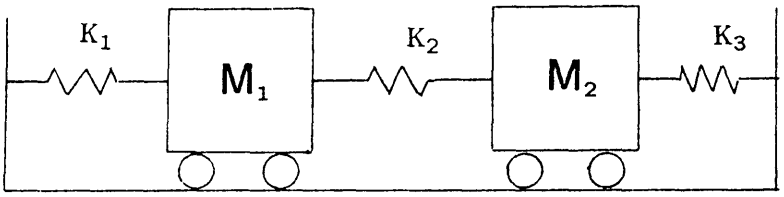


Fig. 5.3.3 Two-car-three-spring system.

To simplify notation without lossing the point, we take $M_1 = M_2 = 1$, $K_1 = 1$. The dynamics of this system is given by:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + B \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -(1+K_2) & -\epsilon_1 & K_2 & 0 \\ 0 & 0 & 0 & 1 \\ K_2 & 0 & -(K_2+K_3) & -\epsilon_2 \end{bmatrix} , B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

We assume that the observations are $x_1(t)$ and $x_3(t)$, and controls are f_1 and f_2. In order to avoid trivial solutions, we assume that the friction factors ϵ_1 and ϵ_2 are negative so that the open loop is unstable. The stiffness factors K_i are such that car no. 1 is sluggish while car No. 2 is rigid. Thus we choose different sampling rates for x_1 and x_3.

We want to design a hybrid controller to stabilize the system. For three sets of sampling rates, the design procedure of this section is carried out. The results are listed in Table 5.3.2.

Table 5.3.2: Robust stabilization

case	sampling rates:	γ_{opt}	Performance	γ_{opt}	
No.	T_1, T_2; T_1', T_2'			(scaled)	
1	0.45, 0.15; 0.3, 0.15	0.802	stable, poor	0.548	
2	0.6, 0.2; 0.4, 0.2	1.131	stable, poor	0.706	
3	1.0, 0.25; 0.5, 0.25	1.504	stable, poor	0.723	
4	1.6, 0.4; 1.2, 0.4	2.965	unstable	1.485 *	

*: After scaling, the design gives stable closed loop.

5.4 *A Design method for sampled data systems*

Stabilization of multirate sampled systems has been formulated and solved as a robust stabilization problem. We have seen that the perturbation approach to sampling is effective because it is possible to shape the relevant loop transfer functions in such a way that the "disturbing" effect of sampling is absorbed. In this section we will use the same idea to solve the design problem of a sampled data system. Clearly, the perturbation due to sampling not only threatens stability, but also influences performance. Therefore the design must first of all ensure stability and also then make the performance under the worst influence of perturbation as good as possible.

We will consider the optimal regulation problem, i.e. we are interested in minimizing the error signal e in Fig. 5.3.1. As in last section, the perturbation is lumped into a singal block, so Fig. 5.3.1 can be transformed into Fig. 5.2.1, with a P matrix mapping from $[n, r, u]^T$ to $[n', e, e']^T$. Define

$$W_1 = \begin{bmatrix} T\cdot W & \\ & \beta w_o \end{bmatrix}, \text{ and } W_2 = \begin{bmatrix} T^{-1} & 0 \\ 0 & \alpha w_i \end{bmatrix},$$

where W is the bound defined in (5.3.1), and T is an appropriate scaling discussed in the previous section. α and β are the balance factors introduced in section 5.2; $w_i(s)$ and $w_o(s)$ are weighting functions designated to reflect the design requirements, and will be discussed later in this section. Since R and T are fixed, we need to substitute appropriate α, β, $w_i(s)$ and $w_o(s)$ into the cost function

$$|W_1 \cdot F_l[P, C] \cdot W_2| = |F_l[\hat{P}, C]|, \tag{5.4.1}$$

Weights selection is a complicated problems, and a systematic way for their choices is still under investigation. We can only consider the most important aspects of their section in a sampled data system design framework. Hence we will confine ourselves to simplest class weights.

Weight selection

The weight w_i at the input is used to characterize the class of input signals that the regulation is primarily designed for. In a sampled data system, the bandwidth of these signals should not be higher than those of the samplers. Therefore, w_i should be chosen as a low-pass filter, if no other specific requirement is posed. Thus we take

$$w_i(s) = \text{diag}\ [\ w_i^1, \cdots, w_i^n]$$

with $w_i^k(s) = \frac{c_k \omega_k}{s+\omega_k}$ to reflect the bandwidth of a particular loop. If there are other considerations, then the $w_i(s)$ defined above should be used as the basic profile on which further shaping can be made. Similarly, w_o is chosen as

$$w_o = \text{diag}[\ w_o^1, \cdots, w_o^n]$$

with $w_o^k(s) = \frac{c_k' \omega_k'}{s+\omega_k'}$ where ω_k's are chosen to reflect the bandwidth in which the error of tracking should be minimized. For the same reason, we should not let the bandwidth to be too wide as to impose an impossible target for a given set of sampling rates. Sometimes it is necessary to place particular emphasis on certain of the input or output signals. This is accomplished by the constant factors c_k and c_k'. How these factors are chosen is a matter for the designer to decide.

Computation of α and β

Suppose that we have chosen the shapes of the weights w_i and w_o. The next thing to do is to calculate α and β according to the procedure outlined in section 5.2. We should note that the process of computing α and β is basicly a trade off between robust stability and performance. Thus if the bound on the perturbation is conservative, this process could result in too much weight on the robustness and hence an unsatisfactory design. This calls for an examination of the basis that sampling be treated as perturbation.

First we note that the LTI bound on the error Δ_T operator is tight in the sense that the bound is achieved. Thus the focus point is on the issue of if our way of formulating robust design is conservative. It is the author's opinion that the worst case performance defined in (5.2.4) is not only difficult, but also, after some reflection, obscure in meaning in our present context. Because by (5.3.4) we implicitly assumed that the perturbation class consisted of members that can be individually identified. Therefore there was the notion of "the worst perturbation". But the nature of Δ_T indicates that there is not an LTI operator that behaves like it. We can only use the collective effect of a perturbation class to characterize Δ_T. Therefore the definition of (P) is justified.

A more fundamental query is, however, whether it is justified to presume that the influence of sampling on performance is always undesirable. The fact that the output of a hybrid controller can change instantaneously can certainly be advantageous. But to take this advantage it is necessary to use the exact discrete time models of both the plant and the controller. For multirate sampled data systems this is very difficult. Therefore, if the basic philosophy is approximate as opposed to exact, sampling effect has to be regarded as undesirable and should be decoupled from the performance output.

Combining these considerations, the following design procedure is proposed:

(1) Choose bandwidth for the input and output weights, multiply the terms with constant factors according to how much emphasis one wants to put on them;

(2) Multiply the robust function with the scaling matrices T and T^{-1} ;

(3) Use the TSOD search methods to find the balance factors α and β;

(4) Solve the H^{∞} minimization problem (5.3.1).

(5) Model reduce the resultant controller if necessary.

Example 5.4.1: We use the same system as in example 5.3.1. The purpose of this example is to demonstrate the steps that lead to a reasonable design. What is particularly interesting is the use of the balance factors α and β.

Experience has shown that the Nyquist frequency is a wise choice for the bandwidth of the weights w_i and w_o, thus they can be chosen according to the sampling rates of the loops.

We have done two designs here, in both cases the shape of the weighting functions are determined according to the discussion in this section. In the first design, we choose the balance by "intuition". The results are shown in Fig. 5.4.1a. The optimal γ_{opt} for the first design is bigger than one, thus one would not be able to tell how the individual objectives behave. More importantly, from γ_{opt} alone we can not decide if the design gives a stable closed loop. We want to point out particularly that though an optimal γ_{opt} bigger than one dose not necessarily incur instability, it does make the performance vulnerable to perturbation.

The scaling procedure of 5.2 ensures that the least amount of sacrifice is made in the trade off between robust stability and performance. Fig. 5.4.1b shows the simulation results of a design that corresponds to an optimal γ_{opt} less than 1. It is evident that the overall performance is improved a great deal, although less weights is put on tracking error minimization.

5.5 *Concluding Remarks*

It is difficult to evaluate the design procedure for there isn't a clearly defined benchmark against which we can make a judgement. But the following points are common to the evaluation of any design methodology.

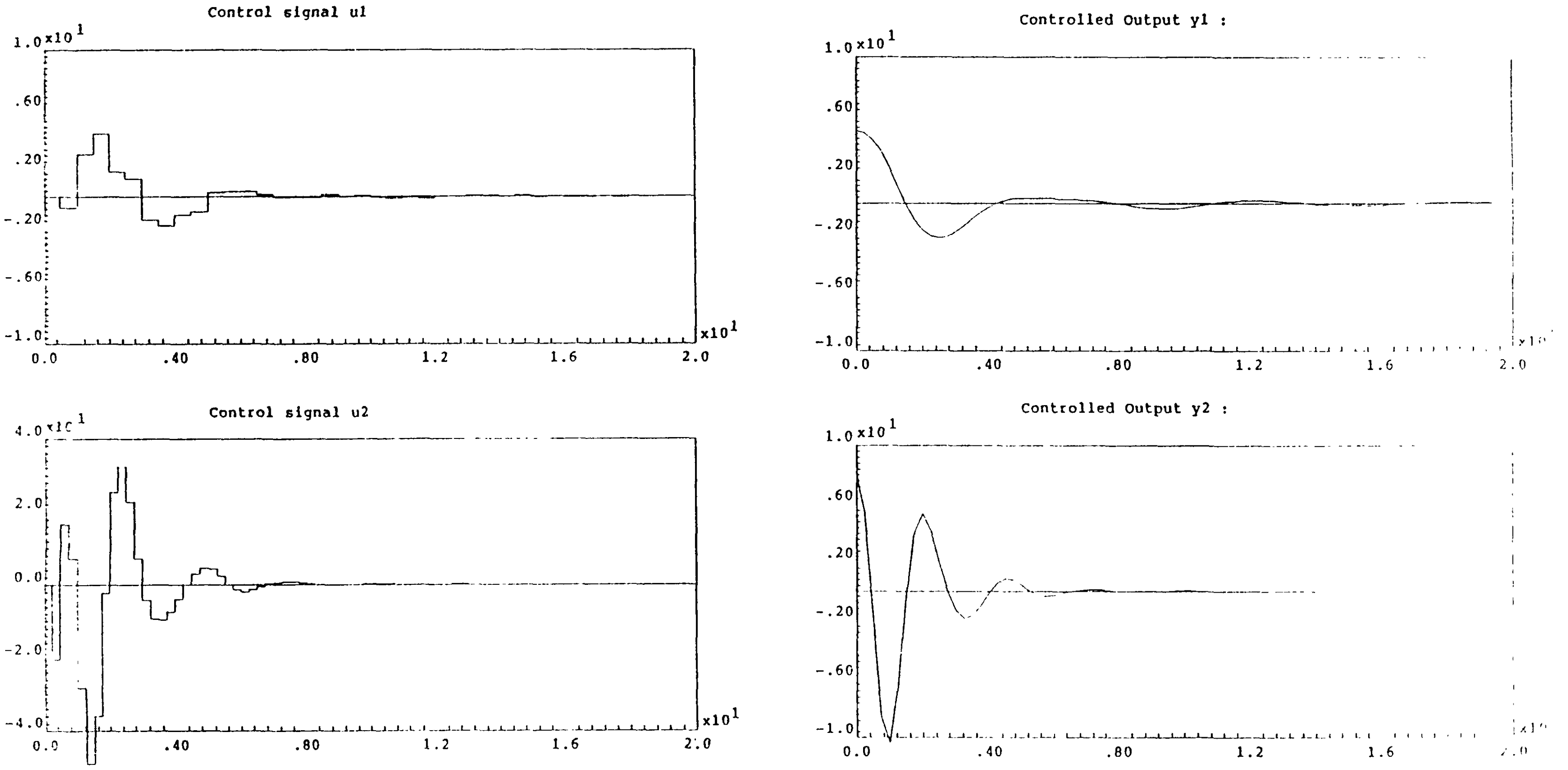
(1) Effectiveness. It is difficult to decide for there is not an objective criterion for this. But from the simulations of the examples, we tend to believe that the method we proposed in this chapter offers reasonably good designs. The central point is whether H^{∞} method can give good design, which is not completely answered yet. Since this method gives stable designs and offers the freedom of shaping the loop transfer functions we are interested in, our own verdict is that it is effective.

(2) Ease of use. This includes whether it is easy to formulate a design task into a form that can be fed into a computer, and whether it is easy to transfer design specification into data used by the design programme. Since this method is based on continuous time approach, there is no need to discretized the plant model. All design specification are stated in Laplace frequency domain, which is understood by control system design engineers. The only things one needs to decide are the sampling rates and the shape the closed loop transfer function. It is straightforward to transfer these information into the standard configuration of H^{∞} optimal design.

(3) Systematic approach to sampled data design. The information about the shape of closed loop leads to the weighting functions for the part of the transfer function representing performances. The bounds on the error of approximation to SAH operators are used for robustness weighting functions, which is the other part of the transfer function we minimize in H^{∞} norm. The controller will be in state space form, and can be easily transformed into

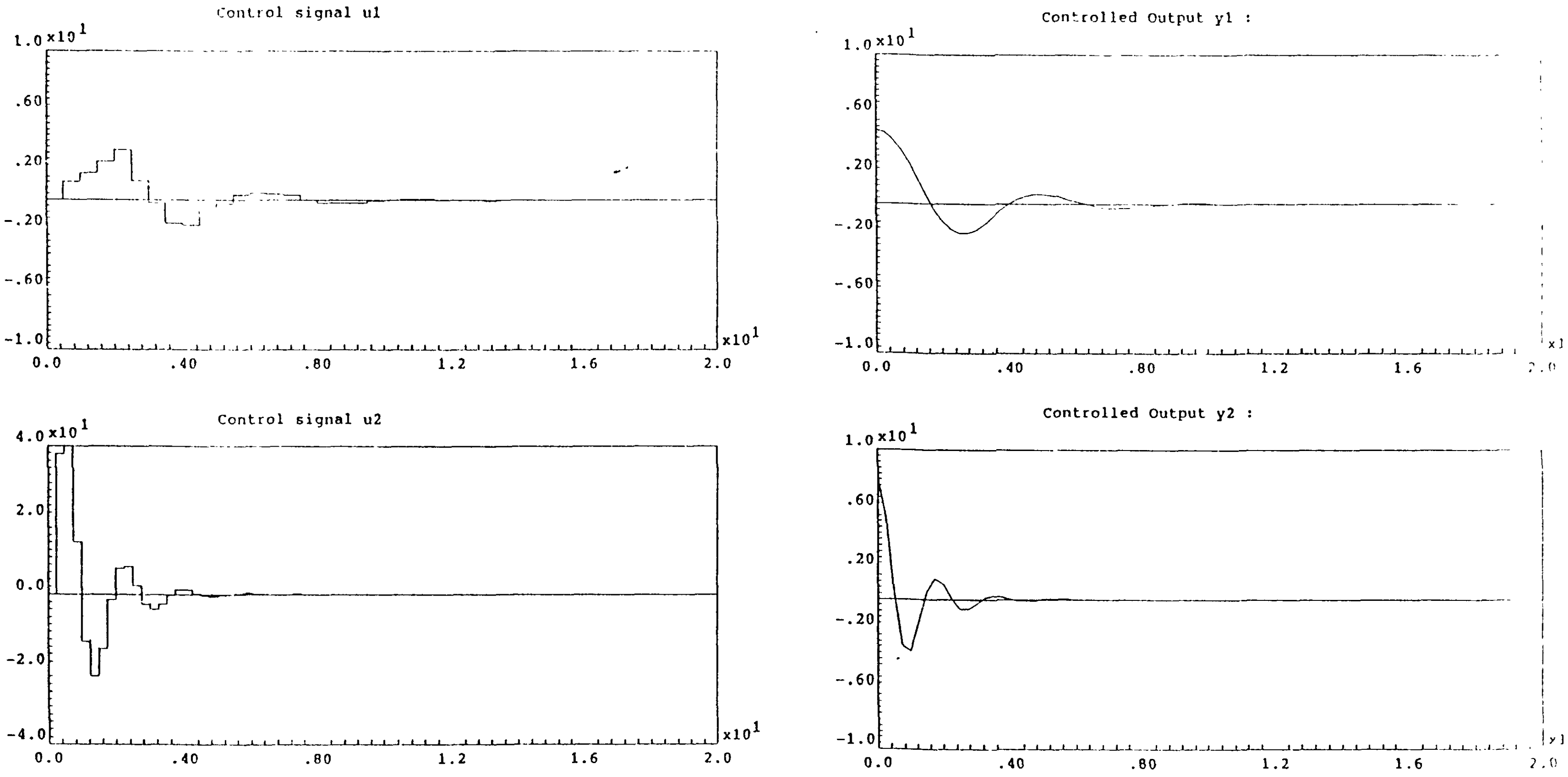
Simulation results of example 5.4.1:

(5.4.1a) Intuitively selected balance



Simulation results of example 5.4.1:

(5.4.1b) With the computed balance factors α and β



realizable computer software.

(4) Complexity of controller. In terms of MacMillan degree, the resultant controller is as complicated as the P matrix in the standard configuration for H^∞ minimization [Limebeer et al, 1986]. In order to have the freedom to shape the loop transfer functions and to accommodate the sampling effect explicitly, a P matrix of high MacMillan degree is inevitable. If we accept the problem formulation as being sensible, then a high degree controller is the necessary cost. This is one side of the picture. On the hand, we can model reduce the resulting controller. For example, the controller obtained form Example 5.3.1 has 12 states, while a 6 state optimal Hankel approximation gives nearly identical performance. Basicly, we have two choices: one is to design a controller for a simplified problem formulation, and hence to get a simple controller; the other is to include all important features in the problem formulation, but model reduce the resultant controller. Since a systematic model reduction procedure will have a better chance to preserve the fundamental characteristics of a system, our intuition is that we should adopt the latter of the two choices.

(5) Computation effort. This is not an important issue in an off-line design context. The core of the design procedure is an H^∞ minimization problem. Compared to this the pre-processing and post-processing are neglectable. The computational issue of H^∞ optimization is beyond the scope of this thesis, apart from the observation that it is improving continuely.

The ultimate test has to be by realistic designs. The author believes that all the signs show that it is a promising direction, alone which more research is undoubtably needed. Here are some immediate questions that need be looked into.

Appendix A.

Robustness scaling

Since the scaling matrices have finite elements, we can solve for them from an optimization procedure. Since

$$|\Gamma_{R(T)}| > 0,$$

there must exist a solution T that minimizes the Hankel norm. The mapping from T to R(T) is complicated, and it is impossible to calculate the derivatives analytically. It is easy to compute the derivatives numerically, but to do so we must show that $|\Gamma_{R(T)}|$ is differentable with respect to the scaling matrix T. The mapping T $\mapsto$ $|\Gamma_{R(T)}|$ involves the following calculations. First, transfer the model matching problem

$$\inf_{Q \in H^{\infty}} |T(T_{11} - T_{12}QT_{21})T^{-1}|$$

into the general distance problem

$$\inf_{Q \in H^{\infty}} \left| R(T) - \begin{bmatrix} Q & 0 \\ 0 & I \end{bmatrix} \right| .$$

To do so one needs to solve two Riccati equations so that the stabilizing Q parameterization is such that both TT_{12} and $T_{21}T^{-1}$ are parts of inner matrices. A state space realization can be found for R(T):

$$R(T)(s) = B(T)\,[sI - A(T)\,]^{-1}C(T) + D(T).$$

Then controllability and observability grammians P(T) and Q(T) can be solved. Finaly, the Hankel norm of R(T) is given by

$$\lambda^{\frac{1}{2}}_{\max}\,[P(T)Q(T)].$$

We will first show that the solutions of the two Riccati equations are differentiable functions of T, hence so are A(T), B(T) and C(T). Then we have to show the solutions of the two Lyapunov equations, i.e. the grammians are differentiable functions of A(T), B(T) and C(T). Then it is a known facts that the eigenvalues of a matrix are analytic functions of the elements of the matrix. We have the following general result on the continuity property of the solution of Lyapunov equations.

Proposition A.1: Let A(t), B(t) and Q(t) be continuous nxn matrix functions on [-δ, δ] (hence bounded), for some $\delta>0$. Assume that $\lambda_i[A(t)] \neq \bar{\lambda}_j[B(t)]$ for i,j = 1,$\cdots$, n, and for t$\in$[-δ, δ]. Then the Lyapunov equation

$$A(t)X + XB(t) + Q(t)t = 0 \qquad \text{(A.1)}$$

has a unique solution X(t) = Y(t)t, for some Y(t) bounded on [-δ, δ].

Proof:

The conditions on the eigenvalues of A(t) and B(t) ensure that there exists a Y(t) for each t$\in$[-δ, δ] satisfying

$$A(t)Y(t) + Y(t)B(t) + Q(t) = 0 \, .$$

Hence the unique solution to (A.1) is Y(t)t. Y(t) is bounded since Q(t) is bounded. □
This proposition above all says that the solution of a Lyapunov equation isn O(t) if the free term is O(t).

Lemma A.1: Let A(t), B(t) and Q(t) be the same as in proposition A.1. In addition, assume that they also differentiable at t=0. The the solution is also differentiable at t = 0.

Proof:

It is easy to see

$$A(t)\,[x(t) - X(0)] + [X(t) - X(0)]\,B(t) + [\cdots]O(t) = 0,$$

where [$\cdots$] is a bounded matrix. Therefore the conclusion follows. □

Lemma A.2: Let A(t), Q(t) be nxn matrix functions continuous in t $\in$[-δ, δ], and differentiable at t=0. Let also R = B$\cdot$B^T. Suppose that the pair (A(t), B) is stabilizable in [-δ, δ], fro some $\delta>0$. Then the unique stabilizing solution to

$$X(t)A(t) + A^T(t)X(t) + XRX + Q(t) = 0$$

is continuous and differentiable at t = 0.

Proof:

The Riccati equation has a unique solution for each $t \in [-\delta, \delta]$. Also we have by hypothesis that $A(t) - A(0) = O(t)$, and $Q(t) - Q(0) = O(t)$. Hence

$$[X(t) - X(0)][A(0) + RX(t)] + [A^T(t) + X(0)R][X(t) - X(0)] + O(t) = 0 .$$

Since A(t) is continuous in t, hance the stabilizing solution for A(0) is also stabilizing for A(t) when t is sufficiently small. Thus the above Lyapunov equation satisfies the conditions for lemma A.1, so we have that $[X(t) - X(0)] = O(t)$, or that X(t) is continuous at t=0 and $\operatorname{limit}_{t \to 0} \frac{1}{t}[X(t) - X(0)]$ exists. □

To simplify presentation, Doyle's notation of Riccati equation is used here. For the Riccati equation

$$XA + A^TX + XRX + Q = 0,$$

denote the unique (if exists) stabilizing solution by

$$\text{Ric.}\begin{bmatrix} A & R \\ -Q & -A^T \end{bmatrix}.$$

If in the standard formulation we have minimal realization:

$$\begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix} P(s) \begin{bmatrix} T^{-1} & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{bmatrix},$$

then the two Riccati equations we are concerned are

$$X(T) = \text{Ric.}\begin{bmatrix} A - B_2D_{12}^TC_1 & B_2B_2^T \\ -C_1^TD_{12}^{\perp}(D_{12}^{\perp})^TC_1 & -(A - B_2D_{12}^TC_1)^T \end{bmatrix}$$

and

$$Y(T) = \text{Ric.}\begin{bmatrix} (A - B_1D_{21}^TC_2)^T & C_2^TC_2 \\ -B_1(D_{21}^{\perp})^TD_{12}^{\perp}B_1^T & -(A - B_1D_{21}^TC_2) \end{bmatrix}.$$

Note that B_2 and C_2 are not changed by T. If we let T be in the form of

$$T = \text{diag}\,(1, 1, \cdots, 1 + t_i, \cdots, 1) \qquad \text{(A.2)}$$

where $t_i \in [-\delta, \delta]$ and $0 < \delta < 1$, then the coefficients of both Riccati equations satisfy the conditions of lemma A.2, i.e. they are differentiable functions of t_i, hence X(T) and Y(T) are differentiable at $t_i = 0$. Define

$$F = -(D_{12}^T C_1 + B_2^T X)$$

and

$$H = -(B_1 D_{21}^T + Y C_2^T) ,$$

then in the state space realization of R(T) all the elments are differentiable functions of t_i in $[-\delta, \delta]$. Lemma A.1 can then be applied to show that the associated controllability and observability grammians are differentiable as well. It is a well known facts that compositions of differentiable functions are differentiable. Thus we conclude with a

Theorem A.3 : For a scaling T defined as

$$T = \mathrm{diag}(1+t_1, \cdots, 1+t_n) ,$$

the Hankel norm of R(T) is differentiable at T = I, i.e. the matrices $\frac{\partial}{\partial t_i} |\Gamma_{R(T)}|$ exist.

Having established the differentiability of $|\Gamma_{R(T)}|$ with respect to the scaling matrix T, we can use the following numerical procedure to optimize the scaling.

(0) Let $T = I_n$ and an index $k = 0$;

(1) Formulating the P matrix, with scaling T absorbed in; $k = k + 1$;

(2) Use finite difference method to compute the derivatives

$$d_i = \frac{\partial}{\partial t_i} |\Gamma_{R(\hat{T}_i)}| , \; i = 1, \cdots, n;$$

where

$$\hat{T}_i = \mathrm{diag}[\, 1, \cdots, 1 + t_i, \cdots, 1]$$

(4) Do one dimensional search on the steepest descent direction

$$T_k = I_n + \mathrm{diag}[\, d_1, \cdots, d_n \,]\cdot t$$

where t is such that $d_i \cdot t + 1 > 0$, and it minimizes $|\Gamma_{R(T_k)}|$.

(5) If d_i's are small enough, stop. Otherwise go back to (1) with the $T = T_k$ computed in (4).

This procedure will converge, for in each step the Hankel norm of R(T) is reduced and this norm is bounded from below. However, there is no guarantee that the convergence is towards a global minima. More is yet to be understood of the relation between $|\Gamma_{R(T)}|$ and T.

Appendix B

On the Q-parameterization

We have seen in chapter three that for the calss of separable hybrid controllers K, the input-output relation can be expressed as:

$$K = S_1 \cdot C \cdot S_2, \qquad \text{(B.1)}$$

where C is an LTI operator. Proposition 3.0.1 asserts that when C ranges through all real rational functions, (B.1) covers all the digital realizations whose computing units satisfy the following conditions:

(1) The A_{ij} matrices in (4.1.1) do not have zero eigenvalues;

(2) $\ln(A_{ij})$ are real metrices.

Without these conditions, the C matrix in (B.1) can well be complex irrational function of s. A typical element of C would be

$$C_{ij}(s) = R_{ij}(s) \cdot \exp(\text{-}skT)$$

where $R_{ij}(s)$ is complex rational, T is the state updating rate and k is an integer. We may assume that there is no delays in the hybrid controller for they do not seem to give rise to a better controller. There is, however, no justification for an *a priori* assumption that $C_{ij}(s)$'s are real rational.

But in all the formulations of this chapter, we assumed that everything is real rational. In particular, C is supposed to be real rational. Thus, the optimal design is only optimal when confined to a subclass of separable hybrid controllers. The purpose of this appendix is to show that the loss of generality in the formulations of this chapter do not incur a suboptimal design as far as the non-delay class is concerned.

Let CH^{∞} denote the set of all proper complex rational matrices which do not have poles on the closed right half complex plan. With the normal definition for addition and multipication, CH^{∞} is ring with much the same properties as RH^{∞}. In particular, any $C \in CH^{\infty}$ has right-coprime and left-coprime factorizations. We note that RH^{∞} forms a subset of CH^{∞}, therefore we see that (5.1.5) gives the complete parameterization of all stabilizing C(s)'s in CH^{∞}, for $Q(s) \in CH^{\infty}$. Since the plant is real rational, hence all the matrices in (5.1.5) are real rational apart from Q(s). This implies that in the derived general distance problem (5.1.9) R(s) is real rational. But it is known that the infimum for (5.1.9) with $Q \in RH^{\infty}$ is not bigger than the same infimum taken with $Q \in H^{\infty}$, and hence not bigger than the one with $Q \in CH^{\infty} \subset H^{\infty}$ [Francis]. Thus we conclude that the formulation in this chapter will give optimal controllers to the H^{∞} criterion, if we preclude any pure delay in the controllers.

References

1] Amit, N., "Optimal Control of Multirate Digital Systems", Stanford Univ., NASA Grant NSG 4002, 1980

2] Araki, M. "Multivariable Multirate Sampled Data Systems: State-Space Description, Transfer Characteristics, and Nyquist Criterion" IEEE. AC Vol. AC-31, No. 2, Feb. 1986.

3] Astrom, K. J. "Zeros of Sampled Systems" Automatica. Vol. 20. No.1, 1984. (D14)

4] Ahmed K. El-Sakkary. "The Gap Metric: Robustness of Stabilization of Feedback Systems" IEEE Tran. Vol. AC-30, March 1985

5] Boykin, W. H. "Analysis of Multiloop, Multirate Sampled-Data Systems" AIAA J. , Vol. 13, April 1975. (D5)

6] Boykin, W. H. "Multirate Sampled-Data Systems Analysis via Vector Operators" IEEE Tran. automatic control, Aug 1975. (D6)

7] Byrnes, C. I. "A Several Complex Variable Approach to Feedback Stabilization of Linear Neutral Delay-Differential Systems" Math. Systems Theory 17, 1984 (M10)

8] Callier, F. "An Algebra of Transfer Functions for Distributed Linear Time-Invariant Systems" IEEE Tran. CAS-25. Sep. 1978. (G8)

9] Callier, F. M. and Desoer, A. "An Algebra of Transfer Functions for Distributed Linear Time-invariant Systems." IEEE AC- 25, No. 9, Sept. 1978

10] Chang B. C. and J. B. Pearson, "Optimal disturbance Reduction in Linear Multivariable Systems." IEEE Trans. AC- 29, No.10, Oct 1984.

11] Chatelin, F., *Spectral Approximation of linear operators*, Academic Press, 1983

12] Coffey, T. C. "Stability Analysis of Multiloop, Multirate Sampled Systems" AIAA J. Vol. 4, No. 12, Dec. 1966. (D3)

13] David, P. "Controlability and Stabilizability in Multi-pair Systems" Siam J. Control and Opti. Vol.18. Sep 1980.(M8)

14] Davis, J. H. "Fredholm Operators, Encirclements, and Stability Criteria" SIAM J. Control. Vol.10 1972(M4)

15] Davis, J. H. "Mean-Square Gain Criteria for the Stability and Instability of Time-Varying Systems" IEEE Tran. AC-27 April 1972. (M3)

16] Davis, J. H. "Stability Conditions Derived From Spectral Theory: Discrete Systems With Periodic Feedback" SIAM J. Control. Vol.10. Feb. 1972. (M2)

17] Desoer C. A., Ruey-Wen Liu and R. Saeks, "Feedback System design: The Fractional Represention Approach to Analysis and Synthesis." IEEE Trans Vol, AC-25 No. 3, June 1980.

18] Doyle , J. C. and Stein, G. "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis." IEEE Trans. AC-26, No. 1, Feb 1981.

19] Duren, P. L., *Theory of H_p Spaces, Academic Press, New York*, 1970

20] Francis, B. A., *A Course in H^∞ Control Theory*, McGraw-Hill Inc., 1986

21] Francis B. A., Helton, J. M. and Zames, G. "H^∞-Optimal Feedback Con- trollers for Linear Multivariable Systems." IEEE Trans, Vol. AC-29, No. 10, Oct 1984.

22] Francklin, G. F. and Powell, J. D., *Digital Control of Dynamic Systems*, addison-Wesley Publishing Co. , 1980

23] Glasson, D. P. "A New Technique for Multirate Digital Control Design and Rate Selection" J. Guidance, Vol. 5, July-Aug. 1982. (D12)

24] Glasson, D. P. "Development and Applications of Multirate Digital Control" Control Systems Magazine, Nov. 1983. (D13)

25] Houpis, C. H. "Refined Design Method for Sampled-Data Control Systems: the Pseudo-Continuous-Time (PCT) Control Systems Design" IEE Proc., Vol. 132, Pt. D, No. 2, March 1985. (D18)

26] Jury, E. I. "A Note on Multirate Sampled-Data Systems" IEEE Tran. on Automat and Contr. June 1967. (D4)

27] Kalman, R. E. "A Unified Approach to the Theory of Sampling Systems" J. F. I, May 1959. (D2)

28] Kamen, E. W. "A Transfer-Function Approach to Linear Time-Varying Discrete-Time Systems" SIAM J. Contr. & Opt. Vol. 23, No. 4, July 1985 (M13)

29] Kato, T., *Perturbation Theory for Linear Operators*, Springer-Verlag, 1986

30] Kranc, G. M. "Input-Output Analysis of Multirate Feedback Systems" IRE Tran. Automatic Contr. Nov. 1957. (D1)

31] Kubrusly, C. S. "Mean Squar Stability for Discrete Bounded Linear Systems in Hilbert Space" SIAM J. Contr. & Opt. Vol. 23, No. 1, Jan. 1985 (M14)

32] Levan, N. "The Stabilizability Problem: A Hilbert Space Operator Decomposition Approach" IEEE Tran. CAS-25 Sep 1978.(M5)

33] Levan, N. and Rigby, L. "Strong Stabilizability of linear Contractive Control Systems on Hilbert Space" SIAM J. Con. and Opt. Vol. 17, No. 1, 1979

34] Liepa, P. E. "Feedback Systems in a General Algebraic Setting" IEEE Tran. CAS-25. Sep. 1978. (G7)

35] Litkouhi, B. "Multirate and Composite Control of Two-Time-Scale Discrete-Time Systems" IEEE Tran. AC-30, July 1985. (D17)

36] Lu, C. H. "Optimal Design of Multirate Digital Filters with Application to Interpolation" IEEE Tran. CAS-26, Mar. 1979. (D8)

37] Mita, T. "Optimal Digital Feedback Control Systems Counting Computation Time of Control Laws" IEEE Tran. AC-30, June 1985. (D16)

38] Moroney, P. "The Digital Implementation of Control Compensators: The Coefficient Wordlength Issue" IEEE Tran. AC-25, Aug. 1980. (D9)

39] Murray, J. "Time-Varying Systems and Crossed Products" Math. Systems Theory 17, 1984 (M12)

40] Oz, H. "Some Problems Associated with Digital Control of Dynamical Systems" J. Guid. and Contr. Vol. 3, Nov./Dec. 1980. (D10)

41] Rattan, K. S. "Digitalization of Existing Continuous Control Systems" IEEE Tran. AC-29, March 1984. (D15)

42] Rosenbrock, H. H., *Compter-Aided Control System Design*, Academic Press, 1974

43] Rudin, W. R., *Real and Complex Analysis*, McGraw-Hill Inc., 1970

44] Safonov, M. G., *Stability and Robustness of Multivariable Feedback Systems*, The MIT press, 1980

45] Safonov, M. G. "Stability of Interconnected Systems Having Slope-Bounded Nonlinearities" Oct. 1983, Report. (M9)

46] Vidyasagar, M., Schneider, H. and Francis, B. A., "Algebraic and Topological Aspects of Feedback Stabilization." IEEE Trans. AC-27. No. 4, Aug 1984

47] Walton, V. M. "State Space Stability Analysis of Multirate Multiloop Sampled Data systems" AAS/AIAA Astrodynamics Specialist Conference, Aug 1981. (D11)

48] Whitbeck, R. F. "Digital Control Law Synthesis in w' Domain" J. Guid and Contr. Vol. 1, Set./Oct. 1978. (D7)

49] Whitbeck, R. F. "Frequency Response of Digitally Controlled Systems" J. Guidance and Control, Vol. 4, July/Aug 1981. (D11)

50] Willems, J. C. "Stability, Instability, Invertibility and Causality" SIAM J. Control. Vol. 7. Nov. 1969. (M1)

51] Youla, D. C., Jabr, H. A. and Bongiorno, J. J., JR. "Modern Wiener-Hoph Design of Optimal Controllers—Part 2: The Multivaraible Caces." IEEE Trans. on Auto. Contr., Vol. AC-21, No. 3, June 1976.

52] Zames, G., "Feedback, Minmax Sensitivity, and Optimal Robustness." IEEE Trans Vol AC-28, No. 5, May 1983.

Reference-Appendix

a] Chu, C. C. and Doyle, J. C., "Computational Issues in μ-thesis," IEEE 2$^{\mathrm{nd}}$ CASCAD, Santa Barara, CA, Mar 1985.

b] El-Sakkary, Ahmed K., "The Gap Metric: Robustnes of Stabilization of Feedback Systems." IEEE Trans., Vol AC-30, No. 3, March 1985

c] Limebeer, D. J. N. and Halikias, G. D. "A Controller Degree Bound for H^{∞} Optimal Problems of the Second Kind," To appear in SIAM J. of Control

d] Safonov, P. M., Jonckheere, E. A., Verma, M. and Limebeer, D. J. N. "Synthesis of Positive Real Multivariable Feedback Systems," Int. J. of Control, Vol. 45, 1987

e] Slepian, D., "On Bandwidth," Proceeding of the IEEE, Vol. 64, No.3, March 1976

g] Thompson, P. M., "Conic Sector Analysis of Hybrid Control Systems," Ph.D Thesis, MIT 1982